Master thesis : Multivariate Statistics for the Joint Analysis of Quantitative Maps

Auteur : Salvoni, Geoffrey
Promoteur(s) : Phillips, Christophe; 3457
Faculté : Faculté des Sciences appliquées
Diplôme : Master en ingénieur civil biomédical, à finalité spécialisée
Année académique : 2017-2018
URI/URL : http://hdl.handle.net/2268.2/4535

University of Liège - Faculty of Applied Sciences
University College of London - Functional Imaging Laboratory (FIL) -
The Wellcome Centre for Human Neuroimaging

Multivariate Statistics for the Joint Analysis of Quantitative Maps

---

Graduation Studies conducted for obtaining the Master's degree in Biomedical Engineering by **Geoffrey Salvoni**, under the supervision of Dr. Martina Callaghan and Pr. Christophe Phillips.

Academic Year 2017-2018

[This page intentionally left blank]

Abstract

Life expectancy increase makes the need for understanding the mechanisms underlying ageing more essential. The ongoing development of magnetic resonance imaging (MRI) provides researchers with an innovative tool to study the brain non-invasively. In particular, quantitative MRI produces high-resolution quantitative maps of parameters dependent of relevant biological measures such as myelin, iron and water concentrations. In addition, a voxel-based quantification (VBQ) procedure corresponding to a spatial normalisation and a smoothing of the maps facilitates group analyses by creating images that are spatially comparable among participants.

On one hand, this work aims at studying the correlation with age in the brain microstructures of a group of healthy volunteers (from 19 to 75 years old) using multivariate statistics. To do so, a joint analysis of the quantitative maps of the entire cohort was carried out using permutation tests implemented in the PALM (Permutation Analysis of Linear Models) toolbox and the results were then compared to univariate methods applied in the Statistical Parametric Mapping (SPM) framework. Whole-brain voxel-wise analysis showed extensive correlations with age in the gray and the white matter in line with histologic reports. Nonetheless, a severe loss of sensitivity in the joint analysis was observed due to the lack of signal present in the longitudinal relaxation rate (R_1) modality.

On the other hand, the VBQ approach was analysed in order to evaluate the pertinence of this normalisation method combining weighting and smoothing. The age effect embedded in the parameter maps was removed while an age-study was performed on the weights, supposed to be independent of ageing. Voxel-wise analysis revealed correlations with age around the ventricles and in the corpus callosum, corresponding to morphological effects, but strong robustness in the cortex. The effect size proved to be tiny compared to the real age effects though.

Keywords: ageing, quantitative, univariate, multivariate, permutation tests, PALM, VBQ, R_1, R_2^*, MT, PD^*

[This page intentionally left blank]

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Martina Callaghan for accepting me in her group, for her support and confidence in my work from the very beginning, for her constant availability, her insightful revision of the report as well as for all the instructive meetings. Her guidance was precious in the realisation of this thesis.

Besides, I would like to deeply thank Christophe Phillips for giving me the chance to have an experience abroad, for the numerous meetings and the valuable help I received to achieve this work.

Then, I am grateful to the Physics group for sharing the office together in a friendly and motivating atmosphere, and more especially to Nadege Corbin and Julio Acosta-Cabronero for answering my questions. In addition, I would like to thank Guillaume Flandin and John Ashburner from the FIL for helping on the understanding of various concepts.

Finally and more generally, I would like to thank my family, my girlfriend and my friends for their continuous support throughout all my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Quantitative magnetic resonance imaging (qMRI) offers a new non-invasive perspective in brain studies. Structural MRI data has a lack of inter-site comparability among the images of a cohort, leading to biased morphometric analyses, something overcame by qMRI as it provides absolute measures for a proper comparison across locations and time points (Weiskopf et al., 2013). Particularly, a quantitative multi-parameter mapping (MPM) approach was developed in order to produce high-resolution quantitative maps of the longitudinal relaxation rate (R_1), the effective proton density (PD^*), the magnetization transfer saturation (MT) and the effective transverse relaxation rate (R_2^*) (Weiskopf et al., 2011, 2013; Helms et al., 2008). Their sensitivities to important tissue properties involved in ageing such as water, myelin and iron contents make them very suitable for age-related studies.

It is known that the brain changes constantly during the course of life. However, the underlying processes involved in these modifications are unclear (Peters, 2006). As brain ages, previous studies reported strong demyelination, iron accumulation and atrophy. These effects usually trigger regions that are known to be targeted by neurodegenerative diseases such as multiple sclerosis, Alzheimer's disease, Parkison's disease and Huntington's disease (Zecca et al., 2004; Connor and Menzies, 1996), making ageing the highest risk factor of neurodegenerative diseases (Callaghan et al., 2014). Since only a small percentage of elderly people affected by cognitive impairment exhibit pure pathologies (Fotuhi et al., 2009), it is first essential to investigate age-related differences in the case of healthy ageing.

To do so, a cross-sectional whole-brain analysis of the four parameter maps acquired for a large population of healthy participants was realised (Callaghan et al., 2014). In order to perform a group analysis, there is a need to spatially normalise the different maps so as to generate brain shapes comparable across subjects. After a spatial normalisation, artefacts might still affect the images. Therefore, a smoothing procedure is necessary. One way is to simply apply a Gaussian smoothing. Additionally, Lee et al. (2009) introduced a tissue-specific, smoothing-compensated procedure (T-SPOON) for diffusion tensor imaging (DTI) data leading to less errors in the normalised and smoothed maps. Later, Draganski et al. (2011) introduced a voxel-based quantification (VBQ) approach combining weighting and smoothing that proved to lead to better cross-participants comparisons and to be well suited for quantitative MRI. Its success motivated one part of our

study in which the VBQ procedure was analysed in order to investigate whether it is an appropriate way of smoothing the data.

In the VBQ analysis, the maps of each single quantitative parameter were analysed independently and the results were then assessed using statistical parametric methods within the statistical parametric mapping (SPM) framework. However, the need of a larger set of assumptions while dealing with multiple testing as well as the different assumptions that might follow each modality make classical inference more likely to fail to generate a statistical framework (Winkler, 2017). By contrast, nonparametric methods, offering a viable and flexible alternative (Winkler et al., 2014), were studied.

Nonparametric methods have the advantage over parametric methods of using directly the data so as to find the statistic distribution under the null hypothesis rather than assuming any distribution a priori (Phillips, 2018). The results' sensitivity might decrease but it is very likely to produce more reliable results since no distribution is assumed. More specifically, permutation tests are a type of nonparametric method whereby the data under study is shuffled several times under the null hypothesis and the resulting statistical value for each permutation is computed to obtain a distribution (Winkler et al., 2014). Therefore, it is conceptually simpler, needs fewer assumptions and copes with multiple comparisons (Nichols and Holmes, 2002), making these methods applicable to a wider range of situations. The increasing use of permutation tests motivated the other part of our study in which quantitative maps were studied through multivariate statistics in a joint analysis by performing permutation tests embedded in the Permutation Analysis of Linear Models (PALM) toolbox. The main goal is to investigate what a joint analysis, i.e. considering all four modalities in the same analysis, can bring compared to the study of each modality separately.

First of all, Chapter 2 will give a literature review of the main concepts under which this thesis rely on. Then, Chapter 3 will describe the experiments that were carried out and for which the associated results will be presented in Chapter 4. Chapter 5 will discuss the results. Finally, Chapter 6 will describe the main conclusion of this work and give some future perspectives.

Chapter 2

Related literature and theoretical focus

2.1 Introduction

In this section, the principal theoretical notions that were helpful in the realisation of this thesis are introduced. The goal is to get acquainted with some concepts that are sufficient to understand the experiments undertaken in the project.

2.2 Quantitative maps

Magnetic resonance imaging, commonly known as MRI, is an imaging technique based on the application of various magnetic fields that will influence the behaviour of molecules present in the body. Depending on the acquisition protocol set up on the machine as well as the diverse tissues properties, a bunch of various images can be acquired. Nowadays, thanks to the huge ongoing improvement of MR technology, the qualitative aspect obtained from these shots is reliable. However, the quantitative comparison between a cohort of subjects is much more difficult. This is the reason why the acquired data can be transformed into quantitative maps. They provide maps of MR parameters that are comparable across sites and time points. One approach to obtain quantitative images is the multi-parameter mapping (MPM) protocol. Using this protocol, typically four maps can be obtained from raw images (Figure 2.1).

During the acquisition protocol, a large number of images are acquired. First of all, a sequence mapping the transmit B_1 field is applied. The goal is to calibrate the flip angle because a certain angle α is desired but, in practise, it will not be constant across the brain due to the inhomogeneities in the magnetic field. The low spatial frequencies must therefore be mapped. Then, the map of the static B_0 field is acquired so as to correct distortions due to field inhomogeneities. Finally, multi-echo 3D volume datasets targeting different microstructural properties of the tissues are acquired. T_1-weighted and PD-weighted images at different echo times (TE) along with MT images are obtained. Biophysical models are then used to combine together the images extracted from the protocol to build quantitative maps of tissue properties.

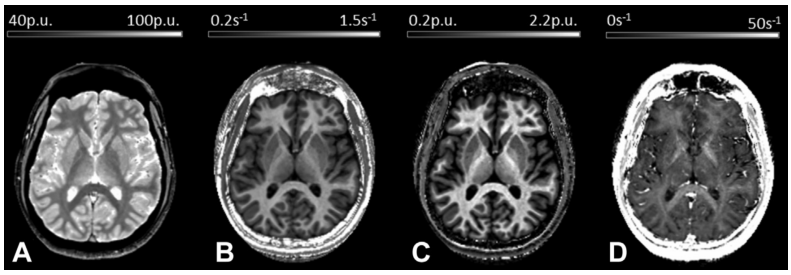

Figure 2.1: Example of quantitative maps (Callaghan et al., 2014): effective proton density, PD* (A); longitudinal relaxation rate, $R_1 = 1/T_1$ (B); magnetization transfer, MT (C), and transverse relaxation rate, $R_2^* = 1/T_2^*$ (D). Units: $p.u.$ = per unit and s = second.

The easiest is the map of transverse magnetization relaxation rate (R_2^*), which only requires the PD-weighted images acquired from the acquisition protocol. At each voxel, a plot of the signal intensity with respect to time for each of these weighted images can be constructed (Figure 2.2). Knowing the echo time (TE) of the protocol, the model:

$$S(TE) = S_0 e^{(-TE \cdot R_2^*)} \tag{2.1}$$

enables to find the R_2^* value of the considered voxel. S_0 represents the intersection with the signal intensity axis.

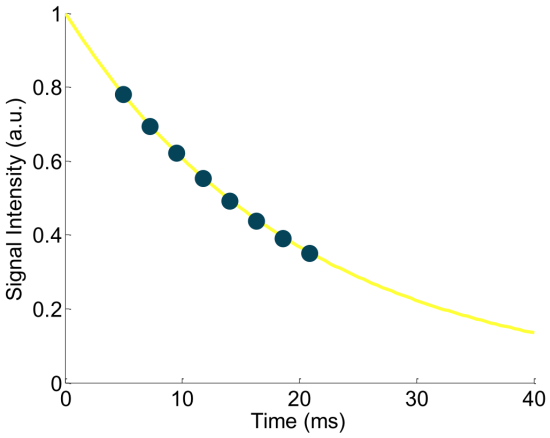

Figure 2.2: Voxel-wise plot of the signal intensity with respect to time obtained from the PD-weighted images (Callaghan).

Regarding R_1, the mean PD-weighted image, the mean T_1-weighted image as well as the B_1 map are required. In the same way as above, knowing the repetition time (TR) and using the data points of the mean images enables to build the model:

$$S(\alpha) = A\alpha \frac{TR \cdot R_1}{\frac{\alpha^2}{2} + TR \cdot R_1} \tag{2.2}$$

where α is the flip angle and A is the effective proton density PD* (see Figure 2.3). It can be thought that the flip angle, a parameter entered by the human, is known. However,

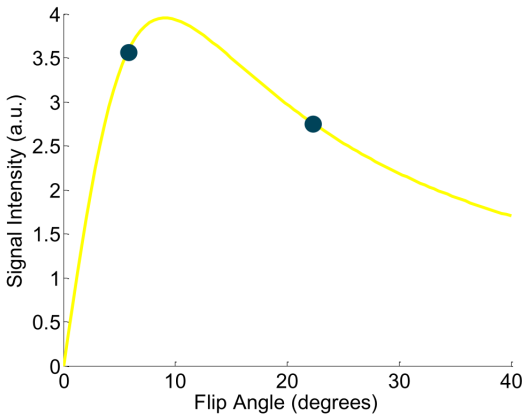

Figure 2.3: Voxel-wise plot of the signal intensity with respect to the flip angle (Callaghan).

this flip angle will not be homogeneous across the whole brain as already mentioned. Therefore, the B_1 map is necessary to account for these variations. In addition, since equation (2.2) contains the PD^* parameter, the same model along with the same B_1 map and the R_1 map are used to get the effective proton density map.

Finally, for the MT map, the R_1 and B_1 maps as well as the mean MT-weighted image are needed and the model is analogous to the previous one:

$$S = A\alpha \frac{TR \cdot R_1}{\frac{\alpha^2}{2} + TR \cdot R_1 + \delta} \tag{2.3}$$

with the δ term to determine.

Another useful advantage of such maps, besides offering a quantitative aspect, is that they are sensitive to some important tissue properties among which are the myelin and the iron contents, two measures indicative of ageing (see Section 2.3). The R_2^* and MT maps are directly sensitive to the concentration of iron and macromolecules, e.g. myelin, respectively while the PD^* map is sensitive to water content. By contrast, R_1 gives information about all of these. It helps to understand why these maps are very appropriate for the study of ageing.

2.3 Biology underneath ageing

The processes occurring with ageing in the human brain are relatively complex, some of which are still misunderstood. However, the influence of both myelin and iron has been demonstrated.

Myelin sheaths, produced by oligodendrocytes, are thin sheaths wrapped around axons to form an insulating layer (Figure 2.4). They favour the conduction of action potentials along the axon. Myelinated axons are abundant in the white matter, giving its whitish aspect, while they are not present in gray matter (Reece and Jackson, 2011).

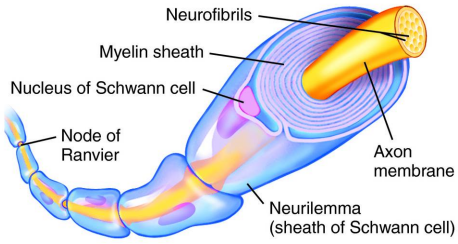

Figure 2.4: Representation of myelin sheaths. Source : Beyond the Dish, *Inhibition of Gli1 Enhances Remyelination Abilities of Endogenous Stem Cell Populations*, `https://beyondthedish.wordpress.com/tag/myelin-sheath/`, page visited on February 28th 2018.

Regarding the oligodendrocytes, they are highly dependent on the iron concentration. Indeed, a sufficient level of iron is required to enable a proper work of these molecules. This is the reason why iron has shown to be abundant in the brain. Its concentration must neither be insufficient nor abnormally high, which might lead to some irreversible cognitive and motor impairment or oxidative injury respectively (Connor and Menzies, 1996). Iron is therefore a key co-factor in the production of myelin sheaths.

It has clearly been demonstrated that iron accumulates with age in the brain regions that are most susceptible to neurodegenerative diseases, along with a loss of myelinated sheaths (Callaghan et al., 2014; Zecca et al., 2004). It is thought that the oxidative stress induced by the iron accumulation are likely to cause neurodegeneration (Zecca et al., 2004). Histologic reports (Hallgren and Sourander, 1958) revealed a concentration increase in the putamen (Figure 2.5), the motor cortex and the prefrontal cortex. These observations are consistent with the fact that white matter is the brain structure subjected to the major impairment in elderly people (Marner et al., 2003).

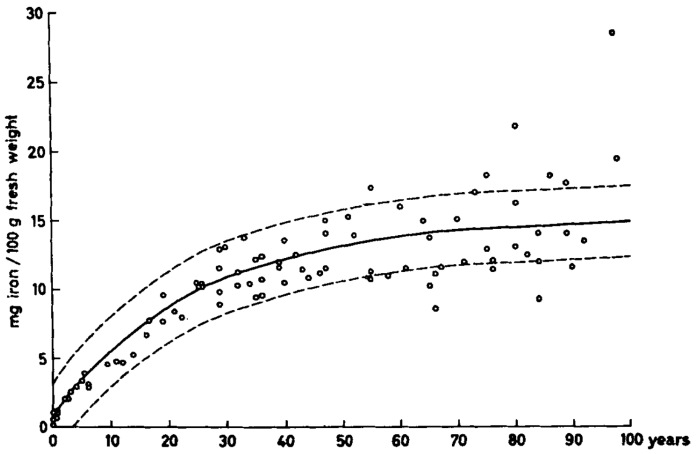

Figure 2.5: Increase of iron concentration with age in the putamen (Hallgren and Sourander, 1958).

In the same way, demyelination, i.e. loss of myelin, was reported in the course of ageing. It leads to reductions in the conduction of action potentials which might further result (Marner et al., 2003) in cognitive decline. Histologic reports (Lintl and Braak, 1983) showed reductions of myelin in the line of Gennari (Figure 2.6) in normal individuals.

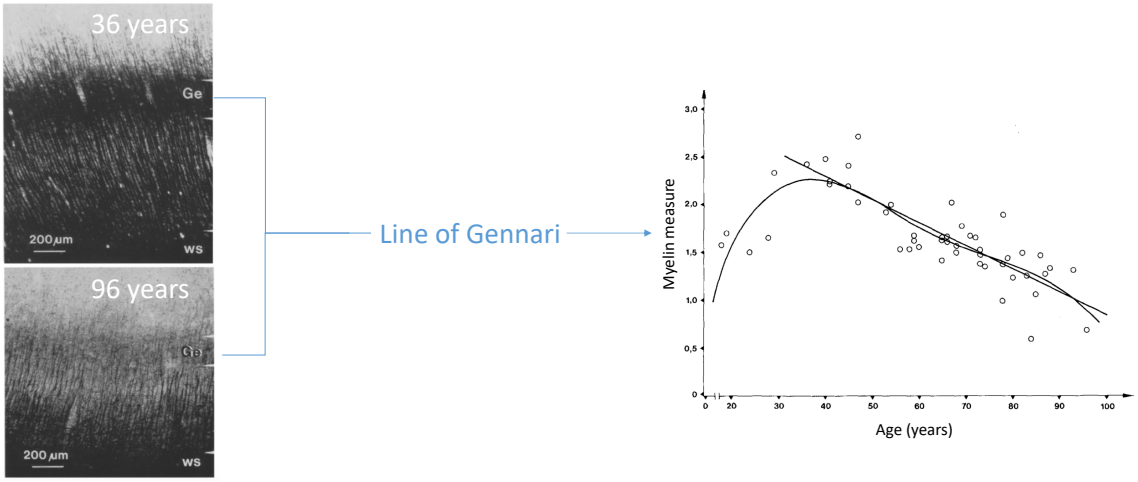

Figure 2.6: Reduction in myelin content of Gennari's line located in the human striate cortex (Lintl and Braak, 1983).

Furthermore, some brain regions exhibit significant atrophy while ageing. Studies using MRI showed a cerebral cortex thinning, volumetric reductions of most subcortical structures and an expansion of the ventricles (Fjell et al., 2009). Figure 2.7 illustrates a brain slice of a young and of an old person. Ventricles expansion can be clearly seen.

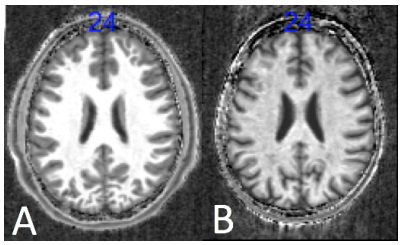

Figure 2.7: Quantitative MT maps of a 21 years old subject (A) and of a 73 years old subject (B). The maps normalised in MNI space are shown.

Demonstrating variations of ageing markers histologically in the brain has the disadvantage of being invasive as it is necessary to have access to the structure under study of the subjects. Due to this invasiveness, the reports are limited to specific regions of interest in the brain. The use of statistical inference on quantitative maps enables to study age-related differences non-invasively in the whole brain in a fast and reliable way.

2.4 Parametric inference

Callaghan et al. (2014) analysed the quantitative maps of a population of 138 subjects to study age-related differences of pertinent biological measures, mainly those described in the previous section, in the brain during normal ageing. To do so, they performed a whole-brain voxel analysis. On one hand, voxel-based quantification was used to assess the differences among the quantitative MR parameters. On the other hand, voxel-based morphometry (VBM) was used to determine the differences in volume or shape of the tissues among the cohort. Based on histologic results, age was assumed to be correlated with reductions in myelin content, modifications in iron and water concentrations as well as with brain atrophy, and that such correlations had to be highlighted in the MPM maps.

In this study, parametric inference was applied on each and every map separately, considering them as independent. Then, the statistical analyses were conducted using the general linear model (GLM) framework embedded in SPM8. The model consisted of four regressors: age, gender, total intracranial volume (TIV) and the type of scanner. One-tailed or two-tailed t-tests were mainly used to assess either the decrease, increase or difference of a parameter with age.

Figure 2.8 represents the regions of the brain subjected to either a decrease of R_1 or MT or an increase in R_2^*. Positive correlations of R_2^* with age were observed especially in the putamen, the pallidum, the caudate nucleus, the red nucleus as well as in extensive cortical regions while negative correlations of MT and/or R_1 were found in the thalamus, the borders of the corpus callosum, the optic radiation and some cortical regions. In general, they found that the results were concordant with histologic reports and presented high specificity for tissue properties. Similar experiments were reproduced and the obtained maps for all quantitative parameters separately are presented in Section 4.1.

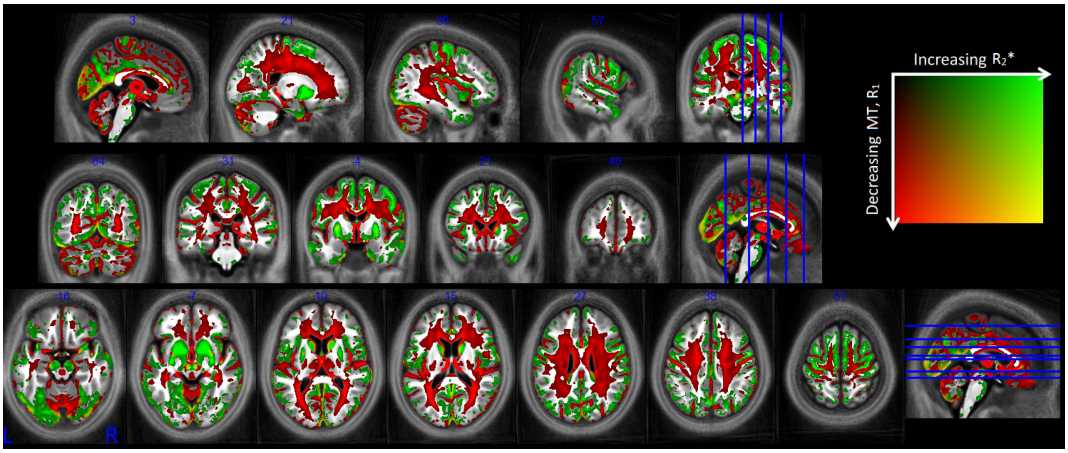

Figure 2.8: Statistical parametric maps showing a whole brain pattern of ageing, at $p < 0.001$ uncorrected for display purposes. R_1 and MT decreases are associated with myelin reductions while iron increases are estimated from increased R_2^*. The t-scores are represented using the square color and some slices of the brain are presented (Callaghan et al., 2014).

In the current section, the four maps were considered as being independent from each other. What if they were somehow related? Studying all the maps as a whole, in one statistical analysis might highlight properties that can not be seen using parametric statistics. This is what will be investigated in this thesis. To do so, nonparametric inference can be applied to compare the results obtained with common parametric statistical tests.

2.5 Nonparametric inference

Nonparametric statistics are techniques that directly use the data in order to build the statistical distribution under the null hypothesis rather than rely on a certain number of assumptions. Here, permutation tests will be especially studied. Briefly, it consists in shuffling the data several times under the null hypothesis and computing the resulting statistical value for each permutation to obtain a distribution. One major advantage is that few assumptions are required.

Winkler (2017) wrote a thesis in which he proposed a general framework allowing permutation inference applied on a GLM. It is thus useful to slightly develop some notions encountered in this framework.

As a reminder, a GLM can be expressed at each voxel by $\mathbf{Y} = \mathbf{M}\psi + \epsilon$ where $\mathbf{Y}$ is the observed data for the whole population under study, $\mathbf{M}$ is the design matrix that gather the regressors of the model, ψ corresponds to the regression coefficients to determine and ϵ are the errors, supposed to follow a normal distribution in the parametric case but not in the nonparametric. Figure 2.9 is a graphical representation of the GLM. It is then

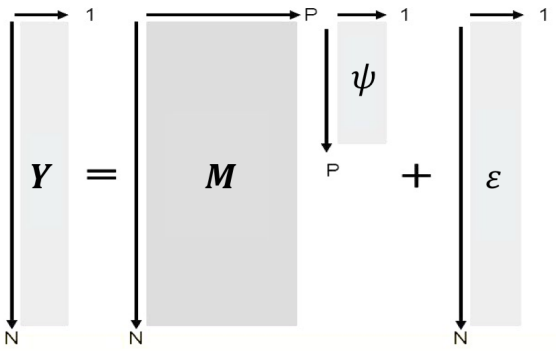

Figure 2.9: General linear model. N is the number of subjects and P the number of regressors.

interesting to partition the model to separate the regressors of interest ($\mathbf{X}$), for the studied null hypothesis, from the nuisance regressors ($\mathbf{Z}$) such that the GLM becomes:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon \tag{2.4}$$

The reason is that the nuisance regressors are likely to modify the assumptions of permutation tests.

From equation (2.4), the statistics used for inference is not necessarily a F- or a t-test as it is common in parametric testing. Any pivotal statistics is convenient here. It means that it is independent of any unknown parameters. Examples are multiple and go from the student's t test to the Pearson's correlation coefficient.

The corresponding p-value is determined by permuting the data several times and to fit the model for every rearrangement. Then, a new realisation of the statistic, called T_j^* with j the permutation index, is computed for every fit. Once all the permutations have been executed, the distribution of T^* under H_0 constructed with the p-value is given by:

$$p = \frac{1}{J} \sum_{j=1}^{J} I(T_j^* \geq T_0) \tag{2.5}$$

where J is the total number of permutations, $I(\cdot)$ is the indicator function and T_0 the observed value. The presence of the indicator function makes the p-values both discrete and multiple of $1/J$. Also, it is worth noting that the minimum value that can be reached is $1/J$ and not 0 because the observed statistic without any permutation is accounted for in the distribution. Therefore, the total number of permutations performed will affect the p-values of the results.

Besides, a corrected p-value distribution accounting for the multiplicity of the tests across an image can be obtained. Indeed, for each rearrangement, the maximum value of T_j across the image, T_j^{max}, is recorded. The FWER(family-wise error rate)-corrected p-value is simply the proportion of T_j^{max} greater than T_0.

An important aspect that governs permutation inference is the principle of exchangeability, which is the manner under which the data are shuffled under the null hypothesis. It is based on the assumption that the joint distribution of the data is unchanged if they are permuted. This assumption, one of the few required for nonparametric tests, can be stipulated in terms of exchangeable errors (EE) or independent and symmetric errors (ISE). In the first case, the distribution of the errors does not change after permuting the data. In mathematics, it can be written that: $\epsilon = \mathbf{P}_j \epsilon$, meaning that the errors can be multiplied by a permutation matrix, i.e. a binary matrix in which each row and each column have respectively only one 1, without affecting their joint distribution. In the ISE case, the principle is the same but with a sign-flipping matrix instead: $\epsilon = \mathbf{S}_j \epsilon$. $\mathbf{S}_j$ is a diagonal matrix with only -1 or +1 depending whether the sign of the observation must be flipped or not. The maximum number of permutations or sign-flippings with n observations is given by $n!$ or 2^n respectively. Both cases can be considered at the same time, the data being both permuted and sign-flipped. An example is shown in Figure 2.10.

In the case of this work, there is a need for multiple testing due to the presence of multiple modalities, the quantitative maps. One way consists in the combination of multiple univariate hypotheses that are analysed separately, the results then being combined together to test a joint null hypothesis. Each individual analysis is called a partial test and exchangeability is assumed only for the observations within each partial test and not between the tests. It is important to say that each partial test will correspond to each quantitative parameter for the purpose of this thesis.

Fisher (1932) proposed a joint analysis for such partial tests. It is based on the idea that the probability of rejecting the joint null hypothesis can be treated as the intersection of the probabilities of each partial test, the intersection being represented as their product: $\prod_k p_k$. From this, he built a statistic for the global null hypothesis given by: $T = -2\sum_{k=1}^{K} \ln(p_k)$, K being the number of partial tests. He showed that T follows a χ^2 distribution with $2K$ degrees of freedom, and from which an uniformly distributed significance level $P_{\texttt{Fisher}}$ can be computed: $P_{\texttt{Fisher}} = 1 - \chi^2$ (Winkler et al., 2016b). A bunch of other methods have been developed since. However, Fisher's still seems to be one of the most popular, with Tippett's (Tippett, 1931), due to their independence assumption about the partial tests as well as their ease of use.

Testing each partial null hypothesis using simultaneous rearrangements is called nonparametric combination (NPC). Figure 2.11 summarizes the two phases of this technique. First of all, a fixed number of permutations is performed independently on each modality. The resulting statistic is then recorded voxel-wise for each shuffling, allowing the empirical null distribution to be estimated for each partial test. From this, the statistic is transformed into a p-value and the extremum statistic among the entire set of voxels is stored. As a second phase, it is possible to combine the various empirical p-values for each shuffling into a joint statistic using Fisher's method for example. Since it is generated from permutations that were performed beforehand, its distribution is directly determined. Therefore, the p-value of the joint analysis is determined.

Furthermore, the algorithm can be modified in two ways so that the whole process happens in a single phase. First, u-values are used instead of p-values to make it more suitable for imaging applications. They simply represent parametric p-values. Second, the empirical distribution is transformed into a known parametric distribution. The main reason for these modifications is the ability to process the algorithm in one phase, i.e. without the need to retrieve the data between both phases. One remarkable feature of NPC is the implicit accountancy for the dependence structure among the partial tests by the synchronised permutations (Winkler, 2017).

Further and more detailed notions and concepts can be found in Winkler et al. (2014, 2016b), but those presented here above are sufficient to give insights about the results and the methods applied in the following. Permutation tests have the advantages of only requiring weak assumptions, either of exchangeable and/or independent and symmetric errors, providing greater or equal power compared to parametric inference but also controlling over the errors of type I, i.e. the false positives, in a better way. Yet, the computational complexity of carrying out such tests is likely to become very important when the number of permutations increases. Due to the high potential of nonparametric inference, Winkler created a toolbox called PALM [1] for Permutation Analysis of Linear Models, allowing to perform permutation tests for statistical inference.

[1] `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM`

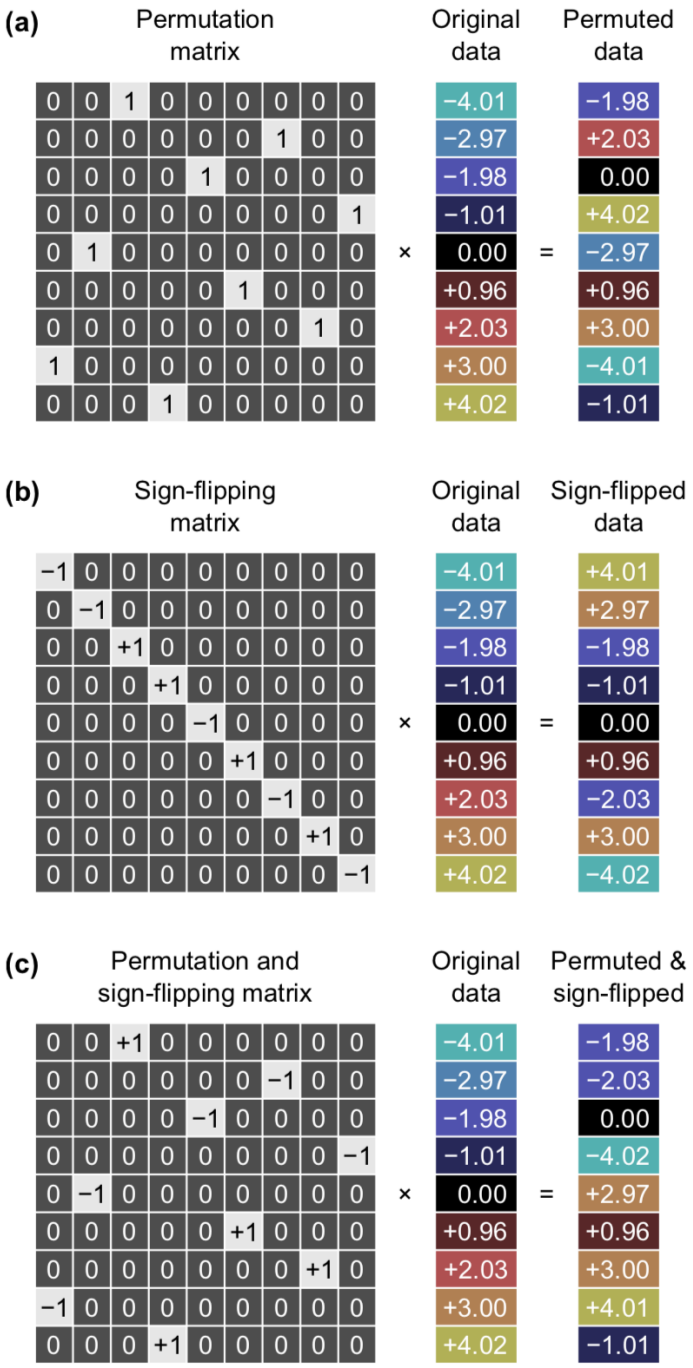

Figure 2.10: Examples of a permutation matrix, sign-flipping matrix or both, respectively in a), b) and c). A) corresponds to permuting the data and b) to change the sign, or not, of the data (Winkler, 2017).

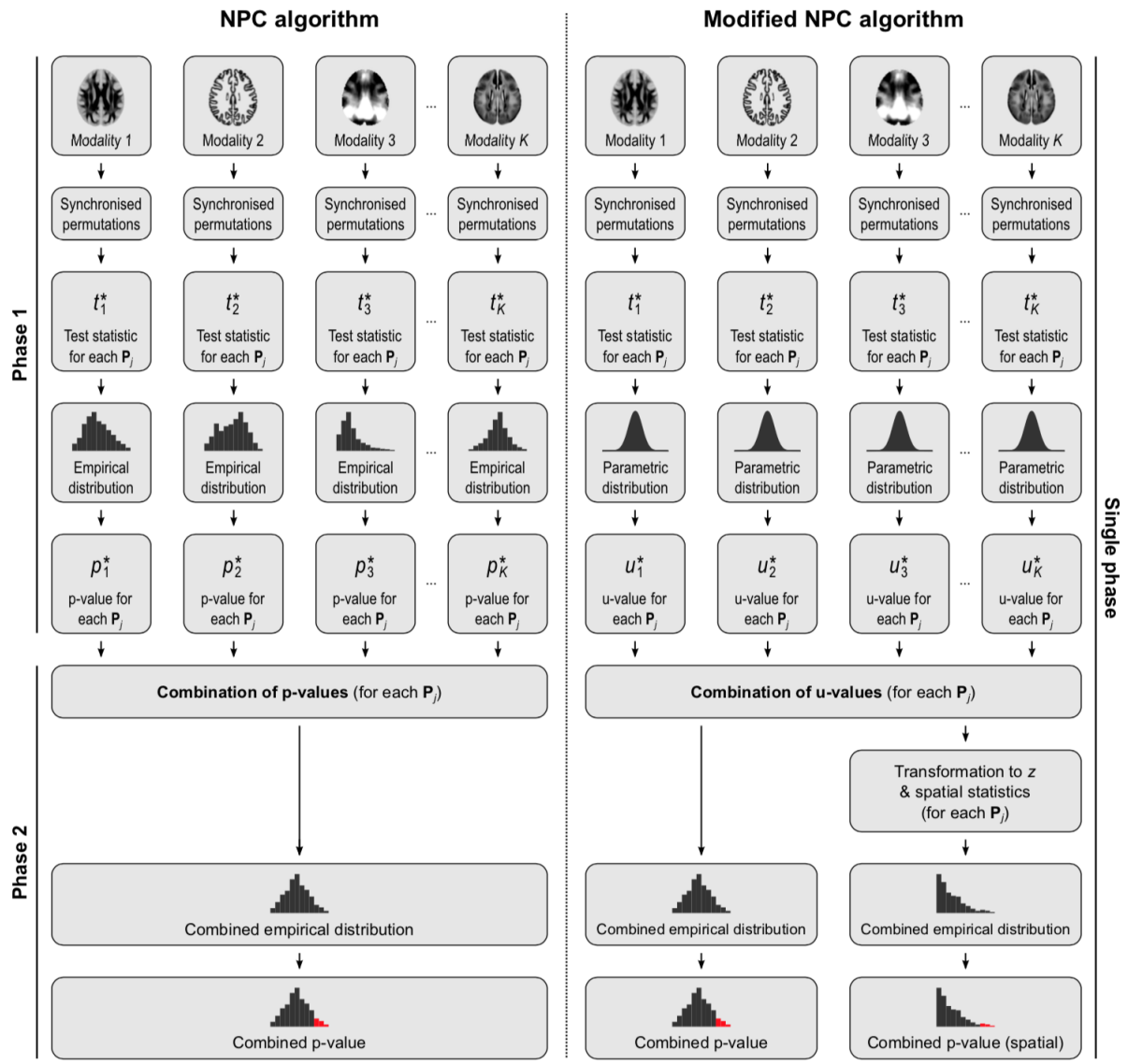

Figure 2.11: Representation of the original and the modified NPC algorithms. The original combines nonparametric p-values. However, it requires huge amount of data storage space, especially as far as imaging applications are concerned. The modified algorithm enables the procedure to run in only one phase instead of two, without the need to extract data to get the distribution of the partial tests (Winkler, 2017).

Chapter 3

Methods

3.1 Data

The data used throughout the project are the quantitative maps produced in Callaghan et al. (2014). The cohort includes 138 participants, aged between 19 and 75 years old, among which 49 are men, with a mean of 46.6 and a standard deviation of 21 years. The subjects were examined on two 3T whole-body MR systems at the Functional Imaging Laboratory (FIL), London. Each MR system was equipped with a standard 32 channel head coil for receive and RF body coil for transmission.

Then, the statistical parametric mapping framework SPM8 was used to process the data. On one hand, the unified segmentation approach was used for each participant to segment the MT maps into gray matter and white matter probability maps. This approach (Ashburner and Friston, 2005) consists in a generative model based on a mixture of Gaussians allowing to perform image registration, tissue classification and bias correction, and is applied to each individual separately. First, the MT-weighted, PD-weighted and T_1-weighted images acquired from the protocol were co-registered, i.e. they were realigned in order to correct for motion correction. Second, physical models from the MPM protocol (Section 2.2) were applied to the aligned images. The quantitative maps were produced. Third, the quantitative MT maps were segmented into gray and white matter tissue class images. On the other hand, inter-subject registration of the tissue class images of all subjects was performed to transform them non-linearly to Montreal Neurological Institute (MNI) space, a common group space, using the diffeomorphic registration algorithm Dartel from SPM (Ashburner, 2007). It basically estimates a smooth and continuous mapping for the voxels from native space to MNI space.

Finally, due to the small misalignments that might still arise after spatial normalisation, the data were smoothed. To do so, two possible ways are the application of either a Gaussian smoothing or a VBQ normalisation. Better results in terms of voxels correspondence were revealed using VBQ normalisation. Indeed, Figure 3.1 shows that when comparing images of a group, the maps of two (or more) participants have shapes that are more similar using VBQ rather than using a normalisation and a Gaussian smoothing. Therefore, a VBQ analysis combining smoothing and weighting as described in Draganski et al. (2011) was used. More details regarding the VBQ approach can be found in Section 3.5 where methods will be presented to evaluate whether the way of smoothing the

data using VBQ is appropriate.

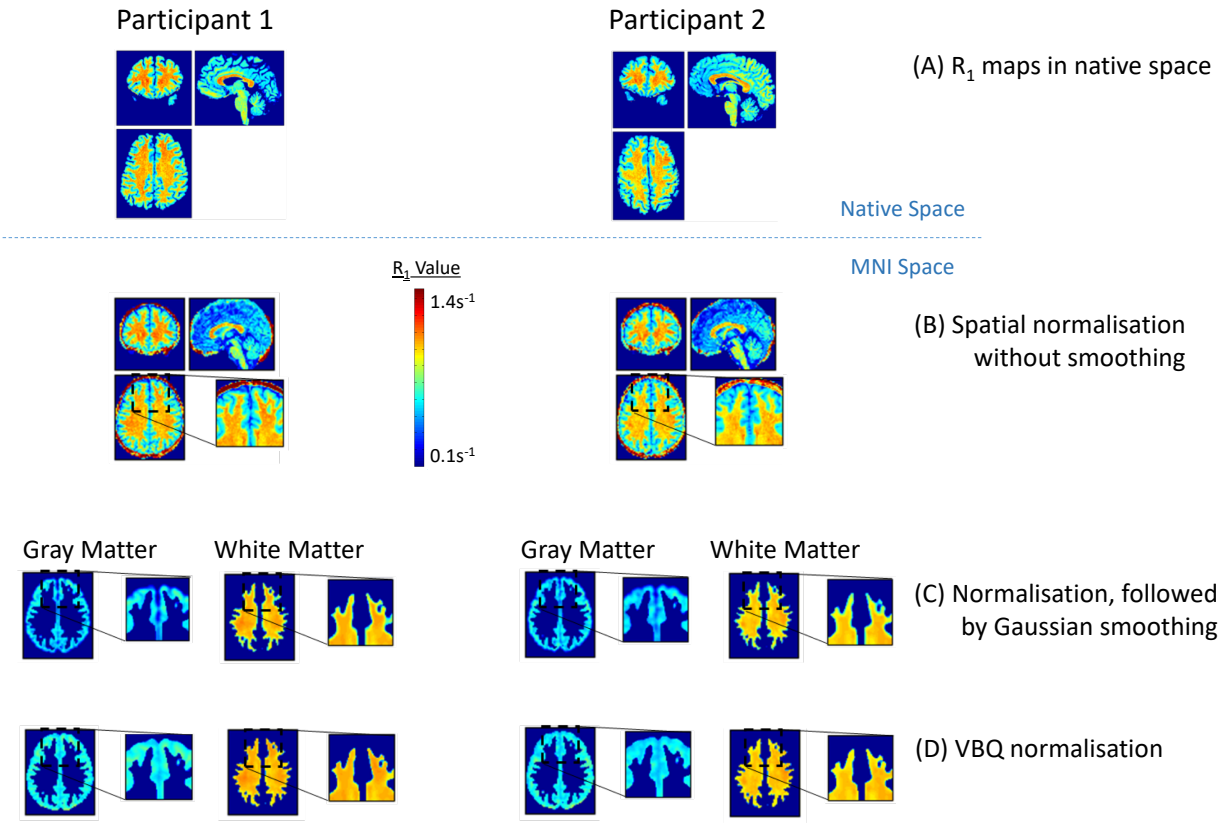

Figure 3.1: Normalisation of R_1 maps from native space to MNI space by applying (B) a spatial normalisation without smoothing, (C) a normalisation followed by a Gaussian smoothing or (D) a VBQ normalisation (Cercignani, 2017).

3.2 Univariate analysis

In this section, a parametric and a nonparametric approach are carried out in order to study each quantitative map separately. Then, the results from both methods are compared to assess their correspondence. This is an important preliminary step before investigating the complex multivariate case so as to ensure that PALM is reliable in the case of the considered study design.

3.2.1 Parametric approach

First of all, the quantitative maps were analysed in the univariate case, i.e. each modality was tested individually with its own hypothesis. The statistical inferences were carried

out using the GLM framework embedded in SPM12. The design matrix was composed of four regressors: age, gender, total intracranial volume and the type of scanner. It is essential to remark that only the age regressor will be of interest throughout this thesis. On one hand, one-tailed t-tests were used to test the decrease with age of MT and R_1. The reason is that both are sensitive to myelin and it is know from previous research that a demyelination in the course of life is observed in the brain. On the other hand, two-tailed t-tests were used to investigate the difference with age of PD* and R_2^* since the effect of these parameters could not be assumed beforehand. The results were assessed after family-wise error (FWE) correction using random field theory (RFT) at a statistical threshold of $p < 0.05$. These are mainly the same statistical tests as those in Callaghan et al. (2014) regarding the discrepancies occurring with ageing. In addition, they were executed voxel-wise across the gray matter (GM) and the white matter (WM) sub-spaces separately using explicit masks defining GM and WM voxels (Callaghan et al., 2014).

3.2.2 Non-parametric approach

As a second step, permutation tests were applied using PALM. To do so, the images had to be transformed from 3D images to a 4D volume in order to be processed by the toolbox. The results are expected to be similar to those obtained from SPM. The goal is therefore to assess whether PALM works in the univariate approach with this study data before enlarging the analysis to multivariate statistics. Testing one single modality, a high number of permutations was used a priori, with the hypothesis of exchangeable errors, along with the contrasts described above. In accordance with Nichols and Holmes (2002), 1000 permutations were chosen as it leads to a robust outcome with significant minimum p-values. For the nonparametric approach, only the transverse relaxation rate R_2^* maps were studied, assuming that if the activations are comparable with SPM for one modality then they will be for all of them.

3.2.3 Cross-methods variability

Finally, the results were compared. One way was to observe the resemblance between the patterns of significant voxels in the respective outputs. The analogy can be assessed both qualitatively and quantitatively. The qualitative aspect was done by visually evaluating the similarity in the significantly detected regions whereas the quantitative aspect consisted first in assessing the similarity between the patterns using a similarity measure called the dice coefficient. The measure is defined as:

$$DC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{3.1}$$

where $|A|$ and $|B|$ are the number of significant voxels in image A and B respectively, and $|A \cap B|$ is the number of significant voxels detected simultaneously in both images. The coefficient is comprised between 0 and 1. In the case where both images detect exactly the same voxels: $|A| = |B| = |A \cap B|$ and $DC = 1$. On the contrary, the intersection is empty whenever the images are completely dissimilar. DC is therefore equal to 0. Obviously, the higher the coefficient the better the results' consistency.

Another way of comparison was to look at the statistical values strictly speaking. One issue is that the outcome of the softwares is different: SPM gives t-scores while PALM generates p-values as final outputs. Comparing directly the values is therefore unfeasible. Three ways are thus conceivable.

First, both outputs can be transformed into z-scores. To do so, p-values are converted into z-scores via the inverse complementary error function while t-scores are modified into z-scores using the SPM12 function `spm_t2z`. However, going from one score to the other introduces assumptions and numerical approximations likely to induce inaccurate results. Furthermore, due to the fact that p-values of permutation tests are discrete and at least equal to the inverse of the total number of permutations, the order of magnitude between the z-scores obtained for PALM and SPM will be different.

Second, the t-scores from SPM can be transformed into p-values using the SPM12 function `spm_Tcdf`. Yet, the issues described previously arise once again.

Third, considering the fact that PALM creates statistical maps of t-scores before performing any permutations, they can be compared with those from SPM. One could argue that generating the maps beforehand might not reflect the true results of the nonparametric analysis. However, it is thought that if the t-scores obtained at the beginning are similar then the p-values outputs would be of confidence.

Even though each of these methods seems to have at least one drawback, the third method was chosen as it prevents from performing any conversion from one score to the other. To compare both softwares, a Bland-Altman plot was used. It offers a similarity perspective between the statistical t-scores obtained at each voxel between two results. It is a way to assess the agreement between two methods. It is worth noting the difference existing between the notions of agreement and correlation. A good correlation might correspond to a bad agreement. A perfect agreement is represented by the equation: $y = x$. In this case, the correlation is also perfect. However, by modifying the scale of measurement, one might have an identical correlation but with a different agreement (Fevre). To generate such a plot, the difference between the two methods was plotted with respect to their average and the density of the points was represented.

3.3 Multivariate analysis

Once PALM and SPM showed to reasonably converge in the univariate study, it was extended to a multivariate analysis within PALM. A multivariate analysis is the natural extension of the univariate implementation in PALM in the nonparametric context. The NPC algorithm is more appropriate than classical permutations in this case because it is more powerful and relies on fewer assumptions (Winkler et al., 2016b). While generating the first tests with all four modalities, it was noticed that the computational time was particularly huge when the number of permutations was increasing. Therefore, acceleration methods available in PALM were used. Also, some tests were carried out in order to choose the minimal number of permutations to perform inducing reliable inference.

3.3.1 Acceleration methods

When the data to permute become relatively cumbersome, which is the case here, the time it takes to generate the exhaustive set of permutations is impracticable. Indeed, the total number of possible permutations is 1.6×10^{185}. However, more permutations do not intrinsically increase the power. Instead, it allows to decrease the resampling rate as well as to generate smaller p-values, which may not be needed in the case of FWE correction (Winkler et al., 2016a). So, the first obvious way to reduce the computational time is to perform fewer permutations. But still with hundreds or thousands of rearrangements, the processing time is high (2 hours for the GM tissue class with 250 permutations, 11 hours and 30 minutes for 1000 permutations). Therefore, it is possible to use a tail approximation. It consists in modelling the tail of the distribution by fitting a generalised Pareto distribution (GPD). The GPD has a cumulative distribution function (CDF):

$$F(z) = \begin{cases} 1 - (1 - \frac{kz}{a})^{\frac{1}{k}}, k \neq 0 \\ 1 - e^{\frac{-z}{a}}, k = 0 \end{cases} \tag{3.2}$$

where a is the scale parameter, k is the shape parameter and where the range of the exceedances z is $0 \leq z < \infty$ for $k \leq 0$ and $0 \leq z \leq a/k$ for $k > 0$. The permutation p-value of the approximation is given by:

$$P_{gpd} = \frac{N_{exc}}{N}(1 - F(x_0 - t)) \tag{3.3}$$

where N_{exc}/N compensates for the fact that the CDF is estimated only on the tail (Knijnenburg et al., 2009). Figure 3.2 illustrates P_{gpd} for different values of the test statistic x_0 as shown in Knijnenburg et al. (2009). It can be seen that the GPD approximation is quite close to the CDF of the F distribution for the range of x_0.

In PALM, the tail is defined by a certain threshold entered as an argument. It this case, a threshold of $p < 0.05$ was chosen for the modelling. It means that the tail approximation will only be applied on the p-values below the threshold. A representation can be seen in Figure 3.3.

Other approximations, e.g. the gamma approximation, are also implemented. However, the tail approximation is preferred in the case of multiple modalities due to its availability for NPC and for its generalisability. Also, in the case of a FWE correction, the fit of a tail is quick, the resample risk is small and the effects revealed might be stronger (Winkler et al., 2016a). Another advantage of using the tail approximation is that if a good fit is not found, then no approximation is made to prevent from introducing erroneous modelling.

One further modification to increase the computational time is to prevent PALM from generating maps of uncorrected p-values. The advantages are that it not only reduces the memory required by such memory intensive methods but also avoids the computation of a high number of p-values.

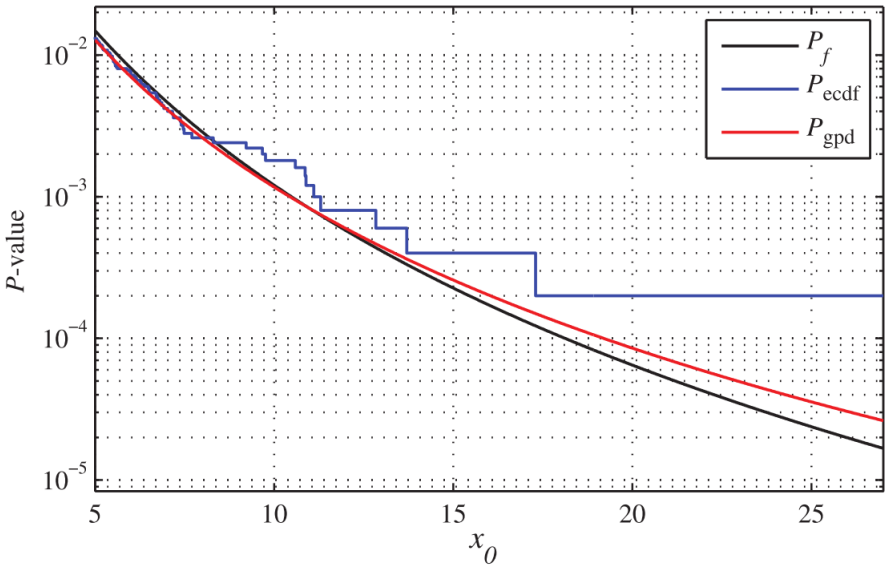

Figure 3.2: P-value as a function of the test statistic. P_f = p-value obtained from the CDF of the F distribution, P_{ecdf} = p-value of the ECDF approximation and P_{gpd} = p-value of the GPD approximation. ECDF stands for Empirical Cumulative Distribution Function. It is an other approximation of p-values including a step function (Knijnenburg et al., 2009).

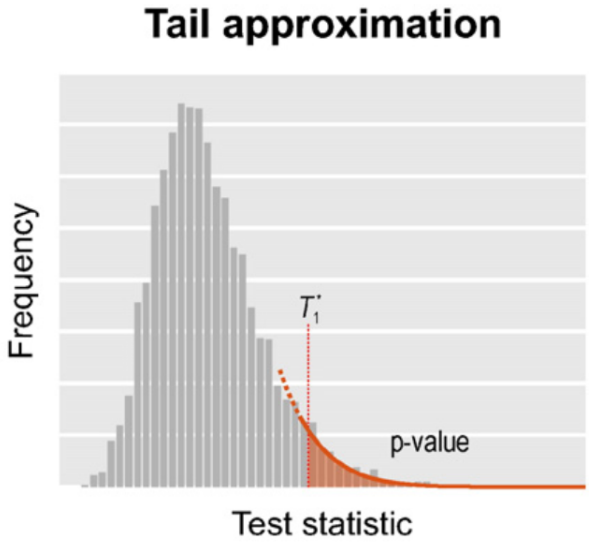

Figure 3.3: The tail approximation consists in approximating the p-values below a certain threshold by fitting the tail of the distribution using a generalised Pareto distribution (Winkler et al., 2016a).

The combination of these two acceleration methods along with the use of faster computing server available at the FIL, tremendous reduction of execution time are obtained (from 2 hours to 1 hour and 20 minutes for the GM tissue class for 250 permutations). One

advantage is that a greater set of permutations can be considered for the same amount of time.

3.3.2 Number of permutations

It is known that the higher the number of permutations, the better. Nevertheless, there is a trade-off between the number of permutations and the computational time as already mentioned. This is the reason why it is interesting to investigate the minimum number of permutations required to have stable and faithful results.

To do so, the NPC algorithm for the analysis of the four modalities while testing a positive correlation with age was performed with 50, 100, 250, 500, 750, 1000 and 2000 permutations. The percentage of voxels detected under the statistical threshold of $p < 0.05$ was estimated. Also, each result was compared with the one obtained with 2000 permutations, considered as the reference. In this way, the dice coefficient between the number of voxels detected FWE $p < 0.05$ for one analysis and for 2000 permutations was determined. It enables to assess whether increasing the number of permutations improves the robustness of the results.

3.3.3 Statistical inference

One can wonder which type of statistical test is more appropriate in the case of a multivariate test. Evaluating either a positive or a negative correlation with age is unlikely to give interesting results since it is known that some parameters are sensitive only to an increase with age while others only to a decrease. Therefore, a F-test was applied so as to observe any difference with age in the brain microstructures.

3.3.4 Combination and correction over modalities and over contrasts

Whenever a nonparametric test contains at least two modalities, the permutations are performed independently on each modality and the empirical p-values obtained are then combined (principle of the NPC algorithm as described in Section 2.5). In the same way, to perform a F-test, the contrast file entered in PALM is in the form:

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{matrix}$$

The second column is associated with the age regressor. The algorithm processes them independently as two t-tests respectively evaluating the effect of an increase and a decrease with age. The combined statistic, through the use of Fisher, is computed at each voxel and then the extremum combined statistic across the image is stored.

Furthermore, it is necessary to correct for the multiplicity of both modalities and contrasts. For the contrasts, it consists in converting their statistics into a z-statistic. For the modalities, it consists in generating a FWER-adjusted p-value, which is the proportion of the extremum statistic across all tests greater than the unpermuted statistic (Winkler,

2017).

Appendix A contains the line of command with the specific input arguments to run PALM as described here above. In addition, a SPM batch was created to use nonparametric inference in a similar way as it was done in this thesis. Appendix B describes how to use the batch interface.

3.4 Comparison

The main question to be answered is to evaluate whether the use of a multivariate approach is worth it. It is particularly interesting to look at what is gained and what is lost by the use of a joint analysis. Several methods to try answering these questions are presented.

A useful measure consists in computing the percentage, or the number, of significant voxels compared to a reference. For example, it can be the number of significant voxels in an analysis with respect to the number of voxels in the mask of the corresponding tissue class. Each percentage used will be specified.

As presented in Section 3.2.3, the dice coefficient was also computed as the similarity measure of reference.

In addition, Bland-Altman plots were created. This time, it investigates the similarity between the statistical p-values obtained at each voxel between two multivariate results. In addition, the average bias and the limits of agreement can be determined. The bias is the mean of the difference values while the limits of agreement are defined as: bias $\pm$ 1.96σ, σ being the standard deviation of the difference values, all of which being superimposed on the graph.

Finally, the matching between a pair of images is determined in which one image is defined as the prediction and the other as the reference. Since, it is not known which is the true pattern, each image was considered as the reference and as the prediction once. The matching is evaluated in terms of the overlap between the images. The overlap is computed in terms of voxel and cluster matching. More importantly for this study, the false positive and true positive rates for each matching give more interpreting results. For the voxel matching, the true positive rate ($\mathrm{TP_r}$), or sensitivity, is the percentage of voxels in the prediction's regions of interest (ROI) matching those in the reference. The false positive rate ($\mathrm{FP_r}$), or 1-specificity, is the percentage of voxels in the prediction's ROI that do not match any voxels in the ground truth. They can be written as:

$$\begin{cases} \mathrm{TP_r} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \\ \mathrm{FP_r} = \dfrac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \end{cases} \tag{3.4}$$

For the cluster matching, they are defined respectively as the percentage of clusters in the prediction matching those in the reference and the percentage of clusters in the prediction

matching any in the reference.

Two main axes of comparison were analysed. On one hand, the multivariate results from PALM at the $p < 0.05$ FWE corrected level and the mass univariate results from SPM will be studied. To be consistent with the corrected level fixed at 0.05 in the multivariate case, the univariate results are analysed at $p < \frac{0.05}{4} = 0.0125$ FWE corrected level. Since a test with $p < 0.05$ has a 5% chance of finding at least one false positive, then generating four tests, due to the presence of four modalities here, at the same threshold would lead to 4×5 = 20% chance of false positive. Therefore, by reducing the threshold for each univariate analysis, their union has also a 5% chance of false positive. The main objective is to evaluate what can be brought by a multivariate approach. On the other hand, results of permutation tests containing a varying number of modalities are contrasted. In this case, one wants to understand the influence of adding or removing one modality to the multivariate analysis. What each modality brings to the joint analysis?

3.5 VBQ analysis

To go from the raw MPM quantitative maps to the sub-space maps ready to be analysed, a voxel-based quantification analysis was performed. The goal of this approach is to achieve an optimal parameter (MT, R_1, R_2^*, PD^*) assignment to the particular tissue class (Draganski et al., 2011). It is essential to transform the quantitative maps from native space to MNI space. To do so, the data must be spatially normalised and smoothed by a Gaussian kernel. A better way for voxels correspondence between subjects is to apply a VBQ normalisation though. However, it is interesting to investigate whether this way of normalisation and of smoothing is appropriate.

To do a VBQ analysis, several steps were carried out. It is worth noting that the associated processing pipeline can be found in the hMRI toolbox[1]. First, the participant-specific tissue class images had to be transformed non-linearly to MNI space using the diffeomorphic registration algorithm Dartel from SPM (Ashburner, 2007). In this case, the images were modulated by the Jacobian determinants of the deformations so as to account for local compression and expansion due to both linear and non-linear transformation while going from native space to MNI space (Draganski et al., 2011). Warped, modulated and aligned tissue class images were therefore produced.

Second, another Dartel procedure was applied to the MPM images, without the need to scale by the Jacobian determinant, to generate warped and aligned images.

Finally, a combined weighting/smoothing/masking procedure was applied:

$$\text{signal} = \frac{g * (ws(\phi)[\text{eTPM} > 0.05])}{g * w}[g * w > 0.05] \tag{3.5}$$

where $w = |D\phi|t(\phi)$ are the weights in standard space with $|D\phi|$ being the Jacobian determinant of the deformation ϕ and $t(\phi)$ being the warped tissue class image, and $s(\phi)$ is the warped parameter map. Both are the outputs of the two Dartel procedures described

[1] `https://github.molgen.mpg.de/hMRI-group/Toolbox`

previously. The goal of the weighting is to up weight the voxels guaranteed of belonging to a particular tissue class. Then, $g *$ represents a convolution with a 3 mm FWHM (full width at half maximum) Gaussian smoothing kernel. The goal of the smoothing is not only to improve the spatial overlap by blurring over minor anatomical differences and registration errors but also to suppress noise by averaging over several voxels (Phillips, 2018). Finally, two masking thresholds at 0.05 are applied. On one hand, the product of w and $s(\phi)$ is masked by the non-participant specific TPM embedded in SPM, ahead of the smoothing procedure. On the other hand, the final ratio is masked by the denominator, a participant-specific tissue map. The latter was unpublished and its purpose is not well understood. It might induce an undesired age effect because of its participant-specific characteristic. However, the way SPM is implemented to estimate the model parameters by the use of an external masking, in this case being a relatively conservative mask including the voxels present in at least 20% of the subjects, prevent this term from having any effect. This triple procedure aims at enhancing the specificity for the particular tissue class. However, there is a risk of misalignment from one subject to the other, even though the normalisation tries to prevent it. If this is the case, some spurious effects associated to atrophy may appear in the age studies described in the previous sections. Several subsequent modifications of equation (3.5) will thus be done in order to determine whether spurious effects are present and where they would come from.

The VBQ procedure (equation (3.5)) can be written as:

$$\text{signal}_i = \frac{g * (|D\phi|_i t_i(\phi) s_i(\phi))}{g * |D\phi|_i t_i(\phi)} \tag{3.6}$$

where both maskings have been omitted and where the subscript i means that each parameter is participant-specific, $i \in \{1, ..., 138\}$. When studying age-related effects on such images, it is thought that the age effect is embedded in the parameter maps $s(\phi)$ rather than in the weights w. Therefore, performing an age analysis where the weights are kept participant-specific but where the parameter maps are taken from only one subject aims at verifying that no age effect is present in the weights. The equation then becomes:

$$\text{signal}_i = \frac{g * (|D\phi|_i t_i(\phi) s(\phi))}{g * |D\phi|_i t_i(\phi)} \tag{3.7}$$

It is interesting to capture the variations in the patterns of detected voxels in this case while selecting the parameter maps of a young participant of 22 years old, of an old participant of 74 years old and of the mean map of the entire cohort. The modified procedure was tested for each quantitative parameter. In addition, the evolution with age in a ROI in the body of the corpus callosum was reported for both equations (3.6) and (3.7) as well as for the parameter maps, the tissue class images and the Jacobian determinants for all subjects. In this way, the correlation with ageing is assessed for each of them.

In the same idea, the weights can be modified in several ways so as to investigate the effect of the different elements present in the formula. Subsequently, the various analyses

were conducted:

$$\text{signal}_i = \frac{g * (t_i(\phi)s_i(\phi))}{g * t_i(\phi)} \tag{3.8}$$

$$\text{signal}_i = \frac{g * (t_i(\phi)s(\phi))}{g * t_i(\phi)} \tag{3.9}$$

$$\text{signal}_i = \frac{g * (|D\phi|_i s(\phi))}{g * |D\phi|_i} \tag{3.10}$$

$$\text{signal}_i = \frac{g * (|D\phi|_i s_i(\phi))}{g * |D\phi|_i} \tag{3.11}$$

where in the first and the second equations, the Jacobian was removed meaning the first Dartel procedure was conducted without scaling by the Jacobian while the third and fourth studies use only the Jacobian as weights.

Additional analyses were to study the influence of modifying the size of the smoothing kernel. In addition to the reference of 3 mm FWHM, 0.5 and 6 mm were applied. In the first case, the smoothing effect is strongly reduced since the size of a voxel in MNI space was fixed to 1 mm. Therefore, the bottom of the Gaussian will be close to the voxel size. In the latter, the kernel is doubled to reinforce the effects of the smoothing procedure.

Futhermore, for the sake of completeness, the VBQ approach was compared with a tissue-specific, smoothing-compensated (T-SPOON) procedure based on which VBQ was built. The method was introduced in order to compensate for the problems of simple voxel-based analyses leading to poor tissue specificity from image registration and smoothing (Lee et al., 2009). In this case, the weights w are replaced by masks. To do so, the modulated, warped and aligned tissue class images are thresholded at 20%, to be in accordance with Lee et al. (2009) and to get a participant-specific mask. Using the same notations as previously, the T-SPOON formula can be written as:

$$\text{signal}_i = \frac{g * ([|D\phi|_i t_i(\phi) > 0.2]s_i(\phi))}{g * [|D\phi|_i t_i(\phi) > 0.2]} \tag{3.12}$$

This procedure was also studied by taking the parameter map from only one young participant.

Finally, the dice coefficients comparing some of these different analyses were then computed to assess the similarity of the results from a quantitative point of view.

For all the here above described tests, a F-test was used. The results were then assessed after FWE correction at a statistical threshold of $p < 0.05$. To be consistent with Callaghan et al. (2014), a smoothing kernel of 3 mm FWHM and a voxel size of 1 mm were used. Any modification with respect to these default parameters will be specified accordingly.

Chapter 4

Results

4.1 Univariate analysis

4.1.1 Parametric approach

In this section, the univariate analyses from Callaghan et al. (2014) were conducted again and the results were replicated for comparison. It is an essential step in order to ensure the analyses' reproducibility provided that the SPM and the Matlab versions were updated.

Magnetization transfer

The MT results obtained show significant negative correlations with age (Figure 4.1) mainly in frontal and parietal regions. The thalamus, the borders of the corpus callosum, some cortical regions and the optic radiation were identified.

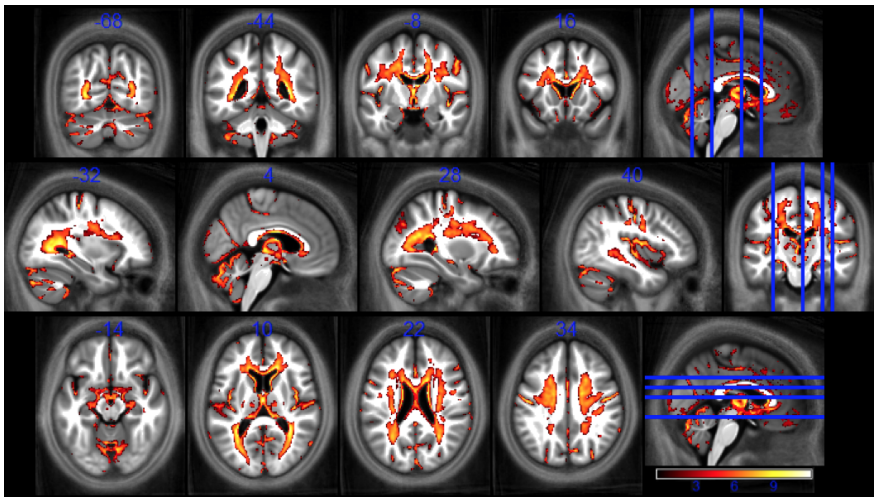

Figure 4.1: Statistical parametric maps of regions with a decrease in MT with age, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The t-scores are represented using the colobar and some slices of the brain are presented in the sagittal, coronal and axial orientations.

Effective transverse relaxation rate

Regarding R_2^*, mainly positive correlations were identified with age and were located in regions including the putamen, the pallidum, the caudate nucleus and some cortical regions (Figure 4.2). Negative correlations within the medial part of the ventral surface of the frontal lobe were also seen but in a much smaller amount.

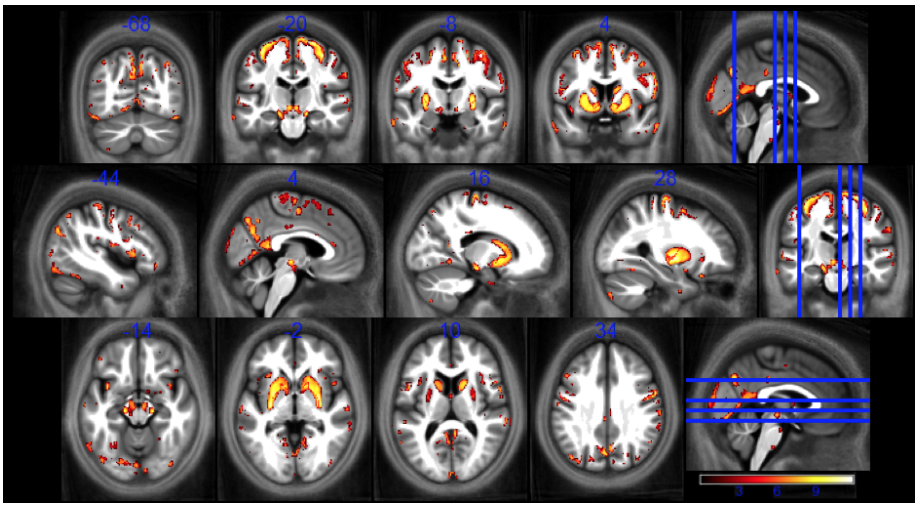

Figure 4.2: Statistical parametric maps of regions with an increase in R_2^* with age, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The t-scores are represented using the colobar and some slices of the brain are presented in the sagittal, coronal and axial orientations.

Longitudinal relaxation rate

In the R_1 map (Figure 4.3), negative correlations with age were identified in the optic radiation and around the corpus callosum.

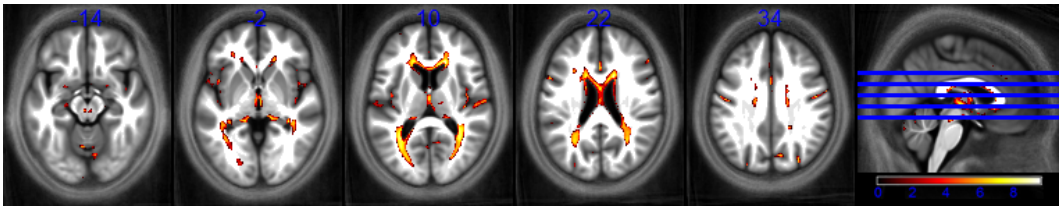

Figure 4.3: Statistical parametric maps of regions with a decrease in R_1 with age, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The t-scores are represented using the colobar and some slices of the brain are presented in the axial orientation only.

Effective proton density

Proton density correlations with age were found in both directions in different structures (Figure 4.4). In the case of a negative correlation, the putamen, the pallidum, the caudate nucleus and the red nucleus were identified. In the case of a positive correlation, significant voxels were observed in the optic radiation and in superior regions of the white matter.

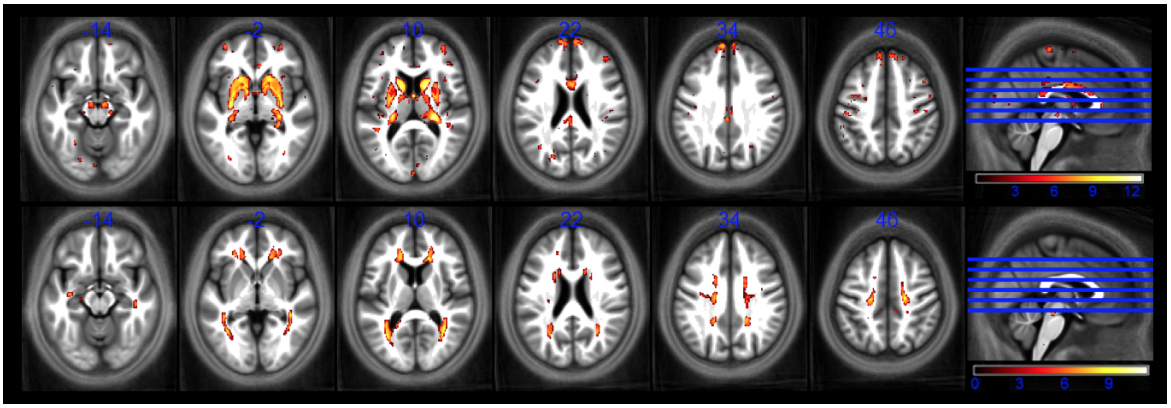

Figure 4.4: Statistical parametric maps of regions with a decrease (top) and an increase (bottom) in PD* with age, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The t-scores are represented using the colobar and some slices of the brain are presented in the axial orientation only.

4.1.2 Nonparametric approach

From a qualitative point of view, the regions displaying a positive correlation with age for R_2^* using permutation tests (Figure 4.5) are quite similar to those obtained using parametric techniques (Figure 4.2). The main regions are detected, with the putamen, the caudate nucleus and the pallidum being significantly correlated. However, some cortical regions that were observed in the sagittal view are not significant anymore (see arrows, Figure 4.6).

4.1.3 Cross-methods variability

As discussed in the previous section, the patterns of significant voxels between the parametric and the nonparametric methods for R_2^* were relatively comparable except for small discrepancies in the cortex. To quantify this correspondence, the dice coefficients were computed in order to give more information about the overlap of the two results. For the GM, the measure was of 0.9243 which is very high. By contrast, the WM coefficient was of 0.0042 which is surprisingly very small. Figure 4.7 shows the statistical maps in the WM for both softwares. Even though both found significant voxels in the brainstem, the clusters were not located at the same location and PALM did not find any other regions compared to SPM with the exception of a small number of isolated voxels.

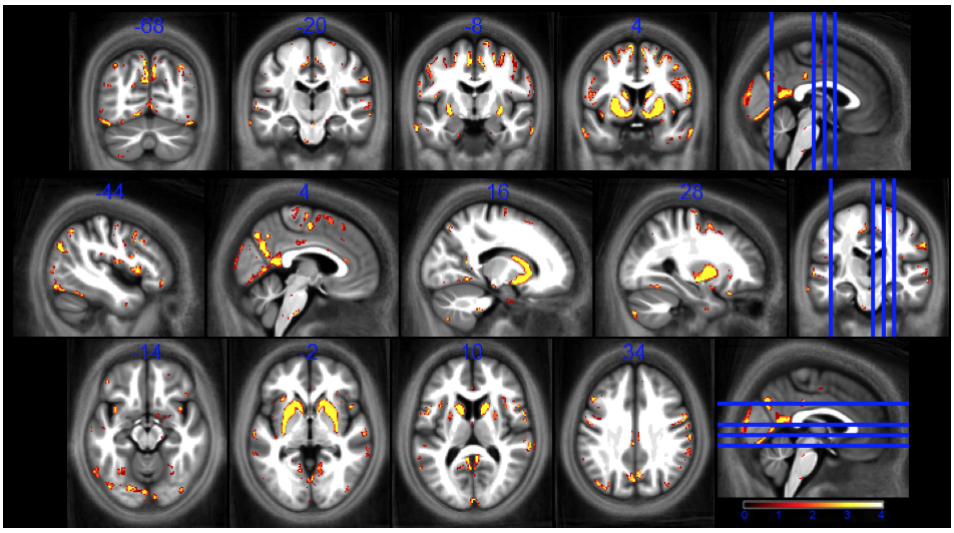

Figure 4.5: Statistical parametric maps of regions with an increase in R_2^* with age obtained with permutation tests, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. For visual purposes, the negative logarithm of the p-values ($-\log_{10}(p)$) are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

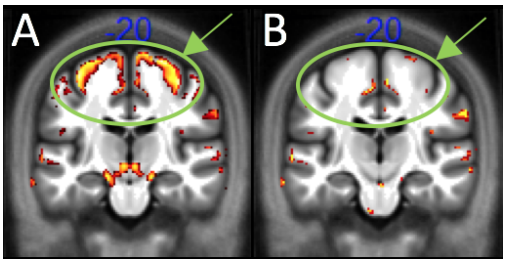

Figure 4.6: Statistical parametric maps of regions with an increase in R_2^* with age obtained (A) parametrically with SPM12 and (B) nonparametrically with PALM, at $p < 0.05$ FWE corrected. The results were overlaid on the mean MT map of the cohort in MNI space.

Figure 4.8 shows the results of the Bland-Altman plot of the t-scores from the two approaches. First of all, it can be observed that the bias, i.e. the mean of the differences in the y axis, is slightly below zero. It means that there is a small bias towards larger t-scores to SPM. Furthermore, the histogram of the differences does not spread far way (between -0.2 and 0.2). Also, it can be noticed that the average of the t-scores is around 2, corresponding to moderately significant voxels.

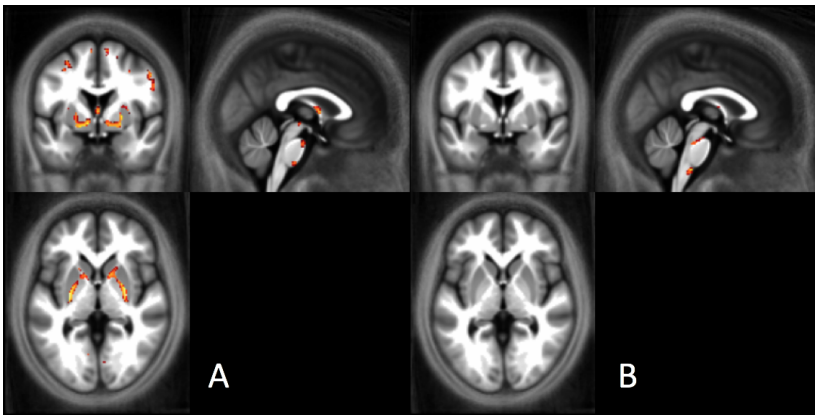

Figure 4.7: Statistical parametric maps of regions with an increase in R_2^* with age in the WM obtained with SPM12 (A) and with permutation tests using PALM (B), at $p < 0.05$ FWE corrected. The results were overlaid on the mean MT map of the cohort in MNI space.

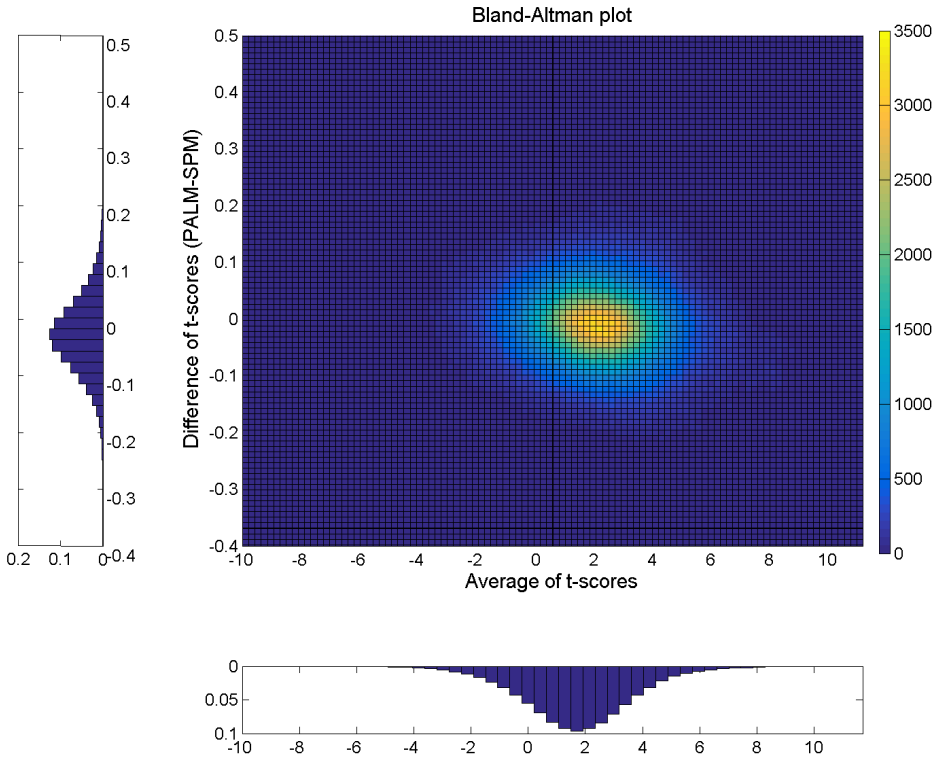

Figure 4.8: Bland-Altman plot comparing the t-scores from SPM and PALM. The difference is computed as the statistical score obtained in PALM minus the statistical score from SPM, and only the voxels detected in both measures were considered. The density of points is shown via a heat map. The histograms of the points for each axis are also represented.

4.2 Multivariate analysis

In this section, the multivariate analysis carried out in PALM is presented. In particular, the number of permutations varies and the results' robustness is evaluated. Also, the statistical maps of the joint analysis are represented.

4.2.1 Number of permutations

Results suggest that the percentage of significant voxels tends to stabilise when the number of permutation increases. Indeed, Figure 4.9 shows that for small number of rearrangements, the percentage varies from one test to the other. Increasing the set of shufflings leads to percentages converging just above 4%.

A similar trend can be observed in Figure 4.10. Small number of permutations differ from the results obtained with 2000 permutations. However, as far as the number of permutations is made bigger, a dice coefficient close to 1 is reached, showing the consistency of the results as the permutations increase.

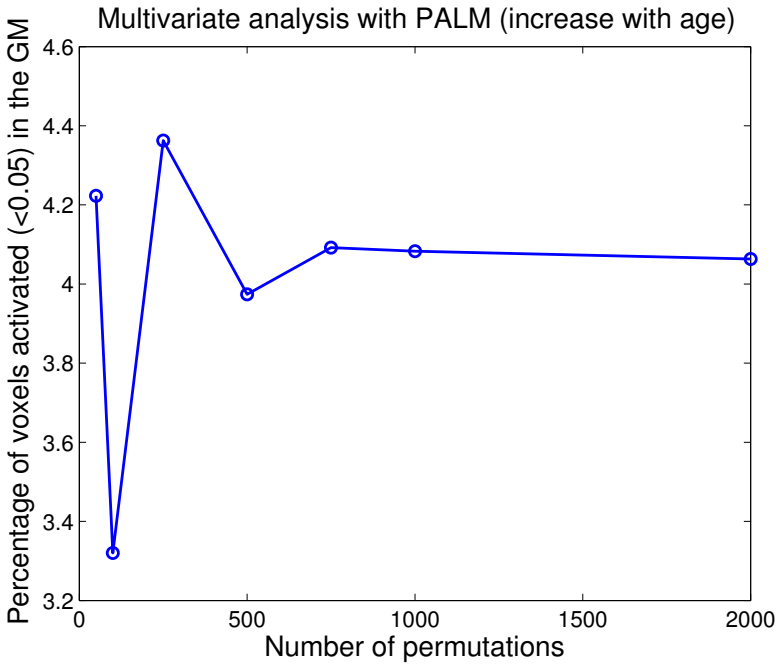

Figure 4.9: Percentage of significant voxels FWE $p < 0.05$ in the GM for a multivariate analysis with the four modalities as a function of the number of permutations. The statistical test corresponds to a t-test depicting the positive correlation with age.

4.2.2 Joint analysis: correlation with age

The nonparametric procedure implemented in PALM between the set of modalities showed extensive correlations with age in the putamen, the pallidum, around the corpus callosum, the caudate and the red nuclei, the optic radiation, the thalamus and the insula (Figure 4.11). Some cortical regions were also detected.

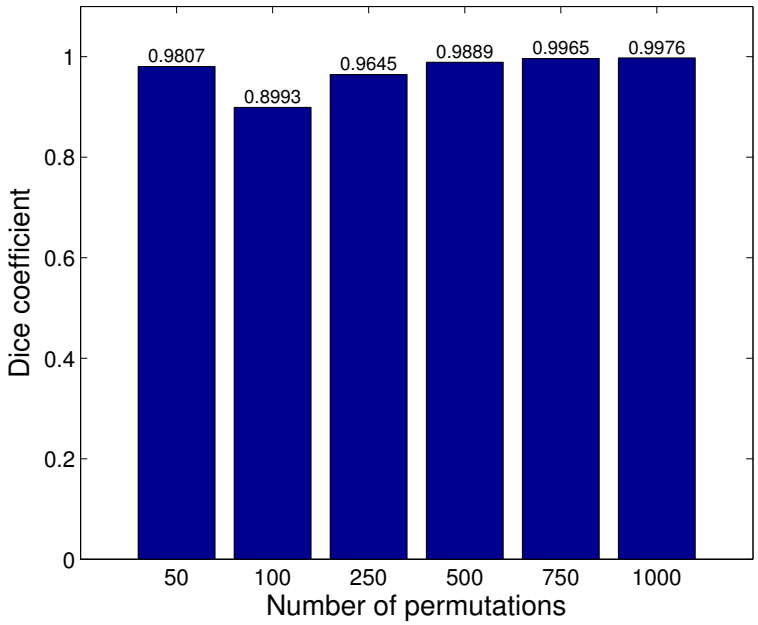

Figure 4.10: Dice coefficient between a multivariate analysis with n, $n = \{50, 100, 250, 500, 750, 1000\}$, permutations and the same analysis with 2000 permutations, taken as the reference.

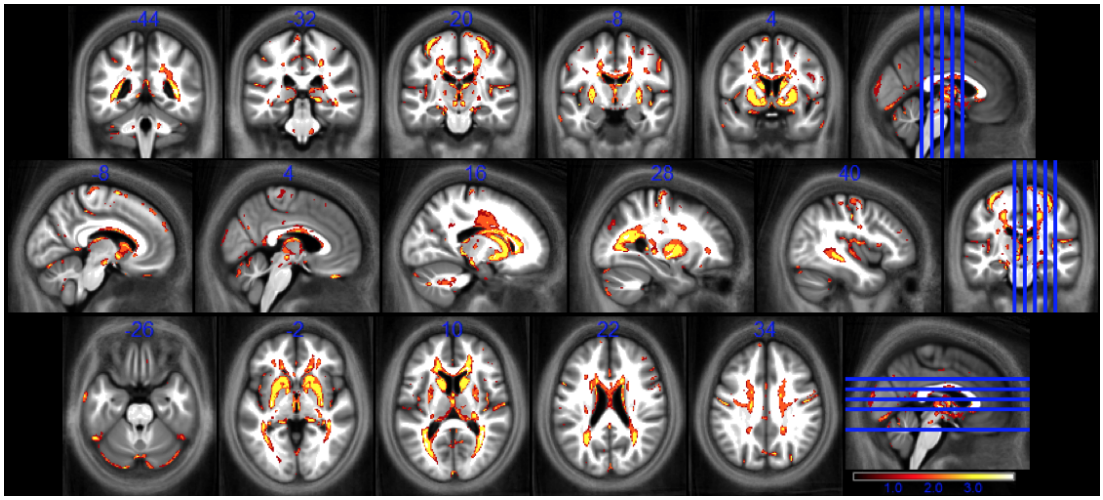

Figure 4.11: Statistical parametric maps of regions showing a correlation (both positive and negative) with age of the joint analysis of the four modalities, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. For visual purposes, the negative logarithm of the p-values ($-\log_{10}(p)$) are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

4.3 Comparison

In this section, the comparison is made between, on one hand, the mass univariate analyses and the joint multivariate analysis and, on the other hand, between multivariate analyses with varying number of modalities. The focus is given on the technical comparability and the power/sensitivity trade-off rather than on the biology. Therefore, all presented results are only considering the gray matter class, showing a better cross-software similarity in the univariate case.

4.3.1 Mass univariate analyses vs. joint multivariate analysis

Significant voxels and dice coefficients From Table 4.1, it can be observed that the joint analysis is definitely different from the union of the univariate tests, which has twice as much significant voxels. In addition, as observed in Section 4.1.1, R_1 has few significant voxels while MT has more voxels than the joint analysis and than any other single modalities. However, despite this trend, the dice coefficient of MT with PALM is lower than the dice coefficient between PALM and PD* or R_2^* (Figure 4.12). Logically, R_1 shows a much lower similarity. In addition, it was discovered that 3.59% of the significant voxels in the multivariate analysis were not present in any of the separate modalities. These voxels were located at borders and do not introduce additional clusters.

R_2^*	R_1	MT	PD*	PALM
3.18%	0.63%	4.96%	2.20%	4.28%
		Union: 8.97%		

Table 4.1: Percentage of significant voxels in each univariate modality at $p < 0.0125$ FWE corrected, in their union, and in the multivariate analysis with four modalities at $p < 0.05$ FWE corrected. The percentage is determined with respect to the number of voxels included in the GM mask.

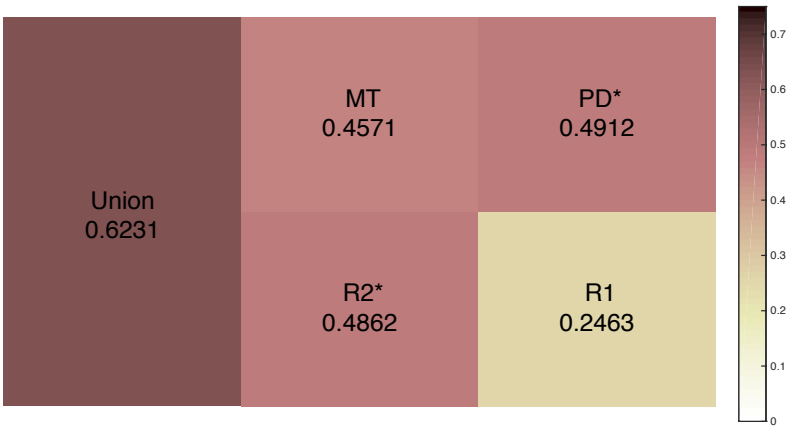

Figure 4.12: Dice coefficient between each univariate analysis and the multivariate analysis with all four modalities as well as between their union and the multivariate analysis.

True positive and false positive rates Then, the univariate cases and the nonparametric analysis including all four modalities were compared in terms of true positive and false positive rates (Table 4.2). Given that the true pattern is not know, each analysis is taken as the reference and as the prediction once. One general observation is that the false positive rate at the voxel level is very low. Then, the table illustrates that the overlap changes depending on the reference chosen. For example, the comparison between R_1 and PALM shows that the true positive rates are very high when R_1 is the reference meaning that what is found in R_1 alone is very likely to be found by the multivariate case while the inverse is very unlikely to happen.

Prediction	**Reference**	**Voxels**		**Clusters**	
		TP_r	FP_r	TP_r	FP_r
Union	PALM	0.96	0.0070	0.92	0.72
PALM	Union	0.46	0.0002	0.28	0.08
R_2^*	PALM	0.42	0.0019	0.40	0.56
PALM	R_2^*	0.57	0.0035	0.44	0.60
R_1	PALM	0.14	0.0001	0.15	0.01
PALM	R_1	0.96	0.0053	0.98	0.84
MT	PALM	0.49	0.0041	0.45	0.69
PALM	MT	0.42	0.0031	0.31	0.55
PD*	PALM	0.37	0.0008	0.30	0.43
PALM	PD*	0.72	0.0039	0.57	0.70

Table 4.2: Voxel and cluster matching results between each single modality and the multivariate analysis, as well as the union of the modalities and the multivariate case. Each method was considered as the prediction and as the reference once. The numbers are comprised between 0 and 1.

Significant voxels of various combinations In Tables 4.3 and 4.4, respectively the union and the intersection of the significant voxels for each univariate modality, for each pair of modalities, for each triplet and for all of them were determined. For each of these combinations, the intersection of the significant voxels with those found in the nonparametric joint analysis was evaluated and various ratios were computed. As could be expected, the more modalities are considered the more the number of significant voxels in their unions. Consequently, the intersection with the multivariate analysis is made slightly higher. By contrast, the more modalities the lower the number of significant voxels in their intersection and the lower the intersection with the joint test.

Combination	$\mathbf{SV}_i$	$\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}$	$\frac{\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}}{\mathbf{SV_i}}$	$\frac{\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}}{\mathbf{SV}_{\mathrm{PALM}}}$	$\frac{\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}}{\mathbf{SV}_{\mathrm{union}}}$
R_2^*	32 315	5171	16.00%	11.89%	5.68%
MT	50 400	13 954	27.69%	32.08%	15.32%
R_1	6390	92	1.44%	0.21%	0.10%
PD*	22 389	3242	14.48%	7.45%	3.56%
$R_2^* \cup$ PD*	42 670	19 752	46.29%	45.41%	21.68%
$R_2^* \cup$ MT	81 322	20 280	24.94%	46.63%	22.26%
$R_2^* \cup R_1$	38 111	5389	14.14%	12.39%	5.92%
MT $\cup R_1$	51 345	19 269	37.53%	44.30%	21.15%
MT $\cup$ PD*	71 682	17 948	25.04%	41.27%	19.70%
$R_1 \cup$ PD*	28 103	3419	12.17%	7.86%	3.75%
$R_2^* \cup$ PD* $\cup R_1$	48 231	20 468	42.44%	47.06%	22.47%
$R_2^* \cup$ PD* $\cup$ MT	90 770	35 785	34.42%	82.28%	39.28%
$R_2^* \cup R_1 \cup$ MT	81 727	25 747	31.50%	59.20%	28.26%
PD* $\cup R_1 \cup$ MT	72 129	23 498	32.58%	54.03%	25.79%
$R_2^* \cup$ PD* $\cup R_1 \cup$ MT	91 090	41 928	46.03%	96.40%	46.03%
PALM	43 492				

Table 4.3: The union of the number of significant voxels in the GM detected in each combination of modalities in the univariate case as well as in the multivariate analysis in PALM (second column). For each case, the intersection with those found in PALM was determined (third column) and the ratio with respect to the number of significant voxels in each case (fourth column), in PALM (fifth column) and in the union of all voxels (sixth column) was computed.

4.3.2 Varying the number of modalities in the multivariate analysis

Bland-Altman plot A nonparametric analysis was carried out using from one to four modalities with R_2^* as starting point. The goal is to look at the effect of increasing the number of modalities. Consecutively, PD*, R_1 and MT were added. The effect of adding a modality in the analysis was evaluated using Bland-Altman plots. An inclined square appears in each case (Figure 4.13). If the difference for methods A and B is plotted as $B - A$, then the top corner of the square corresponds to the lost effect or the sensitivity loss by adding the new modality from A to B, the bottom corner corresponds to the new information gained from adding this new modality, the left and right corners give the consistency between A and B for finding voxels as either being significant or not. It can be shown (Figure 4.14, top left plot) that adding the proton density to the R_2^* study will induce some loss and some gain of information. However, R_1 (top right plot) leads to a significant loss of information while MT (bottom plot) brings a comparable gain. It is thus beneficial to include a modality with a lot of effects.

Combination	$\mathbf{SV}_i$	$\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}$	$\frac{\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}}{\mathbf{SV_i}}$	$\frac{\mathbf{SV_i} \cap \mathbf{SV}_{\mathrm{PALM}}}{\mathbf{SV}_{\mathrm{PALM}}}$
R_2^*	32 315	5171	16.00%	11.89%
MT	50 400	13 954	27.69%	32.08%
R_1	6390	92	1.44%	0.21%
PD*	22 389	3242	14.48%	7.45%
$R_2^* \cap$ PD*	12 034	11 339	94.22%	26.07%
$R_2^* \cap$ MT	1393	1155	82.91%	2.66%
$R_2^* \cap R_1$	594	128	21.21%	0.29%
MT $\cap R_1$	5445	5223	95.92%	12.01%
MT $\cap$ PD*	1107	752	67.93%	1.73%
$R_1 \cap$ PD*	676	85	12.57%	0.19%
$R_2^* \cap$ PD* $\cap R_1$	441	413	93.65%	0.95%
$R_2^* \cap$ PD* $\cap$ MT	200	172	86.00%	0.40%
$R_2^* \cap R_1 \cap$ MT	54	26	48.15%	0.06%
PD* $\cap R_1 \cap$ MT	178	150	84.27%	0.34%
$R_2^* \cap$ PD* $\cap R_1 \cap$ MT	28	28	100 %	0.06%
PALM	43 492			

Table 4.4: The intersection of the number of significant voxels in the GM detected in each combination of modalities in the univariate case as well as in the multivariate analysis in PALM (second column). For each case, the intersection with those found in PALM was determined (third column) and the ratio with respect to the number of significant voxels in each case (fourth column) and in PALM (fifth column) was computed.

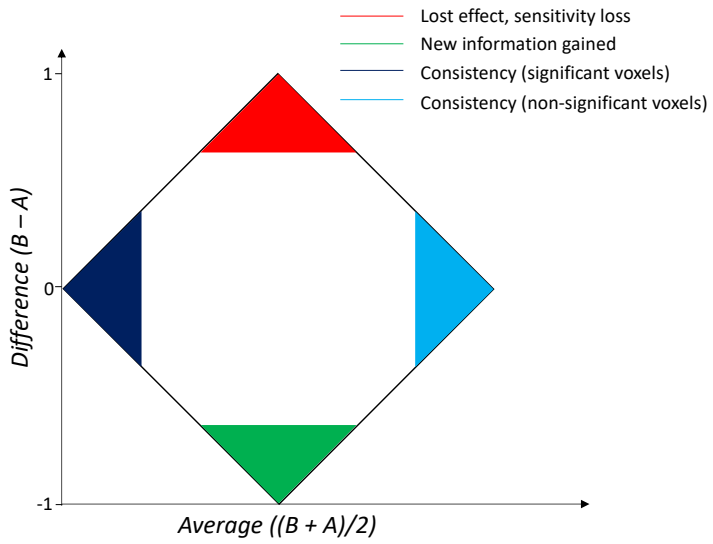

Figure 4.13: Graphical representation of the Bland-Altman plots results.

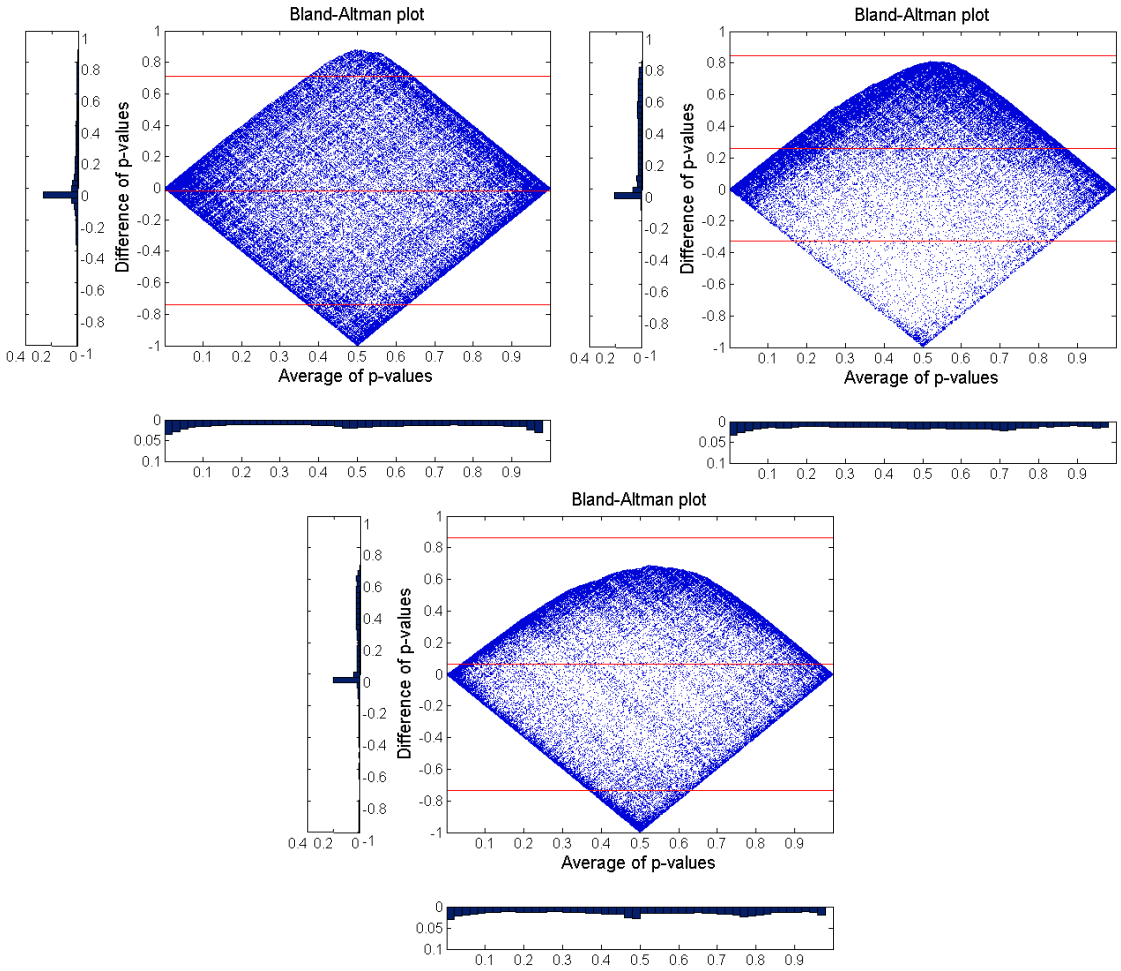

Figure 4.14: Bland-Altman plots comparing the p-values from R_2^* and R_2^*, PD* (top left); R_2^*, PD* and R_2^*, PD*, R_1 (top right); and R_2^*, PD*, R_1 and R_2^*, PD*, R_1, MT (bottom). The statistical score from the analysis with the smallest number of modalities is subtracted from the one with the highest number of modalities. The voxels with a p-value of 1 in both results were removed from the plot.

4.4 VBQ analysis

Changes in the original VBQ formula (equation (3.6)) were applied and subsequent age studies were conducted. In addition, T-SPOON was tested to compare the statistical maps with VBQ. Finally, the dice coefficients between some of the analyses with a fixed parameter map were determined.

Non participant-specific parameter map

Results of the VBQ procedure considering the MT map of a young participant to which all participant-specific weights were applied (Figure 4.15) show that a significant correlation with age was observed around the ventricles as well as along the caudate nucleus. The patterns look very similar when testing the three other parameter maps (Figures 4.16, 4.17 and 4.18). However, when testing the direction of the correlation with ageing, a positive correlation was obtained for MT, R_1 and R_2^* while PD* was negatively correlated.

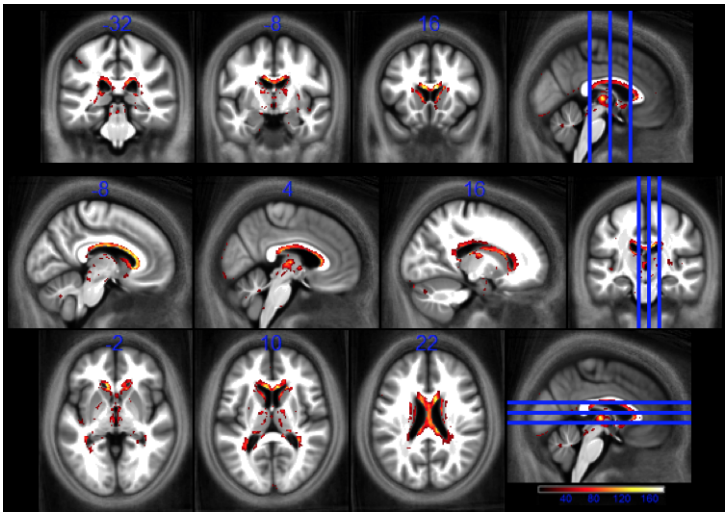

Figure 4.15: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis (equation (3.7)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

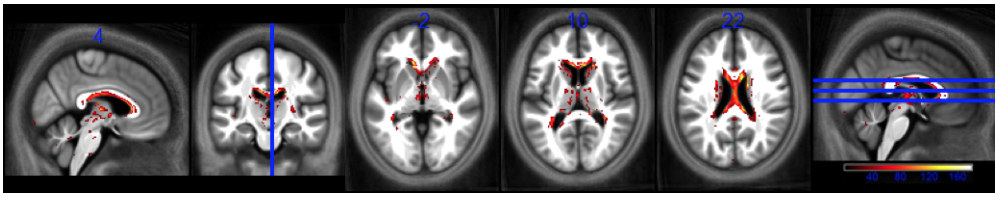

Figure 4.16: Statistical parametric maps of regions showing a correlation of PD* with age in the case of the modified VBQ analysis (equation (3.7)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

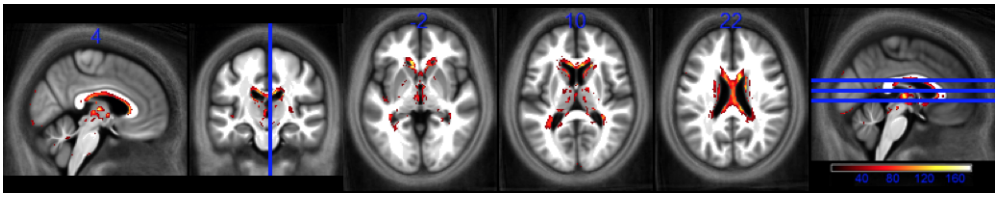

Figure 4.17: Statistical parametric maps of regions showing a correlation of R_1 with age in the case of the modified VBQ analysis (equation (3.7)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

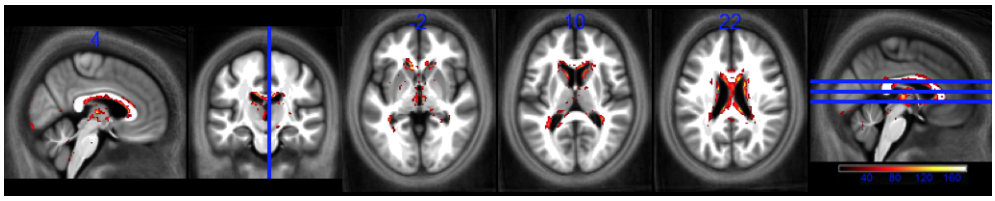

Figure 4.18: Statistical parametric maps of regions showing a correlation of R_2^* with age in the case of the modified VBQ analysis (equation (3.7)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

Figure 4.19 indicates that considering the parameter map $s(\phi)$ of a young participant, of an old participant or of the mean map of the entire cohort does not change the patterns found in the analysis.

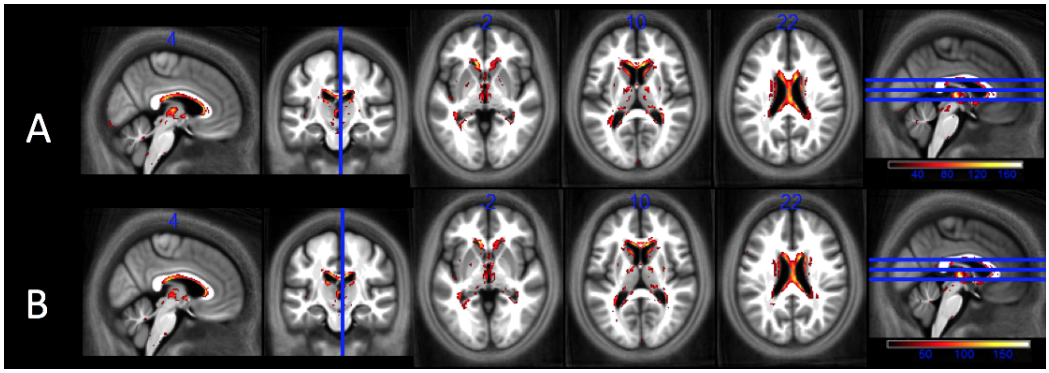

Figure 4.19: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis (equation (3.7)) for the mean map of the cohort (A) and for an elderly participant (B), at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

The plot representing the evolution with age in a ROI in the body of the corpus callosum (Figure 4.20) demonstrated a negative correlation of MT with age (left, blue line) for the traditional study. When considering only the parameter map MT of a young participant (left, red line) the correlation was made positive as described above but the variance of the points is small. In the same way, the parameter maps of all the subjects were negatively correlated with age (right, blue line), while the tissue classes (right, red line) and the Jacobian (right, green line) were independent of ageing.

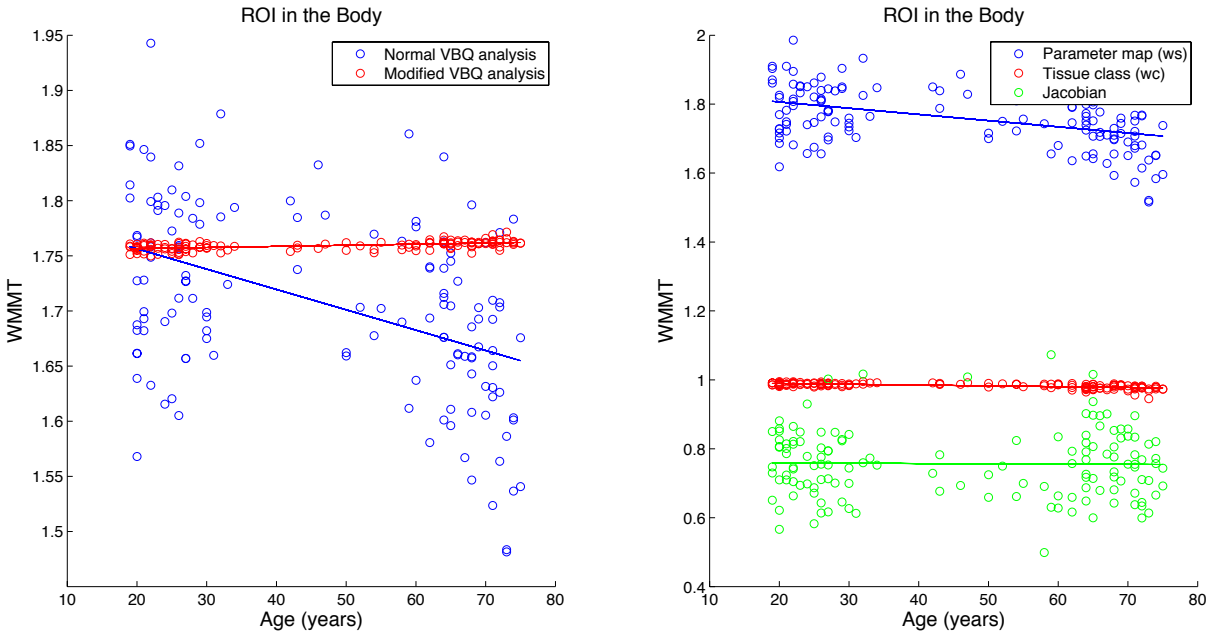

Figure 4.20: Evolution with age in a white matter region of interest in the body of the corpus callosum for (left) equations (3.6) and (3.7) and for (right) the parameters maps ($s_i(\phi)$), the tissue classes ($t_i(\phi)$) as well as for the Jacobian determinants ($|D\phi|_i$).

Smooting kernel

Modifying the full width half maximum of the smoothing kernel (Figure 4.21) shows that the higher the kernel size the more extended the detected clusters. The same regions are detected though.

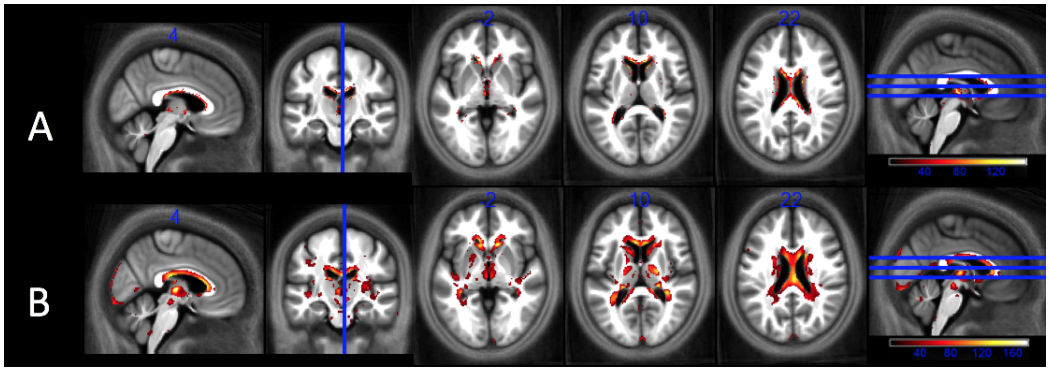

Figure 4.21: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis (equation (3.7)) for a young participant with a 0.5 mm (A) and a 6 mm (B) FWHM isotropic Gaussian kernel, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

Removing the Jacobian determinants from the weights

By removing the Jacobian determinants from the weights, significant voxels were observed in the ventricles and in the caudate nucleus for a young participant parameter map (Figure 4.22) and strong correlations were detected in frontal and parietal regions, in the thalamus, in the borders of corpus callosum and in the optic radiation for participant-specific parameter maps (Figure 4.23).

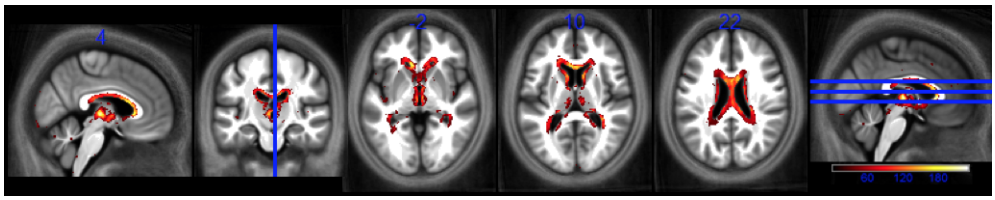

Figure 4.22: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis without Jacobian (equation (3.9)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

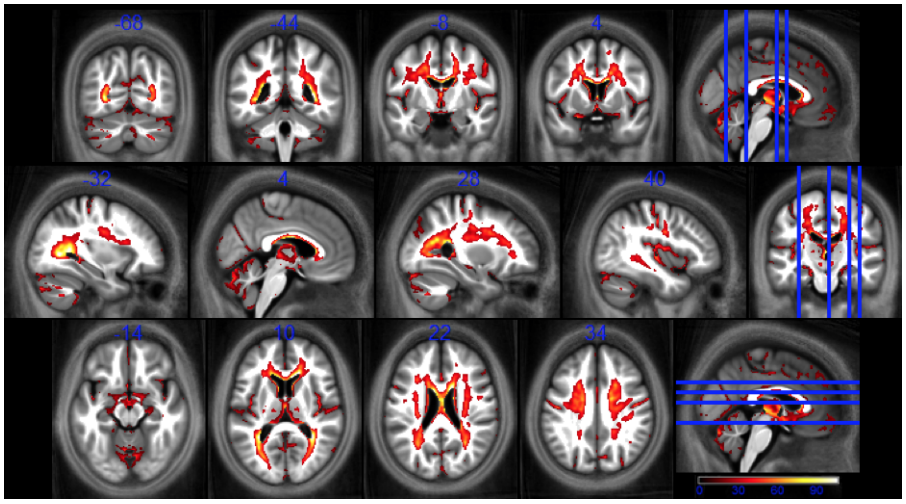

Figure 4.23: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis without Jacobian (equation (3.8)), at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

Removing the tissue class images from the weights

Considering only the Jacobian determinants in the weights of the VBQ approach for a young participant parameter map (Figure 4.24) indicated a more extended correlation around the ventricles, in the corpus callosum, in frontal regions and in the midbrain.

In the case of participant-specific parameter maps (Figure 4.25), similar regions as found in Figure 4.23 were detected, i.e. in frontal and parietal regions, in the thalamus, in the borders of corpus callosum and in the optic radiation.

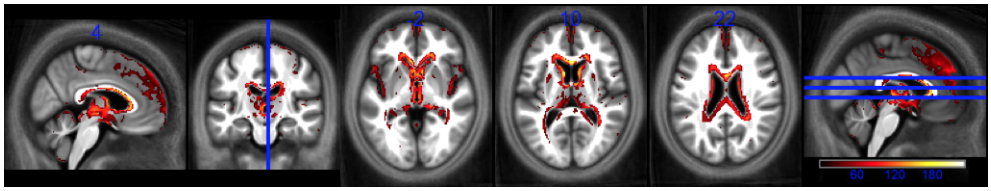

Figure 4.24: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis without tissue class (equation (3.10)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

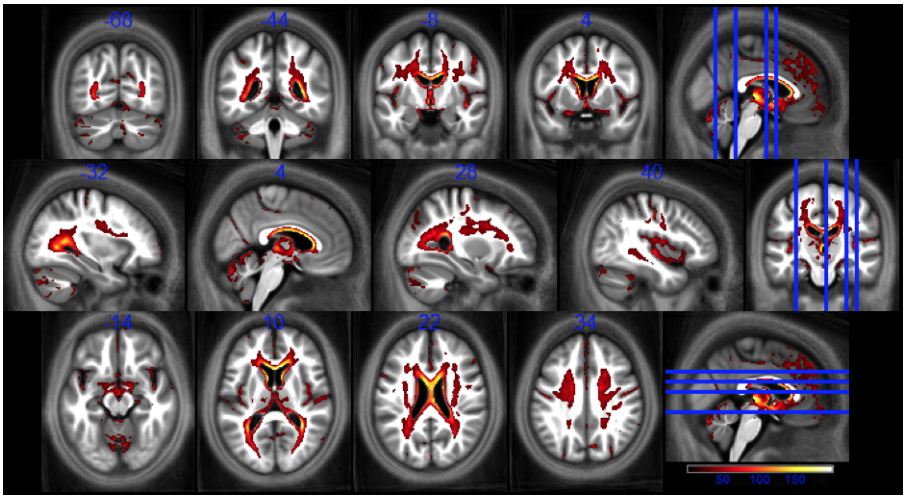

Figure 4.25: Statistical parametric maps of regions showing a correlation of MT with age in the case of the modified VBQ analysis without tissue class (equation (3.11)), at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

T-SPOON procedure

Statistical maps generated from the T-SPOON approach (Figure 4.26) suggest that correlations were displayed in the thalamus, the borders of the corpus callosum, some cortical regions and the optic radiation as it was the case using VBQ but T-SPOON found additional voxels in these structures.

In the case where the parameter map was not participant-specific (Figure 4.27), the regions were less widespread but the ventricles and the corpus callosum were still identified.

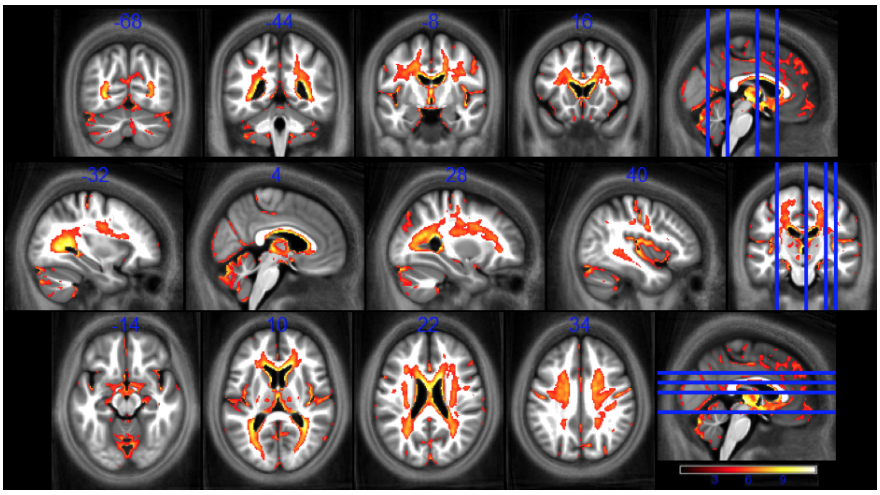

Figure 4.26: Statistical parametric maps of regions showing a correlation of MT with age in the case of the T-SPOON approach (equation (3.12)), at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The t-scores are represented using the colobar and some slices of the brain are presented in the coronal, sagittal and axial orientations.

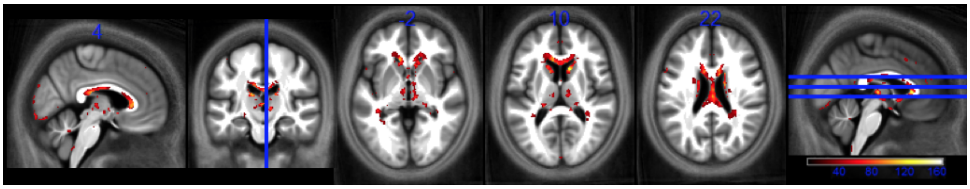

Figure 4.27: Statistical parametric maps of regions showing a correlation of MT with age in the case of the T-SPOON approach (equation (3.12)) for a young participant, at $p < 0.05$ FWE corrected. The results (of both the GM and WM analyses) were overlaid on the mean MT map of the cohort in MNI space. The F-scores are represented using the colobar and some slices of the brain are presented in the sagittal and axial orientations.

Similarity measure

The dice coefficients were calculated between different VBQ analyses presented above in the case of a fixed parameter map, for the GM and the WM separately. In the GM, the measures range from 0.1588 to 0.6071 (Figure 4.28). The highest similarity is reached between the analyses using only the tissue classes as weights and the original weights with a smoothing kernel of 3 mm FWHM, both with the parameter map of a single young participant. In the WM (Figure 4.29), the range is much narrower (from 0.4321 to 0.6757) with greater similarities but the highest number is in the same order than for the GM.

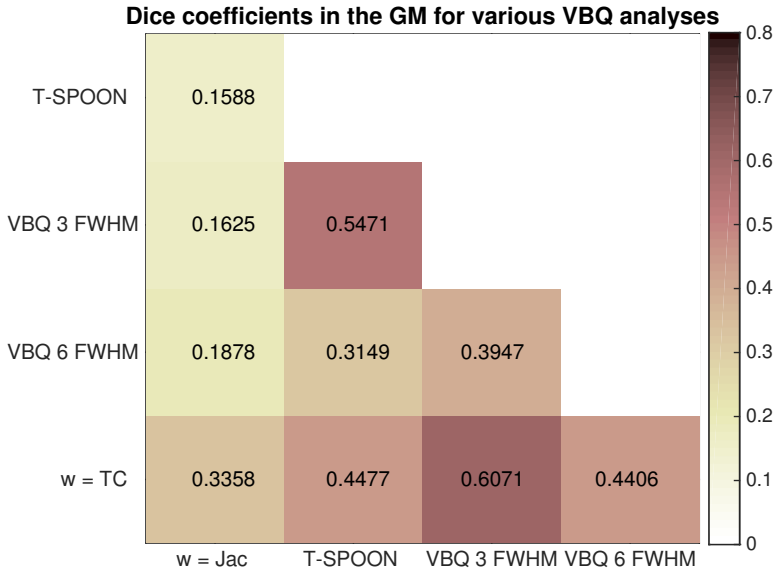

Figure 4.28: Dice coefficients between some VBQ analyses in the GM in the case of the parameter map MT of a young participant. T-SPOON: T-SPOON procedure, VBQ 3 FWHM: use of a smoothing kernel of 3 mm FWHM, VBQ 6 FWHM: use of a smoothing kernel of 6 mm FWHM, w = TC: keep only the tissue class images in the weights and w = Jac: keep only the Jacobian determinants in the weights.

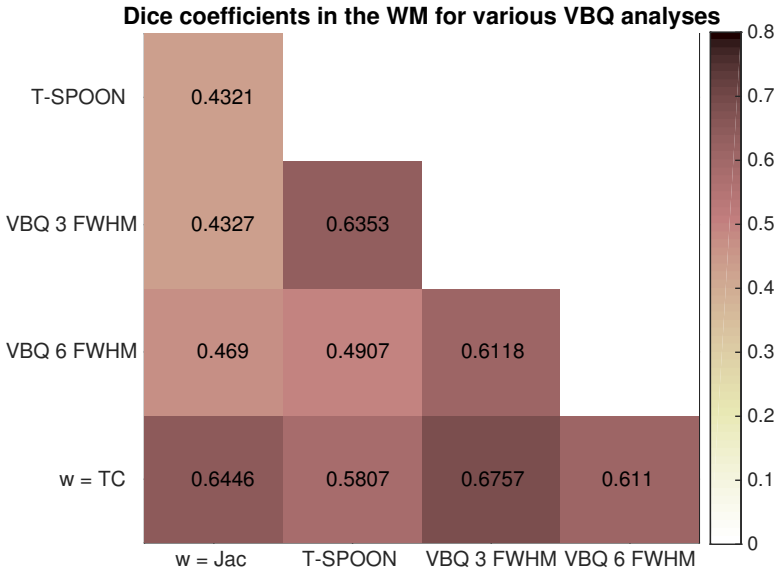

Figure 4.29: Dice coefficients between some VBQ analyses in the WM in the case of the parameter map MT of a young participant. T-SPOON: T-SPOON procedure, VBQ 3 FWHM: use of a smoothing kernel of 3 mm FWHM, VBQ 6 FWHM: use of a smoothing kernel of 6 mm FWHM, w = TC: keep only the tissue class images in the weights and w = Jac: keep only the Jacobian determinants in the weights.

Chapter 5

Discussion

5.1 Univariate analysis

The univariate analyses demonstrate the regions of significant voxels observed with SPM12 are consistent with those in Callaghan et al. (2014). Yet, small discrepancies arise likely due to the fact that statistical inferences from SPM12 were used on data processed in SPM8. Updates from one version to the other can lead to variations.

In addition, PALM's results show to be in agreement with the results of SPM. The high dice coefficient found in the GM means that the clusters and voxels found in SPM and PALM are highly concordant. The low similarity in the WM however suggests that both methods are not exactly comparable. Such difference may be due to the assumptions underlying parametric inference. Indeed, if these assumptions prove to be verified, the use of parametric statistics is better than the use of nonparametric statistics leading to a sensitivity difference.

Furthermore, a small bias towards large t-scores to SPM was observed in the Bland-Altman plot. It might originate from the fact that the parametric approach in SPM provides maximally efficient estimators and hence has higher sensitivity than PALM and therefore higher t-scores. In other words, it makes assumptions rendering it more sensitive as long as these assumptions hold. In addition, numerical approximations might be involved. The spreading of the values in the histogram was relatively narrow, suggesting that the methods can be considered as essentially equivalent. It is important to note that small differences are expected while comparing two packages as cross-software variability was demonstrated in Bowring et al. (2018). Therefore, one would not expect to get identical results.

5.2 Multivariate analysis

Dealing with multivariate analysis, it could be inferred that the results converge starting from 750 permutations. But as the higher the total number the better, and in accordance with Nichols and Holmes (2002), it was chosen to use 1000 permutations for all analyses conducted in this thesis as it leads to robust and accurate measures. There is a trade-off between the computational time and the accuracy while determining the number of per-

mutations to perform. To ensure the results' reliability, it would be advised to increase the number of permutations for further analyses if the use of extensive multi-threaded computing and of GPUs is possible. One issue preventing the speed to be reduced is that parallel computing cannot be applied to PALM. Since the permutations are performed independently, processing samples of rearrangements in a substantial number of processors is likely to decrease the wall clock time along with the scalability of the problem. Modifications in the codes would be required to make PALM parallelisable. However, the complexity of the codes composing the toolbox makes the realisation of these modifications challenging.

Regarding the regions correlated with ageing, it can be deduced that the extensive regions present in all single modalities are not present in the joint study as one could expected. The principal structures known histologically to be affected by ageing are present though, all of which were also detected in the univariate analyses. From a qualitative point a view, the statistical map of the multivariate analysis (Figure 4.11) can be compared with each single modality independently and the structures correlating with age be depicted as it has been done. Nonetheless, answering the question whether performing a joint analysis is worth it, does not seem to be an easy task. Indeed, the significant voxels that are detected can be observed in each univariate or multivariate statistical map. However, since it is not known how the four modalities are exactly linked in order to generate the output of the joint analysis, it is challenging to determine their exact influence and contribution. Some further analyses, more focused on a quantitative aspect were conducted to try to find an appropriate answer to that question.

The way PALM is implemented led to additional issues. It is known a priori that some modalities will either be positively correlated or be negatively correlated with ageing but not both. However, when testing different contrasts in PALM, a specific hypothesis cannot be applied to a specific modality. For example, the results of a positive correlation with age for R_2^* and of a negative correlation with age for MT cannot be combined. On the contrary, both hypotheses will be tested and so the combination will correspond to any difference with age.

5.3 Comparison

The interpretation of a switch from a univariate to a multivariate analysis is not straightforward. It has been shown that the joint analysis is not the union of the modalities, with only half as many significant voxels. There are a lot of voxels found in each separate analysis that are not recovered in the joint analysis. For example, only half the voxels identified in a univariate analysis of MT survive in the multivariate analysis. This observation suggests a significant sensitivity loss using the multivariate statistics as applied in this project.

Also, 3.59% of the significant voxels detected in the multivariate analysis were not found in any univariate age-study. These might be of real interest because they are likely to indicate what is really gained from performing a joint analysis. However, they are only voxels located at different borders across the brain. This new discovery is of little relevance because it corresponds to only 1500 voxels that belong to regions of interest

previously detected. It would have been much more interesting if a new cluster appeared. In this case, a region in the brain correlated with age that was not detected previously would have appeared thanks to the application of the joint analysis.

The true and false positive rates between the mass univariate analyses and the multivariate analysis illustrate the poor overlap between the joint analysis and the modalities MT, PD* and R_2^* because of medium rates (between 0.4-0.7) that are similar no matter of the reference used. However, a high percentage of voxels and clusters found in R_1 are discovered in the joint analysis, meaning that even though R_1 detects very few significant voxels they are likely to be present in the nonparametric output. In the same way, a high number of voxels and clusters found in the joint analysis are in the union. This observation is concordant with the small number of voxels (3.59% from the previous paragraph) found in the multivariate study but in any of the univariate cases.

In addition, it could be deduced from the union and intersection of various combinations (Tables 4.3 and 4.4) that the different modalities do not share a lot of voxels in common. It is good to note that only the significant voxels were considered. Therefore, a same cluster might be detected but only the voxels inside would differ. A further analysis on the voxels in the important clusters would have to be conducted to answer that question. Table 4.4 showed that all four modalities have only 28 significant voxels in common, which is very small. However, a voxel being significant in more than one modality is more likely to be found in the joint analysis.

The analyses conducted to answer the question whether applying a joint analysis is advantageous showed empirically that adding a modality without effect decreases significantly the sensitivity of the output. It is especially the case with R_1. Winkler et al. (2016b) showed theoretically that a loss of power of the NPC algorithm of 30 to 40% can be observed if only one modality out of four has signal. In the same way, the more modalities containing signal are added the higher the power. The sensitivity loss is so severe in multivariate statistics, especially when a modality has very few effects, that the use of PALM over SPM would not be recommended in similar studies. Furthermore, nonparametric approaches are known to be less powerful than parametric approaches when the assumptions of the latter are true. These assumptions provide the parametric analyses with additional information that the nonparametric analyses must discover. The NPC algorithm implemented in PALM has the processing advantage of testing each modality separately and of combining the results afterwards. While performing a multivariate analysis, there is a goal of finding how the modalities are somehow related to produce the final results. Therefore, such an advantage can in this way be viewed also as a caveat.

From what has been said, it is relatively challenging to know the exact contribution of each modality to the multivariate output. Still, a ranking of the information brought by each one of them independently on the final results can be suggested. First, it is clear that R_2^* and MT are the modalities giving the most information. Yet, R_2^* is richer in information than MT as it contains less significant voxels but has a higher dice coefficient with the multivariate analysis. Then, PD* has the highest dice coefficient with the multivariate case but has fewer significant voxels than MT and R_2^*. This is mainly due to the detection of the pallidum. Finally, R_1 is the least informative as it contains little signal. The ranking presented here is specific to this study and will change from one project to the other

depending on the data. Indeed, for example R_1 is characterised by a lack of information while usually it is used as a marker of reference for ageing due to its sensitivity to both myelin and iron. R_1 is known to be related to MT and R_2^*, respectively sensitive to the concentration of iron and myelin, across the brain through a robust and stable model using a single set of global coefficients (Callaghan et al., 2015).

5.4 VBQ analysis

The results of the voxel-based quantification procedure for this study design suggested that the detected spurious effects are much less widespread that the real effect of ageing. The same effects are captured whatever the parameter map considered and whatever the single participant under study. Indeed, using the parameter map of either a young or an elderly participant or even of the mean map of the cohort leaded to the same regions detected. In the same way, the spurious effects are present whether the quantitative MR parameter is R_2^*, PD^*, R_1 or MT. However, despite the presence of these morphological effects, the evolution with age of the modified VBQ analysis illustrates that the effect size is very small compared to the real age effect. Indeed, the red dots present a tiny variance and the slope of the increasing line is much less than the decreasing slope of the normal VBQ analysis. Here, only linear age-related differences were considered. However, non-linear trajectories were demonstrated (Callaghan et al., 2014; Carey et al., 2018).

In addition, the smoothing kernel significantly influences the extent of the regions detected. The higher the kernel the more widespread the effects. The spurious effects are also present when modifying the weights of the VBQ procedure. It was demonstrated that the significant clusters are generally identical even though the low dice coefficients between the different modifications of the VBQ formula indicate some differences in terms of the voxels in these clusters. Excluding the Jacobian determinants leaded to an increase in the effects which might be due to an higher effect size, indicating an emphasis on the underlying morphological effects. Finally, the T-SPOON procedure was tested and detected more activations than the VBQ procedure, indicating that the masks are pushing the significance higher than the weights.

In brief, these analyses showed that the age effect is mainly embedded in the parameter maps $s(\phi)$ and that the regions grossly affected by atrophy are identified in all different approaches that were undertaken. Indeed, the findings are not all consistent with the atrophy analysis of Callaghan et al. (2014). However, the areas grossly changed such as the ventricles and the corpus callosum are detected. Surprisingly, the cortex is robust to these effects while it is known to be influenced by major atrophy. It would thus be mindful to proceed to a similar analysis in all studies in order to see which regions are affected by the atrophy phenomenon. To do so, a voxel-based morphometry (VBM) (Ashburner and Friston, 2000) approach is recommended.

Chapter 6

Conclusion

In this thesis, a nonparametric multivariate analysis using permutation tests from PALM was performed on quantitative MRI data of healthy participants in order to analyse age-related effects in the brain microstructures. Permutation tests appeared to work and give sensible results. However, it was demonstrated that the lack of signal in some data introduced a severe sensitivity loss in the final results. Due to the lack of relations between the modalities provided by PALM, interesting future perspectives would be to experiment further the use of other multivariate statistics such as the CVA (Canonical Variate Analysis) or the RSA (Representational Similarity Analysis) approaches, both of which compute weights related to the importance each modality brings to the output, before hastily concluding the rejection of multivariate statistics.

Then, the validity of the voxel-based quantification (VBQ) normalisation approach was tested. It was shown that spurious effects very likely related to atrophy and morphological changes were introduced when only the weights in the standard MNI space were made participant-specific, i.e. with the parameter map of a single participant. However, their effect size was small and localised in specific regions such as the ventricles and the corpus callosum. In addition, VBQ is well understood by researchers, reported properly in the literature and extensively used. Still, future research might look at ways of further improving the approach to get rid of these undesired effects. In the meantime, it is recommended to perform a priori a VBM study to investigate the morphological effects influencing each study.

Appendices

Appendix A

PALM multivariate analysis code

The use of PALM is relatively straightforward in the sense that calling the toolbox with the command `palm` followed by the desired options to be considered is the only line of code to perform the analysis. The line of code used to carry out the multivariate analysis including all four modalities as described earlier is the following:

```
palm -i R2s_4Dimg.nii -i R1_4Dimg.nii -i MT_4Dimg.nii -i PDs_4Dimg.nii
-d design_matrix.mat -t contrast.con -m mask4D.nii -accel tail tail_threshold
-n 1000 -npc -npccon -nouncorrected
```

The various options entered are:

- `-i <file>`: insertion of the inputs. In this case, the 4D images for the entire cohort of each modality were inserted either for the GM or for the WM. The order is not important in this case.
- `-d <file>`: insertion of the design matrix. It corresponds to the design matrix inserted previously in SPM including the four regressors of interest of the study.
- `-t <file>`: insertion of the contrast file. In this file, two lines of contrasts were included: the first corresponds to the evaluation of a positive correlation with age whereas the second corresponds to a negative correlation with age.
- `-m <file>`: insertion of the mask corresponding to the tissue class under study. As mentioned earlier, this study analysed both tissue classes separately. A mask of the specific tissue class is thus included.
- `-accel tail tail_threshold`: addition of a tail approximation along with the tail value. The threshold was fixed to 0.05 as it is the value under which the p-values were considered as significant FWE corrected.
- `-n 1000`: specification of the number of permutations. As proven in this thesis, 1000 permutations showed to be of a good trade-off between time efficiency and accuracy.

- `-npc -npccon`: application of the NPC algorithm with the Fisher method, and the correction of both modalities and contrasts. The option `-npc` is a shortcut for `-npcmethod Fisher -npcmod`.
- `-nouncorrected`: do not save the uncorrected p-values. Performing various analyses, it was observed that due to the huge data (138 subjects $\times$ 4 modalities), the generation of the uncorrected results might take as long as a few hours. In addition, the interest was on corrected p-values.

Appendix B

Batch - Multivariate statistics

The multivariate statistics batch has been created in order to allow the user to apply permutation tests implemented in PALM in a simple way (Figure B.1). In this Appendix, further information are given about how to use it. It is important to note that the possibilities with PALM are much more extended than what can be done using this interface. The batch was created with the aim of realising analyses comparable to those presented in this thesis. For any other objectives or the use of any other options, please refer to the website (`https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM`) as well as to Winkler's papers (Winkler et al., 2014, 2016a,b; Winkler, 2017).

B.1 PALM directory

Select the directory where the PALM toolbox is located.

In order to practise on PALM, the toolbox has first to be downloaded (see: `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM/UserGuide` or `https://github.com/andersonwinkler/PALM`).

B.2 Data

Select the data to be permuted.

The data to be selected must be 4D images in order to be processed by the toolbox. If several 4D images are entered, they must be of the same size. The batch `Util` → `3D to 4D File Conversion` in SPM creates a 4D image from concatenating a number of 3D volumes.

B.3 Mask

There is the possibility to associate a mask with the data previously selected.

No mask

Select if no mask is associated with the data to be processed.

Mask file

Select to insert a mask file in the analysis.

In this configuration, two cases are conceivable. First, only one mask is associated with the data. Second, one mask is joined to each input data. In the latter case, one has to make sure that the order of selection of the masks is the same as the order of selection of the data so as to respect the matching.

B.4 Design matrix

Select the design matrix related to the inputs.

Several design matrices can be selected. In this case, the data will be permuted independently for each design matrix and the results will either be combined or be generated separately. In order to be readable by PALM, the format must either be `csv` or the vest format from FSL. Refer to: `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/GLM/CreatingDesignMatricesByHand` for creating a design matrix compatible with PALM.

B.5 T-contrast file

Select the t-contrasts file to carry out the statistical inference.

The number of files to be selected must either be 1 or be equal to the number of design matrices selected. In the same way as for the design matrices, the files must be in the `csv` or the vest format from FSL. Also, more that one contrast can be inserted into one single contrast file. Refer to: `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/GLM/CreatingDesignMatricesByHand` for creating a t-contrast file compatible with PALM.

B.6 Number of permutations

Specify the total number of rearrangements to perform on your data.

A default value of 1000 is present in accordance with what was demonstrated in this thesis. However, depending on the type of inputs or studies it may not be sufficient or may be reduced.

B.7 Assumptions

Choose the assumption on the errors and therefore the type of shufflings.

An important aspect governing permutation inference is the principle of exchangeability: the manner under which the data are rearranged under the null hypothesis. It is based on the assumption that the joint distribution of the data is unchanged if they are permuted. This assumption can be stipulated in terms of exchangeable errors (EE) and/or independent and symmetric errors (ISE).

EE

The errors are assumed to be exchangeable: the distribution of the errors does not change after permuting the data. Only permutations will be carried out.

ISE

The errors are assumed to be independent and symmetric: the errors do not change after changing their signs. Only sign-flippings will be done.

Both

The errors are assumed to be exchangeable, and independent and symmetric: do both permutations and sign-flippings.

B.8 Algorithm

Choose the options constituting the algorithm to perform the permutation tests.

B.8.1 NPC

Two permutation methods can be applied for your analysis. On one hand, the modified Non-Parametric Combination (NPC) (Section 2.5) corresponds to the combination of the joint analysis of various partial tests. This algorithm particularly finds its interests when multiple inputs are to be permuted. On the other hand, simple rearrangements can be executed.

B.8.2 Correction over modalities

Once the algorithm has been selected, it is possible to correct over multiple modalities. In this case, the permutations are performed independently on each input and the empirical p-values obtained are then combined. Performing a joint analysis means the inputs have to be combined. Otherwise, outputs will be generated for each input separately.

B.8.3 Correction over contrasts

In the same way, when several contrasts are present in one t-contrast file, they can be combined. A correction over contrasts implies that each contrast is tested independently by computing the combined statistic at each voxel and the results are combined afterwards. Otherwise, the p-values of each contrast are produced separately.

B.9 Options

Various useful options can be added to execute the multivariate analysis. Some aims at decreasing the computational time while others have rather a visual purpose.

B.9.1 Acceleration methods

It is possible to use a tail approximation in order to reduce the computational time. It consists in modelling the tail of the distribution by fitting a generalised Pareto distribution (GPD) (Section 3.3.1).

No tail

No acceleration method will be used in the process.

Tail value

Specify the value of the tail.

The tail value corresponds to the p-value below which the fit is to be used. The default value set in PALM is 0.1 but can be modified. The higher the value the greater the modelling.

B.9.2 Uncorrected

The p-values generated by PALM are corrected and uncorrected p-values. However, one can decide to generate only corrected p-values. The time to compute the uncorrected might be huge when the data is cumbersome.

B.9.3 Save as -log10(p)

PALM offers the opportunity to produce outputs displaying not p-values but $-\log_{10}(p)$ instead. Since p-values are likely to be very small, the contrast will not be seen easily on the brain images. Therefore, taking the negative logarithm will emphasize the smallest p-values, i.e. with the highest activation.

B.10 Output prefix name

Specify the prefix name of the outputs.

All the outputs will be in the form: `prefix_*` depending on the image generated.

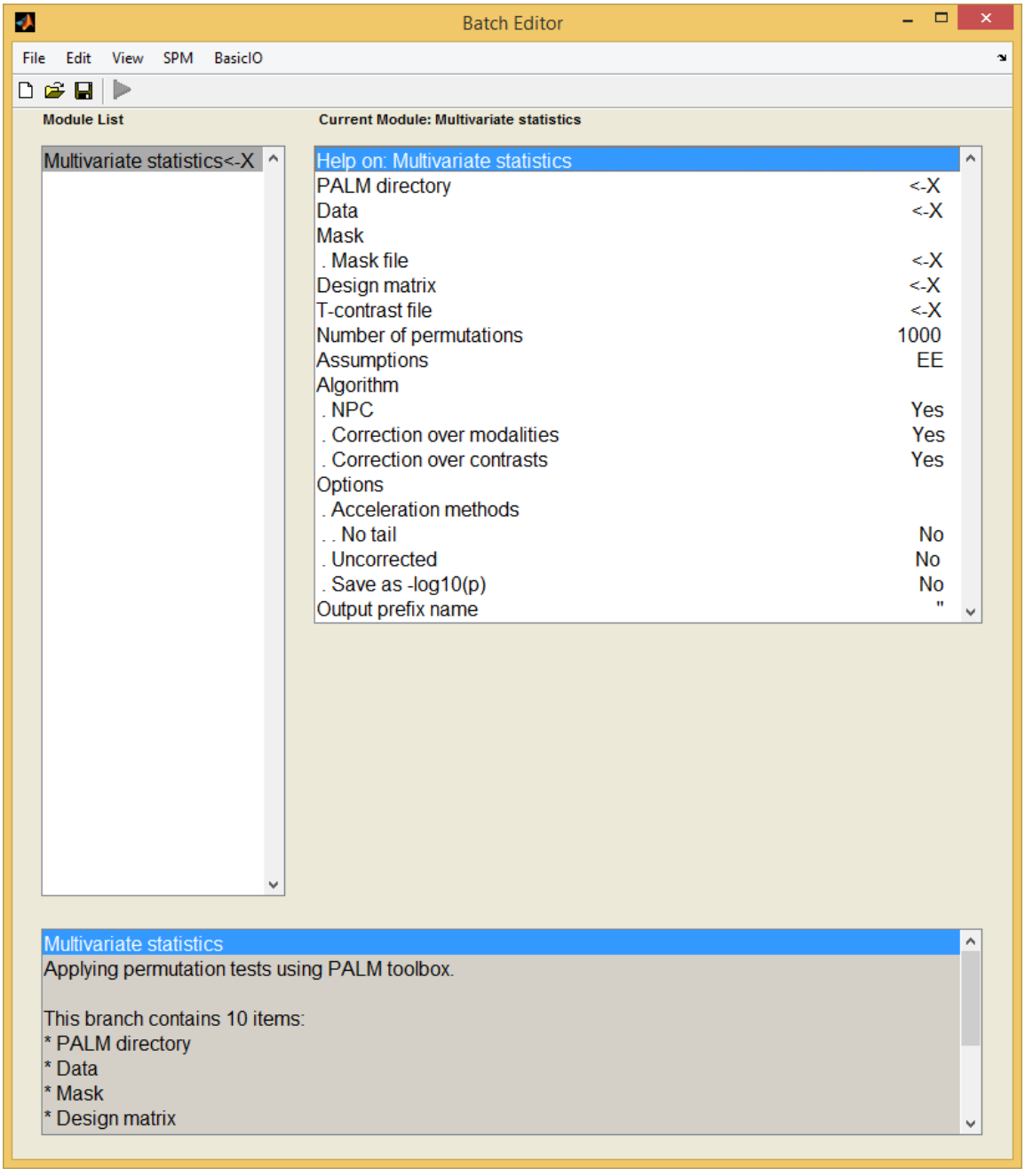

Figure B.1: Permutation tests implemented in PALM using SPM user interface.

Bibliography

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm.

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometry - The methods. *NeuroImage*, 11(6 I):805–821.

Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3):839–851.

Bowring, A., Maumet, C., and Nichols, T. (2018). Exploring the Impact of Analysis Software on Task fMRI Results. *bioRxiv*.

Callaghan, M. F. Multi-Parameter Mapping, University College of London.

Callaghan, M. F., Freund, P., Draganski, B., Anderson, E., Cappelletti, M., Chowdhury, R., Diedrichsen, J., FitzGerald, T. H., Smittenaar, P., Helms, G., Lutti, A., and Weiskopf, N. (2014). Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. *Neurobiology of Aging*.

Callaghan, M. F., Helms, G., Lutti, A., Mohammadi, S., and Weiskopf, N. (2015). A general linear relaxometry model of R1 using imaging data. *Magnetic Resonance in Medicine*, 73(3):1309–1314.

Carey, D., Caprini, F., Allen, M., Lutti, A., Weiskopf, N., Rees, G., Callaghan, M. F., and Dick, F. (2018). Quantitative MRI provides markers of intra-, inter-regional, and age-related differences in young adult cortical microstructure. *NeuroImage*, (October):1–12.

Cercignani, Mara, D. N. G. T. P. S. (2017). *Quantitative MRI of the Brain, Principles of Physical Measurement*. CRC Press, Taylon and Francis Group, 2 edition.

Connor, J. R. and Menzies, S. L. (1996). Relationship of iron to oligodendrocytes and myelination. *Glia*, 17(2):83–93.

Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R. S., Helms, G., and Weiskopf, N. (2011). Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *NeuroImage*.

Fevre, E. Bland-Altman plot - Measuring agreement between methods, Edinburgh.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 4 edition.

Fjell, A. M., Walhovd, K. B., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Holland, D., Brewer, J. B., and Dale, A. M. (2009). One-Year Brain Atrophy Evident in Healthy Aging. *Journal of Neuroscience*, 29(48):15223–15231.

Fotuhi, M., Hachinski, V., and Whitehouse, P. J. (2009). Changing perspectives regarding late-life dementia. *Nature Reviews Neurology*, 5(12):649–658.

Hallgren, B. and Sourander, P. (1958). The effect of age on tetrahydrobiopterin metabolism in the human brain. *Journal of Neurochemistry*, 3(Pergamon Press Ltd., London):41–51.

Helms, G., Dathe, H., and Dechent, P. (2008). Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magnetic Resonance in Medicine*, 59(3):667–672.

Knijnenburg, T. A., Wessels, L. F., Reinders, M. J., and Shmulevich, I. (2009). Fewer permutations, more accurate P-values. *Bioinformatics*, 25(12):161–168.

Lee, J. E., Chung, M. K., Lazar, M., DuBray, M. B., Kim, J., Bigler, E. D., Lainhart, J. E., and Alexander, A. L. (2009). A study of diffusion tensor imaging by tissue-specific, smoothing-compensated voxel-based analysis. *NeuroImage*, 44(3):870–883.

Lintl, P. and Braak, H. (1983). Loss of intracortical myelinated fibers: A distinctive age-related alteration in the human striate area. *Acta Neuropathologica*, 61(3-4):178–182.

Marner, L., Nyengaard, J. R., Tang, Y., and Pakkenberg, B. (2003). Marked loss of myelinated nerve fibers in the human brain with age. *Journal of Comparative Neurology*, 462(2):144–152.

Nichols, T. and Holmes, A. (2002). Nonparametric Permutation Tests for functional Neuroimaging: A Primer with Examples. *Hum Brain Mapp*, 15(1):1–25.

Peters, R. (2006). Ageing and the brain. *Postgraduate medical journal*, 82(964):84–8.

Phillips, C. (2017-2018). STAT0722-1 Introduction to medical statistics, University of Liege.

Reece, Urry, C. W. M. and Jackson (2011). *Campbell Biology*. Pearson Education, Inc., 9 edition.

Tippett, L. (1931). *The methods of statistics*. Williams and Northgate, London.

Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., and Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *NeuroImage*, 54(3):2116–2124.

Weiskopf, N., Suckling, J., Williams, G., Correia M., M. M., Inkster, B., Tait, R., Ooi, C., Bullmore T., E. T., and Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: A multi-center validation. *Frontiers in Neuroscience*, 7(7 JUN):1–11.

Winkler, A. M. (2017). *Statistical analysis of areal quantities in the brain through permutation tests*. PhD thesis.

Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E., and Smith, S. M. (2016a). Faster permutation inference in brain imaging. *NeuroImage*, 141:502–516.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*.

Winkler, A. M., Webster, M. A., Brooks, J. C., Tracey, I., Smith, S. M., and Nichols, T. E. (2016b). Non-parametric combination and related permutation tests for neuroimaging. *Human Brain Mapping*, 37(4):1486–1511.

Zecca, L., Youdim, M. B., Riederer, P., Connor, J. R., and Crichton, R. R. (2004). Iron, brain ageing and neurodegenerative disorders. *Nature Reviews Neuroscience*, 5(11):863–873.