

Developing Neuroimaging Biomarkers for Neurodegeneration and Ageing Using Machine Learning Methods

Zobair Arya

Oxford Centre for Functional MRI of the Brain,
Nuffield Department of Clinical Neurosciences

&

Jesus College

University of Oxford



A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2019

Abstract

Developing Neuroimaging Biomarkers for Neurodegeneration and Ageing Using Machine Learning Methods

Zobair Arya

Jesus College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2019

In this thesis, we present three novel methods based on machine learning for use with MRI-derived neuroimaging data, all with the aim of aiding biomarker development for neurodegeneration.

Resting-state functional MRI (rfMRI) can potentially detect early functional changes in disease. Therefore, the first method is a novel supervised learning algorithm for use in classifying rfMRI data into two groups. The main advantage over existing rfMRI-based classification approaches is that the entire voxel by time data is fed in without any prior decomposition or parcellation of the data into brain regions, and it does not require any prior knowledge of potential discriminatory networks. We show that the algorithm can give interpretable results for simulated data, performs better than two existing approaches for a Parkinson’s disease dataset and gives results consistent with those reported previously for an anaesthesia dataset.

Ageing is strongly linked with neurodegeneration. Hence, the second method is an errors-in-variables model for estimating the brain age of individuals. It takes a fundamentally different approach compared to existing methods, which gives it the advantages of being able to capture inverted U-shaped brain ageing trajectories, provide visualisable trajectories and potentially model interactions with disease. Using simulated data, we show that it is effective at estimating brain age, and can match and in some cases outperform current approaches. Using the UK biobank dataset, we show that the difference between chronological age and predicted brain age for an individual correlates with variables that are consistent with those reported previously. We also show that it can potentially provide information that is complementary to existing methods.

The final method is a tool for decomposing multimodal neuroimaging data into spatial maps and associated trajectories. The main differences compared to existing methods are that the decomposition is driven by a variable of interest, which in our case was age, and each brain region is associated with only one trajectory as opposed to a mixture of different trajectories. This means that the components do not have to be associated with age post-hoc and the results are potentially more interpretable. Using simulated data, we show that, compared to two existing methods, our method can more accurately reproduce trajectories and associated spatial maps. Using the UK biobank dataset, we show that our method estimates trajectories and associated regions that are in agreement with previously published studies.

Acknowledgements

Firstly, I would like to thank my supervisor Mark Jenkinson for his guidance throughout my DPhil project. I could not have imagined having a better mentor and I am grateful to him for all his support. I would also like to thank my co-supervisor Clare Mackay for her valuable input and advice. In addition, I am thankful to Fidel Alfaro-Almagro, Ludovica Griffanti, Verena Heise, Paul McCarthy, Ricarda Menke, Stephen Smith and Katie Warnaby for providing advice and access to preprocessed data. My DPhil studies were funded by the EPSRC and MRC, and to them I am also thankful. Finally, I am grateful to my family and friends for their support, especially my father and Jade Bailey.

I wish to dedicate this thesis to my late mother who has always been a source of inspiration.

Contents

1	Introduction	16
1.1	Biomarkers for neurodegeneration	16
1.2	MRI modalities	18
1.2.1	Structural MRI	18
1.2.2	Diffusion MRI	20
1.2.3	Functional MRI	21
1.3	Layout of thesis	22
2	Iterative dual linear discriminant analysis: A novel classification algorithm for resting-state fMRI	24
2.1	Introduction	24
2.2	Methodology	38
2.2.1	Linear discriminant analysis	38
2.2.2	Iterative dual linear discriminant analysis	38
2.2.3	Adapting IDLDA to high dimensional and low sample size data	41
2.2.4	Datasets	45
2.2.4.1	Simulated data	45
2.2.4.2	Parkinson's disease data	47
2.2.4.3	Amyotrophic lateral sclerosis data	47
2.2.4.4	Anaesthesia data	48
2.2.5	Classification	48
2.2.6	Comparing to other methods	51

2.2.6.1	Independent components analysis + dual regression .	51
2.2.6.2	Network based + negative/positive connectivity index	52
2.2.6.3	Network based + all correlations	53
2.3	Results	53
2.3.1	Simulated data	53
2.3.2	Parkinson's disease data	56
2.3.3	Amyotrophic lateral sclerosis data	60
2.3.4	Anaesthesia data	62
2.4	Discussion	66
2.5	Conclusion	71
3	An errors-in-variables approach to estimating brain age	72
3.1	Introduction	72
3.2	Methodology	79
3.2.1	The Model	79
3.2.2	Optimisation	82
3.2.3	Simulated data	83
3.2.3.1	Simulation method	83
3.2.3.2	Estimating σ_m	84
3.2.3.3	Sensitivity to σ_m and ν	85
3.2.3.4	Effects of various data attributes	86
3.2.3.5	When the trajectory is not known	87
3.2.3.6	Sensitivity to initialisation	88
3.2.3.7	Non-Gaussian noise	89
3.2.3.8	Comparison to other methods	89
3.2.4	UK Biobank data	91
3.3	Results	94
3.3.1	Simulated data	94
3.3.1.1	Estimating σ_m	94
3.3.1.2	Sensitivity to σ_m and ν	95

3.3.1.3	Effects of various data attributes	97
3.3.1.4	When the trajectory is not known	101
3.3.1.5	Sensitivity to initialisation	105
3.3.1.6	Non-Gaussian noise	106
3.3.1.7	Comparison to other methods	107
3.3.2	UK Biobank data	110
3.4	Discussion	122
3.5	Conclusion	130
4	A method for decomposing multimodal neuroimaging data into tra-	
	jectories of ageing	131
4.1	Introduction	131
4.2	Methodology	138
4.2.1	The model	138
4.2.2	Simulated data	142
4.2.2.1	Simulation method	142
4.2.2.2	Influence of simulation and estimation parameters . .	144
4.2.2.3	Determining the correct number of components . . .	145
4.2.2.4	Deriving continuous component maps	146
4.2.2.5	Comparisons to other methods	147
4.2.3	UK biobank data	148
4.3	Results	149
4.3.1	Simulated data	149
4.3.1.1	Influence of simulation and estimation parameters . .	149
4.3.1.2	Determining the correct number of components . . .	153
4.3.1.3	Deriving continuous component maps	157
4.3.1.4	Comparisons to other methods	159
4.3.2	UK biobank data	161
4.3.2.1	Single modality	161
4.3.2.2	Multimodal	168

4.4	Discussion	175
4.5	Conclusion	185
5	Final discussion and future work	187
5.1	Summary of main thesis contributions	187
5.2	Future work	188
5.2.1	Modelling errors associated with variables other than age . . .	188
5.2.2	Regional and multimodal brain ages	189
5.2.3	Investigating the clinical relevance of biomarkers	191
A	Model derivatives	192
B	The chosen UK biobank IDPs	194
C	Full spatial maps	207

List of Figures

2.1	Visual representation of the Iterative Dual LDA metric	39
2.2	Flowchart of the proposed IDLDA algorithm	41
2.3	A diagram of nested cross validation	49
2.4	Output $\mathbf{w}$ and $\mathbf{v}$ images for the $g = 2$ simulated dataset	54
2.5	Output $\mathbf{w}$ and $\mathbf{v}$ images for the $g = 5$ simulated dataset	54
2.6	Output $\mathbf{w}$ and $\mathbf{v}$ images for the $g = 9$ simulated dataset	55
2.7	Sensitivity to initialisation as a function of λ	55
2.8	The λ value at which the algorithm stops being sensitive to initialisation	56
2.9	LOO CV AUC results as a function of added noise	56
2.10	k -fold CV AUC results for the PD dataset	57
2.11	AUC as a function of iteration number for the PD dataset	58
2.12	Histogram of the optimum λ values	58
2.13	The objective function value as a function of iteration number for the PD dataset	59
2.14	Output $\mathbf{w}$ and $\mathbf{v}$ images for the PD dataset	59
2.15	Output $\mathbf{w}$ and $\mathbf{v}$ images for the ALS dataset with $\mathbf{r}_1$	60
2.16	Output $\mathbf{w}$ and $\mathbf{v}$ images for the ALS dataset with $\mathbf{r}_2$	61
2.17	k -fold CV results for the ALS dataset	62
2.18	Classification accuracy for the anaesthesia data.	63
2.19	Classification accuracy when applying ICA+DR to the anaesthesia data	63
2.20	Output $\mathbf{w}$ and $\mathbf{v}$ images for the anaesthesia dataset	64

2.21	Sensitivity to initialisation on the anaesthesia data	65
2.22	The spatial maps corresponding to $\mathbf{w}^+$ and $\mathbf{v}^+$	66
2.23	The spatial maps corresponding to $\mathbf{w}^-$ and $\mathbf{v}^-$	66
2.24	Interpreting the sign of weights	67
3.1	A graphical model illustrating our brain age formulation	79
3.2	The functions we used when simulating data	83
3.3	The age distribution of the sampled UK biobank subjects.	93
3.4	The estimation error as a function of SNR for $N = 100$	94
3.5	The estimation error as a function of SNR for $N = 1300$	95
3.6	How the brain age error is affected when an incorrect σ_m is given to the optimisation	96
3.7	How the brain age error is affected when an incorrect ν is given to the optimisation	97
3.8	The brain age and model errors as a function of the number of features	98
3.9	An illustration of how having more features improves brain age pre- diction.	99
3.10	The brain age and model errors as a function of the number of subjects	100
3.11	The fractional improvement in brain age error from the initial state as a function of the SNR ratio	101
3.12	How ε_{BA} and ε_f are affected by different fitting functions	101
3.13	The change in the brain age and model errors with the number of interior knots	102
3.14	The change in the brain age and model errors with the polynomial degree	103
3.15	An illustration of how our method shifts the chronological ages be- longing to a sigmoid trajectory	104
3.16	An illustration of how our method shifts the chronological ages be- longing to an exponential trajectory	105
3.17	Sensitivity to initialisation	106

3.18	The model and brain age errors for data with non-Gaussian noise . .	107
3.19	The mean estimated absolute brain age delta for UK biobank training subjects	111
3.20	The mean estimated absolute brain age delta for UK biobank test subjects	111
3.21	Estimated brain age deltas for training subjects from our method plotted against chronological age	112
3.22	Estimated brain age deltas for test subjects from our method plotted against chronological age	113
3.23	Estimated brain age delta for GPR plotted against chronological age	114
3.24	Three IDPs plotted against the estimated brain ages	115
3.25	The mean absolute corrected delta for training subjects	116
3.26	The mean absolute corrected delta for test subjects	116
3.27	Correlation of deltas across methods	117
3.28	Manhattan plot for deltas from a linear trajectory	119
3.29	Manhattan plot for deltas from a polynomial trajectory	119
3.30	Manhattan plot for test subject deltas from a linear trajectory	121
3.31	Manhattan plot for test subject deltas from a polynomial trajectory .	121
3.32	Manhattan plot for deltas from GPR	122
4.1	A diagram of the decomposition method	139
4.2	The simulated trajectories	142
4.3	The spatial maps associated with the simulated trajectories	143
4.4	Change in dice index with various parameters	150
4.5	Change in estimated trajectory accuracy with various parameters . .	151
4.6	The estimated trajectories for simulated data with $p = 0.005$	152
4.7	The estimated binary spatial maps associated with the trajectories in figure 4.6	153
4.8	The estimated binary spatial maps when $p = 0.05$	153
4.9	The ‘complete’ hierarchical clustering dendrogram for simulated data	154

4.10	The ‘ward’ hierarchical clustering dendrogram for simulated data . .	155
4.11	The ‘complete’ hierarchical clustering dendrogram for simulated data that contains non-trajectory voxels	156
4.12	The ‘ward’ hierarchical clustering dendrogram for simulated data that contains non-trajectory voxels	157
4.13	The multimodal dendrogram for simulated data	158
4.14	The estimated multimodal spatial maps associated with figure 4.13 .	158
4.15	The resulting spatial maps when our method was rerun using k-means	159
4.16	The multimodal dendrogram when the p threshold was lowered to 0.1	159
4.17	The multimodal dendrogram after pruning	160
4.18	The estimated multimodal spatial maps associated with figure 4.17 .	160
4.19	The trajectories corresponding to the spatial maps in figure 4.19. . .	161
4.20	Continuous component spatial maps obtained using the multivariate Gaussian approach.	162
4.21	Continuous component spatial maps obtained using the squared Eu- clidean distance approach.	162
4.22	The estimated subject weightings plotted against age for informed multimodal PLS and joint ICA	163
4.23	The estimated informed multimodal PLS component spatial maps associated with the trajectories in figure 4.22a.	163
4.24	The estimated joint ICA component spatial maps associated with the trajectories in figure 4.22b.	164
4.25	The UK biobank unimodal dendrograms	164
4.26	The estimated unimodal subject weightings from the training set . . .	165
4.27	The estimated unimodal subject weightings from the test set	165
4.28	The fitted trajectories from figure 4.26	166
4.29	The spatial map associated with the second unimodal VBM component	167
4.30	The spatial map associated with the third unimodal VBM component	168
4.31	The spatial map associated with the first unimodal VBM component	168

4.32	The spatial map associated with the second unimodal FA component	169
4.33	The spatial map associated with the third unimodal FA component	169
4.34	The spatial map associated with the first unimodal MD component	170
4.35	The spatial map associated with the fourth unimodal MD component	170
4.36	The spatial map associated with the second unimodal L1 component	171
4.37	The spatial map associated with the third unimodal L1 component	171
4.38	The UK biobank multimodal dendrogram	172
4.39	The estimated multimodal subject weightings	173
4.40	The fitted trajectories from figure 4.39a	173
4.41	The spatial maps corresponding to the second multimodal component	174
4.42	The spatial maps corresponding to the fourth multimodal component	175
4.43	The spatial maps corresponding to the third multimodal component	176
4.44	The spatial maps corresponding to the fifth multimodal component	177
4.45	The subjects weights from informed multimodal PLS	178
4.46	The spatial maps corresponding to the first multimodal component from informed multimodal PLS	179
4.47	The spatial maps corresponding to the second and third multimodal components from informed multimodal PLS	180
4.48	The subjects weights from joint ICA	181
4.49	The spatial maps corresponding to the first and the second multi- modal components from joint ICA	182
4.50	The spatial maps corresponding to the fourth and fifth multimodal components from joint ICA	183
5.1	A diagram of a proposed study investigating the usefulness of region- and modality-specific brain ages.	190
C.1	The VBM unimodal binary maps	208
C.2	The VBM unimodal inverse distance maps	208
C.3	The FA unimodal binary maps	209

C.4	The FA unimodal inverse distance maps	209
C.5	The MD unimodal binary maps	210
C.6	The MD unimodal inverse distance maps	210
C.7	The L1 unimodal binary maps	211
C.8	The L1 unimodal inverse distance maps	211
C.9	The multimodal binary maps for component one	212
C.10	The multimodal inverse distance maps for component one	212
C.11	The multimodal binary maps for component two	213
C.12	The multimodal inverse distance maps for component two	213
C.13	The multimodal binary maps for component three	214
C.14	The multimodal inverse distance maps for component three	214
C.15	The multimodal binary maps for component four	215
C.16	The multimodal inverse distance maps for component four	215
C.17	The multimodal binary maps for component five	216
C.18	The multimodal inverse distance maps for component five	216
C.19	The spatial maps for informed multimodal PLS component one. . . .	217
C.20	The spatial maps for informed multimodal PLS component two. . . .	217
C.21	The spatial maps for informed multimodal PLS component three. . .	218
C.22	The spatial maps for informed multimodal PLS component four. . . .	218
C.23	The spatial maps for informed multimodal PLS component five. . . .	219
C.24	The spatial maps for joint ICA component one.	219
C.25	The spatial maps for joint ICA component two.	220
C.26	The spatial maps for joint ICA component three.	220
C.27	The spatial maps for joint ICA component four.	221
C.28	The spatial maps for joint ICA component five.	221

List of Tables

2.1	Summary of studies focusing on classifying subjects based on rfMRI .	35
3.1	Brain age errors across different methods on simulated data	108
3.2	Brain age errors from cross validation across different methods on simulated data	109
3.3	The top 20 non-imaging UK biobank variables that significantly cor- related with the brain age deltas from a linear trajectory	118
3.4	The top 20 non-imaging UK biobank variables that significantly cor- related with the brain age deltas from a polynomial trajectory	118
3.5	The top 20 non-imaging UK biobank variables that significantly cor- related with the test subject brain age deltas from a linear trajectory	120
3.6	The top 20 non-imaging UK biobank variables that significantly corre- lated with the test subject brain age deltas from a polynomial trajectory	120
3.7	The top 20 non-imaging UK biobank variables that significantly cor- related with the brain age deltas from GPR	123
4.1	The correlation values for the unimodal test subject weights	166
4.2	The number of voxels in the binary maps of the unimodal components	167
4.3	The number of voxels in the binary maps of the multimodal components	172
4.4	The correlation values for the multimodal test subject weights	175

List of Abbreviations

AAL	Automated anatomical labelling template
ACC	Accuracy rate
AD	Alzheimer’s disease
ADHD	Attention deficit hyperactive disorder
ADNI	Alzheimer’s Disease Neuroimaging Initiative
ALS	Amyotrophic lateral sclerosis
ApoE	Apolipoprotein
AUC	Area under the curve
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BGN	Basal ganglia network
BMD	Bone mineral density
BMI	Body mass index
BR	Brain region
CCA	Canonical correlation analysis
CNN	Convolutional neural networks
CSF	Cerebrospinal fluid
CV	Cross validation
DMN	Default mode network
dMRI	Diffusion MRI
DNA	Deoxyribonucleic acid
DR	Dual regression
DTI	Diffusion tensor imaging
EIV	Errors-in-variables
EPI	Echo planar imaging
FA	Fractional anisotropy
FC	Functional connectivity
FDR	False discovery rate
FIX	FMRIB’s ICA-based Xnoiseifier
fMRI	Functional magnetic resonance imaging

FMRI	Oxford Centre for Functional MRI of the Brain
FSL	FMRI Software Library
FTD	Frontotemporal dementia
FWHM	Full width at half maximum
GLM	General linear model
GPR	Gaussian process regression
HC	Healthy control
HIV	Human immunodeficiency virus
HRF	Haemodynamic response function
ICA	Independent component analysis
IDLDA	Iterative dual linear discriminant analysis
IDP	Imaging-derived phenotype
JHU	John Hopkins University
KNN	K-nearest neighbour
KRR	Kernel ridge regression
L1	Axial diffusivity
LDA	Linear discriminant analysis
LOO	Leave one out
LOSO	Leave one subject out
MCI	Mild cognitive impairment
MD	Mean diffusivity
MDD	Major depressive disorder
MDL	Minimum description length
MLR	Multiple linear regression
MMSE	Mini mental state examination
MNI	Montreal Neurological Institute
MRI	Magnetic resonance imaging
NB+AC	Network based + All correlations
NB+NPI	Network based + Negative/positive connectivity index
NG	Non-Gaussian
OPDC	Oxford Parkinsons Disease Centre
PCA	Principal components analysis
PD	Parkinson's disease
PE	Parameter estimate
PET	Positron emission tomography
PLS	Partial least squares

Poly3	Polynomial with degree three
RBF	Radial basis function
RD	Radial diffusivity
RF	Reliability filter/Random forests
rfMRI	Resting-state fMRI
ROC	Receiver operator characteristic
ROI	Region of interest
RVR	Relevance vector regression
SNR	Signal-to-noise ratio
SVD	Singular value decomposition
SVM	Support vector machine
SVR	Support vector regression
SZ	Schizophrenia
TBSS	Tract-based spatial statistics
TFCE	Threshold-free cluster enhancement
tfMRI	Task fMRI
UK	United Kingdom
VBM	Voxel-based morphometry

Chapter 1

Introduction

1.1 Biomarkers for neurodegeneration

Neurodegenerative diseases are a group of disorders affecting the central nervous system. They are characterised by the progressive loss of function or structure of neurons. Alzheimer's disease (AD) and Parkinson's disease (PD) are two of the most common diseases which stem from neurodegeneration. They have, as of yet, no cures.

For any possible therapies to come into fruition, biomarkers specific to each disease need to be identified so as to provide a tool for early diagnosis. Early stage diagnosis is essential so that suitable patients can be identified for clinical trials, treatments can be monitored and pathogenesis is better understood [1]. For example, patients with vascular dementia exhibit many of the same early cognitive symptoms as AD, but the two diseases are caused by very different processes [2]. Without a biomarker specific to the diseases, it is possible that vascular dementia patients may be included in clinical trials for an AD therapy, thereby decreasing the likelihood of success.

The current gold standard of diagnosis is through autopsy after patient death, which does little to aid drug development and the better understanding of disease progression. However, various biomarkers are under active investigation. Genetic biomarkers that are currently being investigated include the Apolipoprotein (ApoE)

[3] and α -synuclein protein encoding [4] genes for AD and PD respectively. Biochemical biomarkers such as a decrease in β -amyloid1-42 peptides in the cerebrospinal fluid (CSF- β 42) in AD [5] also exist.

Ageing is the largest risk factor for most neurodegenerative diseases [6] and normal ageing is often mistakenly assumed to be the cause of initial symptoms of AD patients [7]. Therefore, developing models of healthy ageing, understanding how neurodegenerative diseases interact with ageing and identifying subjects that deviate from healthy ageing trajectories can aid in understanding pathogenesis and developing complementary biomarkers. Neuroimaging methods are natural tools for investigating this, since they can be used for obtaining valuable information about structural and functional changes in the brain, which is the primary organ affected by these diseases. Positron emission tomography (PET) of the brain, used to measure metabolic processes and the levels of specific chemicals such as β -amyloid and tau in AD [8], has already been shown to provide a successful biomarker. However, it is expensive, invasive and the type of information that can be gathered in a single scanning session is limited by the radioactive tracer.

Magnetic resonance imaging (MRI), on the other hand, is a non-invasive and cheaper tool that can potentially provide alternative or complementary biomarkers. The power of MRI lies in the fact that it can be used to manipulate the ‘spin’ (an intrinsic quantum property of particles) of hydrogen nuclei, belonging mainly to water and to an extent fat, in different ways. This means that the signal that the nuclei subsequently emit can be made sensitive to different properties of tissue based on how their spins have been manipulated. This gives rise to different MRI modalities, which can be obtained in a single scanning session. The three modalities that we focus on in this thesis are structural MRI, diffusion MRI and functional MRI (fMRI).

1.2 MRI modalities

1.2.1 Structural MRI

The aim of structural MRI scans is to establish the overall anatomy of the brain. A scanning sequence is chosen by the investigator based on the regions and features that they are interested in. A T1-weighted sequence, which in general is the best at differentiating between grey matter, white matter and CSF in a given amount of time, is typically used in studies investigating ageing and AD. In PD on the other hand, a T2* or susceptibility weighted scan is used since they are sensitive to iron content and can show abnormalities in the substantia nigra [9].

One common approach in AD and ageing literature is to segment regions of interest and measure the rate of atrophy of these regions. For example, in AD it has been used to show that the entorhinal cortex, hippocampus and temporal lobe are particularly vulnerable in AD [10]. In ageing meanwhile, it has been used to show that the volumes of the frontal and temporal cortex, putamen, thalamus, and accumbens show the largest changes [11]. Regional segmentation approaches are advantageous in that they provide easily interpretable information about well known structures. However, they rely on the expertise of the user or the quality of an atlas, and do not give information about local changes within the region as well as is suffering from partial volumeing issues. Partial voluming refers to the fact that voxels in MRI scans are of a finite resolution. This means that boundaries of regions and tissue cannot be determined in a completely accurate way, since boundary voxels are a mixture of adjacent tissue.

An alternative approach for measuring grey matter changes is voxel-based morphometry (VBM). To obtain a typical VBM image for a subject, a grey matter segmentation of the structural image is first carried out. The partial volume estimate image of the grey matter, where each voxel has a number between zero and one indicating the fraction of grey matter in the voxel, is then nonlinearly registered to a template (usually a study specific one). A perfect registration would mean that

all useful information would be lost. For this reason, the registered image is subsequently modulated based on how much the local volume had to expand or contract in order to be registered and this is finally smoothed. Therefore, a VBM image is indicative of the local amount of grey matter relative to the template. VBM images belonging to a cohort can be compared on a voxel-wise basis to see how the local amount of grey matter varies across groups or in relation to a variable of interest [12, p.101-103].

VBM has been extensively used in studies of ageing and AD. For example, it has been used to observe local areas of grey matter loss with age in the insula, superior parietal gyri, central sulci, and cingulate sulci [13]. In AD, it has been used to observe grey matter reductions in medial temporal structures [14]. The advantage of VBM is that it is more data-driven compared to regional segmentation and can provide more localised information. However, it can produce erroneous results if there is a systematic misregistration to the template among one of the groups of subjects, which can arise if there is an anatomical abnormality in the group. In this case, group differences will be found that are associated with misregistration as opposed to real grey matter differences. The smoothing step also makes interpretation difficult as the grey matter volume measure for a voxel effectively becomes a weighted average over a local area. This means that, for example, both increased folding and thicker grey matter could result in similar increases in the grey matter volume measure [15].

In addition to partial voluming, structural MRI and all other MRI modalities suffer from motion artefacts. If subject motion is excessive, then the motion artefacts will be strong enough to corrupt the data, or cause spurious group differences if only one subject group is prone to excessive motion. This limits the use of MRI in investigating subjects with certain neurodegenerative diseases, e.g. PD patients that have substantial tremor symptoms.

1.2.2 Diffusion MRI

Diffusion MRI [12, p.43-56] measures the diffusion of water molecules. It is mainly used to probe structural connectivity or properties of tissue microstructure, typically focused on the white matter. In this thesis we focus on the latter. The principle behind diffusion MRI is that the diffusion of water molecules is governed by Brownian motion. Water molecules cannot easily pass through the cell membranes of axons, meaning that the diffusion of water molecules is restricted by bundles of axons. As a result, the water molecules diffuse more in the direction that the axons are oriented than the other directions.

The diffusion tensor imaging (DTI) model [16] is the most widely used method for investigating tissue microstructure. It uses diffusion-weighted scans to calculate the principal direction and magnitude of water diffusion within a given voxel, along with the directions and magnitudes of diffusion along the two axes that are perpendicular to the principal direction. The directions and magnitudes are known as the eigenvectors and eigenvalues respectively. Using these, various metrics can be calculated. The two most common are fractional anisotropy (FA) and mean diffusivity (MD). FA is a value between 0 and 1 that measures how anisotropic the diffusion within a voxel is, while MD is a measure of mean diffusivity within a voxel. Both FA and MD can be considered measures of white matter integrity. For example, axonal loss will lead to less restricted diffusion, meaning that MD will go up. It will also lead to less anisotropic diffusion as water can more easily diffuse in any direction, meaning that FA will go down. Other metrics include axial diffusivity (L1) and radial diffusivity (RD), which represent the amount of diffusion along the principal and perpendicular directions respectively.

One limitation with directly comparing voxel-wise metric maps across subjects arises from the difficulty in aligning white matter tracts, largely due to the fact that they are relatively featureless in T1-weighted structural images. An alternative approach, that avoids the problems of registration based on structural images alone, is tract-based spatial statistics (TBSS) [17]. In TBSS, a skeletonised version of a

template FA image (this can be the mean registered FA image of all the subjects) is created, which only contains voxels at the centre of the major white matter tracts. The subjects' FA images are subsequently projected onto the skeleton by filling each skeleton voxel with the maximum nearby FA value in a subject's image. The maximum nearby FA value is found by searching perpendicular to the voxel's local skeleton structure. The same registration and projection profiles can then be applied to the other DTI metrics as well, and they can be compared voxel-wise across subjects on the skeleton. Therefore, highly accurate alignment of diffusion images is not needed initially, prior to the projection step. The DTI-derived values that we use in this thesis consists of TBSS images.

DTI analyses in AD studies have shown differences in microstructure throughout the white matter [18], including white matter that is part of the limbic-diencephalic network [19]. Ageing studies have also shown widespread differences, with one recent study showing that the strongest association with age was a decrease in MD in the thalamic radiation and association fibres [20].

One limitation with the DTI model is that it only accounts for one fibre bundle within a voxel. This means that areas of crossing fibres are not accurately modelled. Another limitation is the difficulty in the biological interpretation of changes in DTI metrics. For example, although axonal loss can cause an increase in MD and a decrease in FA, similar changes will also be observed if there is a reduction in axon diameters or demyelination instead.

1.2.3 Functional MRI

The aim of functional MRI (fMRI) [21] is to investigate neuronal activity. It is based on the idea that when the activity of neurons increase, their demand for oxygen also increases. To make sure that these neurons are not starved of oxygen, the brain overcompensates and the amount of oxygen that is supplied in response to this extra demand is more than the neurons need. This leads to an increase in blood oxygenation in areas of increased neuronal activity. In fMRI, T2*-weighted

sequences that are sensitive to blood oxygenation levels are used to image the brain every few seconds, and changes in blood oxygenation are taken to represent changes in neuronal activity. The fact that fMRI uses blood oxygenation as a surrogate for neuronal activity is one of its limiting factors.

There are two major variants within fMRI. The first is where the subject performs a predefined task, termed task fMRI, and the aim is to probe how neuronal activity changes in response to the task. The second is where the subject simply lies in the scanner in a so called ‘resting’ state, termed resting-state fMRI (rfMRI), and the aim is to identify regions of the brain that show functional connectivity under rest conditions. The basis of rfMRI is that when resting there is still constant neuronal activity in the brain, with brain regions sharing information with each other. Therefore, by comparing how similar the fluctuations in the signal in different brain regions are, it is possible to gauge how functionally connected they are. In this thesis, we make use of rfMRI data. A detailed review of rfMRI applications and analysis methods is provided in chapter 2.

1.3 Layout of thesis

This thesis comprises three distinct but related projects. Each project has its own dedicated chapter, with each chapter consisting of a literature review, methodology, results and discussion section.

Data from rfMRI has potential to detect neurodegenerative diseases in the early stages of the disease course. However, current approaches can have limitations when used in a biomarker framework. Therefore, in the first project, detailed in chapter 2, we developed a novel algorithm for use in classifying rfMRI data into two groups.

As we pointed to earlier, ageing is strongly linked to neurodegenerative diseases. Identifying subjects that deviate from healthy ageing can potentially single out at-risk subjects. For this reason, detailed in chapter 3, we developed a model that can be used to measure how different the brain age of a subject is compared to their chronological age. In addition, the model estimates trajectories of neuroimaging data

with respect to brain age and, unlike common current neuroimaging approaches to estimating brain age, explicitly models for brain age and is framed such that it can easily be extended to measure interactions of brain ageing with disease progression.

Finally, it is clear from our summary of MRI modalities that each modality provides different but complementary information about the brain. This means that data-driven multimodal analysis methods can potentially be powerful exploratory tools. Hence, in chapter 4, we present a novel method that can be used to decompose multimodal neuroimaging data into spatial maps and associated trajectories. The difference from most common current approaches is that the decomposition is informed on a variable of interest, which in our case was age, and each brain region is associated with only a single trajectory to aid interpretability.

In the final chapter, we suggest some possible future directions for the work in this thesis.

Chapter 2

Iterative dual linear discriminant analysis: A novel classification algorithm for resting-state fMRI

2.1 Introduction

It has been estimated that spontaneous neuronal activity during rest uses over 80% of the brain's energy, suggesting that studying this spontaneous activity may provide important information about the health of the brain [22]. In addition, unlike task fMRI, rfMRI does not require active participation by the subject. This makes it ideal as a tool for investigating brain function in neurodegenerative diseases, where it is likely that subjects may not be suitable for active participation in tasks due to being cognitively impaired.

Task fMRI data has traditionally been analysed using a general linear model (GLM). In this approach, the expected haemodynamic response function (HRF) corresponding to the task is modelled by the experimenter. The extent to which this reference model correlates with the fMRI signal at each voxel is then calculated. Regions showing statistically significant levels of correlation are then considered the regions of activation for the given task [23].

In contrast to the GLM technique, independent component analysis (ICA) is a more exploratory approach. It aims to find a set of statistically independent spatial-maps along with their associated time-courses from the fMRI data. The goal is to express the measured fMRI signal as a linear combination of source signals. These sources can, for example, be different brain networks, motion artefacts or physiological noise [24]. It is the model-free nature of ICA that has made it a prominent tool for analysing rfMRI data. Consider rfMRI data, belonging to a subject s , stored in an $n_v \times n_t$ matrix $\mathbf{X}_s$, where n_v = number of voxels and n_t = number of time-points. Initially, McKeown et al. applied ICA to fMRI data by seeking a decomposition of $\mathbf{X}_s$ such that [25]:

$$\mathbf{X}_s^T = \mathbf{A}_s \mathbf{M}_s^T \quad (2.1)$$

where $\mathbf{M}_s$ is a $n_v \times n_c$ (n_c = number of ICA components) matrix containing spatial maps corresponding to independent source components in each of its columns, and $\mathbf{A}_s$ is a $n_t \times n_c$ matrix that mixes the signals from these components to give the measured signal stored in $\mathbf{X}_s$.

In the above model, the mixing matrix is assumed to be square, i.e. the number of components is equal to the number of observed time-points. The model also does not take into account the fact that observations are noisy, which restricts the ability to carry out statistical significance testing on the component maps. This led Beckmann et al. to present a probabilistic ICA model that includes an additive noise term and allows for non-square mixing of the components [24].

One of the earlier studies that used fMRI and probabilistic ICA was by Greicius et al., in the context of AD [26]. In this study, 13 subjects with mild AD and 13 control subjects performed a sensory-motor task and ICA was used to decompose each subject's dataset into 31 components. A template for the default mode network (DMN) was derived using a separate dataset, and each subject's ICA components were compared to this template using a goodness of fit estimate. A two-sample t -test was finally used to compare the components that best matched the DMN in

the healthy group to those in the AD group. It was found that the AD group had decreased co-activation in the hippocampus and posterior cingulate. In addition, the best goodness-of-fit parameter for each subject was presented as a potential biomarker. Their classification procedure was to produce a scattergram of goodness-of-fit parameters, and use this to select an optimal threshold for classification. This yielded an accuracy of 81%. This study was a powerful proof of concept and showed the potential of rfMRI at discriminating between subjects.

Since the publication of the study by Greicius et al., a different method for obtaining subject-specific spatial maps corresponding to different ICA-derived networks has been proposed, called dual regression (DR) [27]. The standard pipeline involving DR is first concatenating the rfMRI datasets of all the subjects along the temporal dimension to obtain one large voxel by time matrix, and applying ICA to this large matrix, in a process termed ‘group ICA’. The DR step then consists of two stages. The first is to estimate a set of subject-specific time-courses that correspond to the group ICA spatial maps. This is carried out by using the group spatial maps as the design matrix in a GLM like so:

$$\mathbf{X}_s = \mathbf{G}\mathbf{T}_s + \mathbf{E}_1 \quad (2.2)$$

where $\mathbf{G}$ is a $n_v \times n_c$ (n_c = number of ICA components) matrix containing the group spatial maps from group ICA, $\mathbf{T}_s$ is an $n_c \times n_t$ matrix representing the time-courses to be estimated for subject s , and $\mathbf{E}_1$ is the error term. The solution to this, using a least sum of squares solution, is:

$$\hat{\mathbf{T}}_s = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{X}_s \quad (2.3)$$

The second stage of DR is to estimate subject-specific spatial maps by using $\hat{\mathbf{T}}_s$ as the design matrix in a second GLM:

$$\mathbf{X}_s^T = \hat{\mathbf{T}}_s^T \mathbf{M}_s^T + \mathbf{E}_2^T \quad (2.4)$$

where $\mathbf{M}_s$ is a $n_v \times n_c$ matrix representing the subject-specific maps to be estimated. The solution to this is:

$$\widehat{\mathbf{M}}_s = \left(\widehat{\mathbf{T}}_s \widehat{\mathbf{T}}_s^T \right)^{-1} \widehat{\mathbf{T}}_s \mathbf{X}_s^T \quad (2.5)$$

Therefore, column j of $\widehat{\mathbf{M}}_s$ is a subject-specific spatial map corresponding to the brain network stored in column j of $\mathbf{G}$. The subject-specific maps are sometimes referred to as ‘parameter estimates’ (PEs).

If the aim is to compare the subject-specific maps across two different groups, the usual step is to apply some form of univariate (i.e. voxel-wise) statistical testing. Permutation testing is one popular method that determines the statistical significance of how different the voxel values are across the two groups [28]. This is usually carried out in conjunction with threshold-free cluster enhancement (TFCE) [29] in order to increase sensitivity to regions of spatially extended signal.

In the context of PD, one of the first studies to use rfMRI and group ICA to show group differences with healthy controls was by Tessitore et al. [30]. The DMN was the network that was found to have disrupted functional connectivity. However, they did not propose any specific procedure for classification. Szewczyk-Krolikowski et al.’s study on the other hand did propose one [31]. In their study, rfMRI scans were performed on 19 healthy and 19 PD subjects. A basal ganglia network (BGN) template was derived by applying group ICA to a separate group of controls with a dimensionality target of 50, identifying the BGN network from the resulting set of components. The BGN network along with a set of group ICA components they identified to be noise were then used as the design matrix in the first stage of DR to obtain subject specific maps. Noise components are usually included to remove the variance associated with them. These maps were fed into permutation testing with TFCE to obtain significant voxels, with results showing that PD subjects showed reduced functional connectivity with the BGN in a range of areas. Similar to Greicous et al., a post-hoc analysis was proposed as a potential biomarker. The analysis they performed was to use the significant voxels from permutation testing (thresholded at $p < 0.05$) as a mask within which

to average the subject PEs, thus resulting in a single number for each subject. A receiver operator characteristic (ROC) curve was built using these values, and an optimal threshold for discrimination was calculated. This resulted in a classification accuracy of 85% on a validation dataset consisting of 13 PD subjects.

Both the studies by Greicius et al. and Szewczyk-Krolikowski et al. show that ICA-based analysis of rfMRI data can be a powerful exploratory approach, but has non-ideal characteristics for use in a discriminatory framework. The first is the selection of the dimensionality parameter. Greicius et al. chose 31 components while Szewczyk-Krolikowski et al. chose 50 components to decompose their data into. This shows that the choice of this parameter is largely arbitrary, with an increase in the parameter resulting in resting-state networks further splitting, while a decrease results in networks being grouped into larger ones. The user is also required to subjectively choose the component for discrimination, without knowing how well it will discriminate; Greicius et al. chose the DMN and Szewczyk-Krolikowski et al. chose the BGN. It is possible to carry out statistical testing on subject specific maps corresponding to all the different group ICA components to compare the discriminatory power of them. However, if more than one component is shown to be discriminatory, combining them into one classifier becomes complicated. In addition, there is a choice to be made in terms of what components to include in the first stage of DR. Szewczyk-Krolikowski et al. chose to include the BGN plus noise components, whereas others such as Cole et al. [32] have chosen to include all the group ICA components during the DR step (their study was on abstinent smokers). These issues can reduce the robustness and reproducibility of ICA or DR-derived biomarkers.

The challenges with reproducibility of results derived from rfMRI and ICA have been demonstrated in detail by Griffanti et al. [33]. As part of their study, they repeated the post-hoc analyses of Szewczyk-Krolikowski et al. on a dataset consisting of 30 HC and 59 PD subjects. One of their aims was to investigate the influence of the choice of the set of components used in the first stage of DR on the resulting

averaged PEs. They tested a total of 6 different sets of components. As an example, for the right putamen, only 3 out of the 6 component sets resulted in significantly different PE values across the two groups.

In contrast to the hypotheses-driven approach of ICA with DR, network or graph-based analyses have been employed in literature. Typically, when employed in the context of classification, these approaches make use of functional connectivity (FC) matrices as features for a learning algorithm. In order to construct an FC matrix, a set of nodes corresponding to different regions of the brain is first defined. A time-course is then extracted for each node and each node’s time-course is compared using some form of correlation to the time-course of every other node, with the correlation values stored in a node $\times$ node connectivity matrix. Each element in this matrix is referred to as an edge for a pair of nodes and the whole matrix can then be considered as a ‘network’. Various strategies for defining nodes (also referred to as parcellation), extracting time-courses and correlating time-courses exist [34, p.81-106].

Chen et al. [35] used functional connectivity matrices in order to classify 20 AD, 15 Mild Cognitive Impairment (MCI) and 20 HC subjects. They initially mapped their rfMRI data to 116 anatomically defined regions of interest (ROIs) using the Talairach template (details of the specific atlas used were not given) and extracted an average time-course for each ROI. The time-course for each ROI was then correlated with the time-course for every other ROI using the Pearson correlation coefficient (r), resulting in a total of 6670 r values for each subject. Firstly, these r values for each pair of ROIs were compared between the AD and non-AD subjects using a two-sample Wilcoxon rank-sum test. The r values corresponding to pairs with the n_- most negative Wilcoxon z values were averaged and referred to as the reduced connectivity index. Similarly the r values corresponding to pairs with the n_+ most positive z values were averaged and referred to as the increased connectivity index. The two indices for each subject were then fed into linear discriminant analysis (LDA). Various n_- and n_+ were tested. They then repeated the same process, but

this time discriminating between MCI and HC subjects. The classification accuracies were reported in terms of area under the curve (AUC) of the ROC curve. They were 0.87 and 0.95, using the best n_- and n_+ values, for AD vs non-AD and MCI vs HC respectively.

The use of an atlas to split the brain into ROIs can be advantageous in that information from the whole brain is incorporated. Moreover, the nodes become interpretable in terms of familiar structures, meaning that the outputs of classifiers will be more interpretable. One issue however, is that many different atlases exist, there is no consensus on what is the ‘best’ one and a study by Bohland et al. demonstrated that the existing atlases are mutually inconsistent [36]. In addition, as Thirion et al. point out in their paper [37], it is possible for any given atlas to not fit the data well. For example, an atlas may not represent the population being investigated well, which could very well be the case in studies investigating neurodegeneration. Also, an anatomical-based atlas may not represent functional regions very well. It is common for atlases to group the anterior part of the brain into large regions when it is actually possible for these regions to be split into smaller ones with more specific functional roles. Another point relating to this is that atlas-based parcellation assumes hard boundaries between each functional region, which is unlikely given that hierarchical layers of functional organisation are known to exist in the brain [38]. Overall these limitations can result in discriminatory information being lost, hence making atlas-based parcellation non-ideal for use in classifiers.

Others have chosen to construct manual ROIs instead of atlas-based ones. Craddock et al. used this approach to discriminate between 20 subjects with major depressive disorder (MDD) and 20 HC subjects [39]. They first manually constructed 15 ROIs that they felt were most relevant to MDD. Afterwards, they chose the first singular vector from the singular value decomposition (SVD) of all the time-courses within an ROI as the summary time-course for that ROI. They then carried out a pairwise correlation of these time-courses and converted the pairwise r values to z scores using the Fisher transform. Finally, these were fed into a support vector

machine (SVM) classifier. Similar to Chen et al., they chose to perform some form of feature selection before applying SVM. The feature selection that performed best was called a ‘reliability filter’ (RF). In an RF, B bootstrap datasets are generated from the training dataset by sampling N subjects with replacement. An SVM is then trained on 62.3 % of the samples within each bootstrap dataset. For a given feature, i.e. ROI pair, its trained weights from all the bootstrap sets are collected and ordered. If the 2.5 to 97.5 percentile range of the weight values do not include zero, then the feature is included in the feature set to feed into the final SVM classifier. Craddock et al. used a leave-one-out (LOO) cross validation to evaluate the performance of RF + SVM and on average 11 out of the possible 105 ROI pairs survived the RF feature selection. The final classification accuracy was 95%. They also tried a univariate t -test using a threshold of $p < 0.05$ to select features before feeding the connectivity matrices into SVM. On average this method still selected 11 features but the final classification accuracy was 82.5%.

The use of manual ROIs makes it more likely that the regions are suitable for the population being studied. However, drawing manual ROIs requires user expertise, which may not always be available. Also, as already pointed out, anatomical ROIs may not represent functional regions very well. Moreover, only 15 regions of the brain were represented by Craddock et al., and it is common for studies using manual ROIs to not cover the whole brain. Focusing on a few ROIs is beneficial in a hypothesis testing framework but for discrimination it has the obvious limitation of ignoring signal that can be discriminatory.

There have been attempts at building network-based classifiers using more data-driven approaches to parcellation and without using anatomical information. Arbashirani et al. used ICA to construct functional connectivity matrices and discriminate between schizophrenia (SZ) and HC subjects [40]. Before applying ICA, they used principle component analysis (PCA) to reduce the dimensionality of each subject’s fMRI data to 80, and then applied group ICA to this reduced dataset with an automatic dimensionality estimate based on the minimum description length (MDL)

criterion. Subject-wise time-courses were obtained using a back-reconstruction approach, and r values calculated for all pairs of time-courses. In calculating this however, they allowed for any possible lag by shifting one of the time-courses in a pair from -3 to 3 seconds and saving the maximum r value. Their automatic dimensionality estimate gave 20 components and 9 of these were chosen to be non-artifactual, resulting in a total of 36 r values to use as features. However, since their SZ subjects were on medication, they chose to remove motor network and temporal-parietal related features, resulting in 27 r values. A variety of linear and non-linear classification algorithms were used, including LDA, SVM, naive Bayes, random forests, k-nearest neighbours and neural networks. The parameters for each algorithm were optimised using LOO cross validation on the training set, consisting of 16 HC and 16 SZ subjects. The performance of each algorithm was then validated on a test set consisting of 12 HC and 12 SZ subjects. The best classification accuracy was obtained by non-linear methods; KNN, kernel SVM, random forests and neural networks all obtained 96%. On the other hand, linear SVM, naive Bayes and LDA resulted in 83%, 79% and 71% respectively.

The strength in this approach is that their nodes are more likely to be representative of functional regions. The parcellation will also fit their data better. However, as already pointed to, the main disadvantage of using ICA-based discrimination is that relevant components have to be manually and subjectively chosen. In fact, Arbashirani et al. investigated how the classification accuracy would differ if all 20 components were used to derive r values. All the algorithms ended up resulting in different accuracies when both the sensitivity and specificity are taken into account. The largest change in overall accuracy was 21%, obtained using neural networks.

Attempts have been made at using the voxels themselves as nodes. Dey et al. used voxel-wise functional connectivity to discriminate between HC and attention deficit hyperactive disorder (ADHD) subjects [41]. For each subject, they calculated r values for every pair of voxels and stored this in voxel $\times$ voxel node matrix. This matrix was thresholded and binarised, resulting in what is referred to as an

adjacency matrix, where row i of the matrix contains a 1 for every voxel functionally connected to voxel i . For each voxel, they summarised its activity patterns using several graph metrics, and used this as the feature vector for each subject. Therefore, the dimensionality of the feature vector is equal to the number of voxels times the number of summary metrics. The subject by features matrix was then fed into PCA, and the first 100 principle components were kept. The features in this reduced space were finally fed into LDA. They trained their algorithm on 468 HC and 308 ADHD subjects, and validated using a separate set of 93 HC and 78 ADHD subjects. They tried a combination of different graph-summary metrics, with a maximum classification accuracy of 69.6%, obtained using a 3-cycle map metric. The 3-cycle map metric for a node counts how many unique paths consisting of 3 node-steps are possible, where the starting and end points are the node itself. For example, if only three nodes exist in a graph and they are all connected to each other, then each node would have a 3-cycle map of 1.

Although parcellation was not explicitly carried out, they chose voxels to feed into the PCA-LDA algorithm using a method based on parcellation. This first step in this method was to divide the brain volume into small cube-shaped regions, with a typical dimension of $5 \times 5 \times 5$ voxels. They applied their PCA-LDA algorithm to a feature derived from connectivity matrices of a random subset of regions. The feature they chose is the ‘degree’, i.e. number of other nodes a given node is connected to. They repeated this for 500 random subsets and calculated the classification accuracy for each subset. Random subsets that resulted in the top 10% of the classification accuracies were collected, and the percentage occurrence of each region in this collection was calculated. Voxels in regions that occurred 60% of the time were selected as the ‘useful’ voxels to include in their final analysis. One issue with this procedure is the vast number of parameters that need to be optimised, such as the probability of inclusion of a region into a random subset, the percentage occurrence at which to threshold and the size of each cube-shaped region. It is theoretically possible to make use of full voxel by voxel connectivity matrices without prior voxel

selection. However, this would be computationally expensive. For example, just one subject’s dataset containing 10^5 voxels could require approximately 40 GB of memory.

The studies described in this section were chosen so as to be representative of the different approaches researchers have chosen to take. A more comprehensive (although still not complete) list of studies is given in table 1. In summary, current techniques for deriving biomarkers from rfMRI data require some arbitrary or subjective choices to be made, and while tools such as atlases exist, they can be unsuitable for rfMRI. These limitations can both affect the reproducibility and effectiveness of these biomarkers. For this reason, we have developed a novel classification algorithm, which we call iterative dual linear discriminant analysis (IDLDA), for use with rfMRI data. It does not require any parcellation or decomposition and it does not require the user to make any choices regarding what regions or resting-state networks to use. Instead, the whole voxel by time data (after cleaning, motion correction etc.) is fed in. It employs a supervised learning approach to discriminate between two groups of subjects and outputs spatial maps highlighting voxels that may have altered functional connectivity between the two groups. Therefore, IDLDA can potentially be utilised both as a classification tool and as an exploratory tool to find group differences that exist in a cohort. In this study, we apply IDLDA to simulated, Parkinson’s disease, amyotrophic lateral sclerosis and anaesthesia datasets. Its performance is evaluated using cross validation and by comparison to existing methods.

Table 2.1: Summary of studies focusing on classifying subjects based on rfMRI data. AD = Alzheimer’s disease, ADHD = Attention deficit hyperactivity disorder, ASD = Autism spectrum disorder, AUC = Area under curve, BGN = Basal ganglia network, BP = Back projection, CNN = Convolutional neural networks, DAE = Deep auto-encoder, DMN = Default mode network, DR = Dual regression, EP = Epilepsy, fALFF = fractional Amplitude of low-frequency fluctuation, FCN = Functional connectivity network, FTD = Frontotemporal dementia, GICA = Group ICA, GM = Grassman Manifold, GP = Gaussian process, HC = Healthy controls, HMM = Hidden Markov model, ICA = Independent component analysis, kM = k-means, Kt = Kendall tau, LDA = Linear discriminant analysis, LLE = Locally-linear embedding, LOO = Leave-one-out, LR = Logistic regression, MAR = Multivariate autoregression, MCC = Mean clustering coefficient, MCI = Mild cognitive disorder, MDD = Major depressive disorder, MKL = Multi-kernel learning, mRMR = minimum Redundancy maximum relevance, NN = Neural network, PCA = Principle component analysis, PCE = Prenatal cocaine exposure, PD = Parkinson’s disease, PET = Positron emission tomography, RCE = Recursive cluster elimination, RF = Reliability filter, RH = Regional homogeneity, ROI = Manual regions of interest, RS = Random shuffle, SICA = Subject ICA, SN = Saliency network, SR = Sparse regression, SVM = Support vector machine, SZ = Schizophrenia, VBA = Voxel-based analysis, WRST = Wilcoxon rank-sum test.

Study	Year	Subjects	Approach	Features	Classifier	Validation	Accuracy
de Vos et al. [42]	2018	77 AD, 173 HC	Various; best results for FCN + ICA	Best results for partial correlation and partial correlation dynamics	LR	10-fold	AUC: 0.82 for partial correlation; 0.84 for dynamics
Koch et al. [43]	2012	15 AD, 21 HC	VBA + ICA; FCN + ROI	DMN maps from SICA; Within-DMN node correlation	Not specified	LOO	97%
Greicius et al. [26]	2004	13 AD, 13 HC	VBA + ICA	DMN maps from SICA	Goodness of fit	None	81%
Zhou et al. [44]	2010	12 AD, 12 FTD, 12 HC	VBA + ICA	Scores derived from BGN and SN SICA maps	LDA	LOO	92% for multi-class
Fan et al. [45]	2011	31 SZ, 31 HC	VBA + ICA	Maps from GICA + BP	GM + SVM	LOO	87%
Du et al. [46]	2012	28 SZ, 28 HC	VBA + ICA	Maps from GICA + BP	KPCA + LDA	LOO	93%
S-Krolikowski et al. [31]	2014	19 PD, 19 HC	VBA + ICA	BGN maps from GICA + DR	t-test	13 PD	85%
Rahim et al. [47]	2015	21 AD, 73 MCI	VBA + Atlas + PET Prior	ROI to voxel correlation maps	LR	5-fold + RS	85%
Zhu et al. [48]	2008	9 ADHD, 11 HC	RH	RH maps	PCA + LDA	LOO	85%
Craddock et al. [39]	2009	20 MDD, 20 HC	FCN + ROI	Correlation	RF + SVM	LOO	95%
Deshpande et al. [49]	2010	30 PCE, 26 HC	FCN + ROI	Correlation + Effective connectivity	RCE + SVM	10-fold	90%

Anderson et al. [50]	2011	40 ASD, 40 HC	FCN + ROI	Correlation	t-test + Clustering	LOO; 8 ASD, 13 HC holdout	79% for LOO; 71% for holdout
Fair et al. [51]	2012	112 ADHD-C, 80 ADHD-I, 455 HC	FCN + ROI	Correlation	SVM	LOO	77% for ADHD-C vs HC; 81% for ADHD-I vs HC; 69% for multi-class
Arbashirani et al. [40]	2013	16 SZ, 16 SZ	FCN + ICA	Correlation	Variety	12 SZ, 12 HC	96%
Murdaugh et al. [52]	2012	13 ASD, 14 HC	FCN + Atlas; FCN + ROI	Correlation	LR	LOO	96% for FCN + ROI; 78% for FCN + Atlas
Cheng et al. [53]	2012	98 ADHD, 141 HC	FCN + Atlas + RH	Correlation + partial correlation + RH maps + fALFF + spatial correlations	PCA + SVM	LOO	76%
Supekar et al. [54]	2008	21 AD, 18 HC	FCN + Atlas	Graph metrics	MCC	None	75%
Shen et al. [55]	2010	32 SZ, 20 HC	FCN + Atlas	Correlation	Kt + LLE + KM	LOO	92%
Chen et al. [35]	2011	20 AD, 15 MCI, 20 HC	FCN + Atlas	Correlation	WRST + LDA	LOO	AUC: 0.87 for AD vs. Non-AD; 0.95 for MCI vs HC
Bassett et al. [56]	2012	29 SZ, 29 HC	FCN + Atlas	Graph metrics	SVM	2-fold + RS	75%
Liu et al. [57]	2012	24 SZ, 22 HC	FCN + Atlas	Correlation	PCA + SVM	LOO	80%
Lord et al. [58]	2012	22 MDD, 22 HC	FCN + Atlas	Graph metrics	mRMR + SVM	2-fold + RS	99%
Richiardi et al. [59]	2012	22 MS, 14 HC	FCN + Atlas	Correlation	RF	LOO	82%
Tang et al. [60]	2012	22 SZ, 22 HC	FCN + Atlas	Correlation	Kt + SVM	LOO	93%
Venkataraman et al. [61]	2012	18 SZ, 18 HC	FCN + Atlas	Correlation	RF	None	75%
Zeng et al. [62]	2012	24 MDD, 29 HC	FCN + Atlas	Correlation	SVM	LOO	94%
Zhang et al. [63]	2012	100 EP, 80 HC	FCN + Atlas	Graph metrics	SR + SVM	LOO	84%
Wee et al. [64]	2013	12 MCI, 25 HC	FCN + Atlas	ROI mean time-courses	MAR + SVM	LOO	92%

Yu et al. [65]	2013	24 SZ, 25 HC	FCN + Atlas	Correlation	PCA + SVM	LOO	62%
Jie et al. [66]	2014	12 MCI, 25 HC	FCN + Atlas	Graph metrics	MKL	LOO	92%
Challis et al. [67]	2015	17 AD, 40 MCI, 29 HC	FCN + Atlas	Covariance	Kt + GP + LR	10 AD, 10 MCI, 10 HC	75% for MCI vs HC; 97% for MCI vs AD
Dodero et al. [68]	2015	42 ASD, 37 HC	FCN + Atlas	Correlation	GM + SVM	LOO	63%
Suk et al. [69]	2015	12 MCI, 25 HC	FCN + Atlas	ROI mean time-courses	DAE + HMM	LOO	81%
Zhang et al. [70]	2015	36 MCI, 25 HC	FCN + Atlas	Correlation	t-test + LR	LOO	88%
Chen et al. [71]	2016	112 ASD, 128 HC	FCN + Atlas	Correlation in Slow-4 and Slow-5 bands	SVM	LOO	79%
Kim et al. [72]	2016	50 SZ, 50 HC	FCN + Atlas	Correlation	NN	5-fold	86%
Meszlnyi et al. [73]	2017	25 MCI, 24 HC	FCN + Atlas	Dynamic time warping	CNN	7-fold	72%, AUC: 0.75
Dey et al. [41]	2012	308 ADHD, 468 HC	FCN (Voxel-wise)	Graph metrics	PCA + LDA	78 ADHD, 93 HC	70%

2.2 Methodology

2.2.1 Linear discriminant analysis

The algorithm we propose is based on LDA, so we will give a brief overview of standard LDA first. LDA is a supervised learning technique, which when applied to two classes, aims to project a vector of features onto a line, such that the projected between-group mean difference divided by the sum of the projected within-group variances is maximised [74]. Consider n_1 and n_2 training samples belonging to class 1 and class 2 respectively, with each sample having a set of features $\mathbf{x}_s$, $s = 1, \dots, (n_1 + n_2)$. LDA seeks to maximise the following objective function with respect to $\mathbf{w}$:

$$J_{\text{LDA}} = \frac{(\mathbf{w}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^2}{\sum_{s \in C_1} (\mathbf{w}^T (\mathbf{x}_s - \boldsymbol{\mu}_1))^2 + \sum_{s \in C_2} (\mathbf{w}^T (\mathbf{x}_s - \boldsymbol{\mu}_2))^2} \quad (2.6)$$

where $\boldsymbol{\mu}_p$ is the set of mean features in class C_p ($C_1 = \text{class 1}$ and $C_2 = \text{class 2}$).

This can be rewritten as:

$$J_{\text{LDA}} = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (2.7)$$

Here, $\mathbf{S}_B = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$ and $\mathbf{S}_W$ is the sum of each class's covariance matrix. This has an analytical solution:

$$\mathbf{w}^* = \mathbf{S}_W^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (2.8)$$

For discrimination, $\mathbf{w}^*$ is used to project a test sample down to a scalar value and a number of different strategies can then be employed for classifying this sample based on the scalar value.

2.2.2 Iterative dual linear discriminant analysis

Our algorithm is as follows. Consider rfMRI data that has been de-meaned and variance normalised across time for each voxel, belonging to a subject s , and stored in an $n_v \times n_t$ matrix $\mathbf{X}_s$, where $n_v = \text{number of voxels}$ and $n_t = \text{number of time-points}$.

The similarity between two rfMRI time-courses belonging to two different brain regions can be assumed to be a measure of how connected the two regions are. This is due to the fact that if two regions, e.g. region A and region B, are communicating with each other, then a certain amount of region A's intrinsic activity will be encoded in region B's time-course and vice versa.

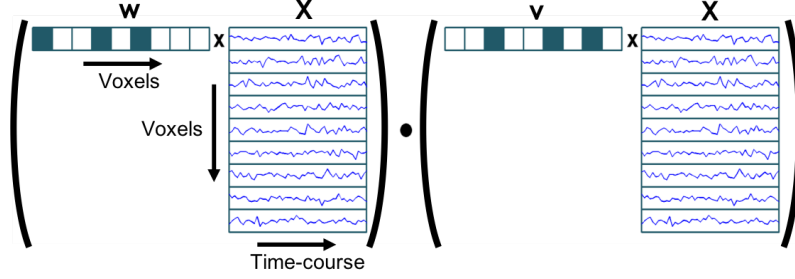


Figure 2.1: A visual representation of equation 2.9

If we define vectors $\mathbf{w}$ and $\mathbf{v}$, both of size $n_v \times 1$ and magnitude 1, as a vector of weights corresponding to the spatial maps of regions A and B, then an expression for this similarity metric for subject s can be:

$$y_s = (\mathbf{w}^T \mathbf{X}_s) (\mathbf{v}^T \mathbf{X}_s)^T = \mathbf{w}^T \mathbf{X}_s \mathbf{X}_s^T \mathbf{v} \quad (2.9)$$

Here $\mathbf{w}^T \mathbf{X}_s$ is an aggregate time-course corresponding to $\mathbf{w}$ and $\mathbf{v}^T \mathbf{X}_s$ is an aggregate time-course corresponding to $\mathbf{v}$. If $\mathbf{w}$ and $\mathbf{v}$ are made to have unit norm, the dot product of the time-courses is similar to a measure of correlation. This is shown visually in figure 2.1.

Now consider an rfMRI dataset for two groups of subjects, with n_1 subjects in group 1 and n_2 subjects in group 2. Analogously to the linear discriminant analysis (LDA) objective function (equation 2.6), we can seek to maximise the following objective function with respect to $\mathbf{w}$ and $\mathbf{v}$:

$$J_{\text{IDLDA}} = \frac{\left(\frac{1}{n_1} \sum_{s \in C_1} y_s - \frac{1}{n_2} \sum_{s \in C_2} y_s \right)^2}{\sum_{s \in C_1} \left(y_s - \frac{1}{n_1} \sum_{s \in C_1} y_s \right)^2 + \sum_{s \in C_2} \left(y_s - \frac{1}{n_2} \sum_{s \in C_2} y_s \right)^2} \quad (2.10)$$

Therefore, we also seek to maximise the between-group mean difference of y_s divided

by the sum of the projected within group variances of y_s . The complication in using standard LDA to maximise this function is that we have a time-course associated with each voxel and we need to maximise two vectors instead of one. However, if we fix $\mathbf{v}$, then:

$$\frac{1}{n_p} \sum_{s \in C_p} y_s = \mathbf{w}^T \frac{1}{n_p} \sum_{s \in C_p} \mathbf{X}_s \mathbf{X}_s^T \mathbf{v} = \mathbf{w}^T \frac{1}{n_p} \sum_{s \in C_p} \mathbf{x}_s = \mathbf{w}^T \boldsymbol{\mu}_p \quad (2.11)$$

where we have defined $\mathbf{x}_s = \mathbf{X}_s \mathbf{X}_s^T \mathbf{v}$ and $\boldsymbol{\mu}_p = \frac{1}{n_p} \sum_{s \in C_p} \mathbf{x}_s$. This means equation 2.10 reduces to equation 2.6 and we can use LDA to optimise with respect to $\mathbf{w}$, having held $\mathbf{v}$ constant. Importantly, equation 2.10 is invariant to swapping $\mathbf{w}$ and $\mathbf{v}$. Therefore, to maximise equation 2.10, we can use an iterative approach as follows.

We initialise $\mathbf{v} = \mathbf{m}_0$, which can be a random vector, and then use LDA to find $\mathbf{w}$ (defined as $\mathbf{m}_1$ in the algorithm description below). This value of $\mathbf{w}$ is then fixed and $\mathbf{v}$ is found using LDA in the next iteration ($\mathbf{v} = \mathbf{m}_2$), and so on. Due to the symmetry of $\mathbf{w}$ and $\mathbf{v}$, each iteration proceeds in the same way - starting with a fixed vector, $\mathbf{m}_{i-1}$, and then solving for the other, $\mathbf{m}_i$. This is repeated until the algorithm has converged. See the full algorithm description below. This is also shown graphically in figure 2.2.

Algorithm 1 IDLDA (Unregularised)

- 1: **for** ($i = 1$ to k) **do**
 - 2: $\mathbf{x}_s = \mathbf{X}_s \mathbf{X}_s^T \mathbf{m}_{i-1}$ and $\mathbf{S}_W = \sum_p \sum_{s \in C_p} (\mathbf{x}_s - \boldsymbol{\mu}_p)(\mathbf{x}_s - \boldsymbol{\mu}_p)^T$
 - 3: $\mathbf{m}_i = \mathbf{S}_W^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$
 - 4: $\mathbf{m}_i = \frac{1}{\|\mathbf{m}_i\|} \mathbf{m}_i$
 - 5: **end for**
 - 6: $\mathbf{w} = \mathbf{m}_k$ and $\mathbf{v} = \mathbf{m}_{k-1}$
-

After running IDLDA on a training set of subjects, the obtained $\mathbf{w}$ and $\mathbf{v}$ vectors can then be used to reduce a test subject's data down to a scalar value using equation 2.9. As in standard LDA, any preferred strategy can be chosen for classifying the test subject based on these scalar values. One simple strategy is to classify the test subject to the group with the closest mean.

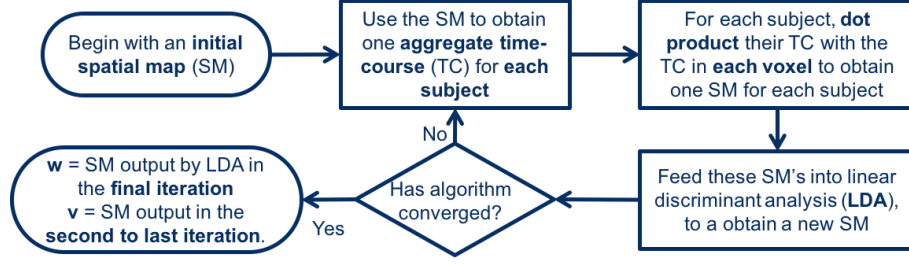


Figure 2.2: Flowchart of the proposed IDLDA algorithm

Two properties of IDLDA are that the sign of $\mathbf{w}$ or $\mathbf{v}$ are arbitrary and the ordering of $\mathbf{w}$ and $\mathbf{v}$ will not always be the same. The former is due to the fact that the projections of samples will be just as discriminatory if $y_s \mapsto -y_s$, while the latter is due to the iterative framework, i.e. we can just as easily set $\mathbf{m}_{k-1} = \mathbf{w}$ to obtain $\mathbf{m}_k = \mathbf{v}$ and vice versa after the algorithm has converged. Both of these properties depend on the initialisation, but will have no bearing on discriminatory power.

2.2.3 Adapting IDLDA to high dimensional and low sample size data

When there are many more voxels associated with each sample than there are samples themselves (as is typical in MRI data), $\mathbf{S}_W$ contains a substantial null space and its inverse is ill-posed. In this case, the iterative process we use to maximise equation 2.10 will fail. To overcome this, a constant λ term can be added to the diagonal elements of $\mathbf{S}_W$ for regularisation. Thus, the third line of the algorithm will change to that shown in algorithm 2.

Algorithm 2 IDLDA (Regularised)

- 1: **for** ($i = 1$ to k) **do**
 - 2: $\mathbf{x}_s = \mathbf{X}_s \mathbf{X}_s^T \mathbf{m}_{i-1}$ and $\mathbf{S}_W = \sum_p \sum_{s \in C_p} (\mathbf{x}_s - \boldsymbol{\mu}_p)(\mathbf{x}_s - \boldsymbol{\mu}_p)^T$
 - 3: $\mathbf{m}_i = (\mathbf{S}_W + \lambda \mathbf{I}_{n_v})^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$
 - 4: $\mathbf{m}_i = \frac{1}{\|\mathbf{m}_i\|} \mathbf{m}_i$
 - 5: **end for**
 - 6: $\mathbf{w} = \mathbf{m}_k$ and $\mathbf{v} = \mathbf{m}_{k-1}$
-

Here $\mathbf{I}$ is the identity matrix, of the same size as $\mathbf{S}_W$. The objective function

will also change to:

$$J_{\text{IDLDA}} = \frac{\left(\frac{1}{n_1} \sum_{s \in C_1} y_s - \frac{1}{n_2} \sum_{s \in C_2} y_s\right)^2}{\sum_{s \in C_1} \left(y_s - \frac{1}{n_1} \sum_{s \in C_1} y_s\right)^2 + \sum_{s \in C_2} \left(y_s - \frac{1}{n_2} \sum_{s \in C_2} y_s\right)^2 + \lambda} \quad (2.12)$$

For this study, we assumed that the algorithm had converged after iteration i if $|1 - \frac{J_{i-1}}{J_i}| \leq 0.005$.

Another issue to contend with is that calculating the LDA output $\mathbf{m}_i$ at each iteration is not trivial due to the fact that $\mathbf{S}_W$ is of size $n_v \times n_v$. Since the number of voxels in rfMRI data is typically of the order $10^5 - 10^6$, storing $\mathbf{S}_W$ is impractical. For this reason, we have developed an efficient calculation, based on a singular value decomposition of the constituents of the $\mathbf{S}_W$ matrix, as follows.

The equation that we want to solve for is:

$$\mathbf{m} = (\mathbf{S}_W + \lambda \mathbf{I}_{n_v})^{-1} \Delta \boldsymbol{\mu} \quad (2.13)$$

where we have defined $\Delta \boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. We can rewrite equation 2.13 as:

$$\Delta \boldsymbol{\mu} = (\mathbf{S}_W + \lambda \mathbf{I}_{n_v}) \mathbf{m} \quad (2.14)$$

For the moment, let $\mathbf{m}$ have the following form:

$$\mathbf{m} = \mathbf{a} + \frac{1}{\lambda} \Delta \boldsymbol{\mu} \quad (2.15)$$

If we substitute equation 2.15 into equation 2.14, we obtain the following:

$$\mathbf{S}_W \mathbf{a} + \lambda \mathbf{I}_{n_v} \mathbf{a} + \frac{1}{\lambda} \mathbf{S}_W \Delta \boldsymbol{\mu} = 0 \quad (2.16)$$

The $\frac{1}{\lambda} \mathbf{S}_W \Delta \boldsymbol{\mu}$ and $\mathbf{S}_W \mathbf{a}$ terms have to generate a vector that is in the space spanned by $\mathbf{S}_W$. This means that for the equation to hold true, $\mathbf{a}$ cannot contain a component that is not in the span of $\mathbf{S}_W$. If it did, the $\lambda \mathbf{I}_{n_v} \mathbf{a}$ term would generate a vector that

contains a component that is not in the span of $\mathbf{S}_W$, and neither of the other two terms would be able to produce a vector that could appropriately combine with this in order to keep the equation equalling to zero. If we define $\mathbf{B}$ as a matrix whose columns contain the eigenvectors that span the space of $\mathbf{S}_W$, then we can say that $\mathbf{a}$ can be obtained by a projection of some other vector $\mathbf{c}$ onto this space, i.e.:

$$\mathbf{a} = \mathbf{B}\mathbf{c} \quad (2.17)$$

In equation 2.14, we can further see that the $\mathbf{S}_W\mathbf{m}$ will produce a vector in the space spanned by $\mathbf{S}_W$, while the $\lambda\mathbf{I}_{n_v}\mathbf{m}$ term will contain any portion of $\Delta\boldsymbol{\mu}$ that is not in this space. As before, we can explicitly state the portion of $\Delta\boldsymbol{\mu}$ that is in the span of $\mathbf{S}_W$ by formulating $\Delta\boldsymbol{\mu}$ as a projection of some other vector $\mathbf{d}$ onto the space spanned by $\mathbf{B}$. This time, we also have to add a vector $\mathbf{e}$ to account for the portion that is not in the span of $\mathbf{S}_W$. Therefore:

$$\Delta\boldsymbol{\mu} = \mathbf{B}\mathbf{d} + \mathbf{e} \quad (2.18)$$

Substituting equations 2.17 and 2.18 into equation 2.16, we obtain:

$$\mathbf{S}_W\mathbf{B}\mathbf{c} + \lambda\mathbf{B}\mathbf{c} + \frac{1}{\lambda}\mathbf{S}_W\mathbf{B}\mathbf{d} + \frac{1}{\lambda}\mathbf{S}_W\mathbf{e} = 0 \quad (2.19)$$

We subsequently take account of the fact that $\frac{1}{\lambda}\mathbf{S}_W\mathbf{e}$ will equal zero, and then rearrange and multiply both sides by $\mathbf{B}^T$:

$$(\mathbf{B}^T\mathbf{S}_W\mathbf{B} + \lambda\mathbf{B}^T\mathbf{B})\mathbf{c} = -\frac{1}{\lambda}\mathbf{B}^T\mathbf{S}_W\mathbf{B}\mathbf{d} \quad (2.20)$$

Hence:

$$\mathbf{c} = -\frac{1}{\lambda}(\mathbf{B}^T\mathbf{S}_W\mathbf{B} + \lambda\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{S}_W\mathbf{B}\mathbf{d} \quad (2.21)$$

The vector $\mathbf{d}$ can similarly be calculated by multiplying both sides of equation 2.18

by $\mathbf{B}^T$ and solving for $\mathbf{d}$:

$$\mathbf{d} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \Delta \boldsymbol{\mu} \quad (2.22)$$

By substituting equation 2.22 into equation 2.21, and equation 2.21 into equation 2.17, and finally equation 2.17 into the equation 2.15, we obtain the following solution for $\mathbf{m}$:

$$\mathbf{m} = \frac{1}{\lambda} (\Delta \boldsymbol{\mu} - \mathbf{B}(\mathbf{B}^T \mathbf{S}_W \mathbf{B} + \lambda \mathbf{B}^T \mathbf{B})^{-1} (\mathbf{B}^T \mathbf{S}_W \mathbf{B}) (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \Delta \boldsymbol{\mu}) \quad (2.23)$$

The number of eigenvectors required to span the space of $\mathbf{S}_W$ is typically on the order of number of subjects, meaning that $\mathbf{B}$ can be stored efficiently. It can also be calculated efficiently like so.

We first define an $n_v \times n_p$ matrix that contains the demeaned intermediate feature vectors (the intermediate feature vectors are $\mathbf{x}_s = \mathbf{X}_s \mathbf{X}_s^T \mathbf{m}_{i-1}$ in the algorithm description) of the subjects in group p :

$$\mathbf{A}_p = [(\mathbf{x}_{1 \in C_p} - \boldsymbol{\mu}_p) \dots (\mathbf{x}_{n_p \in C_p} - \boldsymbol{\mu}_p)] \quad (2.24)$$

The matrix $\mathbf{S}_W$ is then:

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 \quad (2.25)$$

where $\mathbf{S}_p = \mathbf{A}_p \mathbf{A}_p^T$. We can now consider the SVD of $\mathbf{A}_p$:

$$\mathbf{A}_p = \mathbf{U}_p \mathbf{D}_p \mathbf{V}_p^T \quad (2.26)$$

The matrix $\mathbf{U}_p$ contains the eigenvectors of $\mathbf{S}_p$ since:

$$\mathbf{S}_p = \mathbf{A}_p \mathbf{A}_p^T = \mathbf{U}_p \mathbf{M}_p \mathbf{U}_p^T \quad (2.27)$$

where $\mathbf{M}_p = \mathbf{D}_p \mathbf{D}_p^T$. Therefore, $\mathbf{U}_1$ and $\mathbf{U}_2$ are the matrices that we need. We cannot construct $\mathbf{A}_p \mathbf{A}_p^T$ since it is of size $n_v \times n_v$. Instead, to calculate $\mathbf{U}_1$ and $\mathbf{U}_2$, we make note of the fact that $\mathbf{V}_p$ and $\mathbf{M}_p$ contain the eigenvectors and eigenvalues

of $\mathbf{A}_p^T \mathbf{A}_p$, which is of size $n_p \times n_p$:

$$\mathbf{A}_p^T \mathbf{A}_p = \mathbf{V}_p \mathbf{M}_p \mathbf{V}_p^T \quad (2.28)$$

This means that the eigenvectors of $\mathbf{S}_p$ can be obtained by rearranging equation 2.26:

$$\tilde{\mathbf{U}}_p = \mathbf{A}_p \mathbf{V}_p \mathbf{M}_p^{-\frac{1}{2}} \quad (2.29)$$

Here, when an element on the diagonal of $\mathbf{M}_p$ is very small, we set the corresponding element of $\mathbf{M}_p^{-\frac{1}{2}}$ to zero. We can now obtain $\mathbf{B}$ by concatenating $\tilde{\mathbf{U}}_1$ and $\tilde{\mathbf{U}}_2$ into a matrix:

$$\mathbf{B} = [\tilde{\mathbf{U}}_1, \tilde{\mathbf{U}}_2] \quad (2.30)$$

Finally, the $\mathbf{B}^T \mathbf{S}_W \mathbf{B}$ terms in equation 2.23 can be calculated efficiently by substituting $\mathbf{S}_W$ with equation 2.25:

$$\mathbf{B}^T \mathbf{S}_W \mathbf{B} = (\mathbf{B}^T \mathbf{A}_1)(\mathbf{B}^T \mathbf{A}_1)^T + (\mathbf{B}^T \mathbf{A}_2)(\mathbf{B}^T \mathbf{A}_2)^T \quad (2.31)$$

The $\mathbf{B}^T \mathbf{A}_p$ terms are of size number of column vectors in $\mathbf{B}$ by n_p , meaning that the terms in the parentheses can be calculated first and $\mathbf{S}_W$ does not need to be constructed.

2.2.4 Datasets

2.2.4.1 Simulated data

We first tested our algorithm on a simulated rfMRI dataset. This dataset consisted of 59 simulated ‘healthy’ subjects and 59 simulated ‘disease’ subjects. We chose to simulate this number of subjects since one of the real datasets that we used, which we can assume to be a typical study size, contained this many subjects. Each subject’s data contained 91125 voxels and 180 time-points. To simulate resting-state connectivity, the 91125 voxels were split equally between 27 brain regions (BRs).

Each BR was associated with a unique, intrinsic time-course that was randomly generated with no average autocorrelation. These time-courses were stored in a matrix $\mathbf{T}$, of size 27×180 .

We simulated connectivity between BRs by creating a time-course for each BR that was a mixture of its own intrinsic time-course and the intrinsic time-course of all the BRs with which it is connected. This was implemented by defining a 27×27 mixing matrix $\mathbf{M}$ that encodes the connectivity between regions. The rows and columns of $\mathbf{M}$ represent the strength of incoming and outgoing connections respectively. The diagonal elements of $\mathbf{M}$ were all set to 1 while all the off-diagonal elements were set to random values between 0 and 0.3. Thus the final resting-state time-courses for the set of BRs is given by $\mathbf{R} = \mathbf{MT}$, where $\mathbf{R}$ is of size 27×180 .

For the disease case we assumed that 25% of the cells in one BR die, and that this results in a 25% decrease in the intrinsic time-course signal in this region. We assumed that this would then cause a 25% decrease in the strength of the incoming and outgoing connections to and from this region (based on decreases in the number of incoming and outgoing synapses/axons connected to functioning cells). When calculating $\mathbf{R}$ for disease subjects, we multiplied the row and column of $\mathbf{M}$ corresponding to the affected region by 0.75. Finally, we added Gaussian noise to the time-course of each voxel.

We simulated a total of five different datasets, with each dataset simulated in exactly the same way as described above except that a different noise level was added at the last step. We describe the level of added noise by g – the standard deviation of the Gaussian noise $\sigma(n)$ divided by the standard deviation of the true resting-state time-course $\sigma(rs)$. Hence $g = \frac{\sigma(n)}{\sigma(rs)}$. For example, if resting-state activity caused a base-line fMRI signal change of 0.2%, and the noise caused a base-line change of 1%, then $g = 5$.

2.2.4.2 Parkinson’s disease data

We tested our algorithm on a real rfMRI dataset that was provided to us by the Oxford Parkinson’s Disease Centre (OPDC). It consisted of rfMRI scans for 59 Parkinson’s Disease (PD) subjects and 30 healthy control (HC) subjects. These are the same subjects that were studied in [33]. In order to balance the dataset a further 29 HC subjects were added to it. These were chosen at random, while still controlling for age and gender, from the datasets used in [75], [76] and [77]. These extra HC subject were scanned using the exact same scanner and sequence, hence the viability of adding them to the OPDC cohort. This then gives a total of 59 PD and 59 HC subjects. A 3 T Trio Siemens MRI scanner was used to acquire the data, using Echo Planar Imaging (EPI) with a spatial resolution of $3 \times 3 \times 3.5$ mm and a repetition time of 2 s. The subjects were scanned for 6 minutes, resulting in a total of 180 time-points per scan.

Processing of the images was carried out using FSL 5.0.9 [78]. This involved motion correction using MCFLIRT [79], brain extraction, regression of 24 motion parameters, unwarping using fieldmap data, high-pass temporal filtering of 150 s, spatial smoothing using a Gaussian kernel with a FWHM of 6 mm, cleaning using FMRIB’s ICA-based Xnoiseifier (FIX) [80] [81] (trained on 50 of the OPDC cohort subjects) and registering to the 2 mm MNI 152 standard space. These processing steps had already been carried out for us.

Further processing involved masking the data so that only voxels for which all the subjects had data were included in the analysis. An intersection mask, intersected with all the individual subject masks, and further intersected with a standard space mask, was created for this purpose.

2.2.4.3 Amyotrophic lateral sclerosis data

The second real dataset that we evaluated our algorithm on consisted of 58 amyotrophic lateral sclerosis (ALS) and 37 HC subjects. To balance the dataset, we selected a random subset of 21 ALS subjects to exclude from the analysis. This

resulted in a total of 37 ALS and 37 HC subjects. Data was acquired using a 3 T Trio Siemens MRI scanner, using EPI with a 3 mm isotropic resolution and a repetition time of 2 s. The subjects were scanned for 6 minutes, resulting in a total of 180 time-points per scan.

Processing of the images was again carried out using FSL 5.0.9. This involved motion correction using MCFLIRT, brain extraction, unwarping using fieldmap data, spatial smoothing using a Gaussian kernel with an FWHM of 6 mm, high-pass temporal filtering of 150 s, cleaning using FIX and registering to the 2 mm MNI 152 space. These processing steps had again already been implemented for us. We additionally masked the data using the steps described previously.

2.2.4.4 Anaesthesia data

The third and final real dataset we applied our algorithm to was an anaesthesia dataset consisting of 13 subjects that were scanned both pre and post ultraslow propofol administration of 4 $\mu\text{g}/\text{ml}$. Some of this data set has previously been used in [82]. Data was acquired with a 3 T Trio scanner, using EPI with a resolution of $3.5 \times 3.5 \times 3.5$ mm and a repetition time of 3 s. The subjects were scanned at peak propofol dose for 10 minutes, resulting in 200 time-points per scan.

Processing was carried out using FSL 5.0.10 and involved motion correction using MCFLIRT, brain extraction, spatial smoothing using a Gaussian kernel with an FWHM of 5 mm, high-pass temporal filtering of 50 s, cleaning using FIX and registering to a 3.5 mm MNI 152 standard space that had been downsampled from the 2 mm version. These processing steps had also already been implemented for us. The data was again masked in the same way as for the PD case.

2.2.5 Classification

For the PD and ALS datasets, we used a k -fold cross validation (CV) ($k = 10$ for PD and $k = 9$ for ALS), along with the AUC of the ROC curve to evaluate the classification performance. In k -fold CV, the dataset is split into k test sets,

termed folds. The samples/subjects not included in a given set are used to train an algorithm, and the output of the training is used to classify the samples within the test set. This is repeated for all the sets, and the classification metric, in this case the AUC, is averaged across the sets to obtain an overall measure of performance. For binary classification of balanced test data, the AUC metric will be 1 if all the samples in the test sets have been classified correctly, while a value of 0.5 will signal that the performance has been no better, on average, than chance. We repeated the k -fold CV 10 times, each time with a random shuffling of the within-group subject indices, to obtain a measure of variability in performance. To optimise for λ , we used a nested k -fold approach. For each outer fold, an inner k -fold CV ($k = 9$ for PD and $k = 8$ for ALS) was applied to just the training set, and repeated with a range of different λ values. The λ value corresponding to the maximum AUC was then used to train the whole of the training set in a given outer fold, and the output of this training was applied to the corresponding test set. This nested approach ensures that no double-dipping occurs and the final AUC is not an over-estimate. An example of nested CV is shown in figure 2.3.

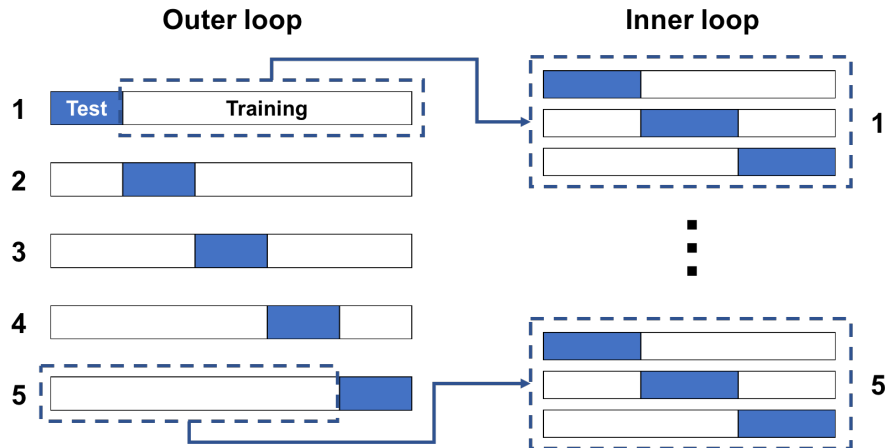


Figure 2.3: A diagram of nested cross validation. In this example, k for the outer loop is 5 and k for the inner loop is 3.

For the anaesthesia dataset, we adopted a leave-one-subject-out (LOSO) CV instead, due to the low number of subjects in each group. We thus tested whether the algorithm was able to correctly predict if a test subject's scans belonged to the pre or post anaesthesia state, after training it using the other subjects. Since for

each CV fold, there are only two test samples, the prediction accuracy rate (ACC) was used as the classification metric as opposed to the AUC. An ACC value of 1 indicates perfect performance whilst a value 0.5 is no higher than chance. As before, a nested loop was utilised to optimise for λ . Each test sample was classified to the group that had the closest training mean.

To decide on the scale of λ , we considered line 2 of the algorithm description. An element in the diagonal of $\mathbf{X}_s \mathbf{X}_s^T$ will be equal to $n_t - 1$, which is approximately equal to n_t when $n_t \gg 1$. This follows from the fact that each time-course in $\mathbf{X}_s$ is demeaned and normalised to have a variance of 1. If we assume that $\mathbf{m}_{i-1}$ is a normalised vector (with a norm of 1) and all its elements are equal to each other, an element in $\mathbf{x}_s$ will then be in the order of $n_v \frac{1}{\sqrt{n_v}} n_t$. This is by virtue of each element in $\mathbf{m}_{i-1}$ having a value of $\frac{1}{\sqrt{n_v}}$, and there being n_v elements. Ignoring the $\boldsymbol{\mu}_p$ term, this will result in an element in the diagonal of $\mathbf{S}_W$ to be in the order of $(n_1 + n_2) n_v n_t^2$. Thus λ will have to be in the order of this to have an obvious regularisation effect. For the OPDC dataset ($n_v = 189108$, $n_1 + n_2 = 118$ and $n_t = 180$), this calculation equals 7.23×10^{11} . We could set $\lambda = \sigma(n_1 + n_2) n_v n_t^2$ and optimise for σ . However, in deriving this we have assumed that the elements in $\mathbf{m}_{i-1}$ are all equal. In the extreme case of $\mathbf{m}_{i-1}$ having one element equal to 1 and all others set to 0, the n_v term would disappear. In addition, we ignored $\boldsymbol{\mu}_p$. As $\|\mathbf{x}_s - \boldsymbol{\mu}_p\|^2 \rightarrow 0$ then elements in $\mathbf{S}_W$ will have a smaller magnitude. Therefore, optimising for σ may result in the full range of possible optimum λ values not getting included. A different approach, which we chose to take here, is to optimise for γ , where $\lambda = \gamma^4$. The γ values we tried were spaced equally up to a value of 3000 to ensure we cover both small λ values and also values of the order of magnitude calculated above. In section 2.3.1 we will learn that it is possible, for some datasets, to set a lower limit on λ in order to remove the algorithm's dependence on initialisation.

For the simulated datasets, we kept λ constant for training phases and so did not use a nested loop to optimise for it. In addition, we used leave-one-out (LOO) CV. We do not believe the method of CV to be an issue for the simulated data. This is

since our main aim with the simulated data was to gain insight into the algorithm and see if the differences it picks out are consistent with those we simulated.

2.2.6 Comparing to other methods

The performance of any classification method is reliant both on the algorithm and on the dataset itself. If the group differences within a dataset are not substantial then even the best algorithm will fail to perform well. Therefore, we compared our performance with that of existing methods that have been reported in literature.

2.2.6.1 Independent components analysis + dual regression

The first existing method we used for comparison was that used by Szewczyk-Krolikowski et al. as described in [31]. We will refer to this approach as ICA+DR throughout the rest of the chapter. The protocol we followed was:

1. FSL MELODIC was used to carry out group ICA by concatenating the time-courses of all subjects. The target dimensionality was set to 50.
2. Subject-specific maps corresponding to the 50 independent components were obtained using DR. All 50 of the group ICA maps were used in the design matrix. The subject-specific maps for the network of interest (the BGN for PD, and the DMN for ALS and anaesthesia) were then used for further analysis. We used the DMN for the two latter datasets since it has been previously been shown to have altered functional connectivity both in ALS [83] and anaesthesia [84].
3. The dataset was split into folds as described in section 2.2.5. Permutation testing with TFCE, implemented in FSL Randomise [28], was applied to the subject-specific maps in the training sets. For the PD and ALS data, the statistical contrast was set up so that voxels where the HC group had significantly higher values than the disease group were found, while for the anaesthesia data both directions were tested.

4. These significant voxels were used as a binary mask within which to average the voxel values from subject-specific network maps. Although Szewczyk-Krolikowski et al. thresholded the TFCE maps at $p < 0.05$, we found that this p value was at times too strict and left very few voxels to average over. Instead, we took advantage of the inner nested loop within each fold to optimise for this threshold.
5. The masks from the training sets were finally used to average over voxels in the subject-specific network maps in the corresponding test sets, resulting in one discriminatory value for each subject. For the PD and ALS data, AUC values were calculated using the test subject values and for the anaesthesia data ACC values were calculated by classifying to the group with the closest mean.

2.2.6.2 Network based + negative/positive connectivity index

The second method we compared IDLDA to was the network based approach used by Chen et al. [35]. This approach will be referred to as NB+NPI and the protocol we followed was:

1. The Automated Anatomical Labeling (AAL) template [85] was used to parcellate each subject's rfMRI data into 116 regions/nodes. The time-courses within a given region were averaged so that there was a single time-course for each region.
2. For a given subject, the Pearson correlation coefficient (r) was calculated for all pairs of node time-courses, resulting in 6670 r values for each subject.
3. The dataset was split into folds as before, and r values for subjects within the training set were fed into a two-sample Wilcoxon rank-sum test, resulting in 6670 z values. The magnitude of a z value corresponding to a pair of nodes signifies how different the pair's r values are between the two groups. Meanwhile, the sign of the z value indicates which of the two groups has the higher or lower correlation.

4. The r values belonging to the node-pairs with the highest n positive and lowest n negative z values were averaged for each subject, resulting in a positive and negative connectivity index respectively. The inner nested loop was used to optimise for the value of n .
5. These two features were fed into standard LDA. The output was then used to obtain discriminant values for the subjects in the test set and AUC/ACC values were calculated.

2.2.6.3 Network based + all correlations

The protocol for this third method is similar to NB+NPI, except that the r values corresponding to the n highest and n lowest z values were not averaged. Instead, they were fed directly into LDA as a vector. When calculating the inverse of the covariance matrix $\mathbf{S}_W$, we used the pseudo-inverse to deal with the fact that there may be a null space. This method will be referred to as NB+AC.

2.3 Results

2.3.1 Simulated data

We evaluated the algorithm on the simulated data for a variety of noise levels, described by g . Figures 2.4, 2.5 and 2.6 shows one slice (containing 9 BRs – arranged in 3×3 blocks) from the final $\mathbf{w}$ and $\mathbf{v}$ images for $g = 2, 5$ and 9 respectively. These were obtained when the whole 118 subjects were used as the training set, with $\lambda = 10^9$ and $\mathbf{m}_0 = \mathbf{r}$. Here we use $\mathbf{r}$ to refer to a normalised random vector with each of its elements sampled from a Gaussian distribution. The region highlighted in the $\mathbf{w}$ image in figure 2.5 shows the voxels in the BR that we simulated to be affected by disease. Our algorithm clearly picks out these voxels in one of the vectors and gives a relatively high weighting to these voxels. Furthermore, the objective function monotonically increased with iterations in all cases, converging after 14, 7 and 8 iterations for the $g = 2, g = 5$ and $g = 9$ datasets respectively.

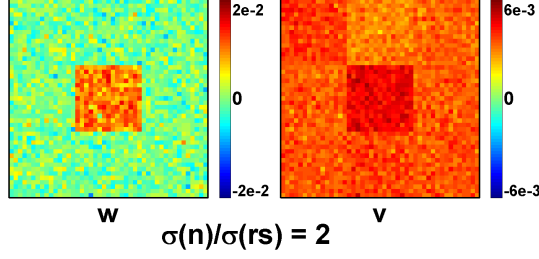


Figure 2.4: Output $\mathbf{w}$ and $\mathbf{v}$ images for the $g = 2$ simulated dataset, obtained using $\lambda = 10^9$.

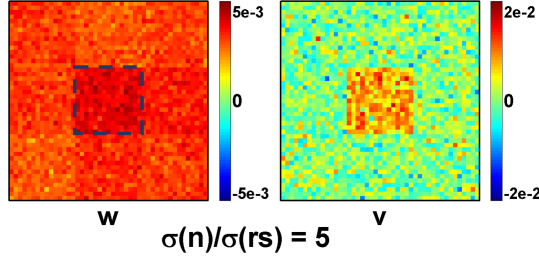


Figure 2.5: Output $\mathbf{w}$ and $\mathbf{v}$ images for the $g = 5$ simulated dataset, obtained using $\lambda = 10^9$. The region highlighted in the $\mathbf{w}$ image shows the voxels in the BR that we simulated to be affected by disease.

The reason we chose $\lambda = 10^9$ can be seen in figure 2.7. To obtain this graph, for a given λ value, we applied the algorithm to the $g = 5$ dataset and initialised it with two different vectors. One of these vectors was a vector of ones that was then normalised. The other was a vector that had the number 10 in voxels that corresponded to the simulated disease BR, and a 1 elsewhere, which was then normalised. Therefore, the first initialisation vector was uninformative, while the second was informative. We then calculated the Pearson correlation coefficient (r) between the vector from the first run that looked most like the $\mathbf{w}$ vector in figure 2.5, and the vector from the second run that looks most like $\mathbf{w}$. The same was done for $\mathbf{v}$. For $\mathbf{v}$, the correlation is close to one for all λ values, while the correlation approximately equals one at around $\lambda = 10^9$ for $\mathbf{w}$. We also found that there was a relationship between the number of voxels and the λ at which the algorithm becomes desensitised to initialisation. The plot for this is shown in figure 2.8.

To assess the classification performance of the algorithm, we re-simulated the data and used a LOO CV with $\lambda = 10^9$, and $\mathbf{m}_0 = \mathbf{r}$. The classification performance was evaluated by calculating the AUC of the ROC curve. The classification accuracy

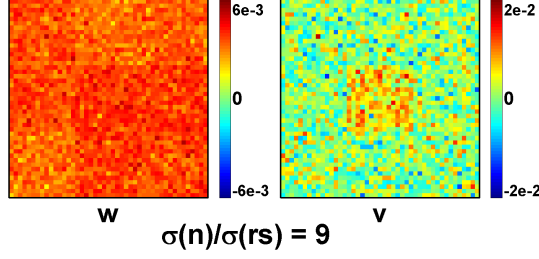


Figure 2.6: Output $\mathbf{w}$ and $\mathbf{v}$ images for the $g = 9$ simulated dataset, obtained using $\lambda = 10^9$.

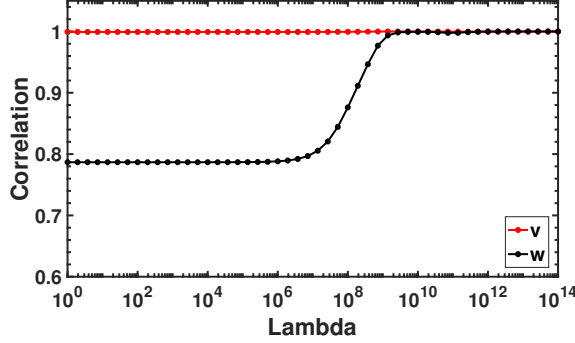


Figure 2.7: Sensitivity to initialisation as a function of λ . A correlation value of 1 means the algorithm is not sensitive to initialisation. For $\mathbf{v}$, the correlation is close to one for all λ values, while the correlation approximately equals one at around $\lambda = 10^9$ for $\mathbf{w}$.

as a function of noise level is shown in figure 2.9. The figure shows that the algorithm performs well, in the region of 0.85 to 1, at low to moderate noise levels.

We applied the ICA+DR approach reported in [31] to the simulated data for comparison. However, DR requires the manual selection of the discriminating component, thus needing knowledge of the ground truth. For an upper bound on performance we used our prior knowledge of the ground truth and selected the ICA component that corresponded to the simulated disease region. For a lower bound we selected a different ICA component chosen at random. The AUC values for these are also shown in figure 2.9. At low to medium noise levels the performance of IDLDA is close to the upper bound, moving towards the lower bound for high noise levels.

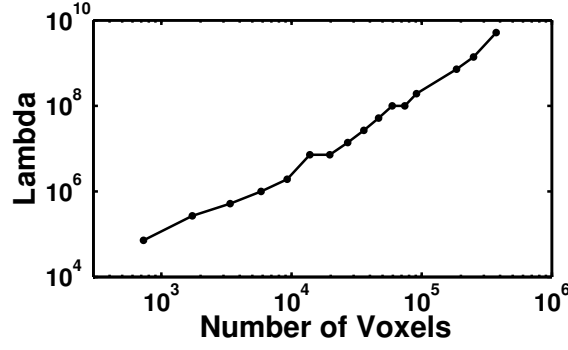


Figure 2.8: The λ value at which the algorithm stops being sensitive to initialisation (using a cut-off of $r = 0.99$), as a function of number of voxels. Datasets were simulated using $g = 5$.

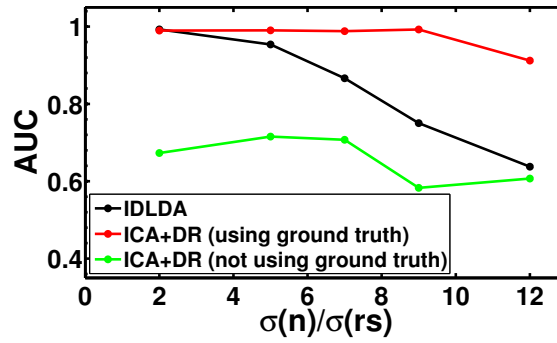


Figure 2.9: LOO CV AUC results as a function of added noise, using $\lambda = 10^9$. At low to medium noise levels the performance of IDLDA is close to the upper bound, moving towards the lower bound for high noise levels.

2.3.2 Parkinson's disease data

IDLDA was applied to the PD data using $\mathbf{m}_0 = \mathbf{r}$ and a minimum γ of 300 (based on figure 2.8) to ensure there is minimal sensitivity to initialisation. In addition, we used a minimum of 8 iterations and a maximum of 20 iterations. The box plot labelled IDLDA in figure 2.10 shows its classification performance. The median AUC of IDLDA (0.70) was better than ICA+DR (0.62), NB+NPI (0.66) and NB+AC (0.67). To test the statistical significance, we fed the AUC values into a two sample Wilcoxon rank-sum test. The resulting p values were 0.0002, 0.02 and 0.03 for IDLDA vs. ICA+DR, IDLDA vs. NB+NPI and IDLDA vs. NB+AC respectively, showing that IDLDA performed significantly better. Furthermore, the interquartile range was 0.02, 0.02, 0.05 and 0.07 for IDLDA, ICA+DR, NB+NPI and NB+AC respectively. This shows that the variability associated with IDLDA is also better

when compared to NB+NPI and NB+AC. It should be noted that since for a given method the AUC values were obtained from cross validation and the shuffling of within group subject indices, the AUC values are not independent from each other. This violates the independence assumption behind the Wilcoxon rank-sum test, and this violation should be taken into account when considering the p values reported here and in section 2.3.3.

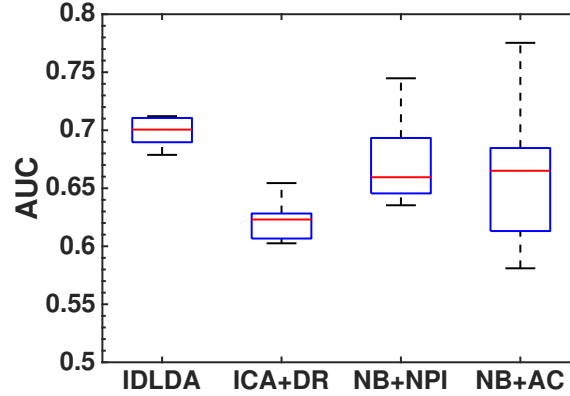


Figure 2.10: k -fold CV AUC results when applying IDLDA and existing algorithms to the PD dataset. The p values under the Wilcoxon rank-sum test were 0.0002, 0.02 and 0.03 for IDLDA vs. ICA+DR, IDLDA vs. NB+NPI and IDLDA vs. NB+AC respectively, showing that IDLDA performed significantly better.

Figure 2.11 shows the mean AUC as a function of iteration number. We obtained this figure by using $\mathbf{m}_i$ and $\mathbf{m}_{i-1}$ to classify test subjects, resulting in AUC values for iteration i . We then averaged AUC values across the random shuffles so that there was one AUC value corresponding to iteration i . This figure shows that the iterative framework does well to increase the generalisability of the discriminatory maps obtained during training. It should be noted that although the full 20 iterations are shown, 70 out of the possible 100 outer training folds had already converged by the time the minimum number of iterations was reached. However, 16 of the training folds had not converged by the time the maximum number of iterations was reached.

The spread of the optimum γ 's that the inner nested loop found is shown in figure 2.12. The most common was $\gamma = 1300$, which corresponds to the $\lambda = 2.9 \times 10^{12}$. This is not too far off the value we calculated in section 2.2.5. However, there is also another peak at the minimum γ . This is most likely a result of cutting the range of

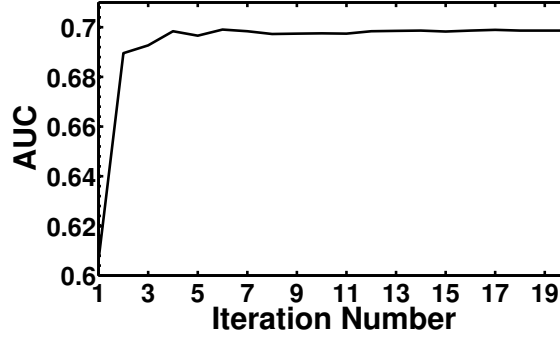


Figure 2.11: AUC as a function of iteration number when applying IDLDA to the PD dataset. The iterative framework does well to increase the generalisability of the discriminatory maps obtained during training.

γ at 300. It is possible that the peak would get reduced in height if smaller values were included.

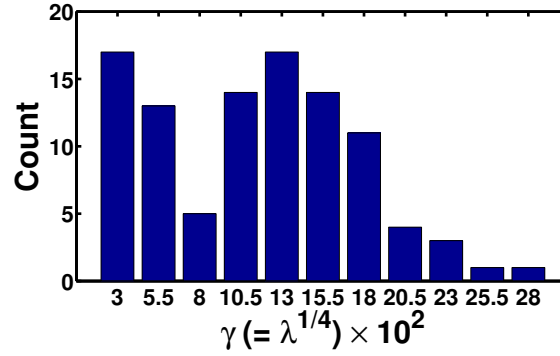


Figure 2.12: Histogram of the optimum λ values found using the nested CV scheme for the PD dataset.

In order to see the training outputs of IDLDA, we trained it using all 118 subjects, $\mathbf{m}_0 = \mathbf{r}$ and $\lambda = 2.9 \times 10^{12}$ (the most common λ picked out during CV). The objective function as a function of iteration number is shown in figure 2.13, and it shows that it converged after 5 iterations.

The $\mathbf{w}$ and $\mathbf{v}$ images output by the algorithm are shown in figure 2.14. The $\mathbf{w}$ image does not contain any obvious structure while the $\mathbf{v}$ image has large weightings in superior and inferior-posterior gray-matter voxels. We also re-ran the algorithm two extra times but each time with a different $\mathbf{m}_0$. The cross-correlations between the output images that were most similar to the $\mathbf{v}$ image shown in figure 2.14 were -0.9976, -0.9997 and 0.9980. For the $\mathbf{w}$ image it was -0.9989, 0.9991 and -0.9999. The negative correlations are due to the the sign ambiguity discussed in section 2.2.2.

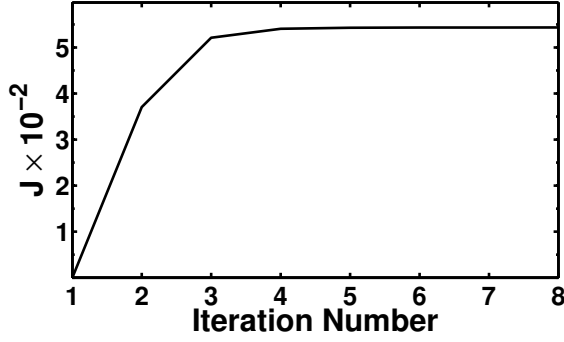


Figure 2.13: The objective function value as a function of iteration number for the PD dataset, using $\lambda = 2.9 \times 10^{12}$. The algorithm converged after 5 iterations.

The ordering ambiguity was also apparent, since one of the $\mathbf{w}$ images in one of the runs corresponded to the $\mathbf{v}$ images in the other two and vice versa. However, since these ambiguities do not affect classification performance, we can conclude that for the PD data the algorithm was not sensitive to initialisation by these three vectors at this λ value.

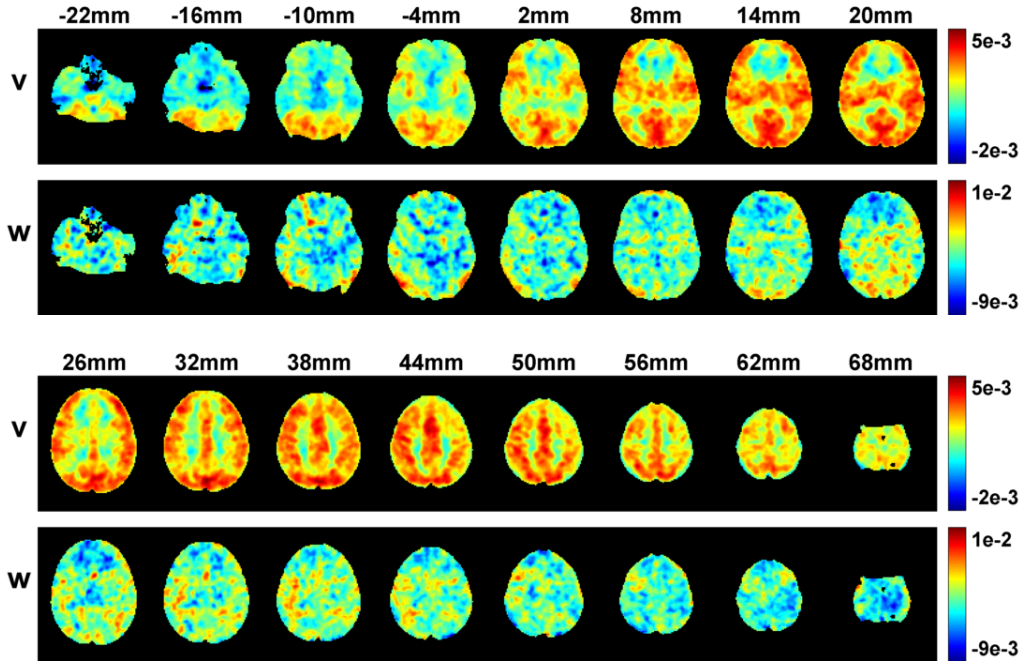


Figure 2.14: Output $\mathbf{w}$ and $\mathbf{v}$ images for the PD dataset, using $\lambda = 2.9 \times 10^{12}$. The images are axial slices in MNI 152 standard space coordinates. The $\mathbf{w}$ image does not contain any obvious structure while the $\mathbf{v}$ image has large weightings in superior and inferior-posterior gray-matter voxels.

2.3.3 Amyotrophic lateral sclerosis data

We first applied IDLDA to the full 74 subjects from the ALS dataset, using $\lambda = 2.9 \times 10^{12}$. This was the value that we learned was optimum for the PD data but, more importantly, at this value our algorithm should be insensitive to initialisation according to figure 2.8. However, we found that the algorithm was still sensitive to initialisation. Figure 2.15 shows the $\mathbf{w}$ and $\mathbf{v}$ images output by the algorithm when $\mathbf{m}_0 = \mathbf{r}_1$. Note that we subscript $\mathbf{r}$ here to distinguish between different random vectors. The $\mathbf{w}$ image has similar properties to that shown in the previous section in that there is no apparent structure with very small clusters of hot-spots. The $\mathbf{v}$ image on the other hand contains large scale structure but, unlike for the PD data, this is concentrated on white matter. The algorithm was initialised a further two times with $\mathbf{r}_2$ and $\mathbf{r}_3$. Both of these converged to different images from $\mathbf{r}_1$, but to the same images as each other. These are shown in figure 2.16 and are clearly different from those in figure 2.15. They still exhibit similar properties however, with one of the vectors containing small clusters while the other contains large-scale structure.

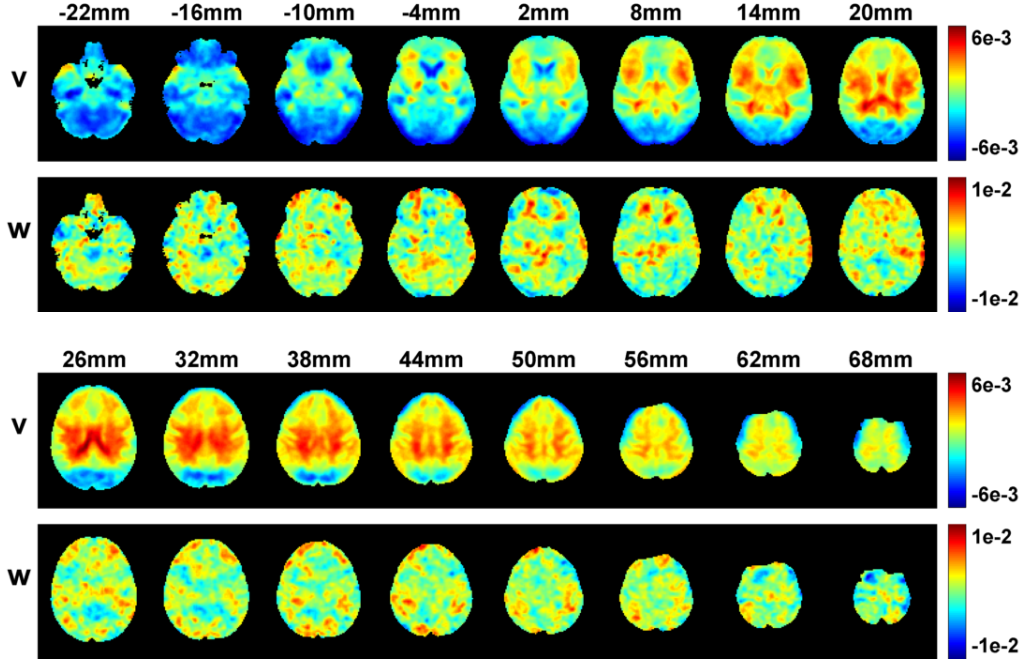


Figure 2.15: Output $\mathbf{w}$ and $\mathbf{v}$ images for the ALS dataset, using $\lambda = 2.9 \times 10^{12}$ and $\mathbf{m}_0 = \mathbf{r}_1$. The images are axial slices in MNI 152 standard space coordinates.

The algorithm converged after 8, 25 and 20 iterations for $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ respectively. The value of J was 0.084 and 0.140 corresponding to $\mathbf{r}_1$ and $\mathbf{r}_2$ or $\mathbf{r}_3$ respectively. This suggests the algorithm thinks the images shown in figure 2.16 are more discriminatory.

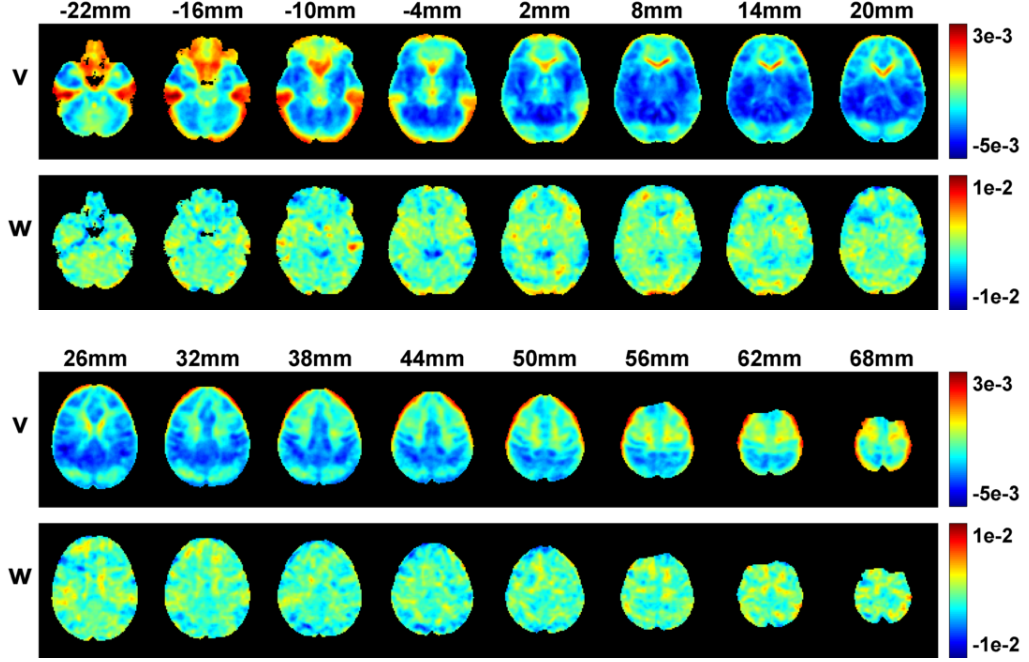


Figure 2.16: Output $\mathbf{w}$ and $\mathbf{v}$ images for the ALS dataset, using $\lambda = 2.9 \times 10^{12}$ and $\mathbf{m}_0 = \mathbf{r}_2$. The images are axial slices in MNI 152 standard space coordinates.

To take into account this dependence on initialisation, we modified our classification protocol. For each training fold, we initialised with two random vectors, and used the $\mathbf{w}$ and $\mathbf{v}$ maps that corresponded to the higher J value. We also ran the algorithm for a maximum of 40 iterations to make sure it converges. The results are shown in figure 2.17. The figure shows that all the algorithms resulted in very low AUC values, suggesting that the discriminatory power of the dataset was very small. Furthermore, the performance of IDLDA compared to the three other algorithms was not significantly different when using the Wilcoxon rank-sum test, with p values of 0.10, 0.82 and 0.24 for IDLDA vs. ICA+DR, IDLDA vs. NB+NPI and IDLDA vs. NB+AC respectively. Due to the very low AUC values, we did not give any more focus to this dataset.

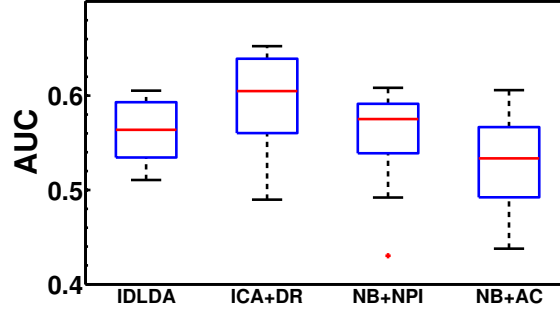


Figure 2.17: k -fold CV results when applying IDLDA and existing algorithms to the ALS dataset.

2.3.4 Anaesthesia data

One reason for the lack of clear structure in the IDLDA maps for the PD and ALS data could be that the discriminatory power of the datasets themselves was not that high to begin with, as corroborated by the AUC values for all of the algorithms. For this reason, we applied IDLDA to the anaesthesia dataset as well. We lowered the lower bound on γ to 55 since there are a lot less voxels for this data. To take account of the fact that this data may be sensitive to initialisation, we initialised the algorithm with 10 different vectors, 9 of them random and one a normalised vector of ones, and used the $\mathbf{w}$ and $\mathbf{v}$ pair corresponding to the highest J value. The classification accuracy for the four different algorithms are shown in figure 2.18. ICA+DR+ and ICA+DR- correspond to a contrast of pre-anaesthesia $>$ post-anaesthesia and pre $<$ post respectively, and the DMN was used as the network of interest. NB+NPI performed the best with 100% classification accuracy, ICA+DR performs the worst and IDLDA is equal with NB+AC at 92%. Since we used a LOSO CV, we could not test for the variability of the results as we did for the PD and ALS data. Hence, it is not clear whether these are significantly different, especially given that the difference is only for a single subject. However, it is safe to say both the IDLDA and NB algorithms performed well on this dataset.

In order to test whether any of the other ICA components can perform as well as IDLDA, we re-ran the ICA+DR algorithm for all 50 of the extracted components. These results are shown in figure 2.19. Since the choice of ICA component was not

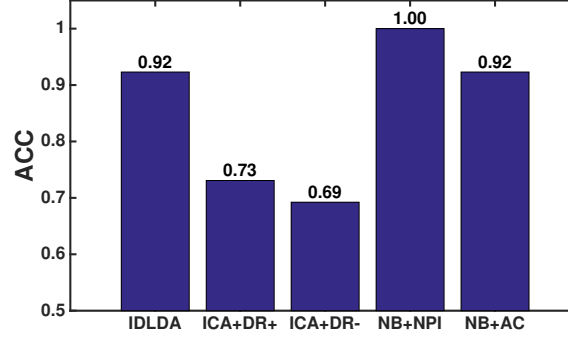


Figure 2.18: Classification accuracy rate (ACC) when applying IDLDA and existing algorithms to the anaesthesia data.

optimised in an inner CV loop, the maximum obtained classification accuracy shown in the figure is likely to be an overestimate. However, the maximum value of 88% percent still is not as good as that achieved by IDLDA.

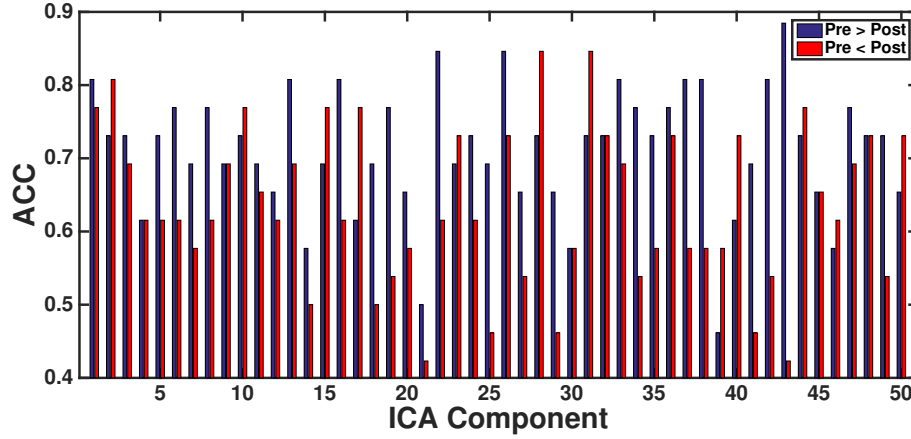


Figure 2.19: Classification accuracy rate when applying ICA+DR to the anaesthesia data, tested for all 50 ICA components. The maximum value of 88% percent is still not as good as that achieved by IDLDA.

Every single fold's inner CV loop chose $\gamma = 55$ as optimum, and so we used this value to apply IDLDA to the whole dataset. We used the same 10 initialisation vectors as for the CV, and the $\mathbf{w}$ and $\mathbf{v}$ maps corresponding to the largest J value are shown in figure 2.20. One of the maps, the $\mathbf{w}$ map, again looks like it has picked out large sections of the gray matter, but there is now clearer localised structure in the corresponding $\mathbf{v}$ map. Before interpreting the maps in more detail, we measured how similar the outputs from each of the 10 initialisations were. This was performed by first relabelling each vector of a $\mathbf{w}$ - $\mathbf{v}$ pair to either a $\mathbf{w}$ or a $\mathbf{v}$ based on which of the two maps in figure 2.20 they looked most similar to, (remember that in the IDLDA

algorithm description, the LDA outputs from the last and second to last iteration are labelled as $\mathbf{w}$ and $\mathbf{v}$ respectively, but the ordering is arbitrary). For example, if for one of the runs the $\mathbf{w}$ looked more like the $\mathbf{v}$ in figure 2.20, and the $\mathbf{v}$ more like the $\mathbf{w}$, then the labels would be swapped around. The set of 10 $\mathbf{w}$ and $\mathbf{v}$ vectors were then cross-correlated, using the Pearson correlation coefficient, not accounting for the sign of the correlation. This resulted in two 10×10 cross-correlation matrices, one for the set of $\mathbf{w}$ vectors and one for the set of $\mathbf{v}$ vectors. These are shown in figure 2.21, with the subscripts corresponding to the which initialisation they belonged to. Finally, to examine the stability of the classification with regards to initialisation, we re-ran the LOSO CV using each of the initialisation vectors on their own. The classification accuracy corresponding to each initialisation is shown in the diagonal of the $\mathbf{v}$ correlation matrix.

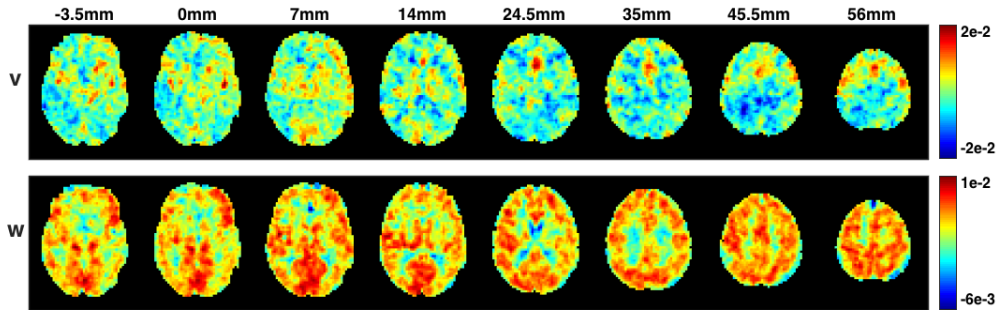


Figure 2.20: Output $\mathbf{w}$ and $\mathbf{v}$ images for the anaesthesia dataset, using $\lambda = 9.2 \times 10^6$. The images are axial slices in MNI 152 standard space coordinates. The $\mathbf{w}$ map has picked out large sections of the gray matter, and there is now clearer structure in the corresponding $\mathbf{v}$ map.

Figure 2.21 shows that the set of vectors labelled as $\mathbf{w}$ were highly variable whereas the set of $\mathbf{v}$ vectors were relatively stable. The classification accuracies were also stable. This suggests that it is the $\mathbf{v}$ map from figure 2.20 that is driving the discrimination. Consequently we chose to focus on interpreting the $\mathbf{v}$ map.

There is still the question of how to interpret the positive or negative weights. For this we thresholded the vectors in figure 2.20 so that each was split into two vectors that contained only the positive weights, or only the negative weights, with the rest of the elements set to zero. This resulted in four vectors: $\mathbf{w}^+$, $\mathbf{w}^-$, $\mathbf{v}^+$ and $\mathbf{v}^-$. The

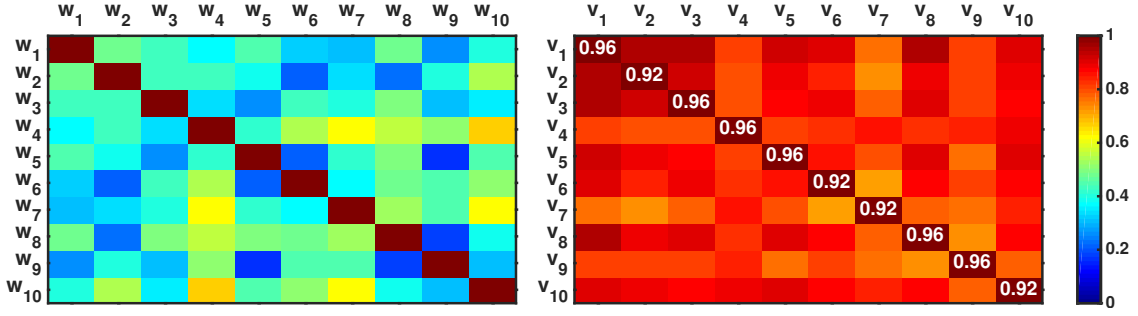


Figure 2.21: The algorithm was trained on the whole anaesthesia data using $\lambda = 9.2 \times 10^6$ ten times, each time with a separate initialisation vector. The output vectors were matched to what they most resembled in figure 2.20 and were subsequently cross-correlated, and the cross-correlation matrices are shown here. The numbers on the diagonal in the matrix on the right show the CV accuracy when only the initialisation vector corresponding to that run was used. Note that we have chosen to present the CV accuracy on the diagonal in order to have a concise figure. They should not be confused with the actual value of the diagonal elements of the cross-correlation matrix, which would always be 1. The $\mathbf{w}$ vectors were highly variable whereas the set of $\mathbf{v}$ vectors were relatively stable. The classification accuracies were also stable, suggesting $\mathbf{v}$ in figure 2.20 is driving the classification.

absolute value of all the vector elements was then taken. The spatial maps for these four vectors are shown in figures 2.22 and 2.23. We projected the data (i.e. calculated y in equation 2.9) using the four possible combinations of the thresholded $\mathbf{w}$ and $\mathbf{v}$. The results for this analysis are shown in figure 2.24. The idea behind this approach is that if e.g. the projected values for a given thresholded and positive-signed $\mathbf{w}$ and $\mathbf{v}$ are all positive but the post-anaesthesia group has lower projected values on average, then the voxels weighted in $\mathbf{v}$ have decreased functional correlation with the voxels weighted in $\mathbf{w}$ under anaesthesia. Similarly, higher projected values would indicate an increase in functional connectivity. The top-left and top-right plots in the figure show that both positively and negatively weighted voxels in $\mathbf{v}$ show a decrease in functional correlation with positively weighted voxels in $\mathbf{w}$. However, it is clear that out of these two, using the positively weighted voxels in $\mathbf{v}$ provides a much better discrimination. This, together with the more perceivable structure shown in the positively weighted voxels, meant we further focused on interpreting these voxels in the $\mathbf{v}$ map. The most clear region that these voxels lie on include the insular cortex, caudate, thalamus, anterior cingulate cortex and the precentral gyrus. Using the negatively weighted voxels in $\mathbf{w}$ gave negative projected values

for some subjects, suggesting they are picking out areas that show anti-correlation. However, the discrimination between the groups is poor as shown in the bottom two plots, hence we did not focus on them.

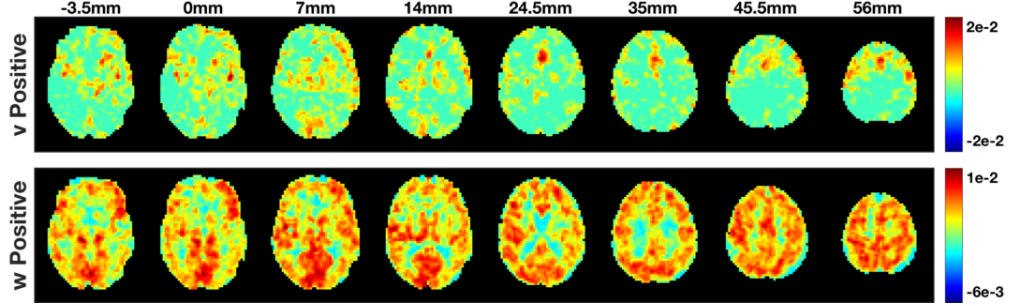


Figure 2.22: The spatial maps corresponding to $\mathbf{w}^+$ and $\mathbf{v}^+$.

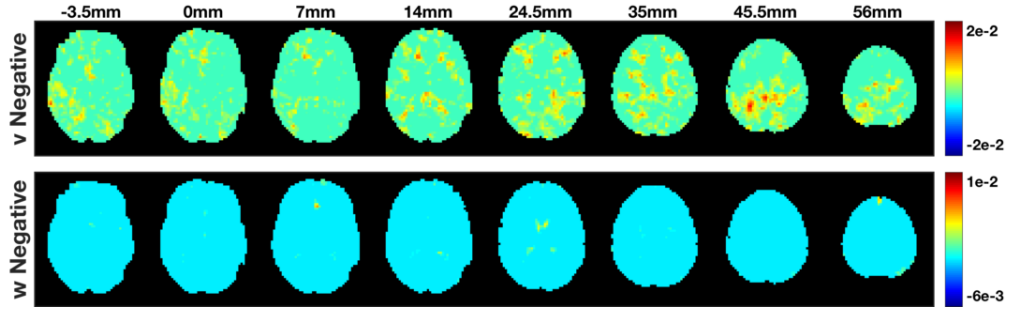


Figure 2.23: The spatial maps corresponding to $\mathbf{w}^-$ and $\mathbf{v}^-$.

2.4 Discussion

Our algorithm is comparable to the DR approach reported in [27], since $\mathbf{x}_s$ in the algorithm description represents a spatial map for subject s , obtained through the dot product of the individual voxel time-courses in $\mathbf{X}_s$ and the aggregate time-course $\mathbf{X}_s^T \mathbf{m}$. However, the major advantage of our algorithm is that it iterates to learn discriminatory information, meaning we do not need to do any ICA decomposition or parcellation of the rfMRI dataset into brain regions, thus removing arbitrary or subjective choices that are usually made with current approaches. For example, in the ICA+DR approach, the choices a user has to make are the dimensionality of ICA decomposition, the independent component(s) of interest, the components to

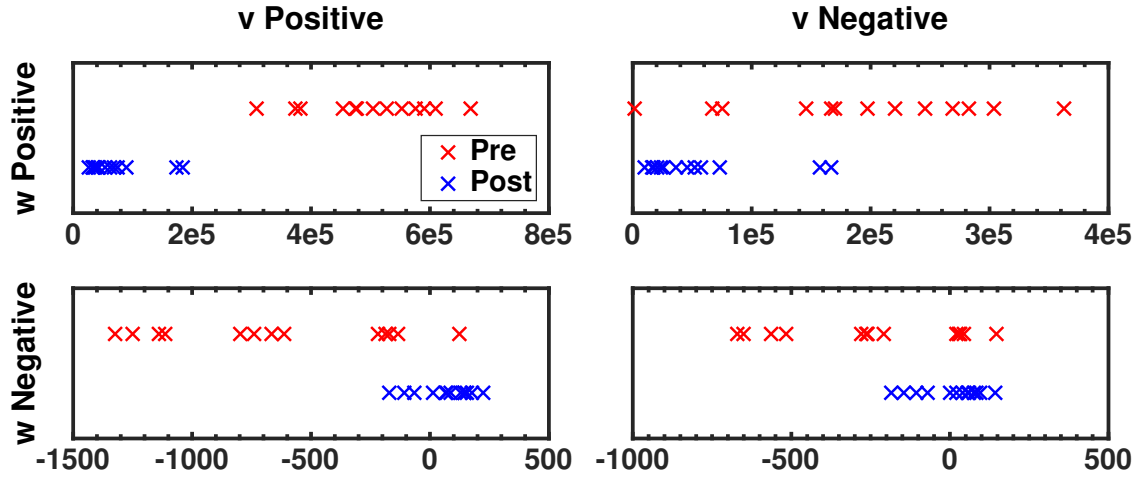


Figure 2.24: The vectors from figure 2.20 were thresholded to obtain four vectors: $\mathbf{w}^+$, $\mathbf{w}^-$, $\mathbf{v}^+$ and $\mathbf{v}^-$, where the $+$ indicates keeping all the positive voxels and $-$ the negative voxels. The absolute value of the all the vector elements was then taken. These four vectors are shown in figures 2.22 and 2.23. The top left plot shows the y values when using $\mathbf{w}^+$ and $\mathbf{v}^+$ to project each subject's data. Since the values are positive but higher in the pre-anaesthesia group, positively weighted voxels in the $\mathbf{v}$ map show a decrease in functional correlation with positively weighted voxels in the $\mathbf{w}$ map under anaesthesia. This combination also provides the best discrimination.

include in DR, the statistical test to use and the p threshold to use. In IDLDA, there is still a λ value that needs to be set. However, a nested CV approach can be used to optimise for this as we have done. Although it is true that this approach can be used to optimise for parameters such as the p value, using it to optimise for further parameters, such as the dimensionality parameter and the choice of component to include in the analysis, becomes infeasible since it would require testing all possible combinations of parameters. It could be argued that the number of random vectors to initialise with and the number of λ values to sample are also choices that a user needs to make. However, both of these simply depend on the computational resources (we discuss the computational demands of IDLDA later) a user has; the choice for both of these should be as many as possible.

IDLDA had a better AUC and was less variable compared to the existing methods that we tested on the PD data, both of which are highly desirable features for a biomarker. For the anaesthesia data, IDLDA's accuracy was high, showing it is capable of obtaining high classification performance if the discriminatory power is there. The performance of NB+NPI was slightly better, but we could not obtain

a measure of variability in this case. This, coupled with the fact that NB+NPI's performance was worse and more variable than IDLDA in the PD data, means that we cannot make a conclusive statement about which of the two approaches is better when considering only classification performance. Nevertheless, it can be argued that our approach may be driven by changes that are more specific to functional connectivity and the cohort being studied, since we did not do any prior parcellation, especially not one based on anatomical regions.

The issue of interpretability of machine learning algorithm outputs is an important one. In the simulated data, the $\mathbf{w}$ and $\mathbf{v}$ vectors are clearly interpretable in figures 2.4 to 2.6. The central square BR in the $\mathbf{v}$ image in figure 2.5 has a high weighting and so do the rest of the square BRs in the $\mathbf{w}$ image. This is saying that the functional connectivity between the regions has been altered, which is what we simulated.

For the PD data, the $\mathbf{v}$ image in figure 2.14 is interpretable in the sense that as it has picked out grey matter voxels, and we expect functional connectivity differences to show up in the grey matter. However, the lack of interpretability of the $\mathbf{w}$ image is non-ideal, but the AUC values in figure 2.10 suggest this is possibly due to the discriminatory power of the data not being high enough for the algorithm to pick out more subtle structure. The ALS vectors were also not interpretable, but again the AUC values were very low for all methods.

In the anaesthesia data on the other hand, the classification accuracy was high and both vectors were interpretable, with $\mathbf{w}$ picking out large scale gray matter and $\mathbf{v}$ picking out areas in the insular cortex, caudate, thalamus, anterior cingulate cortex and the precentral gyrus. The insula [86], thalamus [87], anterior cingulate cortex [88] and precentral gyrus [89] have all previously been shown to exhibit a decrease in functional connectivity with various cortical regions under anaesthesia. This is a very promising result considering the algorithm picked out these regions without any prior parcellation or decomposition. It further shows that IDLDA can potentially be used as an exploratory tool and can be complementary to existing methods.

For example, it can be used as an initial analysis to examine what discriminatory features may be present in a dataset. Regions that are picked out can then be included in further analyses.

In general, all the datasets that we tested showed the same pattern. One of the vectors picked out large scale features while the other focused on more localised regions and, in the cases of the simulated data and anaesthesia, it is the vector focusing on localised regions that seemed to be the most informative about disease.

Although these results show that there is scope for the $\mathbf{w}$ and $\mathbf{v}$ maps to be interpretable, there are still some issues that can make interpretation difficult. For example, as well as being driven by functional connectivity changes across groups, the $\mathbf{w}$ and $\mathbf{v}$ maps could also be driven by changes in the variance of the signal within a region. This is due to this variance change being encoded in the final mixed resting state time courses. We saw this in the spatial maps for the simulated data in figure 2.5, since the central square is also given relatively higher weighting in the $\mathbf{w}$ map. Another issue arises from the fact that the generation of the $\mathbf{w}$ and $\mathbf{v}$ maps is driven simply by the maximisation of equation 2.12 with no additional constraints. This means that regions that show decreased functional correlation, increase functional correlation and even changes in levels of anti-correlation across the two groups can all be encoded in the generated maps. It is not straightforward to disentangle this information by directly looking at the maps. However, it may be possible to extract this information by carrying out a post-hoc analysis like that in figure 2.24.

It should also be noted that if there are regions that do not appear in the $\mathbf{w}$ and $\mathbf{v}$ maps, this does not mean that these regions do not contain discriminatory information. This is due to our method effectively being a rank-one decomposition of the functional connectivity differences. In other words, the $\mathbf{w}$ and $\mathbf{v}$ maps encode one functional connectivity pattern that is most discriminatory based on equation 2.12. However, it may be possible to deflate the data based on these first set of maps and run the algorithm again to obtain a further set of maps that are orthogonal to

the first, and so on.

The fact that the algorithm was not sensitive to initialisation after a certain λ value in the simulated and PD datasets is a positive result, but this property did not generalise to all datasets in terms of the objective function, as demonstrated by the ALS and anaesthesia data. However, we showed that the classification performance and one of the output vectors was still insensitive to initialisation in the anaesthesia data, suggesting this may not be a large issue. We also saw similar results to this in the simulated data. One of the output vectors, the one that looked most like $\mathbf{v}$ in figure 2.5, was insensitive to initialisation across the whole of the λ range that we tested. In addition, $\mathbf{v}$ was the vector that was more informative and focused more on the localised region. For future studies however, we still recommend that the algorithm be initialised with multiple random vectors to be on the safe side.

A limitation of IDLDA when compared with other algorithms is the computational costs. When training the algorithm for a given λ and $\mathbf{m}_0$, the run-time is substantially less than ICA+DR as implemented by FSL. For example, training the full PD dataset with one λ and $\mathbf{m}_0$ takes approximately half an hour, while the ICA+DR protocol runs on the scale of a few hours. However, each training set needs to be split into folds in order to optimise for λ , and the algorithm needs to be initialised at least twice. This means that the total sequential run-time is substantially longer than ICA+DR. The upside is that these processes are highly parallelisable, meaning that training the algorithm can be implemented on a cluster and completed on a time-scale comparable to ICA. The NB+NPI protocol, which can be completed on the scale of a few minutes, is the fastest approach.

For future work, more extensive simulations may be useful. The simulations in this study were limited by the fact that we only considered the case when the functional connectivity of one region to every other region is altered. It would be informative to see how simulating more complicated cases, e.g. altering the functional connectivity of one set of regions to a different set of localised regions, will affect the $\mathbf{w}$ and $\mathbf{v}$ images. Theoretically, we would expect that the first set of regions

would appear in the $\mathbf{w}$ image and the second would appear in the $\mathbf{v}$ image. However, based on the results we have seen for real data, it may also be possible that both of these sets of regions would appear in only one of the images, while the other image highlights the whole of simulated brain. In addition, in figure 2.8, we only considered how the sensitivity to initialisation changes with respect to the number of voxels. Therefore, another possibility for extending the simulation experiments is to measure how the sensitivity to initialisation changes with respect to other variables, such as the number of subjects.

We chose the methods to compare IDLDA to based on how representative of typical rfMRI analyses they were. We also considered ease of implementation and availability of code. Many different protocols exist however, and a separate study conducting more extensive comparisons of rfMRI classifiers may be beneficial. Finally, applying our algorithm to more datasets containing subjects with neurodegenerative diseases would be very valuable as a form of validation, since the neurodegenerative disease datasets that we tested here did not seem to contain much discriminatory power.

2.5 Conclusion

We have presented a novel supervised learning algorithm based on LDA, which we call IDLDA, for the purpose of classifying subjects to one of two groups based on their rfMRI data. Its main strength is that it does not require any prior parcellation or decomposition of the data into brain regions. The output of the algorithm is two vectors, $\mathbf{w}$ and $\mathbf{v}$, which represent spatial maps, and can be used for classifying test subjects. We showed on simulated rfMRI data that the voxels highlighted by the algorithm in these two vectors correspond to voxels that we simulated to have altered functional connectivity in disease. On real PD data, we showed that the performance of IDLDA was significantly better and less variable than two existing methods. On anaesthesia data, we showed that it gave spatial maps that were interpretable and consistent with other results reported in literature.

Chapter 3

An errors-in-variables approach to estimating brain age

3.1 Introduction

The brain undergoes continuous change throughout an adult human's lifespan [90]. This ageing is generally associated with detrimental changes that relate to a loss of function [91] and increased risk of neurodegenerative diseases such as dementia [92]. However, the age at which many age-related conditions become apparent vary largely across a population. This has led to the concept of a 'biological age' that can predict the rate of ageing better than chronological age. The differences in the rates of ageing between individuals are thought to be influenced by biological processes that are themselves affected by both the individual's genotype and environment [93]. Therefore, a marker of biological age that allows the identification of individuals with accelerated ageing will have many benefits. It can potentially help in the early detection of neurodegenerative diseases, aid in understanding how such diseases interact with ageing and assist in identifying factors that contribute to accelerated ageing.

Outside of neuroimaging, there have been studies attempting to identify and validate biological age markers for over three decades [94]. One of the most com-

monly researched biomarkers is the telomere length [95]. Telomeres are complexes at the end of eukaryotic chromosomes, which shorten every time somatic cells divide. Telomere shortening has been associated with increased mortality risk [96], increased risk of some cancers [97], increased risk of coronary heart disease [98] and decreased cognitive function [99]. Patients with Alzheimer’s disease have also been shown to have shorter telomere lengths [100]. One of the clearest markers of biological age is chronological age itself, hence it stands to reason that a biological age marker should correlate highly with chronological age. Although these are all age-related associations, the telomere length has been shown to be a weak predictor of chronological age [101], thus its use as a biological age marker is limited.

A more recent biomarker is the DNA methylation age, often referred to as the epigenetic clock. DNA methylation is when a methyl group is added to a cytosine nucleotide in the DNA molecule [102]. DNA methylation has been shown to change with age and the signature of DNA methylation patterns across specific cytosine-guanine sites is the epigenetic clock. The Horvath [103] and Hannum [104] calculators are the two most popular methods of deriving DNA methylation age, with both showing strong correlations with chronological age. Studies typically use the ‘epigenetic age acceleration’ to draw associations with age-related phenotypes and conditions. The epigenetic age acceleration is derived by either taking the difference between the subjects’ chronological ages and the DNA methylation ages predicted by an epigenetic clock calculator, or by removing the variance associated with the chronological ages from the predicted DNA methylation ages in a GLM [93]. Marioni et al. [105] successfully associated epigenetic age acceleration with poorer cognition, lower grip strength and poorer lung function, while Belsky et al. [106] associated it with poorer balance, motor coordination and facial ageing. Other age-related associations include mortality [71], body mass index (BMI) [107] and frailty [108]. In Alzheimer’s patients, epigenetic age acceleration in prefrontal cortex tissue has been shown to correlate with amyloid load and the presence of plaques [109]. In Parkinson’s disease meanwhile, patients have displayed increased epigenetic age

acceleration compared to controls [110].

Less widely reported markers of biological age based on gene expression profiles [111], proteomics [112], metabolomics [113] and composites of physiological measures [114] have also been reported. All four approaches have resulted in significant associations with age-related phenotypes and clinical outcomes. The review by Jylhava et al. provides a more detailed description of these four methods [94].

The studies and results discussed above indicate that markers of biological age can potentially be very useful in predicting age-related phenotypes and diseases. However, it has also been shown that these biological markers do not necessarily correlate with each other [106], suggesting that different biomarkers relate to different aspects of ageing. It is then fair to assume that a biomarker derived from neuroimaging has potential to provide sensitivity to complementary aspects of biological ageing.

Franke et al. [115] were one of the first to apply the concept of a biological age to neuroimaging data. In the remainder of this study, we refer to neuroimaging-derived biological age as ‘brain age’. They trained a relevance vector regression (RVR) [116] model with a linear kernel on a set of features extracted from T1-weighted scans for healthy subjects, with chronological age as the target variable. When applying this model to a test set of healthy subjects, they found that the model was highly predictive, resulting in a mean absolute prediction error of 5 years. They then hypothesised that the prediction error may in fact be driven in part by real signal, i.e. when a test subject is predicted to have an older age than their chronological age, this represents accelerated ageing and the subject may be more susceptible to age-related conditions. They tested this by applying the model learned using the healthy subjects to Alzheimer’s patients and control subjects from the Alzheimers Disease Neuroimaging Initiative (ADNI) [117] cohort. The finding was that, on average, Alzheimer’s patients had predicted ages that were 10 years older than their chronological age, while for the control subjects the mean difference was zero, thereby providing some validity to the hypothesis. They term the difference

between predicted age and chronological age as the ‘brain age gap estimation’, which is analogous to the epigenetic age acceleration described earlier.

The study by Franke et al. reflects the approach that nearly all brain age related studies have chosen to take. A machine learning regression algorithm is typically trained on a set of healthy subjects to predict chronological age and this is then applied to test subjects to predict their brain ages. The difference between the chronological age and the predicted brain age for the test subjects, also termed brain age gap and brain predicted age difference, is subsequently used to make further inference. The idea behind this is that since training is carried out in healthy subjects, the regression model reflects the typical rate of healthy ageing in the study population. If the number of subjects in the training set is large, then the multivariate trajectory learned during training is assumed to reflect biological ageing. Therefore, if a test subject has a chronological age that is younger than that predicted by the regression model, the subject can be assumed to have been affected by accelerated brain ageing.

Age-related phenotypes that have been associated with non-neuroimaging markers of biological ageing have also been associated with brain age acceleration measures. Cole et al. [118] showed that brain age acceleration derived from T1-weighted images and Gaussian process regression (GPR) is associated with poorer grip strength, lung function, walking speed and fluid intelligence in addition to increased allostatic load and mortality risk. Another important result by Cole et al. was that epigenetic age acceleration was also significantly correlated with mortality, but epigenetic age acceleration and brain age acceleration did not correlate with each other. This supports the notion that neuroimaging markers may provide complimentary information to that from non-neuroimaging markers. Meanwhile, Steffener et al. [119] used regional grey matter volumes and multiple linear regression (MLR) to show that the level of education and physical exercise are associated with decelerated brain ageing. Using T1-weighted images and RVR, Luders et al. [120] showed that meditation is also associated with decelerated brain ageing.

Various neurological and neuropsychiatric conditions have been implicated in accelerated brain ageing. These include Alzheimer’s disease [115], mild cognitive impairment that develops to Alzheimer’s disease [121], schizophrenia [122], major depression [123], alcohol dependence [124], traumatic brain injury [125], Down’s syndrome [126], epilepsy [127] and neurocognitive disorder [128]. Diabetes [129] and HIV [130] have also been shown to be related to brain age acceleration. All of these studies implemented the approach originally reported by Franke et al. [115]. They used a multivariate regression model trained on a healthy cohort to estimate the brain ages of disease subjects in a test cohort, and showed that the mean of brain age minus chronological age for the disease subjects was significantly greater than zero. A more complete list of studies investigating brain age in disease can be found in table 1 of the review by Cole et al. [93].

Support vector regression (SVR) [123, 128], random forests (RF) [128], RVR [115, 121, 120, 129], GPR [118, 130, 126, 125, 127], MLR [119] and convolutional neural networks (CNN) [131] are some of the machine learning algorithms that have been used to train the age predictors. Where kernel methods have been used, i.e. SVR, RVR and GPR, the kernel has typically been a linear one.

PCA component weights from segmented T1 grey matter spatial maps [115, 121, 129], grey matter density maps from T1 [122], segmented grey matter maps from T1 [123, 125, 130, 126, 127], segmented white matter maps from T1 [125, 130, 126, 127], averages for atlas defined T1 grey matter regions [124], cortical thicknesses, cortical surface areas and subcortical volumes from T1 [128], functional connectivity matrices from rfMRI [128], PCA component weights from T1 cortical and subcortical volumes [119] and full T1 scans [131] are example of the types of features that have been employed in literature.

A brain age study by Ronan et al. [132], which is included in the review by Cole et al., stands out in its modelling approach. Instead of trying to predict chronological age, they used chronological age as the independent variable to model white matter volume, employing smoothing splines. They did this separately for obese and lean

subjects, and showed that for a given white matter volume the curve for the obese subjects was on average 10 years younger than that for lean subjects, i.e. the obese subjects underwent age-related changes 10 years sooner than lean subjects. They call this the ‘brain age’ difference since they used neuroimaging data. However, it is interesting that Cole et al. included this in their review, since this is technically the ‘chronological age’ difference. Ronan et al. do not make reference to their model as being a marker of biological ageing, nor do they derive such a measure. They simply have a different definition of brain age, and this inconsistency in the definition of brain age in neuroimaging literature should be kept in mind.

Nonetheless, the work by Ronan et al. does point to a major limitation that current brain age predictors have. They showed that white matter volume has an ‘inverted U’ trajectory when modelled as a function of chronological age. This type of trajectory has been reported elsewhere as well [133]. These trajectories are not invertible, since for a given value of the dependent variable, there are two values for the age. Therefore, current approaches that first model chronological age as a function of features will not be able to model these types of relationships, should some of the features exhibit it.

Another limitation arises from the fact that the brain age prediction models are first trained on a sample containing only healthy subjects. This means that the models are restricted to only identifying subjects that show accelerated ageing with respect to healthy trajectories. Although accelerated ageing has been shown to be an important risk factor for neurodegeneration, various studies have reported that diseases such as Alzheimer’s are not only related to accelerated ageing, and that they have their own distinct disease-related processes as well [134]. This suggests that understanding how brain ageing and the disease-related processes relate to each other is important. However, extending current approaches to model interactions between brain ageing and disease is not straightforward.

Furthermore, the machine learning algorithms currently used are driven by trying to minimise the error between the prediction and the target variable. Since current

approaches use chronological age as the target variable during training, important information can be lost. For example, there may be subjects in the training sample that have been labelled as healthy but go on to develop an age-related disease in the future, and therefore have brain ages that are very different to their chronological age. If their brain ages are driven by consistent features, then the machine learning algorithm may down weight these features so that it can predict a brain age that is closer to the chronological age.

Finally, although the power of current approaches is attributed to the the fact that the algorithms they use are multivariate, this also means that visualising trajectories becomes difficult.

In this study, we present a brain age model inspired by errors-in-variables (EIV) [135]. Models based on EIV, which are most popular in the field of econometrics [136], are regression models that account for the measurement error in the independent variable. In our case we have age as independent variable, and the ‘measurement error’ is the difference between chronological age and brain age, i.e. we say that chronological age is a noisy measure of brain age. We simultaneously model a set of features as functions of brain age. Our model is multivariate in that it uses information from all the features at once, estimates a separate visualisable trajectory for each feature, can model inverted U-shaped trajectories since age is the independent variable, can easily be extended to model disease interactions and takes a fundamentally different approach to estimating brain age that does not rely on minimising chronological age prediction error.

Outside of neuroimaging, a similar formulation to ours exists, reported by Klemra et al. [137]. Although they do not explicitly state that theirs is an EIV-based model, the fundamental principles are the same, i.e. they consider chronological age as a noisy measure of biological age and set biological age as the independent variable. However, to our knowledge, an application of such a formulation to brain ageing has not been reported in neuroimaging literature and, outside of neuroimaging, only applications of a linear form of Klemra et al.’s model has been explored.

We use extensive simulations to investigate the behaviour of our method, apply it to the UK biobank neuroimaging dataset and compare it to a commonly used current approach.

3.2 Methodology

3.2.1 The Model

In our model, there are N subjects, with each having a chronological age a_n . Each subject then has M features, with each feature denoted by y_{nm} . We say that there exists a function that governs the trajectory of each feature with age, but that this trajectory is a function of each subject's brain age a'_n as opposed to the chronological age. Therefore:

$$a_n = a'_n + \delta_n \quad (3.1)$$

$$y_{nm} = f_m(a'_n, \boldsymbol{\theta}_m) + \epsilon_{nm} \quad (3.2)$$

where δ_n is the shift from the brain age, ϵ_{nm} is the error associated with features and $\boldsymbol{\theta}_m$ is the function parameter vector associated with feature m . We assume that δ_n and ϵ_{nm} are independent and come from normal distributions with variances ν^2 and σ_m^2 respectively. Hence, $a_n \sim \mathcal{N}(a'_n, \nu^2)$ and $y_{nm} \sim \mathcal{N}(f_m(a'_n, \boldsymbol{\theta}_m), \sigma_m^2)$. The model is shown graphically in figure 3.1.

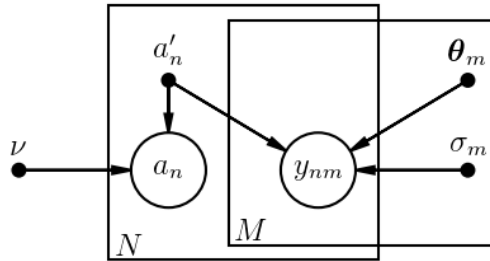


Figure 3.1: A graphical model illustrating our formulation. We observe the chronological age, a_n , and the features y_{nm} . However, we assume that the features are in fact a function of the brain age, a'_n , which is a hidden variable that we infer.

The probability of observing chronological ages in a vector $\mathbf{a}$ (of length n) and

features in a matrix $\mathbf{Y}$ (of size $n \times m$) becomes:

$$p(\mathbf{a}, \mathbf{Y}) = \prod_{n=1}^N \mathcal{N}(a_n | a'_n, \nu^2) \prod_{m=1}^M \mathcal{N}(y_{nm} | f_m(a'_n, \boldsymbol{\theta}_m), \sigma_m^2) \quad (3.3)$$

By taking the natural logarithm and multiplying by -1, we see that maximising equation 3.3 is equivalent to minimising the following cost function

$$J = \sum_{n=1}^N \left(\frac{(a_n - a'_n)^2}{2\nu^2} + \sum_{m=1}^M \frac{(y_{nm} - f_m(a'_n, \boldsymbol{\theta}_m))^2}{2\sigma_m^2} \right) \quad (3.4)$$

with respect to a'_n and $\boldsymbol{\theta}_m$.

The cost function now has a straightforward interpretation. It can be seen from the second term that the model looks to shift each subject's chronological age to get a better model fit for all the features. However, it can potentially overfit and shift them too much. Therefore, the first term in the cost function acts as a regulariser and penalises over shifting. The ratio of ν^2 to σ^2 meanwhile sets the optimal strength of this regularisation.

One major issue in seeking to optimise equation 3.4 is that ν and σ cannot be directly optimised. Two possible options are to estimate them beforehand or to add appropriate priors to the model. We choose to estimate σ_m beforehand to save on computation and let the user set ν based on prior expectations.

In this study, we consider the following forms for $f_m(a'_n, \boldsymbol{\theta}_m)$:

Polynomials

For a polynomial of order P:

$$f(a'_n, \boldsymbol{\theta}) = \sum_{p=0}^P h_p a_n'^p \quad (3.5)$$

where $\boldsymbol{\theta} = [h_0 \dots h_P]$.

Basis splines

The formulation of basis splines (BS) we describe here comes from the formulation described in the technical report by Racine [138]. First a vector $\mathbf{t}^* = [t_0 \leq \dots \leq t_{T+1}]$ is defined that contains the sequence of non-decreasing basis function positions, called knots. The elements t_0 and t_{T+1} are the boundary knots and there are T interior knots. Typically, the boundary knots are set to be the start and end range of the independent variable and the interior knots are spaced equally between these. Since we need to allow for shifts of the independent variable in our model, we use a different approach to define $\mathbf{t}^*$ that we will discuss later.

Next, to make the formulation easier, t_0 and t_{T+1} are prepended and appended to the knot vector $p-1$ times respectively, where p is the order of the BS function. Note that the ‘degree’ of the BS function, which we denote by q , is equal to $p-1$. This new vector is called the augmented knot vector, $\mathbf{t}$, and is indexed by $i = 0 \dots T+2p-1$. A set of functions $B_{i,j}$, where $j = 0 \dots q$, is then recursively defined for each element in $\mathbf{t}$ like so:

$$B_{i,j} = \begin{cases} 1 & \text{if } t_i \leq a'_n < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

$$B_{i,j+1}(a'_n) = \gamma_{i,j+1}(a'_n)B_{i,j}(a'_n) + [1 - \gamma_{i+1,j+1}(a'_n)]B_{i+1,j}(a'_n) \quad (3.7)$$

where

$$\gamma(a'_n) = \begin{cases} \frac{a'_n - t_i}{t_{i+j} - t_i} & \text{if } t_{i+j} \neq t_i \\ 0 & \text{otherwise} \end{cases} \quad (3.8)$$

Finally, the BS function is defined as:

$$f(a'_n, \boldsymbol{\theta}) = \sum_{i=0}^{T+q} \beta_i B_{i,q}(a'_n) \quad (3.9)$$

where $\boldsymbol{\theta} = [\beta_0 \dots \beta_{T+q}]$.

Logistic/Sigmoid

The logistic function is defined as:

$$f(a'_n, \boldsymbol{\theta}) = \frac{h}{1 + e^{-k(a'_n - a_0)}} + c \quad (3.10)$$

where $\boldsymbol{\theta} = [a_0, k, h, c]$.

Exponential

The exponential function is defined as:

$$f(a'_n, \boldsymbol{\theta}) = h e^{k(a'_n - a_0)} + c \quad (3.11)$$

where $\boldsymbol{\theta} = [a_0, k, h, c]$.

For use with constructing the Jacobian, the derivatives of equations 3.4, 3.5, 3.9 and 3.10 with respect to a'_n and $\boldsymbol{\theta}_m$ are given in appendix A.

3.2.2 Optimisation

Equation 3.4 was minimised using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [139, p.55], implemented in the ‘optimize.minimize’ module from SciPy v1.1.0 [140]. For all forms of $f(a'_n, \boldsymbol{\theta})$ except splines, we kept the default SciPy options.

When optimising with splines, we first normalised the data so that each feature has unit variance and zero mean. We then set a parameter tolerance of 10^{-2} , since perturbations of this magnitude to the spline parameters do not cause a substantial change to the trajectory of a normalised feature. Moreover, it translates to only 3.65 days for the brain age parameters. The BFGS implementation in SciPy does not

actually have an option for the parameter tolerance. We instead let the optimisation run for at least 500 iterations, and used a callback function for every iteration after this to check whether the mean absolute change in the parameter values for the last five iterations was less than 10^{-2} .

The brain ages, a'_n , were initialised using the subjects' chronological ages, a_n . The trajectory parameters, θ_m , were initialised by first fitting the trajectory function to the chronological ages for each feature separately, and taking the corresponding trajectory parameter estimates.

3.2.3 Simulated data

3.2.3.1 Simulation method

With the purpose of gaining an understanding of the model, we simulated data to reflect the process shown in figure 3.1. First, a vector $\mathbf{a}'$ of length N was created, containing the brain ages for N subjects. These were uniformly spaced between 0 and 90. Then the same function $f(a'_n, \theta_m)$ was used to create M features for each subject, keeping all the parameters, apart from the offset parameter, in θ_m the same. We simulated data for a variety of functions and these are shown in figure 3.2.

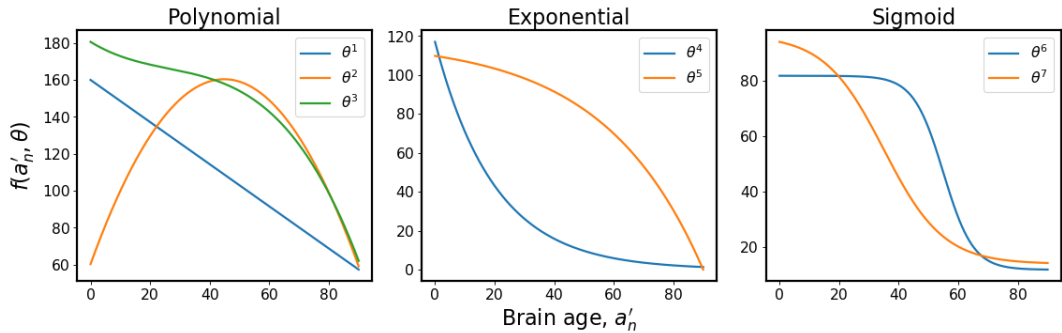


Figure 3.2: The functions we used when simulating data. The vector of parameters, as defined in section 3.2.1, are as follows. Polynomials: $\theta^1 = [102.6, -1.14]$, $\theta^2 = [60.33, 4.46, -0.0497]$, $\theta^3 = [180.6, -0.9933, 0.0253, -0.000321]$. Exponentials: $\theta^4 = [0, -0.05, 117, 0]$, $\theta^5 = [90, 0.03, -117.7, 117.7]$. Sigmoids: $\theta^6 = [55, -0.2, 70.2, 11.7]$, $\theta^7 = [35, -0.1, 82.8, 13.8]$. These seven options will be repeatedly referred to throughout this chapter.

Next, noise was added to the features so that on average they have a target

signal-to-noise ratio (SNR, defined by $\text{variance}(\text{signal})/\text{variance}(\text{noise})$), denoted by s^* . Since in real data the features are likely to be comprised of different modalities, we split the features into three equal sized groups, having SNR's of s^* , $1.5s^*$ and $0.5s^*$ respectively. Noise with variance ν^2 was then added to each subject's brain age to simulate their corresponding chronological age. For a given complete simulated dataset, all the features had the same type of noise, and similarly all the subjects had the same type of noise added to their brain age. The type of noise distributions we investigated were Gaussian with mean 0, Gumbel with mode 0 and scale 3, and Laplace with mode 0 and scale 1, as implemented in Numpy v1.14 [141]. When using a Gaussian, the variance of the distribution was set pre-sampling to give the desired SNR. For the others however, the noise vector was multiplied by a constant factor post-sampling to give the desired SNR.

Unless otherwise stated, the above procedure and Gaussian noise were used in all simulation experiments.

3.2.3.2 Estimating σ_m

We choose to estimate σ_m before fitting equation 3.4 to the data. Therefore, we investigated two methods to do this. The first was fitting a polynomial of degree 6 separately to each feature, so that each feature is modelled as a function of chronological age, i.e. $y_{nm} = f_m(a_n)$. The noise variance parameter for each feature is then estimated using:

$$\hat{\sigma}_m^2 = \sum_{n=1}^N \frac{(y_{nm} - f_m(a_n))^2}{N - 1} \quad (3.12)$$

Polynomials were chosen since numpy has an implementation that can quickly fit them to many features, thereby adding little to the overall optimisation time of our method. We set the degree to 6 since it can, to an extent, model non-linearities but is not so high so as make over fitting likely.

The second method was binning the data for each feature into age groups (again, using the chronological age) so that there are roughly 15 subjects in each bin. The variance of the data in each bin is calculated, and the mean variance across all bins

for a given feature is subsequently the σ_m^2 estimate for that feature.

The latter method is more reliable in that it is likely to give a viable estimate regardless of the shape of the trajectory, whereas the former is susceptible to optimisation issues such as non-convergence. On the other hand, binning will always overestimate the variance to a degree. This is due to the mean of the bin being true for the subjects at the bin centre, but not for the subjects near to the bin edges. This is something the first method tries to avoid by modelling the true mean for all subjects (assuming the trajectory can be modelled well by a polynomial), although this becomes less important with more subjects.

We compared the two methods on simulated data, using the mean estimation error across features as an indicator of performance:

$$\varepsilon_\sigma = \sum_{m=1}^M \frac{(\hat{\sigma}_m - \sigma_m)/\sigma_m}{M} \quad (3.13)$$

where σ_m is the real value with which we simulated the data, while $\hat{\sigma}_m$ is the estimated value. Data was simulated with $M = 10$ and we measured how the performance of the two methods changes with N , s^* , ν and trajectory type. The trajectories we tested were linear, sigmoid and polynomial with degree three, having parameter vectors $\boldsymbol{\theta}^1$, $\boldsymbol{\theta}^3$ and $\boldsymbol{\theta}^7$ respectively (defined in section 3.2.3.1). For this experiment we did not split the features into three SNR groups.

3.2.3.3 Sensitivity to σ_m and ν

Our methods of estimating σ_m prior to optimisation will give incorrect estimates, since they use chronological age. However, the idea is that they will still work well as a rough guess and not affect the accuracy of the optimisation output too much. To test this, we designed an experiment using simulations.

We simulated data and ran our optimisation routine on it, giving it the correct ν and function type, but the incorrect σ_m . The following equation was then used to measure how well our method predicted the brain ages of the subjects in the

simulated data:

$$\varepsilon_{BA} = \sum_{n=1}^N \frac{|\hat{a}'_n - a'_n|}{N} \quad (3.14)$$

where $\hat{a}'_n$ is the predicted brain age and a'_n is the ground truth brain age from the simulation. This represents the mean absolute brain age prediction error per subject. Data was simulated with $M = 10$ and $N = 500$, and we tested how the sensitivity to σ_m varies with s^* , ν and trajectory type (θ^1 and θ^3). Each simulation run was repeated 10 times to get a mean ε_{BA} for a given incorrect σ_m value. As before, the features were not split into three SNR groups.

The above experiment was also carried out to measure how sensitive the optimisation is to ν . This time the correct σ_m and an incorrect ν was given to the optimisation.

3.2.3.4 Effects of various data attributes

Next, we investigated how the performance of our EIV method is affected by N , M , s^* and s_a^* . We defined s^* as the mean SNR of the features. Similarly, s_a^* is the SNR of the chronological ages, defined as:

$$s_a^* = \frac{\text{variance}(a'_n)}{\text{variance}(\delta_n)} \quad (3.15)$$

We gauged the performance using the brain age prediction error, shown in equation 3.14, and the model prediction error, ε_f . To measure ε_f , we first created a vector $\mathbf{x}$ of length 1000 containing uniformly spaced numbers between the minimum and maximum simulated brain ages. The following equation was then used:

$$\varepsilon_f = \sum_{m=1}^M \sum_{i=1}^{1000} \frac{(\hat{f}_m(x_i, \hat{\theta}_m) - f_m(x_i, \theta_m))^2}{M} \quad (3.16)$$

This equation represents the mean (per feature) sum of squared distance between the estimated trajectory, with form $\hat{f}_m$ and parameters $\hat{\theta}_m$, and the simulated trajectory, with form f_m and parameters θ_m .

We first simulated data with $N = 1000$, $\nu = 5$, $s^* = 12$ and Gaussian noise, and tested how the above two metrics change with M for three different trajectories, θ^1 , θ^3 and θ^7 . Afterwards we set $M = 66$ (this value is in the mid-range of the M values we investigated) and tested how ε_{BA} and ε_f change with N . Each simulation run was repeated 10 times to obtain mean performance metrics.

We also probed any interaction between s^* and s_a^* for the same three trajectories. We set $M = 30$, $N = 500$, and simulated data with different s^* and s_a^* by changing σ_m , ν and the range of a'_n . Each simulation run was then repeated 20 times.

All the data in this subsection were simulated using the procedure described in section 3.2.3.1. When running our method on the simulated data, the σ_m values were estimated using the binning method and the correct ν value was given to the optimisation. In addition, we assumed that the trajectory model, including which of the trajectory's parameters are global to all features, is known. For example, if the trajectory is a sigmoid, $N = 500$ and $M = 50$, we told our method to predict 553 values. 500 of these are the subjects' brain ages, 3 are the global trajectory parameters (h, k, a_0) that are the same for all features, and 50 are the local parameters (c) that are different for each feature.

3.2.3.5 When the trajectory is not known

Up to now, when applying our method to simulated data, we assumed that the trajectory is known and gave the correct function type to the optimisation. However, in real situations it is unlikely that the user knows the correct trajectory. Therefore, we investigated how our method performs when the optimisation is carried out using a different function to that used in the simulation.

We simulated data with $N = 1000$, $M = 90$, $s^* = 20$, $\nu = 10$ and a sigmoid trajectory with parameters θ^7 . A polynomial with degree three and splines were considered as functions to use in the optimisation, and their performance was compared to that of a sigmoid trajectory.

In our method, setting the boundary knots to be the start and end range of the

chronological ages is not a viable option. This is due to the possibility that some subjects may have brain ages that go outside this range, which the trajectory needs to be able to model. We tested two different solutions.

The first option was extending the lower and upper boundary knots to be $\min(a_n) - \nu$ and $\max(a_n) + \nu$ respectively. We only extended by one factor of ν to preserve the consistency of the knot spacing as much as possible, while still accounting for any brain ages that go outside of the chronological age range. We call this approach ‘splines extend’. The second option was setting the boundary knots in the same way as the first option, but also adding $\min(a_n)$ and $\max(a_n)$ as extra interior knots to the knot vector. We call this approach ‘splines extra’. We set the degree of the spline to three, and used 5 interior knots (for ‘splines extra’ these are in addition to the two already described).

Using the same simulated dataset, the dependence of the brain age and model errors on degree for polynomials, and number of interior knots for splines, was also measured. From this section onwards, we modified the way we measured the model errors. Before calculating equation 3.16, we normalised the outputs of $\hat{f}_m(\mathbf{x})$ and $f_m(\mathbf{x})$ to have zero mean and unit variance. This is because, as described in section 3.2.2, we normalise the features when fitting with splines. For all experiments in this section, each simulation run was repeated 10 times.

3.2.3.6 Sensitivity to initialisation

In this section, the sensitivity of the optimisation to its initialisation was examined for splines and polynomials. In the experiments before this section, the optimisation had always been initialised once, as described in section 3.2.2. If we define this initialisation vector as $\mathbf{p}_0$, we ran the optimisation an extra nine times, with each run getting initialised by a perturbed version of $\mathbf{p}_0$. The perturbed initialisation vector was created using this equation:

$$\mathbf{p}_{0,\text{perturbed}} = \mathbf{p}_0 + \mathbf{r} \odot \mathbf{p}_0 \quad (3.17)$$

where the vector $\mathbf{r}$ is a random vector of the same length as $\mathbf{p}_0$ and $\odot$ denotes element-wise multiplication. The elements of $\mathbf{r}$ were randomly sampled from $\mathcal{N}(0, \gamma^2)$ and we refer to γ as the magnitude of perturbation. We then calculated two measures of sensitivity. The first was the standard deviation of the converged cost function value across these 10 initialisations (10 since $\mathbf{p}_0$ is included). The second was the mean standard deviation of each subject’s estimated brain age across the initialisations. In addition, to gauge whether our method of creating $\mathbf{p}_0$ is useful, we calculated the fraction of perturbed initialisation vectors that resulted in a lower cost function value than $\mathbf{p}_0$.

This was implemented on data that was simulated with $N = 500$, $M = 30$, $s^* = 20$, $\nu = 10$ and $\boldsymbol{\theta}^7$. The sensitivity measures were calculated for a range of γ and each simulation run was repeated 10 times.

3.2.3.7 Non-Gaussian noise

Real data is unlikely to be compromised of just Gaussian noise. Therefore, in this section we investigated how the performance of our method is affected when the data has non-Gaussian noise. Data was simulated using the procedure in 3.2.3.1, except that the noise added to the brain ages and features came from a Gumbel distribution with mode 0 and scale 3, and a Laplace distribution with mode 0 and scale 1, respectively.

The simulation parameters were $N = 1000$, $M = 90$, $s^* = 20$ and $\nu = 10$. Both sigmoid and exponential trajectories were tested, and each simulation run was repeated 10 times.

3.2.3.8 Comparison to other methods

The final experiment using simulations was to compare the performance of our method to the approach that is commonly used in literature. The current approach involves first training a model to predict chronological age. The model is then applied to test data and the ages that the model predicts become the brain age

estimates for the test subjects.

We first implemented the method by Cole et al. [118]. We split the simulated data into training and test sets and trained a Gaussian process regression (GPR) with a dot product kernel, using the features and chronological ages as the independent and dependent variables, respectively.

GPR with a dot product kernel is designed to model linear relationships. Since many of our simulations contain non-linear trajectories, we also implemented a method that uses kernel ridge regression (KRR) with a radial basis function (RBF) kernel. Again, the features and chronological ages were set as the independent and dependent variables, respectively.

The ‘sklearn’ package [142] was used to implement both of these methods. The training and test sets were created using a 3-fold cross validation. We utilised a 5-fold inner cross validation step, as we described in section 2.2.5 and figure 2.3, to optimise the ‘alpha’ and ‘gamma’ RBF parameters for KRR. This was not done for GPR, since its parameters are automatically optimised during training.

We applied these methods to data that was simulated with $N = 1000$, $M = 90$, $s^* = 20$ and $\nu = 10$, and compared their performance to our method, using polynomials and splines as the fitting functions. All the trajectories in figure 3.2, with both Gaussian and non-Gaussian noise, were tested. Non-Gaussian noise was generated using the procedure from the previous section. All simulation runs were repeated 10 times.

Our method is completely blind to the ground truth brain ages and function parameters, and explicitly has a term for brain age. Therefore, it is valid to directly use the brain age estimates during training for later comparison. However, for the sake of completeness, we also gauged how the performance of our method compares when cross validation is used to derive the brain age estimates. A 3-fold cross validation was used as with GPR and KRR, and the brain age of a test subject was

estimated by minimising the following cost function:

$$J^* = \sum_{m=1}^M \frac{(y_m^* - f_m(a'^*, \hat{\boldsymbol{\theta}}_m))^2}{2\sigma_m^2} \quad (3.18)$$

with respect to a'^* , where $\hat{\boldsymbol{\theta}}_m$ is the vector of trajectory parameters learned during the fitting, y_m^* is the test subject's value for feature m , a'^* is the test subject's brain age and σ_m^2 is the noise variance for feature m . To estimate the brain age of a test subject, it is also possible to create a different cost function that also includes the regulariser term from equation 3.4. However, this would require knowledge of the test subject's chronological age, which current approaches are blind to. Therefore, to enable easier comparisons to current approaches, we chose to construct equation 3.18 instead. The σ_m^2 values were calculated with the binning method described in section 3.2.3.2 using the training subjects included in the fitting, except that the training subjects were binned into age groups based on their estimated brain ages.

3.2.4 UK Biobank data

The UK Biobank is a prospective study aiming to collect data such as biological samples, physical measures and questionnaires for 500,000 subjects [143]. An imaging extension was recently announced that aims to collect MRI, X-ray and ultrasound scans for 100,000 of these subjects [144]. MRI modalities include T1-weighted, T2-weighted, susceptibility-weighted, diffusion-weighted (dMRI), task fMRI (tfMRI) and resting-state fMRI (rfMRI) scans. As of May 2018, scanning and subsequent processing of the raw data such as registration, calculating rfMRI component maps and fitting the diffusion tensor to dMRI had been completed for 14521 subjects. In addition, 2641 imaging-derived phenotypes (IDPs) such as volumes of brain regions, white matter lesion load and connectivity strengths between pairs of regions had been calculated for the subjects. For more details on the scanning protocol and IDP calculation see the paper by Miller et al. [144], and for more details on data processing and quality control see the paper by Alfaro-Almagro et al. [145].

The age range of the scanned subjects as of May 2018 was 44 - 80. This makes the UK Biobank a viable dataset to apply our method to by using the IDPs as features. We used IDP data that had been de-confounded and normalised. De-confounding was performed by creating a confound matrix that contained scanner table position x , scanner table position y , scanner table position z , rfMRI head motion, tfMRI head motion, the squared versions of these confounds, head size scaling, sex and slow date-related drifts. The variance associated to these confounds was then removed using a GLM. Prior to de-confounding, normalisation was performed by feeding the IDPs and confounds through a rank-based inverse Gaussian transformation [144]. The normalisation and de-confounding had already been carried out for us.

There were 2751 subjects that had missing IDP data and we removed these subjects before applying our method. We then correlated the IDPs with the chronological ages of all the subjects to obtain 2641 Pearson correlation coefficient (r) values. The IDPs that had r values in the top 1000 were kept for further analysis. A list of these IDPs is included in appendix B. We also randomly sampled 3036 subjects each time we ran our method. Approximately equal numbers of subjects were sampled from each one-year age group to ensure uniformity. Figure 3.3 shows the distribution of the sampled subjects. Note that the number of subjects that had ages of 47 and 77 is less than the number of subjects in the other one-year age groups. This is because these were the maximum number of subjects available for these two one-year age groups. We could have sampled less subjects from the other one-year age groups so that the distribution was uniform with respect of these two age groups as well. However, this would have resulted in a lot less subjects for us to work with, so we felt that this was a good compromise. This sampling was repeated 10 times to gauge the variability of the results. Before sampling, subjects that had ages of 45, 46, 78, 79 and 80 were removed since there were very few subjects that had these ages, and we wanted to ensure uniformity as much as possible in the age distribution.

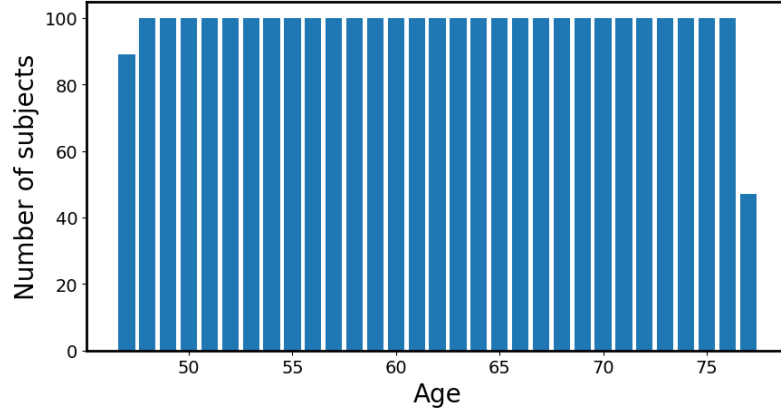


Figure 3.3: The age distribution of the sampled UK biobank subjects.

We applied our method to the sampled sets, testing both linear (polynomial of degree one) and polynomial of degree three curves as the fitting functions. We compared the results for $\nu = 1$, $\nu = 5$ and $\nu = 10$. The GPR and KRR methods described in section 3.2.3.8 were also applied. The fact that it is not possible to obtain ground truth brain ages for the subjects makes validation of the estimated brain ages difficult. We attempted validation by correlating the estimated brain ages with non-imaging variables in the UK Biobank. BMI, alcohol consumption, grip strength and blood pressure are all examples of non-imaging variables. A total of 1296 unique variables that could potentially be interesting in the context of brain age were identified. Many of these variables had been measured over multiple visits. We chose to keep the measurements from all the visits, resulting in 5456 values per subject that we could use for correlation.

Before correlating with the non-imaging variables, the variances associated with chronological age and chronological age squared were removed from the brain ages to account for any bias in the estimations. The non-imaging variables were also normalised and de-confounded in the same way as the IDPs. This time chronological age, chronological age squared, chronological age multiplied by sex and chronological age squared multiplied by sex were included as extra confounds.

For our method, we first included all 3036 subjects in the fitting and used their estimated brain ages for further correlations. We then investigated how these correlations change when the brain ages for test subjects not included in the fitting

are used in the correlations. For this, in the same way as for the GPR and KRR methods, we employed a 3-fold cross validation framework. Equation 3.18 was used to estimate the brain ages of the test subjects.

3.3 Results

3.3.1 Simulated data

3.3.1.1 Estimating σ_m

Figure 3.4 shows the results of comparing the polynomial fitting and binning methods when $N = 100$, while figure 3.5 shows them when $N = 1300$. The first thing to note is that both methods overestimate the value of σ when $\nu = 10$ for both values of N . This is to be expected since they use the chronological age and incorrectly attribute this extra source of error to σ . There is no obvious overestimation for the much smaller $\nu = 1$. The reason for the increase in the estimation error with SNR is the way we define ε_σ . As equation 3.13 shows, we do not plot the absolute error, and normalise it by σ instead.

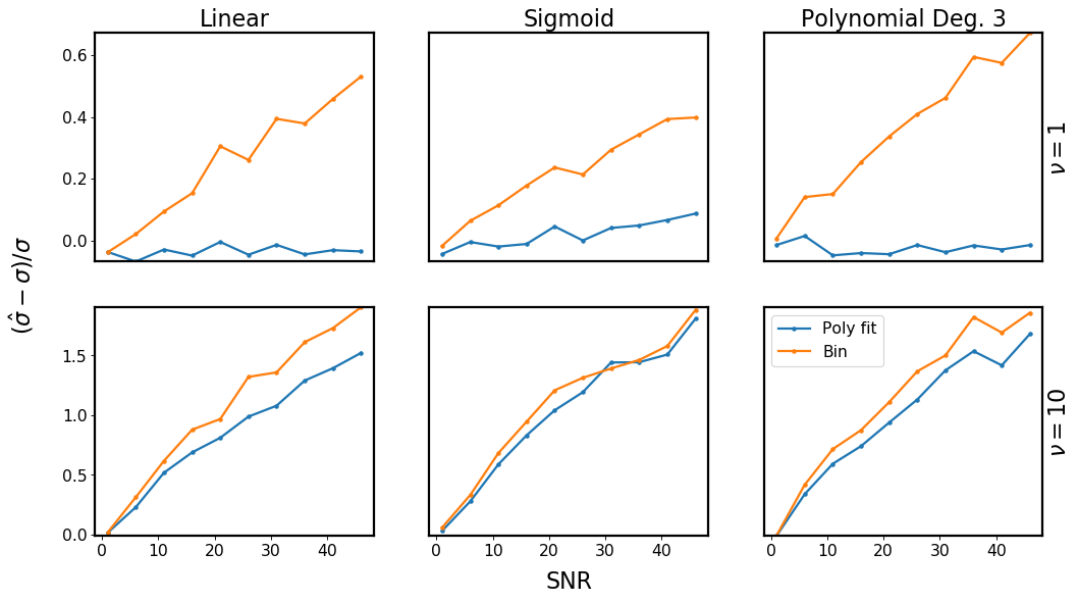


Figure 3.4: The estimation error, ε_σ , as a function of SNR, comparing the polynomial fit and binning methods. $N = 100$.

When both N and ν are small, the polynomial fit method performs better, but

in the other scenarios the differences are not as large. When N is large, the bin method performs slightly better for the sigmoid trajectory and the same for the other trajectories, suggesting it can perform just as well when the trajectory can be modelled by a polynomial, and better when it cannot.

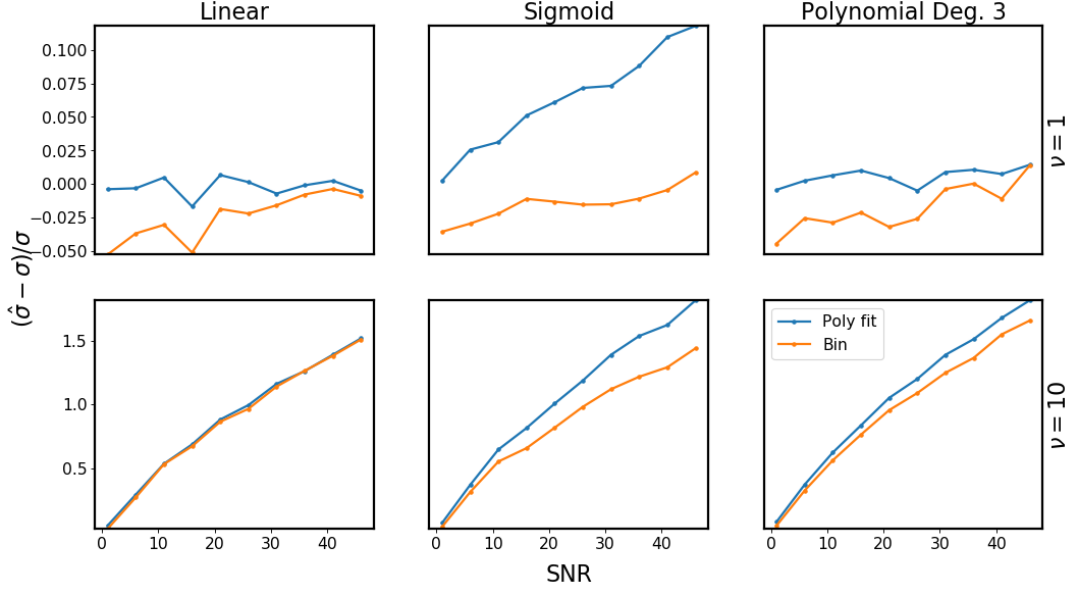


Figure 3.5: The estimation error, ε_σ , as a function of SNR, comparing the polynomial fit and binning methods. $N = 1300$.

We expect our method to be applied to large datasets such as the UK Biobank, and polynomial trajectories are also unlikely to exist in real datasets. Therefore, we concluded that using the bin method is the better option and used it to estimate σ for all other studies.

3.3.1.2 Sensitivity to σ_m and ν

In the previous section we saw that the binning method has the potential to overestimate σ_m by up to 150%, i.e. 2.5 times the real value. In this section we see how detrimental this overestimation is to predicting brain ages.

The results of our sensitivity experiment can be seen in figure 3.6. The x-axis shows how different the value of σ_m that we provided to the optimisation (denoted by $\hat{\sigma}_m$) was from the real σ_m (denoted by σ_m) that we used when simulating the data. The y-axis shows the the mean absolute brain age error per subject, ε_{BA} .

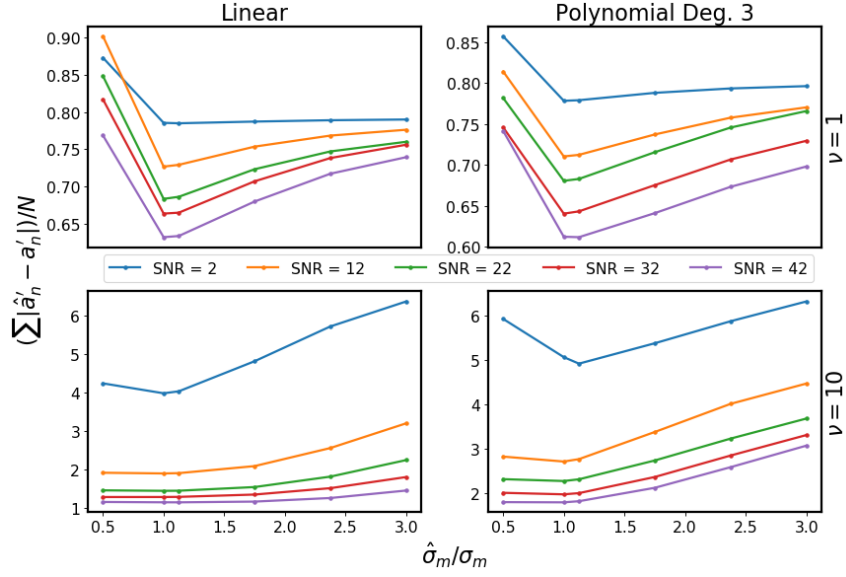


Figure 3.6: These four plots show how the brain age error is affected when an incorrect σ_m is given to the optimisation. The value we gave the optimisation is denoted by $\hat{\sigma}_m$, while the real value used in the simulation is denoted by σ_m .

We expect the initial ε_{BA} before optimisation to be approximately equal to the ν that the data was simulated with. Therefore, a ε_{BA} of just under 3 years when $\nu = 10$, SNR= 12, $\hat{\sigma}_m/\sigma_m = 1$ and the trajectory is a polynomial of degree three constitutes a substantial decrease from the initial error. When we fed in a σ_m that was 2.5 times the real value, this brain age error goes up by only one year and we still have a substantial decrease from the initial error. The same can be said for the other trajectory and SNR's. This suggests that although our binning method can overestimate the value of σ_m , the subsequent improvement in brain age error still warrants its use.

The figure also shows that when $\nu = 1$, the improvement in brain age error, when considered as a fraction of the initial error, is not as large. In later sections we see that it is the ratio of the ν^2 and σ_m^2 in the data that signifies how useful our EIV model will be in predicting brain age.

After measuring the sensitivity to σ_m , we conducted a similar experiment to measure the sensitivity to ν . Figure 3.7 shows the results. When $\nu = 10$ and the SNR is above 12, it seems as if the optimisation becomes completely insensitive to the ν value after it is overestimated, and always returns the optimal solution. A

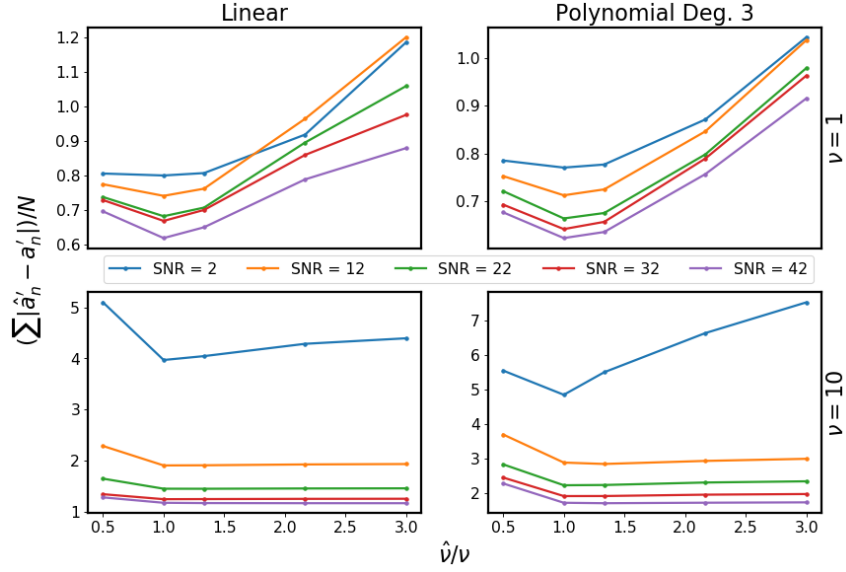


Figure 3.7: These four plots show how the brain age error is affected when an incorrect ν is given to the optimisation. The value we gave the optimisation is denoted by $\hat{\nu}$, while the real value used in the simulation is denoted by ν .

possible reason for this stems from the fact that ν effectively acts as a regulariser. It stops the optimisation from predicting brain ages that are too far from the original chronological ages. Therefore, when it is set below the optimal value, it hinders the optimisation in shifting the brain ages far enough from the chronological ages to obtain an optimal solution. On the other hand, when it is set above the optimal value, the optimisation has enough freedom to reach an optimal solution. It can then be argued that this is a reason to ignore the ν and σ_m terms and simply set the former to be much larger than the latter for all optimisations. However, as can be seen in the figure, this is not the case when $\nu = 1$ or the SNR is 2, and low SNR data is very likely in MRI. Hence we choose to keep ν as a definable parameter.

3.3.1.3 Effects of various data attributes

Figure 3.8 shows how the brain age error, ε_{BA} , and the model error, ε_f , change with the number of features. There is no obvious trend between the model error and the number of features. In contrast to this, there is a clear relationship for all trajectories, showing that the brain age error goes down as the number of features goes up.

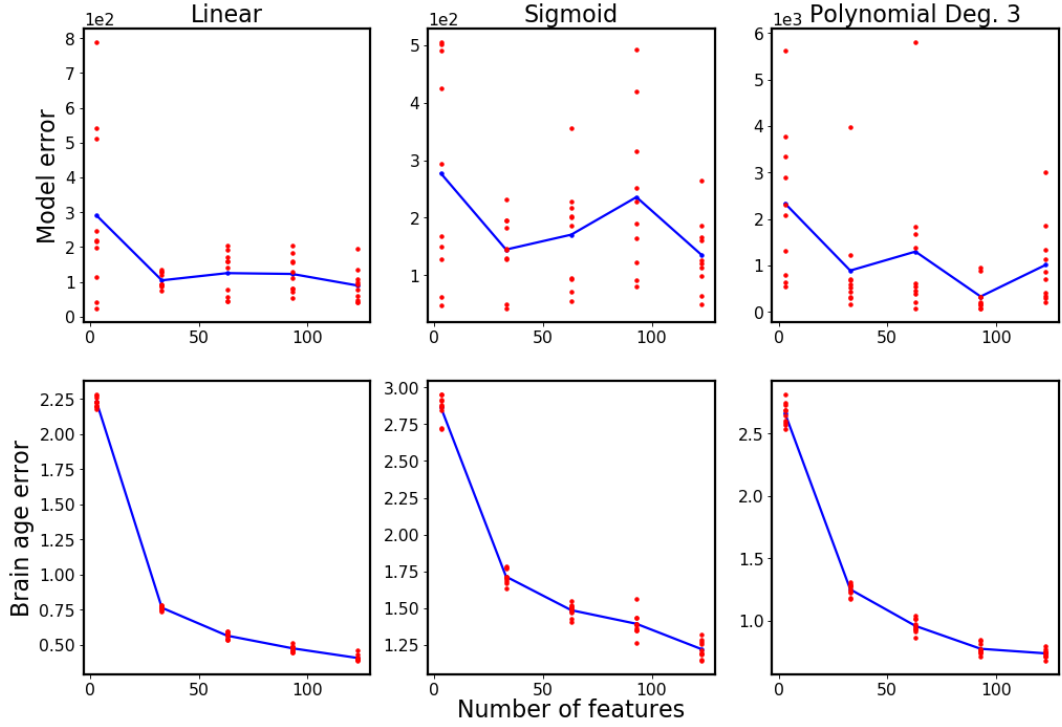


Figure 3.8: The brain age and model errors, defined by equations 3.14 and 3.16 respectively, as a function of the number of features in the dataset. The red points show the values for each of the 10 simulation runs separately, while the blue line shows the mean of the 10 values.

This relationship with M can be understood by considering figure 3.9. The left sub-figure shows one feature plotted against age, taken from data that has been simulated with $\nu = 5$, $s^* = 20$ and a linear trajectory. The black data points are the simulated chronological ages, and the aim of our method is to shift these back to the red data points, which are the simulated brain ages.

The central sub figure shows the case when our method is applied to this feature alone. The optimisation shifts the subjects' chronological ages such that brain ages lie nearly perfectly and erroneously on its predicted model, thus minimising the second term in equation 3.4 by setting it to 0. The total magnitude of the shifts is comparable to the amount that would be needed to shift them back to their correct brain ages, which means that the first term in the cost function, i.e. the regulariser, has no power in stopping this from happening. Therefore, the brain age error does not improve much.

However, when 9 extra features are added, which is the case in the right sub

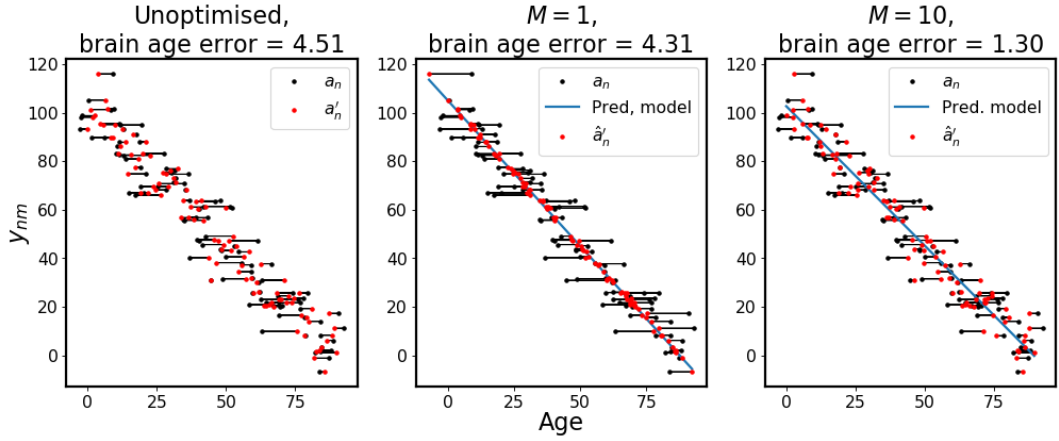


Figure 3.9: An illustration of how having more features improves brain age prediction.

figure, the brain age error drastically improves. This is because if the model tries to shift the chronological ages onto its predicted model for one of the features, the contribution to the second term in equation 3.4 from all the other features stay large. Hence the optimisation has to compromise between all the features to reach a minimum, and the more features there are, the more likely it is that the chronological ages will be shifted back to their true brain ages.

Figure 3.10 shows how the brain age and model error vary with the number of subjects. There is now a stronger trend between the model error and the number of subjects compared to figure 3.8. The relationship with the brain age error, meanwhile, is much weaker, with only the linear and polynomial trajectories showing any trend.

It is important to note that the relationships we have shown here are only valid within the range of parameters that we have simulated. For example, in figure 3.8, the model error for the linear trajectory does show an increase from the first to the second M value, suggesting that the trend saturates very quickly, as opposed to there being no trend at all.

In earlier sections, there were hints that the ratio of s^* (the average SNR of the features) and s_a^* (the SNR of the chronological ages, defined in equation 3.15) plays a role in determining how useful our model is. We simulated data for a range of s^* and s_a^* , and measured how much the brain age error improves from the initial state.

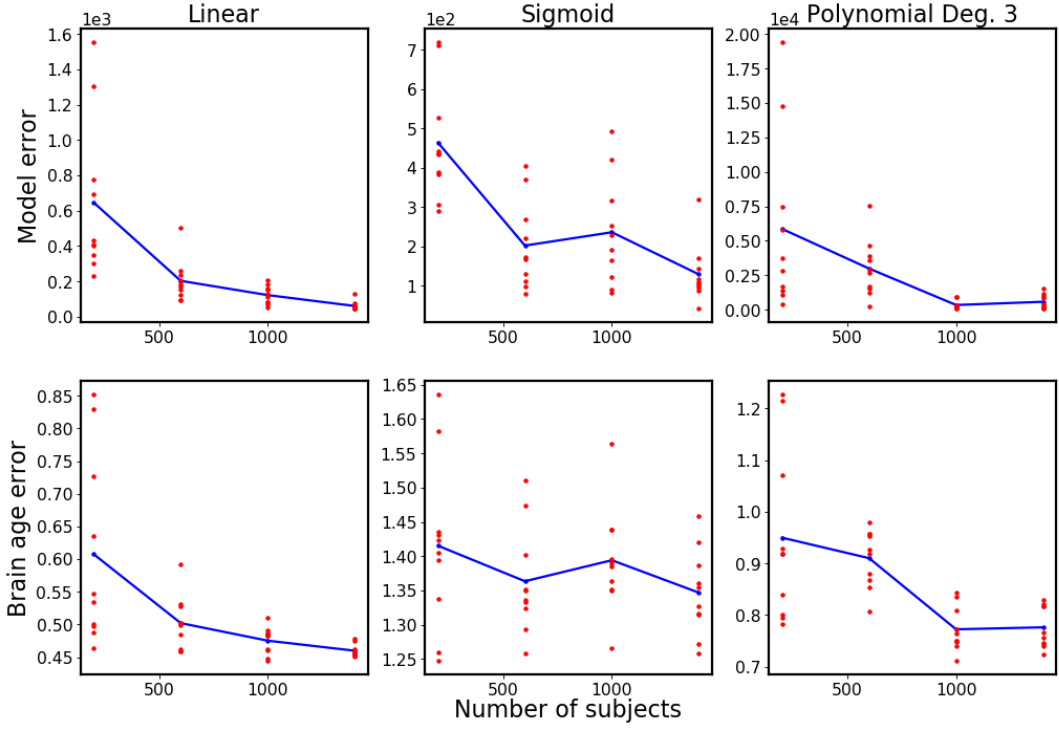


Figure 3.10: The brain age and model errors, defined by equations 3.14 and 3.16 respectively, as a function of the number of subjects in the dataset. The red points show the values for each of the 10 simulation runs separately, while the blue line shows the mean of the 10 values.

Figure 3.11 shows the results.

The linear trajectory exhibits a very clear relationship between the improvement in brain age error and the SNR ratio. This shows that there are cases when our model will not be useful. There is also a point after which the brain age error does not improve no matter how much the SNR of the features improves. To put this into perspective, if there is data for a uniform brain age range of 30 to 90, the variance of the brain age values (the numerator of equation 3.15) will be 300. If the expected brain age error is then 5, $s_a^* = 300/5^2 = 12$. Figure 3.11 shows that for the linear trajectory, the brain age error is halved compared to the initial state (i.e. $\Delta\varepsilon_{BA}/\varepsilon_{BA,initial} = 0.5$) when $\sqrt{s^*/s_a^*}$ is around 0.3. Therefore, by rearranging to obtain s^* , we see that the average SNR of the features would have to be at least $s^* = 0.3^2 \times 12 = 1.08$ in order to half the initial brain age error. This is true for the exact simulation parameters; if there were more features in the dataset, this value would decrease.

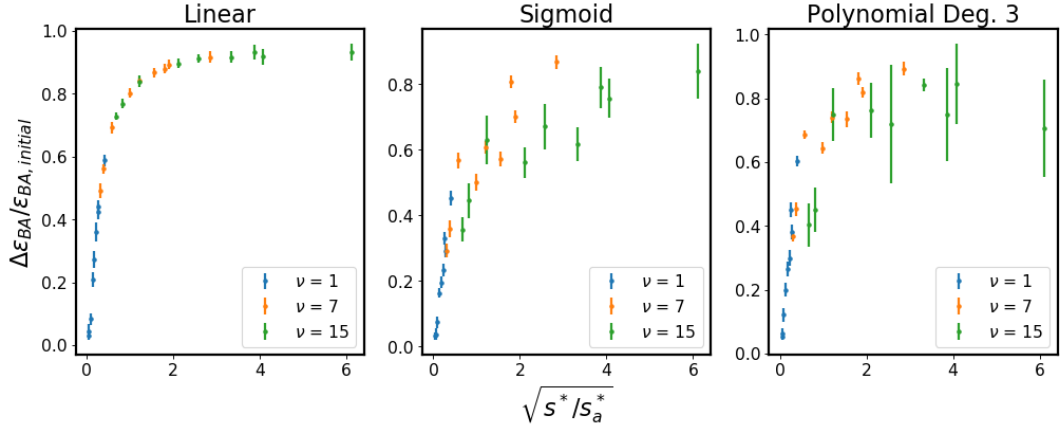


Figure 3.11: The fractional improvement in brain age error from the initial state, defined by $\Delta\varepsilon_{BA}/\varepsilon_{BA,initial}$ where $\Delta\varepsilon_{BA} = \varepsilon_{BA,initial} - \varepsilon_{BA,optimised}$, as a function of the SNR ratio. The average SNR of the features is s^* , and the SNR of the chronological ages is s_a^* , defined by equation 3.15.

There is also a trend for the other two trajectories, but the trend curve is shifted to the right, meaning a higher SNR ratio is required to get the same improvement as the linear trajectory. In addition, the error bars for these two trajectories are larger, especially for the higher ν value. This is likely due to the optimisation being sensitive to both the initialisation and the exact δ_n value for each subject for non-linear trajectories.

3.3.1.4 When the trajectory is not known

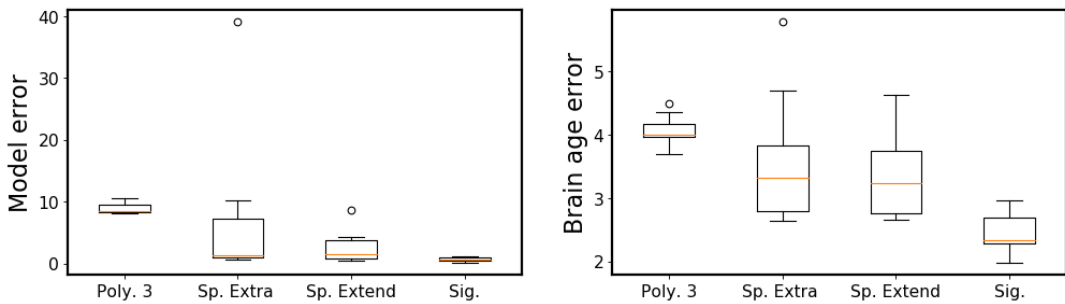


Figure 3.12: How ε_{BA} and ε_f are affected by different functions used in the fitting. The data was simulated using a sigmoid trajectory. Sig. = Sigmoid, Poly. 3 = Polynomial of degree three, Sp. = Splines.

We now present the case when the simulated data is fitted using a different trajectory function to that used in the simulation. Figure 3.12 shows how the model and brain age errors are affected when data is simulated using a sigmoid

trajectory, but fitted using polynomials or splines. For the brain age error, there is very little difference between the two splines methods. Similarly, there is little difference in the medians of the model error for the two splines methods, although ‘splines extend’ appears to be less variable. Compared to the polynomial of degree three, splines resulted in lower median brain age and model errors across the 10 repetitions. However, both in terms of model error and brain age error, the results from the polynomial are a lot less variable.

Neither splines nor polynomials performed as well as using the correct trajectory function, i.e. a sigmoid. However, the data was simulated using $\nu = 10$, meaning that the brain age error improvement that both splines and polynomials provided is substantial.

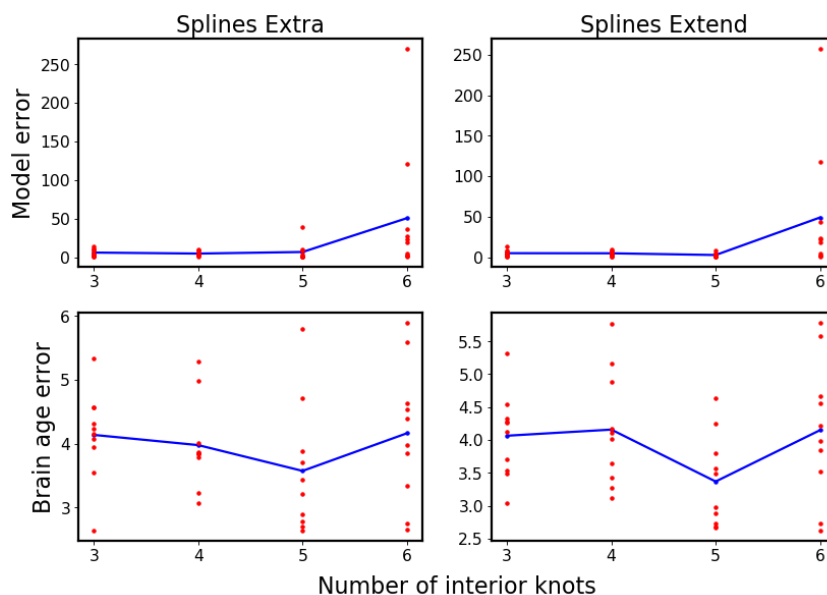


Figure 3.13: The change in the brain age and model errors with the number of interior knots, T , when splines are used to model the trajectory.

Figure 3.13 displays how the number of interior knots, T , used in the spline function affects the model and brain age errors. Both splines options show similar trends. The model error is fairly static until $T = 5$ and increases both in magnitude and variability when T increases to 6. There is not a clear change in the brain age error as T is varied, although it does appear as if the brain age error is smallest when $T = 5$. However, the difference in ε_{BA} at this T value compared to the other

T values is small relative to the overall variability.

The fact that the ‘splines extra’ and ‘splines extend’ methods provided such similar result suggests that using the ‘splines extend’ method may be the better option. This is due to the ‘splines extra’ method having two extra interior knots, thereby increasing the optimisation time.

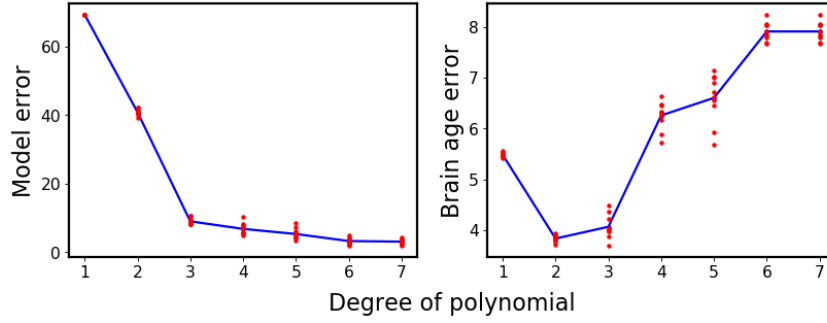


Figure 3.14: The change in the brain age and model errors with the polynomial degree, when a polynomial is used to model the trajectory.

The change in the model and brain age errors as a function of the polynomial degree is shown in figure 3.14. There is a drastic improvement from a degree of one to a degree of three, after which the model error keeps improving, but at a much slower rate. The brain age error is lowest when the degree is 2 and 3. Conversely to the model error, the brain age error worsens after degree three. This illustrates the point that there is not a monotonic function that maps the brain age error to the model error.

To discern how our method shifts the brain ages, we simulated data with $N = 1000$, $M = 30$, $s^* = 10$, $\nu = 10$ and θ^7 . We subsequently applied our method to the data using a polynomial with degree three, ‘splines extend’ and sigmoid as the trajectory functions during the fitting. Figure 3.15 shows the results.

The first thing to note is that all three fitting functions show a slight bias in their brain age estimation, since in the central figures younger subjects are more likely to be shifted to the right, and older ones to the left. A possible reason for this stems from the fact that our method has the freedom to shift subjects in the x -direction. Therefore, if a subject lies on the left of the best fit line, by shifting this subject to the right in the x -direction, it can bring it closer to the best fit line. Similarly, if a

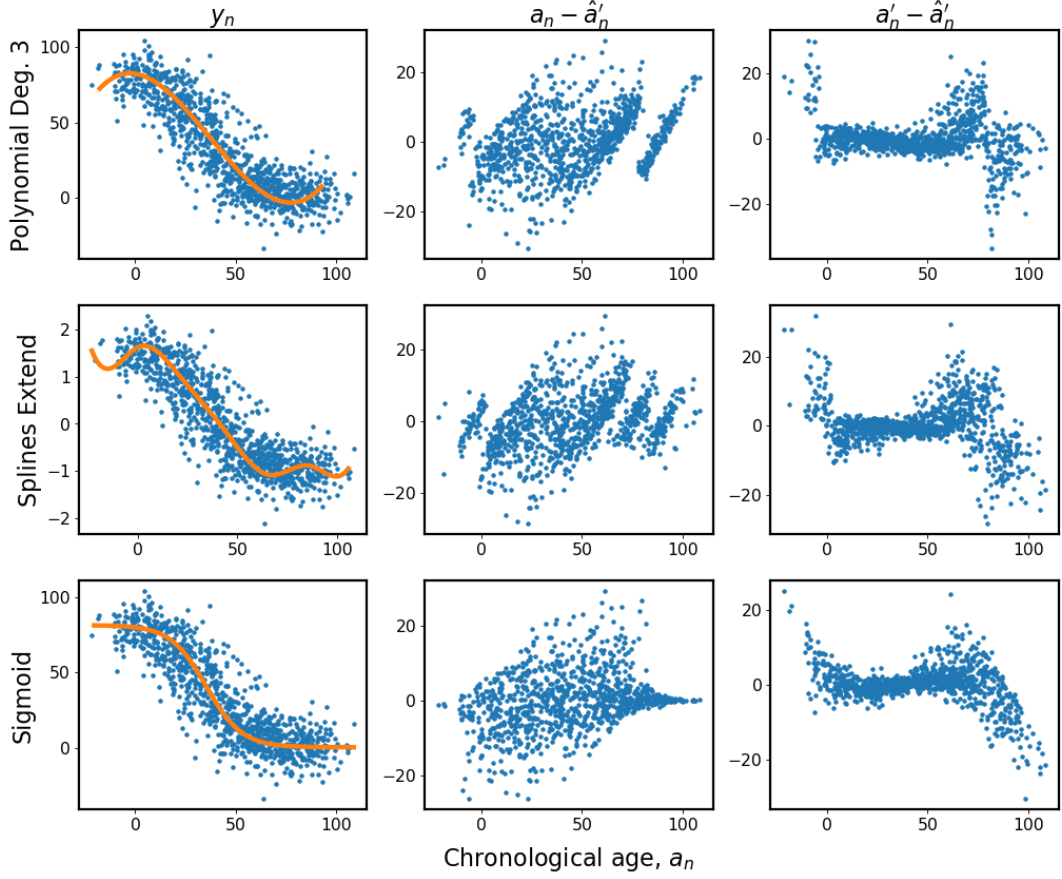


Figure 3.15: An illustration of how our method shifts the chronological ages, shown here for three different types of functions used during the fitting. The data was simulated using a sigmoid trajectory and the leftmost column of figures shows one of the features plotted against chronological age, with the orange line representing the fitted model. The central column of figures shows how much each subject’s chronological age is shifted, plotted against the chronological age itself. Positive and negative values indicate a shift to the left and right, respectively. The rightmost column of figures shows the error between the estimated and the groundtruth brain age for each subject, again plotted against the chronological age.

subject lies to the right of the best fit line, it can bring the subject closer to it by moving it to the left in the x -direction. By doing this, the optimisation can further minimise the second term in equation 3.4.

Another observation is that, in the rightmost column of figures, the brain age error at the edges of the chronological age range are very large. This is due to the simulated trajectory (θ^7 in figure 3.2) not varying much in the regions where $a_n < 10$ and $a_n > 60$. Hence the optimisation does not have enough information about the model in these regions to be able to shift the subjects into the correct positions. This is also a potential explanation for why both splines and the polynomial have

not correctly modelled the tails.

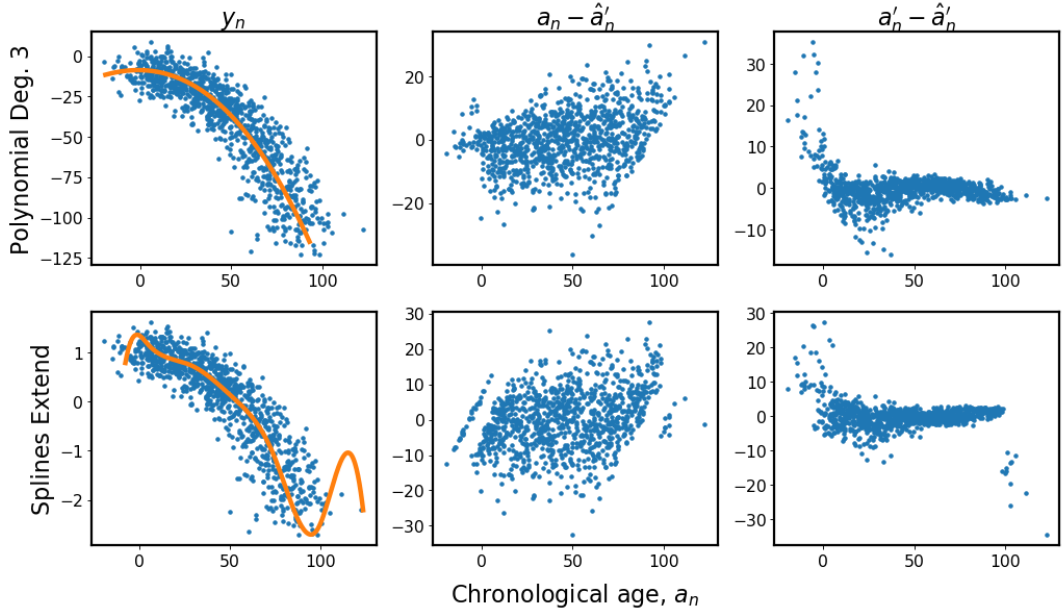


Figure 3.16: The same illustration as figure 3.15, but for the case when the data was simulated using an exponential trajectory.

Figure 3.16 shows the same results, but for the case when the simulated trajectory was an exponential (θ^5). The same behaviours of the brain age error being large in weakly varying regions, and the brain age estimation having a slight bias, can be seen here as well. However, the polynomial has now managed to more accurately model the right tail, where the trajectory is highly varying, while the splines function has not. This suggests that having boundary knots in regions where there is not any data is the main source of badly modelled tails in splines.

3.3.1.5 Sensitivity to initialisation

The results of examining the sensitivity of the optimisation to the initialisation are displayed in figure 3.17. The top row of figures show that for all three fitting functions, all optimisation runs were sensitive to initialisation, with polynomials the least sensitive and ‘splines extra’ the most. This is more evidence that the ‘splines extend’ option is better compared to ‘splines extra’. The bottom row of figures show that these sensitivities translate to sizeable differences in the brain age estimates. A standard deviation of 10^{-1} , i.e. 36.5 days, is acceptable when compared

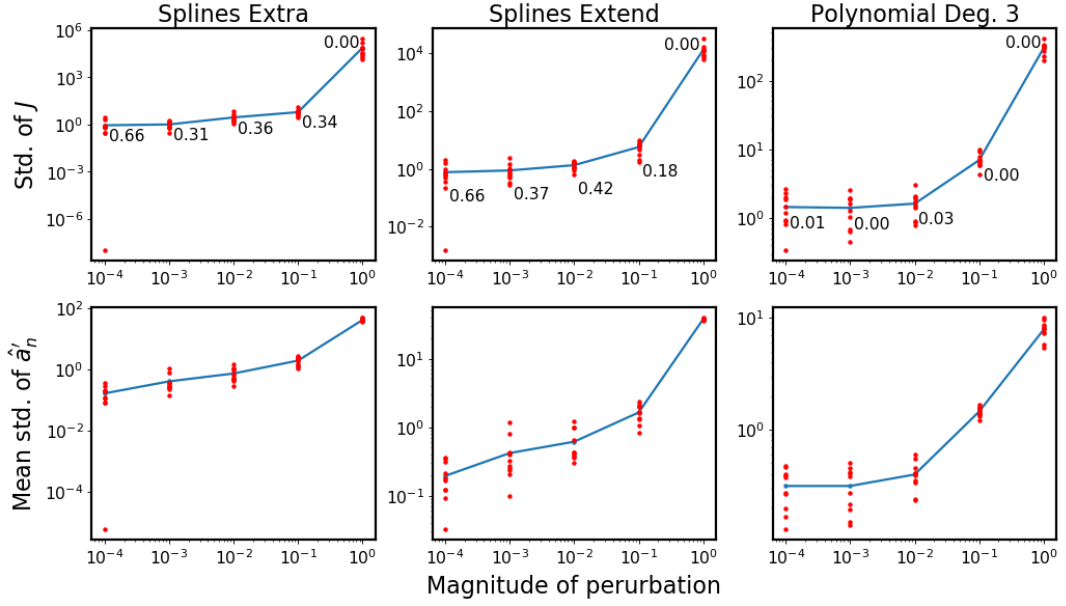


Figure 3.17: Each red data point in the top row of figures shows the standard deviation of the converged J across 10 different initialisations, while the blue line shows the mean of these standard deviations across 10 different repetitions. The numbers next to the data points show the fraction of perturbed initialisation vectors that gave a better J than $\mathbf{p}_0$ for a given magnitude of perturbation. Similarly, a red data point in the bottom row of figures shows the mean standard deviation of each subject’s estimated brain age, while the blue line shows the mean of these.

to typical brain age deltas reported in literature. However, for splines when the standard deviation approaches 1 year, a large fraction of the perturbed initialisation vectors still resulted in better converged cost function values. This means that when using splines with real data, our method should be initialised multiple times. For polynomials on the other hand, the $\mathbf{p}_0$ initialisation vector nearly always resulted in the best cost function value, suggesting extra initialisations in addition to this are not necessary.

3.3.1.6 Non-Gaussian noise

Figure 3.18 shows how the model and brain age errors change when the data is generated using non-Gaussian noise. Only results for the ‘splines extend’ method are shown since in previous sections there was evidence that this is superior to ‘splines extra’. The median model error for the sigmoidal data is higher for non-Gaussian data, while there is no difference in the model error for exponential data.

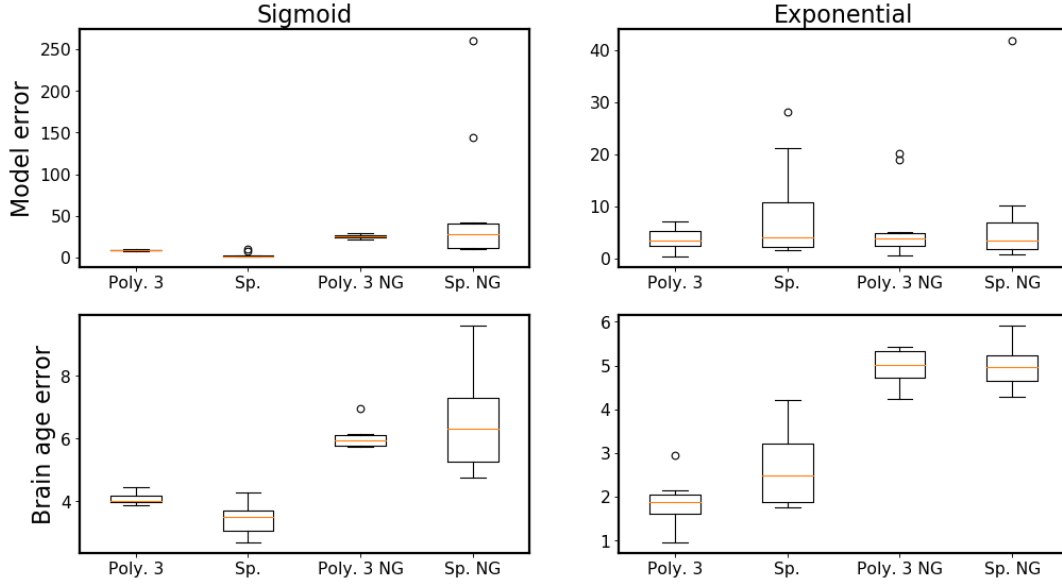


Figure 3.18: The model and brain age errors for data that has been simulated using sigmoid (left column) and exponential (right column) trajectories. Fitting was carried out using ‘splines extend’ (Sp.) and polynomial with degree three (Poly. 3). NG indicates that the results are for data that was generated with non-Gaussian noise. Results for data that was generated with Gaussian noise are included for comparison.

Both the sigmoidal and exponential data resulted in higher brain age errors when the data is non-Gaussian, meaning that our method cannot cope as well with non-Gaussian noise. However, this data was generated with $\nu = 10$ and these brain age errors are still a sizeable improvement, indicating that there is still some merit to using our method on data with non-Gaussian noise.

3.3.1.7 Comparison to other methods

Table 3.1 compares the brain age errors of our method to those obtained from implementing a common current approach used in literature. We do not compare the model errors since current approaches use multivariate methods and have age as the dependent variable. This makes it difficult to extract functions that model each feature as a function of brain age.

The table shows that if knowledge about the real trajectory is available, our method gives the lowest brain age error for all trajectories except linear θ^1 NG. However it is unlikely that this knowledge will be available, so we will not use the

	Splines	Poly3	GPR	KRR	Real
Linear θ^1	1.85 ± 0.50	2.25 ± 0.37	<i>1.13 $\pm$ 0.09</i>	3.66 ± 0.69	0.63 ± 0.09
Sig. θ^6	7.81 ± 1.04	<i>6.37 $\pm$ 0.12</i>	8.43 ± 0.05	6.71 ± 0.13	5.11 ± 0.30
Sig. θ^7	4.21 ± 1.18	4.14 ± 0.17	5.64 ± 0.07	<i>4.05 $\pm$ 0.50</i>	2.51 ± 0.24
Exp. θ^4	<i>3.15 $\pm$ 0.65</i>	3.24 ± 0.19	10.91 ± 0.06	3.88 ± 0.35	-
Exp. θ^5	2.24 ± 0.37	<i>2.03 $\pm$ 0.39</i>	7.34 ± 0.07	4.60 ± 1.18	-
Poly. θ^2	3.53 ± 0.45	<i>2.54 $\pm$ 0.17</i>	22.57 ± 0.04	23.13 ± 0.20	2.54 ± 0.17
Poly. θ^3	2.19 ± 0.76	<i>1.10 $\pm$ 0.13</i>	9.12 ± 0.07	3.17 ± 0.63	1.10 ± 0.13
Linear θ^1 NG	4.93 ± 0.39	4.67 ± 0.20	<i>4.55 $\pm$ 0.25</i>	5.29 ± 0.34	4.56 ± 0.22
Sig. θ^6 NG	7.15 ± 0.38	<i>6.91 $\pm$ 0.25</i>	9.04 ± 0.15	7.55 ± 0.27	6.29 ± 0.35
Sig. θ^7 NG	6.44 ± 1.12	6.12 ± 0.33	6.91 ± 0.22	<i>5.82 $\pm$ 0.37</i>	4.99 ± 0.32
Exp. θ^4 NG	<i>5.12 $\pm$ 0.40</i>	5.40 ± 0.51	11.83 ± 0.09	5.54 ± 0.28	-
Exp. θ^5 NG	5.07 ± 0.39	<i>4.98 $\pm$ 0.43</i>	7.39 ± 0.08	6.70 ± 0.85	-
Poly. θ^2 NG	5.27 ± 0.36	<i>5.21 $\pm$ 0.29</i>	22.77 ± 0.06	23.32 ± 0.25	4.90 ± 0.26
Poly. θ^3 NG	<i>4.96 $\pm$ 0.23</i>	5.12 ± 0.86	9.29 ± 0.05	5.46 ± 0.38	5.17 ± 0.97

Table 3.1: The brain age errors resulting from different brain age estimation methods. The splines and poly3 columns display the results for our method with ‘splines extra’ and a polynomial as fitting functions, respectively. The GPR and KRR columns display the results for a common current approach used in literature, with Gaussian process regression and kernel ridge regression as the prediction models, respectively. The ‘Real’ column displays the results for our method, but with the fitting function being the same as that used in the simulation. Each row represents the trajectory that was used in the simulated data, with the trajectory parameters given in figure 3.2. The best method’s brain age error (not including the ‘Real’ column) for a given trajectory has been bolded and italicised. NG signifies that the data was simulated using non-Gaussian noise.

‘Real’ column to draw any more comparisons.

When the trajectory is linear and noise is Gaussian, the GPR method gives the best brain age error. It also gives the best brain age error when the noise is non-Gaussian, but it and poly3 have values that are within each other’s margins of error. However, GPR struggles with non-linear trajectories, only giving sizeable brain age error improvements from the initial state ($\nu = 10$ in the simulations) for θ^5 and θ^7 . This is as a result of it using a dot product kernel.

The KRR method fares better with non-linear data due to its RBF kernel. It gives the best brain age error for only two of the cases, θ^7 and θ^7 NG, with the former being within the margin of error of poly3.

Both KRR and GPR give very large brain age errors, almost double the initial state, for θ^2 . This is because, as figure 3.2 shows, θ^2 has an inverted U shape. It is not possible to invert these into a single valued function, meaning that current approaches that have age as the dependent variable cannot cope. Overall, out of

	Poly3	GPR	KRR
Linear θ^1	1.94 ± 0.29	1.05 ± 0.10	3.33 ± 0.52
Sig. θ^6	8.41 ± 0.10	8.46 ± 0.08	6.81 ± 0.29
Sig. θ^7	5.25 ± 0.24	5.62 ± 0.10	4.17 ± 0.29
Exp. θ^4	3.25 ± 0.23	10.90 ± 0.05	3.81 ± 0.27
Exp. θ^5	1.90 ± 0.28	7.31 ± 0.04	4.88 ± 1.16
Poly. θ^2	23.50 ± 0.31	22.57 ± 0.04	23.96 ± 2.35
Poly. θ^3	1.12 ± 0.16	9.16 ± 0.03	3.51 ± 0.46
Linear θ^1 NG	4.19 ± 0.28	4.49 ± 0.27	5.23 ± 0.40
Sig. θ^6 NG	9.49 ± 0.20	9.02 ± 0.07	7.52 ± 0.17
Sig. θ^7 NG	6.16 ± 0.44	6.88 ± 0.14	5.70 ± 0.29
Exp. θ^4 NG	5.12 ± 0.63	11.91 ± 0.13	5.74 ± 0.42
Exp. θ^5 NG	5.39 ± 0.44	7.41 ± 0.05	6.92 ± 0.85
Poly. θ^2 NG	26.97 ± 2.68	22.81 ± 0.07	23.43 ± 0.27
Poly. θ^3 NG	4.73 ± 0.96	9.25 ± 0.05	5.08 ± 0.32

Table 3.2: The brain age errors resulting from different brain age estimation methods. This time, the brain ages from our method have been estimated using a cross validation framework. Note that we include this here for the sake of completeness. Our model is blind to the ground truth brain ages and function parameters, and explicitly has a term for brain age. Therefore, it is valid and more in line with our model to use the brain age estimates directly from training, as we have done in table 3.1.

the 14 cases in the table, poly3 had the best brain age error for seven.

These results suggest that when the trajectory is expected to be non-linear poly3 is the better option, but GPR may be better for linear cases. Nevertheless, if there is reason to believe that a dataset consists of linear trajectories, it makes sense to apply our method with a linear model. In this case, as the ‘Real’ column for θ^1 shows, our method may still be preferable to GPR.

In the GPR and KRR approaches, the brain ages are estimated for test subjects within a cross validation framework. For this reason, we repeated the experiment in table 3.1 and estimated the brain ages for poly3 also using cross validation. We again stress that since our model is blind to the ground truth brain ages and function parameters, and explicitly has a term for brain age, cross validation is not necessary. In addition, brain age estimates from training are more in line with the principles of our model, so in general in future studies it should be brain age estimates from training that are used for later analysis.

Table 3.2 shows the results of cross validation. For the sigmoid trajectories, the brain age error from poly3 is now higher. For the others, except θ^2 , the brain age

errors are largely unchanged. The reason that poly3 now fails for θ^2 as well is that the scenario for estimating the brain age of a test subject using equation 3.18 is the same as that for GPR and KRR, i.e. we are trying to estimate the brain age using feature values that all correspond to two brain age values. However, in our simulations all the features have the same trajectory profile, which is unlikely in real data. If any of the features within a given dataset is an invertible function with respect to brain age, equation 3.18 will be minimised when a'^* is in the correct region for the inverted U-shaped features.

3.3.2 UK Biobank data

We applied our method to the UK biobank data, setting a linear (polynomial of degree one) and a polynomial of degree three curve as the trajectory function. For each of these two curves, the optimisation was carried out with $\nu = 1$, $\nu = 5$ and $\nu = 10$. Splines were not investigated, since in the simulations we observed that polynomials were less sensitive to initialisation, gave less variable results, were quicker to optimise and resulted in better looking trajectories. The GPR and KRR methods were also applied but the KRR method gave suboptimal outputs, so its results were discarded. Franke et al. have previously explored the use of non-linear kernels for use with their brain age study [115], and found that it did not improve performance compared to a linear kernel. In addition, the kernels used with regression models in current approaches have nearly always been a linear one. For these reasons, we did not consider an alternative to KRR and compared our method to GPR only.

Figure 3.19 shows the mean absolute predicted brain age delta (δ_n in equation 3.1) for the subjects included in the fitting. Using a linear trajectory resulted in higher mean absolute deltas compared to poly3, meaning the chronological ages of the subjects were shifted more on average to accommodate the model. However, the mean absolute delta was a lot more variable across the 10 different sampled sets for poly3. The effect of the regularisation can also be seen, since for both linear and poly3 the mean absolute delta is lower when $\nu = 1$. There is not much difference

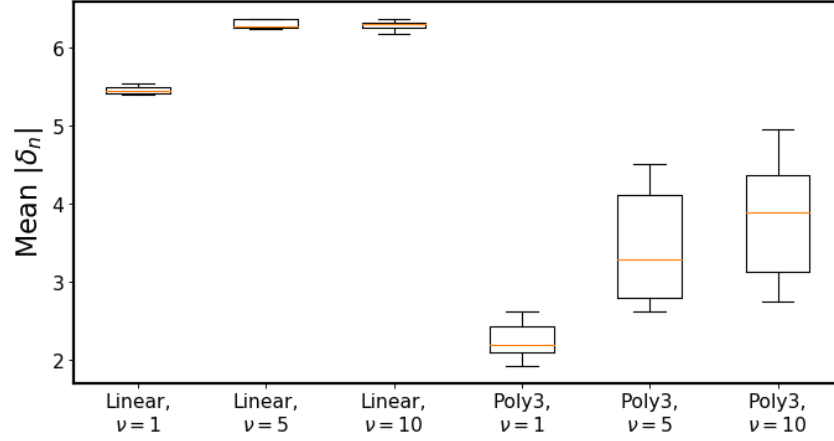


Figure 3.19: The mean absolute brain age delta (chronological age minus estimated brain age) during fitting for the different methods we investigated. Linear - polynomial of degree one, poly3 - polynomial of degree three.

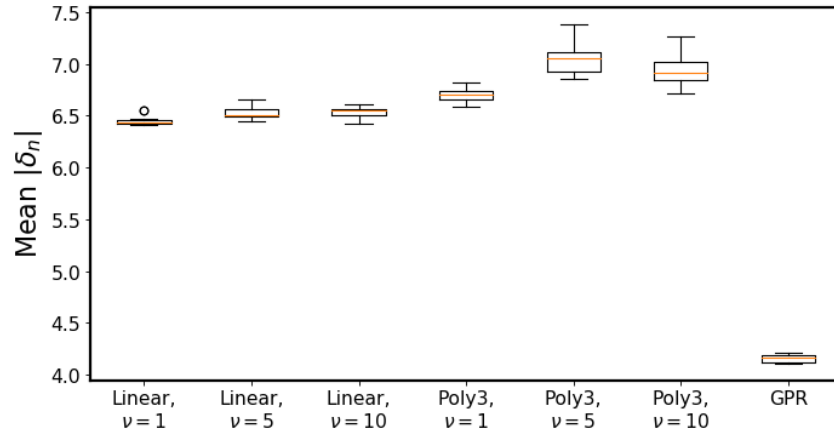


Figure 3.20: The mean absolute brain age delta (chronological age minus estimated brain age) for the test subjects for the different methods we investigated. Linear - polynomial of degree one, poly3 - polynomial of degree three, GPR - Gaussian process regression.

between $\nu = 5$ and $\nu = 10$, especially for the linear trajectory.

Meanwhile, the mean absolute predicted brain age delta for test subjects for the different methods is displayed in figure 3.20. The mean absolute deltas for poly3 is now much higher, most likely due to not using the regularisation term in equation 3.18. They are also much less variable. The mean absolute deltas for the linear trajectories, however, are more similar to those in figure 3.19. The effects of the ν parameter is not as obvious this time, which is again most likely due to not using the regularisation term. The mean absolute delta for the GPR is much smaller compared to our method, showing that the GPR method is much better at

predicting chronological age. However, being the best predictor of chronological age is not important in our case, since the main principle behind brain ageing studies is that chronological age is an inaccurate measure of ageing. Hence being the best predictor of chronological age may not translate to being the best predictor of brain age.

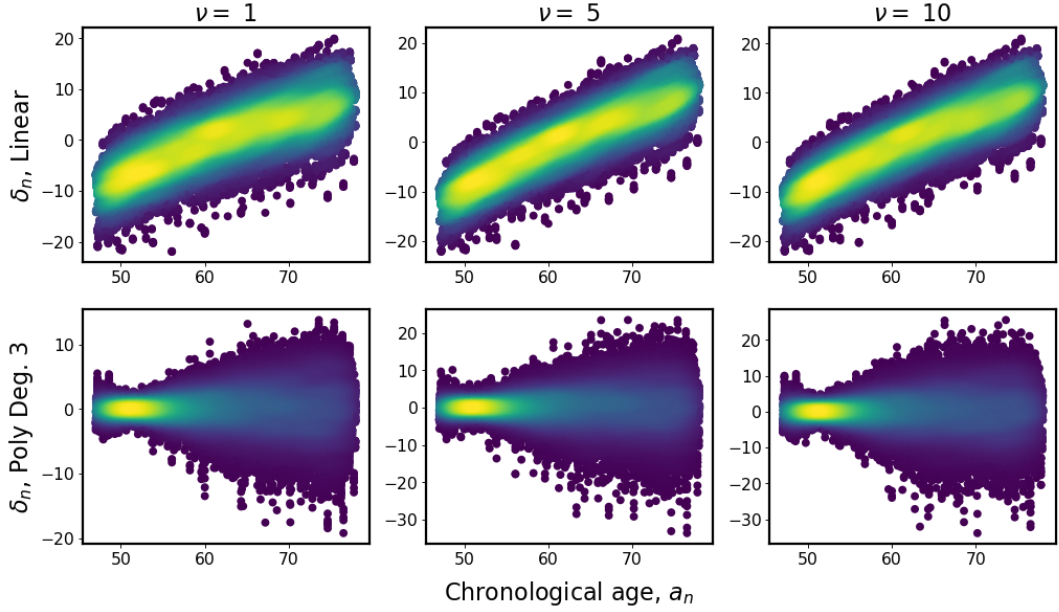


Figure 3.21: Estimated brain age deltas for the subjects included in the fitting from our method (concatenated across all 10 sampled sets) plotted against chronological age.

The estimated brain age training subject deltas for all 10 sampled sets were concatenated and they are plotted against chronological age for the different trajectories and ν values in figure 3.21. The same plots for the test subjects are shown in figure 3.22. For both the training and the test subjects, a bias similar to that we saw in section 3.3.1.4 is apparent for the linear trajectory at all ν values, i.e. younger subjects are more likely to have a negative delta and older subjects a positive delta. This translates to younger subjects being more likely to have an older appearing brain compared to the chronological age. This bias is now much stronger compared to the simulations. Figure 3.23 shows that this exact bias also exists for the GPR method.

The GPR method exhibits a bias due to regression dilution, which is a well-known

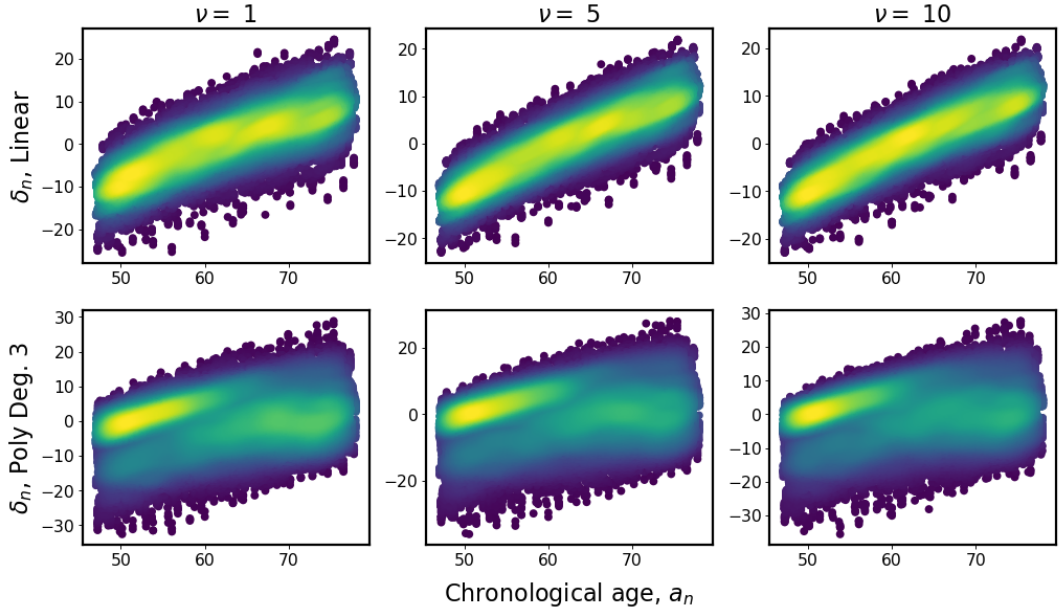


Figure 3.22: Estimated brain age deltas for test subjects from our method (concatenated across all 10 sampled sets) plotted against chronological age.

effect that occurs when there is error that is unaccounted for in the independent variables (for GPR the independent variables are the IDPs). Regression dilution results in a slope that is more shallow than what is needed being fitted to the data, and it is actually one of the main motivations for using errors-in-variables models [135]. For our method however, age is the independent variable, so what we are observing for our method is ‘inverse regression dilution’, i.e. a steeper slope than what is needed is being fitted to the data. One possible reason for this could be the regularisation term. Consider the example where all the IDPs have a $y = a'$ trajectory, where y , the dependent variable, is an IDP and a' , the independent variable, is brain age. When the data is fitted with a linear function, the optimisation can estimate a $y = a'$ trajectory and ‘tighten’ the data around this. It can also estimate a trajectory with a steeper slope or a trajectory with a more shallow slope, and tighten the data around these. The option that will require the overall least amount of shifting is to estimate a steeper slope, so it chooses this in order to be penalised the least by the regularisation. When it does this, the subjects in the younger age range will on average be shifted to the right and the subjects in the older age range will on average be shifted to the left, with the shift being proportional to age. This will

then appear as a bias in the brain age estimation.

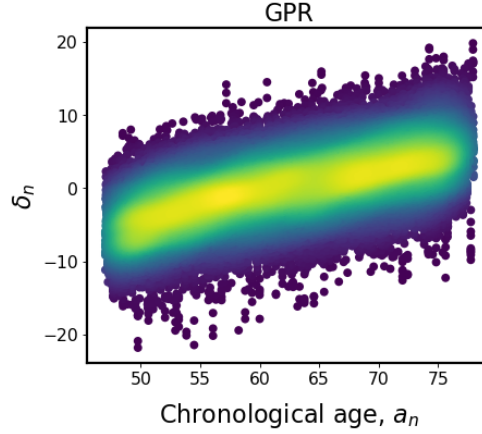


Figure 3.23: Estimated brain age delta for GPR (concatenated across all 10 sampled sets) plotted against chronological age.

There is also a bias for the training subjects in the polynomial trajectory, in that older subjects are more likely to have a larger absolute delta compared to younger ones. It could be argued that the brain age deltas in a population are likely to vary more and get larger as the population ages and so this bias is in part driven biologically. However, as figure 3.24 shows, this bias is more likely due to a fitting artefact. The right hand tail of the trajectories for the polynomial of degree three vary greatly between the 10 sampled sets compared to the linear ones. They are also not realistic, especially for the SIENAX grey matter normalised volume. The deltas for the older subjects need to be larger to accommodate these tails, hence the bias. For the test subjects, there is a combination of this bias and the bias seen for the linear trajectory.

We next correlated the deltas from different methods to each other and to the non-imaging variables. However, we did not want these correlations to be driven by the bias seen for the linear trajectories and GPR. Therefore, we attempted to correct for this by first regressing out chronological age and chronological age squared from the deltas, i.e. setting chronological age as a confound in the correlations. We did this for the polynomial deltas as well. When correlating, we used the delta from $\nu = 5$ since this is a typical brain age delta that has been reported in literature. Figures 3.25 and 3.26 show the effect that the correction had on the mean absolute

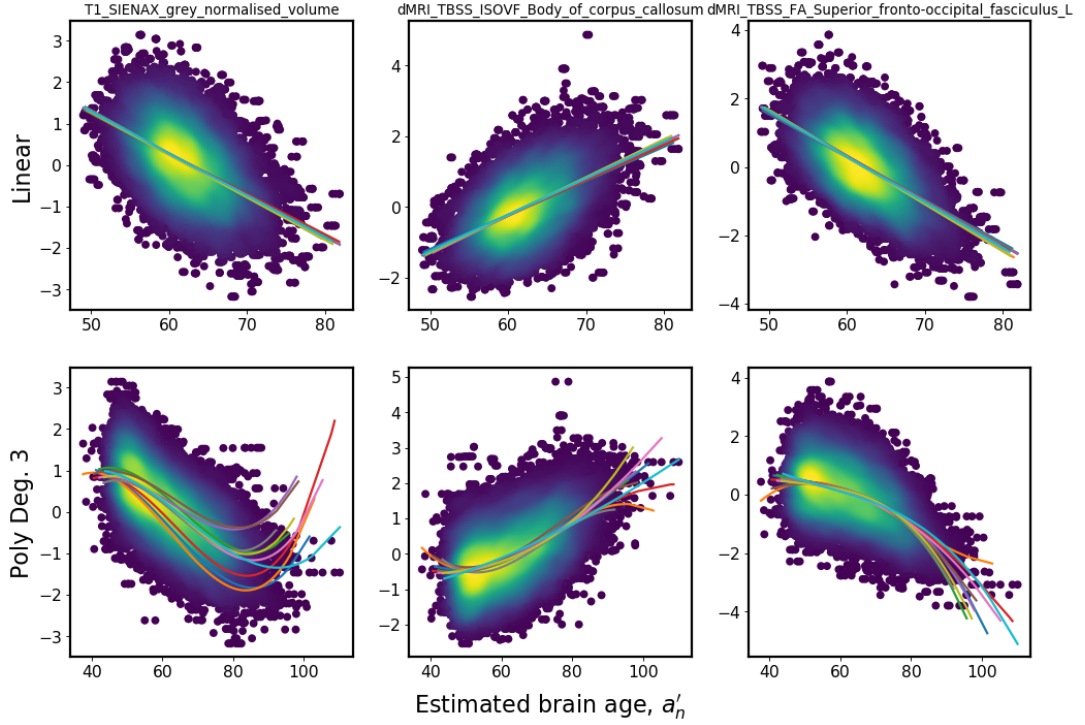


Figure 3.24: The IDPs have been concatenated across the 10 sampled sets, and three of them have been plotted here against the estimated brain ages. The trendlines are the trajectories fitted by our method for each of the 10 sampled set, with $\nu = 5$.

delta for the training and test subjects, respectively. It is vastly decreased for the linear trajectories, decreased slightly for GPR and has not changed for poly3 when compared to figures 3.19 and 3.20. This suggests that the linear bias was larger for our method with a linear trajectory compared to GPR.

The results of correlating the corrected deltas for the test subjects from different methods to each other are displayed in figure 3.27. The deltas from poly3 correlate very strongly with the deltas from the linear trajectory, showing that the corrected deltas were largely independent of the two trajectory functions used in our method. The deltas from linear and poly3 trajectories correlate less strongly with those from GPR, but the correlation is still significant. This suggests that our method and GPR are at least in part driven by the same factors.

Figure 3.28 is a Manhattan plot showing the results of correlating the corrected deltas for the training subjects from our method, using a linear trajectory, with the non-imaging variables for one of the sampled sets. For this set, the p values for 18 of the variables passed the false discovery rate (FDR) correction set at $p_{\text{FDR}} = 0.05$.

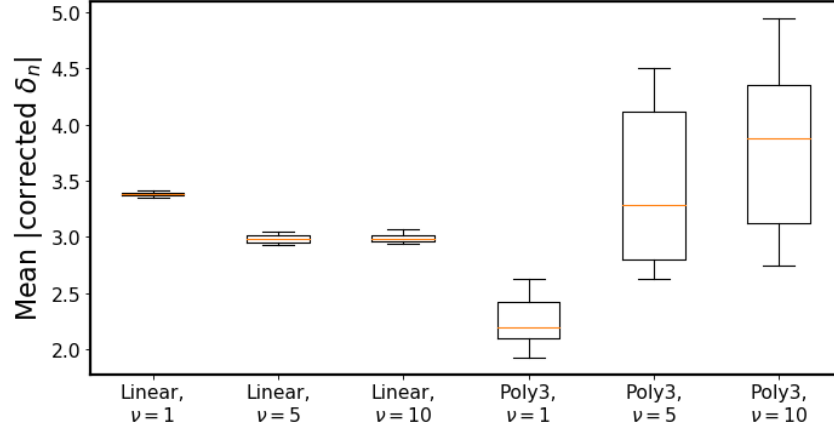


Figure 3.25: The mean absolute delta for subjects included in the fitting after regressing out chronological age and chronological age squared from the deltas.

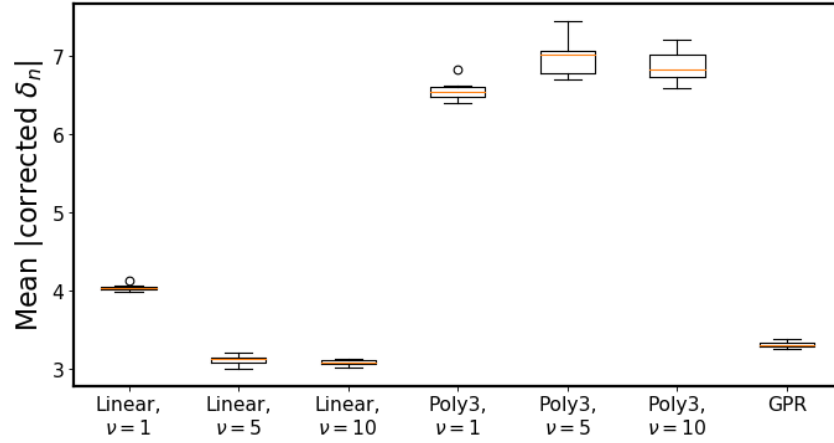


Figure 3.26: The mean absolute delta for test subjects after regressing out chronological age and chronological age squared from the deltas.

However, only two of these consistently passed the FDR threshold for all 10 sampled sets. Table 3.3 ranks the non-imaging variables based on the number of sampled sets in which they passed the FDR threshold. As equation 3.1 shows, the brain age delta that we report in this study is equal to chronological age minus brain age. Therefore, a positive brain age delta for a subject indicates that the subject's brain age is younger compared to their chronological age. A negative correlation of brain age delta with a variable then means that the higher the value for that variable, the older a subject's brain age is likely to be relative to their chronological age. The table shows that the two variables that passed the FDR threshold for all 10 sample sets were the systolic and diastolic blood pressure readings for one of the visits, both having r values of -0.1 when averaged across the 10 sampled sets. This translates

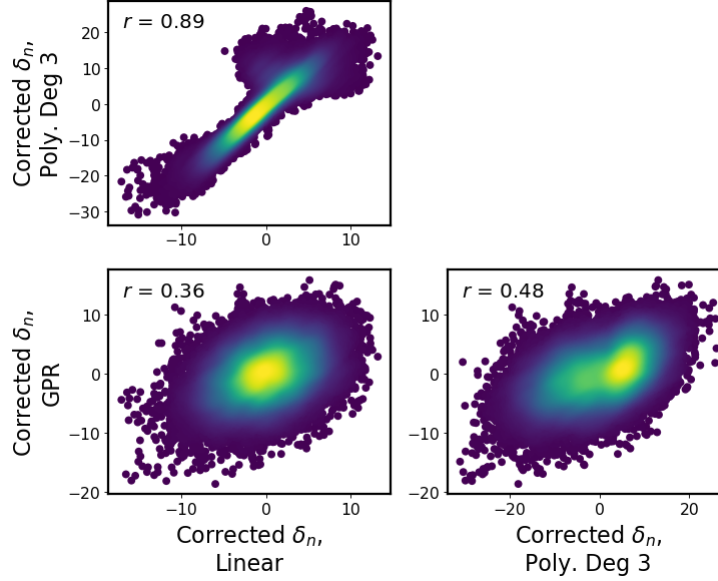


Figure 3.27: The brain age deltas for the test subjects from the different methods were concatenated across the 10 sampled sets and are plotted against each other here. The Pearson correlation coefficient is denoted by r , and has been averaged over the 10 sampled sets. The deltas for the polynomial of degree three and linear trajectories are from when $\nu = 5$.

to subjects that have older appearing brains compared to their chronological age being more likely to have higher systolic and diastolic blood pressures.

It can be seen from figure 3.29 and table 3.4 that the diastolic and systolic blood pressure readings for the second visit were consistently significantly correlated with the deltas for the training subjects when a polynomial of degree three was used as well. This time the blood pressure readings from the first visit were also consistently significant, in addition to the systolic blood pressure reading from the fifth visit. In general, the profile of the Manhattan plot is the same as that from figure 3.28, except that the variables passing the Bonferroni threshold were correlated more significantly in comparison. This suggests that the polynomial of degree three resulted in deltas that were slightly more sensitive to these cardiac measures.

Figure 3.30 and table 3.5 correspond to correlations using the test subject deltas from our method and a linear trajectory. The profile of the Manhattan plot is consistent with the profile for the training subject deltas in figure 3.28, with the systolic and diastolic blood pressures having correlations that are again among the most consistently significant. This time however, neither of the variables passed

	n_{sig}	r	p_{FDR}
Diastolic blood pressure, automated reading visit-2	10	-0.10 ± 0.02	3.78e-04
Systolic blood pressure, automated reading visit-2	10	-0.10 ± 0.02	5.38e-04
Diastolic blood pressure, automated reading	8	-0.09 ± 0.02	1.07e-03
Systolic blood pressure, automated reading	8	-0.09 ± 0.02	9.83e-03
Number of treatments/medications taken visit-3	7	-0.09 ± 0.02	2.24e-02
Diabetes diagnosed by doctor. Uses data-coding 100349 comprises 4 Integer-valued members in a simple list. visit-3	6	-0.08 ± 0.02	3.90e-02
Smoking status. Uses data-coding 90 comprises 4 Integer-valued members in a simple list.	5	-0.06 ± 0.01	4.73e-02
Weight visit-3	5	0.08 ± 0.02	4.97e-02
Duration to complete alphanumeric path (trail 2)	5	-0.10 ± 0.02	6.76e-02
Mean arterial pressure during PWA	4	-0.08 ± 0.03	7.33e-02
Mean arterial pressure during PWA visit-2	4	-0.08 ± 0.03	7.76e-02
Systolic brachial blood pressure	4	-0.08 ± 0.02	8.33e-02
Hip circumference visit-3	4	0.08 ± 0.02	8.79e-02
Time to complete round. Uses data-coding 402 visit-5	4	-0.11 ± 0.02	9.59e-02
Systolic blood pressure, automated reading visit-5	4	-0.07 ± 0.01	1.19e-01
Mean time to correctly identify matches visit-3	4	-0.07 ± 0.02	1.65e-01
Time to answer visit-2	4	-0.11 ± 0.02	1.94e-01
Number of symbol digit matches attempted	4	0.08 ± 0.02	2.06e-01
No-wear time bias adjusted acceleration standard deviation	3	0.08 ± 0.03	9.49e-02
End systolic pressure during PWA visit-2	3	-0.07 ± 0.03	9.86e-02

Table 3.3: The top 20 non-imaging variables that significantly correlated with the brain age deltas for subjects included in the fitting from our method and a linear trajectory. They are ranked by n_{sig} , the number of times that they were significant out of the ten sampled sets. The median FDR p value is denoted by p_{FDR} .

	n_{sig}	r	p_{FDR}
Diastolic blood pressure, automated reading visit-2	10	-0.11 ± 0.01	3.44e-06
Systolic blood pressure, automated reading visit-2	10	-0.11 ± 0.01	2.27e-05
Diastolic blood pressure, automated reading	10	-0.10 ± 0.01	6.71e-05
Systolic blood pressure, automated reading	10	-0.10 ± 0.02	2.87e-04
Systolic blood pressure, automated reading visit-5	10	-0.09 ± 0.01	6.78e-03
Number of treatments/medications taken visit-3	9	-0.10 ± 0.02	1.30e-03
Mean arterial pressure during PWA	9	-0.10 ± 0.02	4.81e-03
Systolic brachial blood pressure	8	-0.10 ± 0.02	5.18e-03
Mean arterial pressure during PWA visit-2	8	-0.10 ± 0.02	6.88e-03
Duration to complete alphanumeric path (trail 2)	8	-0.11 ± 0.01	1.05e-02
End systolic pressure during PWA visit-2	8	-0.09 ± 0.02	1.19e-02
Systolic blood pressure, automated reading visit-4	8	-0.14 ± 0.02	1.30e-02
Systolic blood pressure, automated reading visit-3	8	-0.14 ± 0.03	1.56e-02
Diastolic blood pressure, automated reading visit-5	8	-0.09 ± 0.02	1.65e-02
Systolic blood pressure, automated reading visit-6	8	-0.08 ± 0.01	2.43e-02
Diastolic blood pressure, automated reading visit-3	7	-0.14 ± 0.03	6.96e-03
Diastolic blood pressure, automated reading visit-4	7	-0.13 ± 0.03	1.37e-02
End systolic pressure during PWA	7	-0.09 ± 0.02	1.69e-02
Smoking status. Uses data-coding 90 comprises 4 Integer-valued members in a simple list.	7	-0.07 ± 0.01	1.98e-02
Diastolic blood pressure, automated reading visit-6	7	-0.08 ± 0.02	2.20e-02

Table 3.4: The top 20 non-imaging variables that significantly correlated with the brain age deltas for subjects included in the fitting from our method and a polynomial of degree three trajectory.

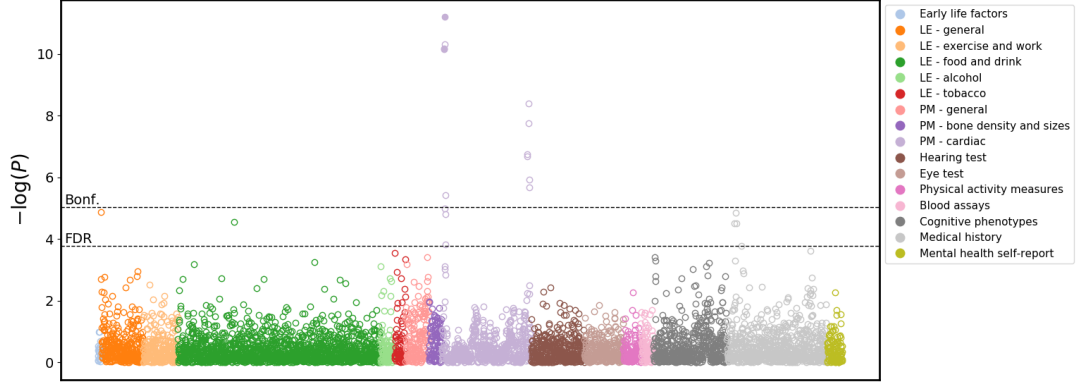


Figure 3.28: A manhattan plot showing the negative logarithm of the p values when correlating the brain age deltas for the subjects included in the fitting with 5456 non-imaging variables from the UK biobank. Multiple comparison correction thresholds for Bonferroni (Bonf.) and false discovery rate (FDR) are shown. The deltas here are for one of the sampled sets and were estimated using our method and a linear trajectory. Two of the data points have been filled in. These belong to variables that passed the FDR threshold for all 10 sampled sets. LE - lifestyle and environment, PM - physical measures.

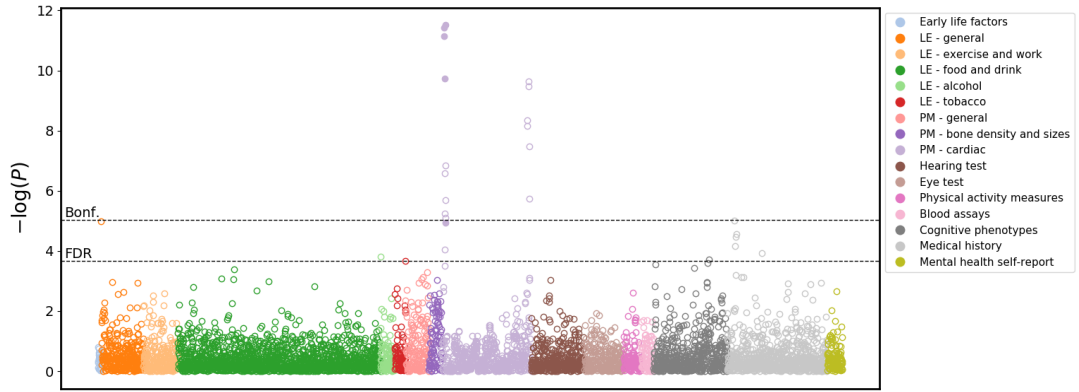


Figure 3.29: Same as figure 3.28, but for deltas for subjects included in the fitting that were estimated using our method and a polynomial of degree three trajectory.

the FDR threshold for all 10 sampled sets, indicating that the deltas for the test subjects lost some sensitivity to these measures compared to those for the training subjects. Figure 3.31 and table 3.6 illustrate that the story for the poly3 trajectory is largely the same. The profile of the Manhattan plot, and variables that appear among the most consistently significant, are largely unchanged when comparing the correlations from test subject deltas to those from training subject deltas, again with a slight loss in sensitivity.

The Manhattan plot and table for the GPR method in figure 3.32 and table 3.7 show its deltas were sensitive to the same cardiac measures as our method.

	n_{sig}	r	$p\text{FDR}$
Systolic blood pressure, automated reading visit-2	9	-0.10 ± 0.02	5.27e-04
Systolic blood pressure, automated reading	9	-0.09 ± 0.02	1.41e-03
Diastolic blood pressure, automated reading	8	-0.08 ± 0.02	3.67e-03
Diastolic blood pressure, automated reading visit-2	8	-0.08 ± 0.02	5.12e-03
Diabetes diagnosed by doctor. Uses data-coding 100349 comprises 4 Integer-valued members in a simple list. visit-3	6	-0.08 ± 0.01	2.69e-02
Number of treatments/medications taken visit-3	6	-0.08 ± 0.01	4.11e-02
Systolic blood pressure, automated reading visit-4	5	-0.13 ± 0.03	4.92e-02
Systolic blood pressure, automated reading visit-6	5	-0.09 ± 0.03	5.53e-02
Mean arterial pressure during PWA	5	-0.09 ± 0.03	6.15e-02
Mean arterial pressure during PWA visit-2	5	-0.10 ± 0.02	6.37e-02
Systolic brachial blood pressure	5	-0.09 ± 0.03	6.57e-02
Systolic blood pressure, automated reading visit-5	5	-0.08 ± 0.03	7.64e-02
Number of self-reported non-cancer illnesses	4	-0.07 ± 0.01	6.18e-02
Weight visit-3	4	0.07 ± 0.01	9.66e-02
End systolic pressure during PWA	4	-0.09 ± 0.03	1.46e-01
Diastolic blood pressure, automated reading visit-6	3	-0.07 ± 0.02	1.16e-01
Duration to complete alphanumeric path (trail 2)	3	-0.09 ± 0.03	1.30e-01
No-wear time bias adjusted acceleration maximum	3	0.08 ± 0.02	1.31e-01
Diastolic brachial blood pressure	3	-0.08 ± 0.02	1.40e-01
End systolic pressure during PWA visit-2	3	-0.08 ± 0.02	1.66e-01

Table 3.5: The top 20 non-imaging variables that significantly correlated with the test subject brain age deltas from our method and a linear trajectory.

	n_{sig}	r	$p\text{FDR}$
Systolic blood pressure, automated reading visit-2	10	-0.11 ± 0.02	4.29e-05
Diastolic blood pressure, automated reading	10	-0.10 ± 0.02	7.32e-04
Systolic blood pressure, automated reading	9	-0.10 ± 0.02	1.03e-04
Diastolic blood pressure, automated reading visit-2	9	-0.09 ± 0.02	2.49e-04
Mean arterial pressure during PWA	8	-0.11 ± 0.02	8.85e-03
Mean arterial pressure during PWA visit-2	8	-0.11 ± 0.02	9.45e-03
Systolic brachial blood pressure	8	-0.10 ± 0.03	1.17e-02
Systolic blood pressure, automated reading visit-6	8	-0.10 ± 0.02	1.21e-02
Diabetes diagnosed by doctor. Uses data-coding 100349 comprises 4 Integer-valued members in a simple list. visit-3	8	-0.09 ± 0.01	1.89e-02
Systolic blood pressure, automated reading visit-5	7	-0.10 ± 0.03	3.97e-03
Systolic blood pressure, automated reading visit-4	6	-0.15 ± 0.04	1.10e-02
Cardiac index during PWA visit-2	6	-0.07 ± 0.02	2.70e-02
Weight visit-3	6	0.08 ± 0.01	2.75e-02
Number of treatments/medications taken visit-3	6	-0.08 ± 0.01	2.80e-02
Diastolic blood pressure, automated reading visit-5	6	-0.08 ± 0.02	3.44e-02
End systolic pressure during PWA visit-2	5	-0.10 ± 0.02	4.63e-02
Hip circumference visit-3	5	0.07 ± 0.02	4.99e-02
Systolic blood pressure, automated reading visit-3	5	-0.13 ± 0.04	5.44e-02
Trunk BMD (bone mineral density)	5	0.11 ± 0.03	6.61e-02
Diastolic brachial blood pressure	5	-0.08 ± 0.02	7.71e-02

Table 3.6: The top 20 non-imaging variables that significantly correlated with the test subject brain age deltas from our method and a polynomial of degree three trajectory.

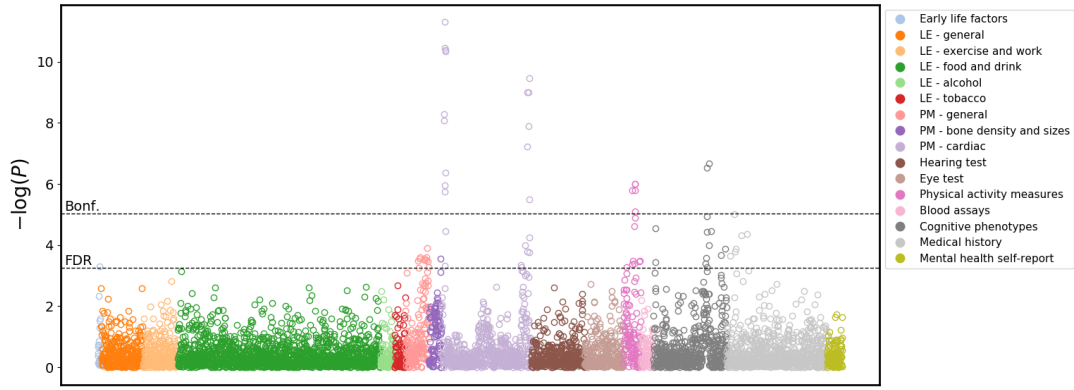


Figure 3.30: Same as figure 3.28, but for test subject deltas that were estimated using our method and a linear trajectory.

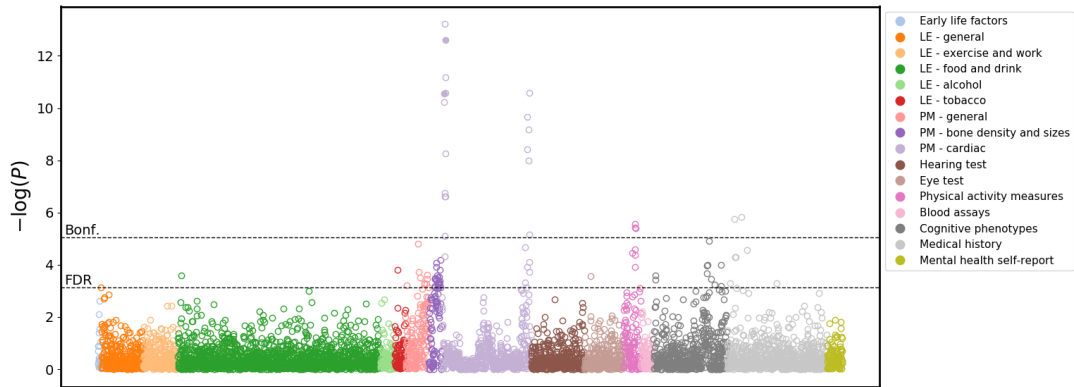


Figure 3.31: Same as figure 3.28, but for test subject deltas that were estimated using our method and a polynomial of degree three trajectory.

This time, the deltas were also sensitive to bone density measurements, with the total bone mineral density (BMD) being consistently and positively correlated with the deltas for all 10 sampled sets. This translates to subjects that have an older appearing brain compared to their chronological age being more likely to have lower total BMD. Weight, BMI and leg fat mass were also significantly correlated for all 10 sampled sets. However, the relationship with these variables is somewhat contrary to expectations, since the table shows that accelerated ageing is associated with lower values for these variables. Note that the absolute mean correlation value for the systolic blood pressure reading was lower compared to total BMD, yet its p value was more significant. This is due to more of the subjects actually having the systolic blood pressure data.

The GPR mean correlation value and median p values for the diastolic and systolic blood pressures in table 3.7 are lower compared to those from our method

with poly3 in table 3.4. This suggests that our method when used with a poly3 trajectory may in fact be more sensitive to these measures. The number of significant correlations, mean correlation value and median p value for the number of treatments/medications taken for GPR are not shown in table 3.7, but they were 1, -0.03 ± 0.02 and $5.74e-01$, respectively. These are much lower than those from our method in tables 3.3 and 3.4. This again suggests that our method was more sensitive to this variable.

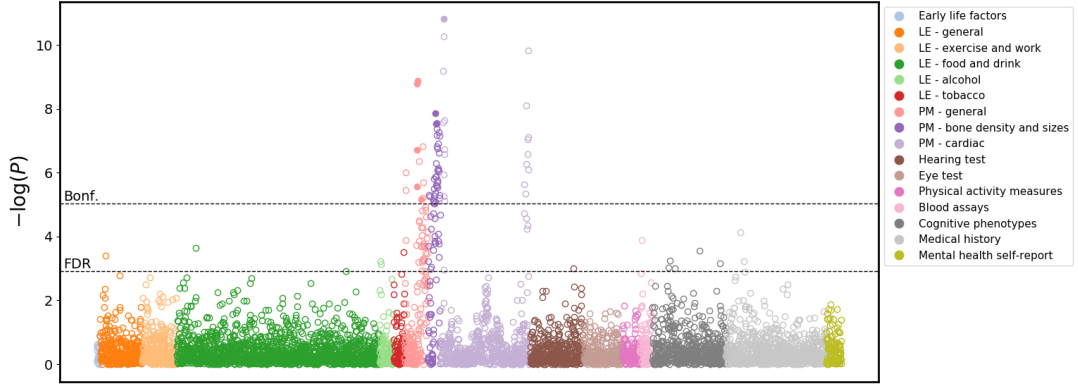


Figure 3.32: Same as figure 3.28, but for test subject deltas that were estimated using GPR.

3.4 Discussion

This study is the first, to the best of our knowledge, that has investigated the application of a brain age model based on errors-in-variables to neuroimaging data. Through simulations, we found that the accuracy of the brain age prediction was dependent on both the number of features and the number of subjects in the dataset, with the dependence on the former being stronger. On the other hand, the accuracy of the estimated brain age trajectories was still dependent on the number of subjects but much less dependent on the number of features. It should be noted that we simulated noise for each feature independently. If the noise was correlated in any way, then more features would have been needed to give the same improvement in brain age error that we observed. Further analysis based on simulations or theory in the future that provides insight into the relationship between brain age accuracy

	n_{sig}	r	$p\text{FDR}$
Systolic blood pressure, automated reading visit-2	10	-0.09 ± 0.02	3.85e-04
Total BMD (bone mineral density)	10	0.16 ± 0.03	9.78e-04
Weight visit-3	10	0.11 ± 0.01	9.84e-04
Weight	10	0.08 ± 0.01	1.11e-03
Arms BMD (bone mineral density)	10	0.15 ± 0.02	1.28e-03
Body mass index (BMI)	10	0.08 ± 0.01	1.90e-03
Body mass index (BMI) visit-3	10	0.11 ± 0.02	3.87e-03
Leg fat mass (right)	10	0.07 ± 0.01	1.05e-02
Systolic blood pressure, automated reading	9	-0.09 ± 0.02	5.81e-04
Total BMD (bone mineral density) T-score	9	0.16 ± 0.03	7.01e-04
Total BMC (bone mineral content)	9	0.15 ± 0.03	1.25e-03
Legs BMD (bone mineral density)	9	0.15 ± 0.03	1.88e-03
Pelvis BMD (bone mineral density)	9	0.15 ± 0.03	2.89e-03
Trunk BMD (bone mineral density)	9	0.15 ± 0.03	3.27e-03
Whole body fat mass	9	0.08 ± 0.01	3.81e-03
Systolic brachial blood pressure	9	-0.10 ± 0.02	5.30e-03
Diastolic blood pressure, automated reading	9	-0.08 ± 0.02	5.42e-03
Femur wards BMD (bone mineral density) T-score (right)	9	0.14 ± 0.03	5.72e-03
Arm fat mass (left)	8	0.09 ± 0.01	3.70e-04
Femur total BMD (bone mineral density) T-score (right)	8	0.15 ± 0.03	2.03e-03

Table 3.7: The top 20 non-imaging variables that significantly correlated with the test subject brain age deltas from GPR.

and the amount of noise correlation would be useful.

Currently, one issue with our method is that two parameters, ν and σ_m , need to be set prior to optimising. However, using simulations we showed that the brain age accuracy is fairly insensitive to these and only rough estimates of them are needed. We put forward two methods to estimate σ_m , concluding that the binning method is the better one. We do not yet have a good approach for automatically estimating ν and rely on the user to set this. One option for future studies could be to first run one of the current approaches, such as GPR, and set ν to the chronological age prediction accuracy that the method achieves. Methods such as GPR are very quick to train, and so would not add too much time to the overall pipeline.

Our formulation gives the user the ability to use specific functional forms should they have prior hypotheses about the shape of the trajectories in their data. For the majority of use cases however, it is likely that functions that can generalise to different trajectory shapes will be required. We investigated two possible options for this, splines and polynomials. We concluded that polynomials were better since,

in simulations, they were less sensitive to initialisation, gave less variable results, were quicker to optimise and resulted in better looking trajectories. Of course, polynomials still require the degree to be set. In simulations we witnessed that a degree of three resulted in brain age accuracies that were comparable to when the ground truth trajectory function was used. Moreover, based on trajectories that have typically been reported in literature for real datasets such as those by Vinke et al. [146], it is unlikely that degrees higher than this will be needed. Nonetheless, we saw in figure 3.24 that when we applied our method to the UK biobank data, the polynomial trajectories suffered from exaggerated and unrealistic tails. It is likely that this artefact was highly detrimental to the brain age estimates and was reflected in the large variability that we saw in figure 3.19. Therefore, our method may greatly benefit from investigations into other generalisable functions or ways to regularise or stabilise the polynomial tails.

In literature, the typical approach to estimating brain age from neuroimaging data is to train an algorithm to predict chronological age in a healthy cohort. The algorithm is then applied to test subjects and brain age deltas are derived by taking the difference between their predicted age and chronological age. In this study, we implemented a reported GPR method with a linear kernel in order to compare the brain age estimations to those from our method. We chose the GPR method for comparison since it is very quick to train, has frequently been used and has been shown to be sensitive to age-related phenotypes [118, 130, 126, 125, 127]. More powerful methods have been reported in literature, such as one based on convolution neural networks [131]. Their advantages include their readiness to model non-linear patterns and a reduced need for preprocessing. However, they require extensive training times and computational resources, and are most suited to spatial maps as opposed to IDPs. For these reasons we did not implement them but their existence should be noted.

In simulations, we showed that brain age estimation accuracies from our method and a polynomial of degree three trajectory compared favourably to those from

current approaches. In addition, when using the brain age estimates directly from training, it was able to capture inverted U-shaped trajectories that current methods could not. However, both our simulations and our method were designed based on the generative model in figure 3.1, which gives an advantage to our method compared to GPR and KRR. In addition, we provided the ground truth ν value to our method, which the GPR and KRR methods were blind to. Nevertheless, table 3.2 illustrated the point that, assuming figure 3.1 is somewhat close to reality, our method can potentially match and in some cases outperform current approaches.

We applied our method to the UK biobank data and as a way of validating the brain age estimates we correlated them with the non-imaging variables. We found that accelerated brain ageing was significantly, and consistently, correlated with higher systolic and diastolic blood pressures. This is consistent with biological age measures derived from gene expression profiles [111]. In general, high blood pressure is associated with negative health outcomes [147], which is compatible with the significant correlation with accelerated ageing that we observed. Other non-imaging variables that were significant for the majority of the sample sets, and for both the linear and poly3 trajectories, were the number of medications taken and diabetes diagnosis. The former is a surrogate for the general health of a subject and it makes sense that it should correlate with accelerated ageing. The latter, meanwhile, is consistent with studies using neuroimaging derived brain age [129] and composite non-neuroimaging biomarkers [148]. The fact that our method resulted in associations that are consistent with biological ageing literature is promising, and indicates that the brain age estimates may be driven in part by real biological signal. In addition, we found that the correlation profiles were largely unchanged when comparing the linear and polynomial of degree three trajectories. This was also the case when comparing the brain ages for the test subjects and those for the training subjects. Although we have reported significant associations, the effect size of the associations were all small. For example, in table 3.4, the correlation values range from -0.07 to -0.14. It is possible to obtain significant associations with weak

effect sizes like these due to the large number of subjects in the UK Biobank.

The GPR method also resulted in the same significant associations with systolic and diastolic blood pressures. In addition to these, BMD, BMI and weight appeared on the list of consistently significant variables for GPR (weight also appeared on the list of significant associations for our method, but was less consistent). Studies based on telomere lengths [149] and metabolics [150] have also associated BMD with accelerated biological ageing. However, the literature is not in agreement with the results for BMI. Our results for GPR showed that accelerated ageing is associated with lower BMI, whereas studies based on gene expression profiles [111] and glycans [112] have shown that the opposite is true. This warrants further investigation, but one possible explanation is that low BMI may be indicative of frailty.

We explicitly have a parameter for brain age in our model. Therefore, it is possible to directly use the brain age estimates for the subjects used during training for further inference. On the contrary, current approaches use a train-test framework. In addition, many brain age studies use the chronological age prediction accuracy on a test set as a way of validating their method. They say that the lower the chronological age estimation error, the better the method is as a marker of biological age. For these reasons, we constructed equation 3.18 to enable brain age estimations for test subjects. However, it is not necessary to rely on this for other studies and there may also exist better ways of applying our model to test subjects. Based on this validation approach, figure 3.20 would indicate that the GPR method is a better method than ours. An argument that we make however, is that it is somewhat contradictory to assume there is a measure of biological age that is different from chronological age, but then go on to validate a method's effectiveness to measure biological age by its ability to predict chronological age. Of course, a marker of biological ageing should still correlate highly with chronological age. But if the marker is predicting chronological age with 100% accuracy, then this is 'too high' as it means it is throwing out all information about biological ageing that is different from chronological age. Reported chronological age estimation errors based

on neuroimaging data for a healthy cohort range from 4.16 [131] to 8.4 [151] years, but no one truly knows what the correct level of estimation error should be.

On the other hand, an argument in favour of using the chronological age prediction error as a method of validation arises from the assumption that the prediction error from chronological age prediction models comes from two sources. The first is from the fact that when the prediction error for an age predictor is calculated, the chronological age is typically used as the ground truth. If the assumption of equation 3.1 is true, then this ground truth is noisy. This would mean that there will be an error that even the best predictor will not be able to remove. The second is from deficiency in the predictor’s model and the error associated with this is reduced as the model is made more accurate. In brain age literature, the prediction error is taken to be the brain age delta (δ_n in equation 3.1). Therefore, for a given dataset there will be an estimation error ‘floor’ that will not be beaten by any method. But when comparing methods, the lower the mean absolute brain age delta $|\delta_n|$ for test samples, the better a model is at reducing the error associated with the second source.

It is not yet clear which of the two arguments is correct. Based on this, along with the possibility that current approaches can reduce their estimation error by down-weighting features that are driving real brain age differences, we believe chronological age prediction error should not be considered heavily as validation for brain age prediction methods.

The brain age deltas from the GPR approach were more sensitive to bone density and weight-related measures. On the other hand, our method was more sensitive to blood pressure measures and the number of treatments/medications taken. The number of treatments/medications taken can be considered as an indication of overall health. This suggests that our method may provide information that is complementary to current approaches. This means that there is useful exploration to be done in comparing or even combining approaches.

Ultimately, one of the aims of brain ageing research is to discover risk factors

and interactions with diseases such as neurodegeneration. Our model provides a framework for these sort of studies in ways that current approaches cannot. For example, disease subjects can be included directly in the fitting and $f_m(a'_n, \theta_m)$ changed to $f_m(a'_n, d_n, \theta_m)$. The variable d_n could be a measure of disease severity, and f_m could be set to $f_m = ag(a'_n) + bh(d'_n) + cg(a'_n)h(d'_n)$, where the $cg(a'_n)h(d'_n)$ term models interaction between brain ageing and disease severity.

As well as identifying factors that can affect healthy ageing (which is one of the aims of brain ageing studies that use healthy subjects as we have done in this study, although we have not carried out any analyses to determine causation), another potential application of brain age is for the brain age delta value to be a direct biomarker of diseases affecting the brain. The major issue with this is that a large variety of diseases show similar or overlapping mean brain age deltas [93]. This makes brain age a very unspecific biomarker. One potential approach to remedy this limitation, which we discuss in more detail in chapter 5, is to consider patterns of brain age deltas across different brain regions and different modalities. Considering only regional and modality specific brain age delta values still has the limitation that any potential biomarker will only be made sensitive to changes caused by a disease that are also affected by ageing. Therefore, a potentially useful biomarker application would be to include these brain age delta values as extra features to a feature set that also contains traditional features, e.g. grey matter volumes or resting state connectivity matrices.

Many measures in neuroimaging contain error. This means that there is scope for errors-in-variables models to be used in applications other than brain age. These applications could range from modelling the error in the covariates in first level fMRI analyses, to e.g. modelling the error in hippocampal volume measurements in a study investigating how hippocampal volume affects memory.

A possible approach to improve our method is to consider a Bayesian formulation of it. By adding appropriate priors for the ν and σ_m parameters, it would be possible to directly optimise for these as opposed to setting or estimating them beforehand.

The issue with the trajectory tails can also be remedied by a Bayesian approach, since the ill-defined tails are in regions where there is not much data and a Bayesian model will set large uncertainties for the part of the trajectory in these regions, thereby not letting it affect the brain age estimations as much. Whether or not a given errors-in-variables model is identifiable, i.e. whether all its model parameters can be accurately estimated given infinite observations and without any prior or additional information, is an important question. The model that we have presented in this study is not identifiable. However, using a Bayesian model along with non-Gaussian distributions for the δ_n parameter, which could very well be the case in reality, could conceivably result in an identifiable model.

One limitation with the UK biobank is the age range of the subjects since it does not capture the whole adult lifespan. Moreover, the distribution of the subjects' ages is not uniform, with many more of the subjects having ages in the centre of the range compared to the edges. This limits the measures of variability that we have reported, since subjects in the edges of the range are much more likely to be repeatedly sampled in the different sampled sets. While this issue will become less important as more subjects are scanned, the issue with the age range will not go away. Therefore, validating the methods in this study in a separate healthy cohort that includes the whole adult lifespan should be a priority. The inverted U-shaped trajectories are also typically seen before the age of 45 [133, 132], meaning that the advantages of our method may become more apparent when applied to a dataset with a longer age range.

A different issue with the experiments on the UK biobank data is the way in which we selected the IDPs to include in the optimisation. We first correlated all the IDPs with the chronological ages of all the subjects, ranked the IDPs by their correlation value and kept the top 1000. This means that the test subjects during cross validation were not totally blind to this feature selection, and the chronological age estimation errors in figure 3.20 may be slightly underestimated. However, given that we still kept around half of the features, and all the methods that we compared

had the same advantage, this is unlikely to be a major issue.

3.5 Conclusion

We have presented a method, based on errors-in-variables, for estimating biological brain age. The advantages of our method over current approaches include the ability to model inverted U-shaped trajectories and it can easily be extended to model the interactions between brain ageing and disease. Using simulated data, we showed that it is effective at estimating brain age, and can match and in some cases outperform current approaches. The method was then applied to the UK biobank data. The brain age estimates from our method were most strongly associated with systolic and diastolic blood pressures, showing that accelerated brain ageing correlates with higher values for these two variables. We also applied a common current approach to the UK biobank data. The brain age estimates from it were more sensitive to bone density and weight-related measures, while our method was more sensitive to blood pressure measures and number of treatments/medications taken, suggesting they may provide complementary information. However, further validation on large datasets that encompass the whole adult lifespan is needed. In addition, our method could benefit from investigations into better generalisable trajectory functions and automated ways of estimating the ν parameter.

Chapter 4

A method for decomposing multimodal neuroimaging data into trajectories of ageing

4.1 Introduction

One possible extension to the work from the previous chapter could be to consider regional brain ages as opposed to a whole-brain brain age. This then raises the question of what is the best way to determine the brain regions that will be used in the analysis. Studies investigating the effects of ageing on neuroimaging data typically make use of atlases or segmentation to first define regions of interest. Trajectories are then estimated for these regions as a second step. The issues with this are that manual segmentation requires user expertise, suitable atlases or automated segmentation tools may not exist for a modality of interest, subregions within a given region may exhibit different trajectories and the scope for novel findings is limited by the initial region definition. For these reasons, we instead consider data-driven approaches that can be applied to any number of modalities with ease and rely on an automated decomposition of the data into latent components. The output of these types of approaches is a set of subject weights and associated spatial map for each

latent component. The collection of prominent voxels in a given component spatial map could then be considered one region. Moreover, the subject weights can be plotted against age to view the trajectory associated with a component. Therefore, in addition to aiding in defining regions for future regional brain age studies, such approaches can be powerful exploratory tools in their own right.

Various multimodal decomposition methods have been reported in literature. These methods are typically applied to data that consists of spatial maps for T modalities and N subjects. Each modality has M_t voxels, where $t = [1, T]$, and the data is stored in $\mathbf{Y}_t$, of size $N \times M_t$.

One of these methods, reported by Calhoun et al. [152], is termed joint independent component analysis (ICA). In joint ICA, the data is concatenated along the voxel dimension to obtain $\mathbf{Y}^c = [\mathbf{Y}_1, \mathbf{Y}_2 \dots \mathbf{Y}_T]$ of size N by total number of voxels across modalities. Following this, it is assumed that this data has been generated through a mixture of L spatially independent components. Therefore:

$$\mathbf{Y}^c = \mathbf{W}^T \mathbf{P} + \mathbf{E} \quad (4.1)$$

where the subject weighting matrix $\mathbf{W}$ is of shape $L \times N$, the independent spatial source matrix $\mathbf{P}$ is of shape $L \times M^c$ and $\mathbf{E}$ is the error term. M^c is the number of voxels in the concatenated data. An element in $\mathbf{W}$ describes the relative amount, shared by all the modalities, of component l that exists in the data for subjects n , where $n = 1 : N$. $\mathbf{W}^T$ and $\mathbf{P}$ are estimated using ICA.

Calhoun et al. applied joint ICA to data comprised of VBM and first-level fMRI contrast maps for a set of 15 healthy controls and 15 schizophrenia patients. They found that for one of the components, the subject weightings were significantly different between the two groups. Furthermore, using the associated component spatial map, they were able to relate structural changes to functional changes, demonstrating the usefulness of multimodal decomposition.

Informed multimodal partial least squares (PLS), as reported by Chen et al. [153], attempts to make use of another variable stored in $\mathbf{a}$, such as age, when

carrying out the decomposition. In this method, the data is again concatenated and modelled like so:

$$\mathbf{Y}^c = \mathbf{W}^T \mathbf{P} + \mathbf{E} \quad (4.2)$$

where the subject score matrix $\mathbf{W}$ is of shape $L \times N$, the loading matrix $\mathbf{P}$ is of shape $L \times M^c$ and $\mathbf{E}$ is the error term. These terms are estimated using PLS with $\mathbf{Y}^c$ as the predictor and $\mathbf{a}$ as the single response variable. In this type of PLS regression, initially the first row in $\mathbf{W}$, $\mathbf{w}_1$, is estimated by projecting $\mathbf{Y}^c$ onto some weight vector $\mathbf{s}_1$, such that the covariance between $\mathbf{w}_1 = \mathbf{s}_1(\mathbf{Y}^c)^T$ and $\mathbf{a}$ is maximised. Following this, $\mathbf{Y}^c$ is deflated based on $\mathbf{w}_1$ and the process is repeated to estimate the rest of the components. The $\mathbf{s}_l$ vectors can finally be rotated back into the non-deflated space and stored in $\mathbf{S}$, of shape $L \times M^c$. Hence, the subject weightings are in $\mathbf{W}$ and the spatial maps used to obtain them are in $\mathbf{S}$.

Chen et al. applied informed multimodal PLS to a set of 29 healthy subjects, which they grouped into young and old based on age. Their data consisted of MRI grey matter structural images and PET images, and they created a binary $\mathbf{a}$ vector whose elements corresponded to group membership. Chen et al. only kept the first component and compared it to the first components from a unimodal decomposition of the two modalities. Using a leave-one-out cross validation scheme, they found that the subject weightings corresponding to the multimodal component separated the young and old groups better than the two unimodal ones.

A method that estimates multimodal spatially independent component maps but does not require concatenation has been reported by Groves et al., called linked ICA [154]. In this approach, modality-specific images are first extracted from the raw data. Each of these image types is then considered a separate modality, and are collected into modality groups that share the same spatial features (e.g. resolution and smoothness). The data is stored in a variable Y_{mtn}^k , where $m = [1, M_k]$; $t = [1, T_k]$; $n = [1, N]$; $k = [1, K]$. Here, K is the total number of modality groups, M_k is the number of voxels in group k , T_k is the number of modalities in group k and

N is the total number of subjects. The data is subsequently modelled as:

$$Y_{mtn}^k = \sum_{l=1}^L M_{ml}^k W_{tl}^k H_{ln} + E_{mtn}^k \quad (4.3)$$

where M_{ml}^k represents the spatial map for component $l = [1, L]$ in group k , with L being the total number of components. W_{tl}^k is the modality weighting for component l in modality t and group k , H_{ln} is the weighting for component l in subject n , and E_{mtn}^k is the error term. A Bayesian approach incorporating factorised Gaussian mixture models is used to fit the data to this model and estimate the spatial sources in M_{ml}^k . The aspect of the model that links different modalities together is the fact that H_{ln} is shared across modalities. Therefore, the output is a set of L component spatial maps, M_{ml}^k , with the weighting for component l and subject n being stored in H_{ln} . This means that each component has different spatial maps but the same subject weighting across modality groups.

Linked ICA was employed by Douaud et al. in order to study the relationship between age and brain structure in 484 healthy subjects [133], and subsequently to classify AD and schizophrenia subjects. The different modalities in the study consisted of VBM images, vertexwise cortical thickness and surface area measures. This dataset was decomposed into 70 components and two of these components showed a significant relationship with age. They further observed that the spatial distribution of one of these components matched regions of accelerated atrophy in AD and regions showing altered development in schizophrenia. A post-hoc classification was then employed by regressing the spatial map of this component into gray-matter images of 58 AD patients and 62 healthy controls as well as 12 schizophrenia patients and 12 healthy controls, to obtain a single measure for each subject. Using leave-one-out cross validation, they obtained accuracies of 72% and 83% respectively.

The last approach that we describe here is multimodal canonical correlation analysis (CCA), as reported by Correa et al. [155]. The aim of this method is to decompose the data into components that best capture information that is joint

between two modalities. The two modalities are modelled like so:

$$\mathbf{Y}_1 = \mathbf{W}_1^T \mathbf{P}_1 + \mathbf{E}_1 \quad (4.4)$$

$$\mathbf{Y}_2 = \mathbf{W}_2^T \mathbf{P}_2 + \mathbf{E}_2 \quad (4.5)$$

where the subject score matrices $\mathbf{W}_t$ are of shape $L \times N$, the loading matrices $\mathbf{P}_t$ are of shape $L \times M_t$ and $\mathbf{E}_t$ are the error terms. These terms are estimated using CCA. The first rows in $\mathbf{W}_1$ and $\mathbf{W}_2$, denoted by $\mathbf{w}_{1,1}$ and $\mathbf{w}_{2,1}$ respectively, are estimated by projecting $\mathbf{Y}_1$ and $\mathbf{Y}_2$ onto some weight vectors $\mathbf{s}_{1,1}$ and $\mathbf{s}_{2,1}$, such that the correlation between $\mathbf{w}_{1,1} = \mathbf{s}_{1,1} \mathbf{Y}_1^T$ and $\mathbf{w}_{2,1} = \mathbf{s}_{2,1} \mathbf{Y}_2^T$ is maximised. The subject weightings corresponding to the next component are estimated by projecting onto vectors that produce the next highest correlation value and so on. However, the weightings are constrained by the need for them to be uncorrelated with the previously estimated subject weightings. The $\mathbf{s}_{t,l}$ vectors are stored in $\mathbf{S}_t$, of shape $L \times M_t$. Hence the subject weightings and the spatial maps used to obtain them, $\mathbf{W}_t$ and $\mathbf{S}_t$, for the first modality are paired with the second modality. This approach was later extended to more than two modalities by Correa et al. [156], which they call multi-set CCA.

Correa et al. utilised multimodal CCA to decompose grey matter and first level fMRI contrast maps belonging to 36 healthy controls and 37 schizophrenia patients. They decomposed the data into 18 component and found that the first pair of grey matter and fMRI subject weightings were significantly different between the two groups. Moreover, the second pair of subject weightings were significantly correlated with the ages of the subjects.

Variants of the above approaches such as parallel ICA [157], coefficient constrained ICA [158], multimodal (two-block) PLS [159] and multimodal (multi-block) PLS also exist [153].

Two-block multimodal PLS is similar to multimodal CCA, except that the decomposition is asymmetric (one of the modalities is assumed to be dependent on

the other) and the covariance between the subject weightings is maximised as opposed to the correlation. Multi-block multimodal PLS extends this to more than two modalities. It assigns one modality as being dependent on the concatenated feature sets of all the other modalities. Hence it is still technically a decomposition of two complimentary data blocks, but the authors choose to call it multi-block to refer to the fact that it is applied to more than two modalities.

Parallel ICA is applied to two modalities and estimates a separate set of independent sources for each one. However, it adds a term that captures the correlation between the subject weights to the ICA objective function. In this sense, it is similar to multimodal CCA, except that there is some statistical independence among the component spatial maps. Vergara et al. later extended this to three modalities [160] and it is straightforward to extend the objective function to incorporate more modalities.

Coefficient constrained ICA is applied to data that is comprised of two groups. It concatenates the modalities and carries out the decomposition such that the subject weights are constrained to be discriminatory between the two groups.

The above approaches all have their own use case and so have advantages and disadvantages. The multimodal CCA and multimodal PLS methods are both computationally efficient and effective at finding information that is joint between modalities. They also provide an insight into how strong the relationships between the estimated components are, since they estimate separate subject weightings for each modality. However, multimodal PLS requires knowledge of what the dependent modality should be. CCA meanwhile, is not inherently suited to data where the number of features vastly outweighs the number of subjects, as is the case in neuroimaging. Therefore, it requires careful regularisation or some form of decomposition such as PCA beforehand. Moreover, if there is interesting information that is unique to only one modality, then multimodal PLS and multimodal CCA may ignore this.

Joint ICA and informed multimodal PLS are again both computationally efficient. Moreover, the independence assumption behind ICA has repeatedly been

shown to capture biologically relevant information in neuroimaging literature. Meanwhile, the informed multimodal PLS method has the ability to produce components relating to a variable of interest. However, they both suffer from disadvantages relating to concatenation. Different modalities can have vastly different numbers of voxels, units and signal-to-noise ratios. This means that the decomposition may be dominated by modalities that are the most varying or have the most features. Issues relating to the former can be minimised by careful preprocessing but issues relating to relative numbers of features will remain.

Parallel ICA overcomes some of the problems relating to concatenation. However, it requires careful tuning of model parameters. Moreover, the correlation term and the ICA term in the cost function may work against each other. This can produce components that are somewhere in between being independent and correlated across modalities, which makes interpretation difficult. The same issue, in addition to the concatenation issue, applies to the coefficient constrained ICA method.

Linked ICA takes into account the differences in attributes that can exist between modalities in its model. Hence it does not suffer from the disadvantages of joint ICA or parallel ICA. However, one of its disadvantage is that it is very computationally demanding. In addition, difficulty in interpreting linked ICA spatial maps and associated subject weights has been reported [161].

Ultimately, in this study, our aim is to utilise a data-driven technique in order to define regions within multiple modalities that have distinct trajectories with respect to age. With the exception of informed multimodal PLS, the major disadvantage with existing decomposition techniques is that it cannot be guaranteed that they will produce age related components. The subject weightings for them have to be associated with age post-hoc and existing studies have typically found only one or two age related components. Informed multimodal PLS can be considered more suited to our aim. However, as we later show in simulations, it is possible for it to combine several distinct ageing components into one component.

For the above reasons, we present a method based on clustering where the es-

estimated components are guaranteed to be age related. In addition, we model each voxel as having a binary membership to one of the components, i.e. that it only belongs to one component. The advantage of this is that the output is more interpretable, since each brain region is described by a single component, as opposed to the complicated mixture of many components provided by existing methods. Although our method is not an explicit decomposition model such as equation 4.1, it is still analogous to the decomposition techniques we have discussed in terms of its outputs. It produces a set of L components, each with its own associated subject weights and spatial map. The subject weights for each component can be plotted against age to view the trajectory associated with the component. We focus on age in this study but our method can be applied to any variable of interest. As with existing methods, L can be set to any arbitrary integer value. We also present an approach for setting its value based on the data. We apply our method to both simulated data and UK biobank data.

4.2 Methodology

4.2.1 The model

We are given spatial maps for T modalities and N subjects, each having age a_n , where $n = [1, N]$. Each spatial map consists of M_t voxels, where $t = [1, T]$, and the data is stored in $\mathbf{Y}_t$, of size $N \times M_t$. We assume that the trajectory of a given voxel as a function of age is governed by only one component, out of a possible L , which may or may not be global to all modalities. Our procedure for estimating these L components and their associated spatial maps is as follows. The procedure below does not only apply to age trajectories. For example, if the data belongs to a set of patients, age can be substituted for a disease measure. A diagrammatic representation of the steps below is shown in figure 4.1.

1. Bin each subject's data into B age bins to obtain $\mathbf{Y}_{b,t}^{Binned}$ of size $N_b \times M_t$, where $b = [1, B]$ and N_b denotes the number of subjects in bin b . Ideally, the

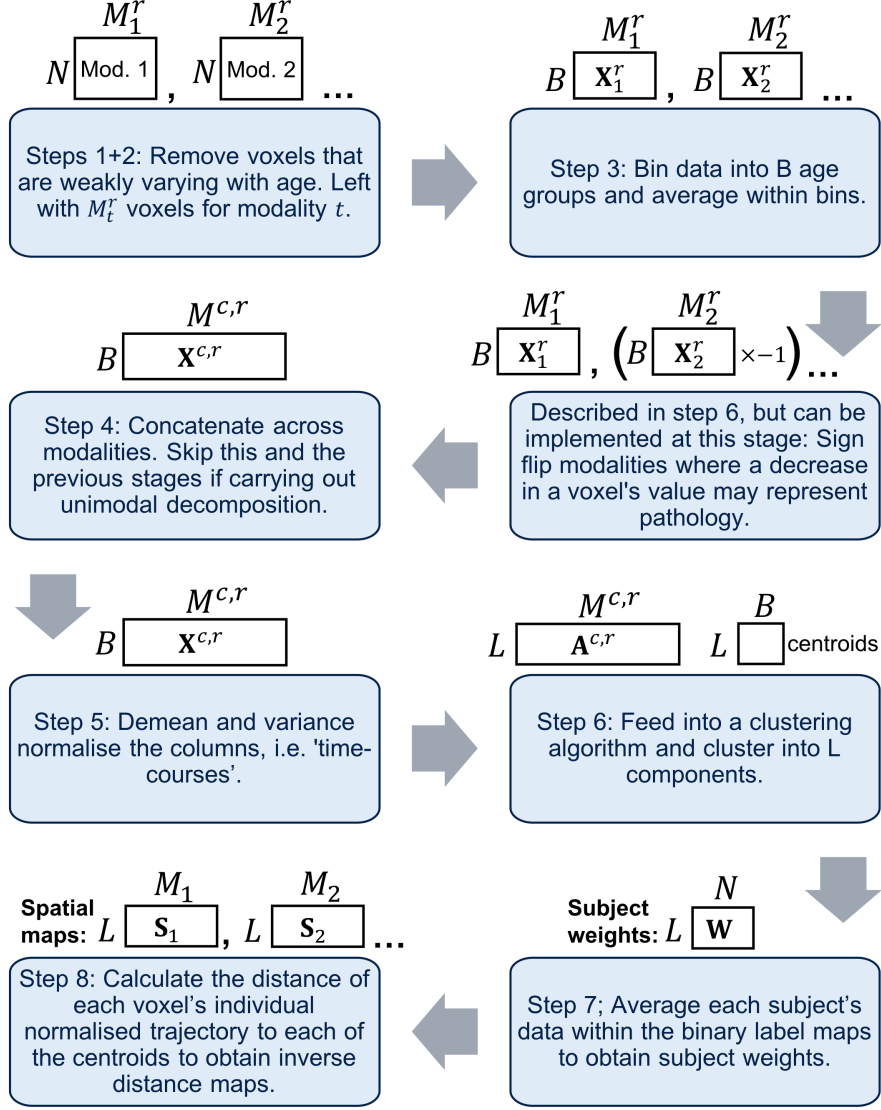


Figure 4.1: A diagram of the steps involved in our method. The step numbers correspond to the step numbers in the main text.

bins would be of equal width, have centres that are equally spaced within the age range of the data and contain roughly equal numbers of subjects. However, this does not have to be the case if the data does not allow it.

2. Remove voxels that are not showing any change with respect to age. We do this by conducting a number of unpaired two-tailed t-tests to compare a given voxel's values in bin b (where there would be N_b values) with the voxel's values in all the other bins. If there were 10 bins, this means that for each voxel we would have to conduct $(10^2 - 10)/2 = 45$ separate t-tests. We subsequently

remove any voxels that do not manage to obtain a p value lower than p_{thresh} for any of their t-tests, i.e. voxels that did not show a significant change across age bins. This leaves a reduced dataset $\mathbf{Y}_{b,t}^{Binned,Reduced}$ of size $N_b \times M_t^{Reduced}$. The p_{thresh} values that we present in the rest of this study have been Bonferroni corrected for the number of bin-wise comparisons.

3. Within each bin and modality, average the reduced dataset over the subjects, and concatenate the resulting $1 \times M_t^{Reduced}$ vectors. This results in the bin-averaged data $\mathbf{X}_t^{Reduced}$ of size $B \times M_t^{Reduced}$. Each column of $\mathbf{X}_t^{Reduced}$ represents a voxel's trajectory with respect to the bin centres.
4. Concatenate the data for different modalities to obtain $\mathbf{X}^{c,Reduced}$ of size $B \times M^{c,Reduced}$, where $M^{c,Reduced} = \sum_{t=1}^T M_t^{Reduced}$. If the aim is to obtain modality specific trajectories and not multimodal ones, then this step can be ignored and the remaining steps carried out for each modality separately.
5. Normalise each column of $\mathbf{X}^{c,Reduced}$ based on what is being investigated. The columns represent voxels, meaning that the values in a given column represent a trajectory for a voxel. In this study, we are interested in identifying trajectories that have different shapes irrespective of the scale or initial offset. Hence we demean and variance normalise the columns.
6. Feed $\mathbf{X}^{c,Reduced}$ into k-means or any other clustering algorithm (with the voxels as the samples and the rows as the features) to assign each voxel to one of L clusters. This results in L binary mask vectors, one for each cluster, of length $M^{c,Reduced}$ that can be reshaped into spatial maps. We store these in $\mathbf{A}^{c,Reduced}$ of shape $L \times M^{c,Reduced}$. In this study, we investigate an approach based on hierarchical clustering to identify the correct value of L to set in the clustering algorithm. If two voxels within the same modality have the same trajectory shape, but opposing directions of change, i.e. one is the negative of the other, then we are interested in assigning these to two separate clusters. In addition, the direction of change that can be considered to represent pathology changes

based on the modality. For example, in VBM data, a decrease in a voxel's value may be a detrimental change while the same would be true for an increase in an MD map's voxel value. If these two voxels have the same trajectory shape, then we are interested in clustering them together. Therefore, prior to concatenation we multiply all the data belonging to modalities such as VBM and FA, where a decrease in a voxel's value is generally considered to represent pathology, by -1.

7. Calculate the subject weightings corresponding to each component. This is achieved by first going back to the original data $\mathbf{Y}_t$ and removing voxels that did not pass the thresholding stage during step 2 to obtain $\mathbf{Y}_t^{Reduced}$ of shape $N \times M_t^{Reduced}$. $\mathbf{Y}_t^{Reduced}$ is concatenated across modalities to obtain $\mathbf{Y}^{c,Reduced}$ of shape $N \times M^{c,Reduced}$, and the columns of this matrix are normalised using the same normalisation factors from step 5. For example, if 2 was subtracted from the values in the first column of $\mathbf{X}^{c,Reduced}$ and the resulting values divided by 10, the same subtraction and division would have to be applied to the first column in $\mathbf{Y}^{c,Reduced}$. A subject weighting matrix $\mathbf{W}$ of shape $L \times N$, which contains an average of each subject's data within the binary masks, is then obtained. It is calculated by first using $\mathbf{W} = \mathbf{A}^{c,Reduced} (\mathbf{Y}^{c,Reduced})^T$, and subsequently normalising each row in $\mathbf{W}$ by the number of voxels in the corresponding binary mask from $\mathbf{A}^{c,Reduced}$. The subject weightings in row l of $\mathbf{W}$, where $l = [1, L]$, can be plotted against their corresponding ages to visualise the trajectory corresponding to component l . The exact same procedure can be applied to a test dataset, $\mathbf{Y}_t^{test}$, to obtain weightings for test subjects.

8. Obtain spatial maps corresponding to each component. The binary masks in $\mathbf{A}^{c,Reduced}$ do not give any indication of the possibility of a voxel belonging to a component that it was not assigned to. Moreover, they do not give any labels to voxels that did not pass step 2. For these reasons, as a final step, for all the voxels and a given component we calculate weightings between 0 and 1

indicating how likely they are to belong to that component. We explore two separate methods for achieving this, which we describe later.

4.2.2 Simulated data

4.2.2.1 Simulation method

We simulated data containing three modalities, $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$, as follows.

For each modality, a set of trajectories comprised of exponential, sigmoidal and polynomial functions were defined as shown in figure 4.2. The ‘measure of change’ can refer to age or a disease measure. For the rest of this chapter, we will refer to the simulated measure of change as a_n . The trajectory with a label of 2 is the same for all three modalities. In addition, for modalities two and three the trajectory with a label of 3 is the same. These are the multimodal trajectories that are common to at least two modalities. Hence, the total number of unique trajectories in the figure is seven.

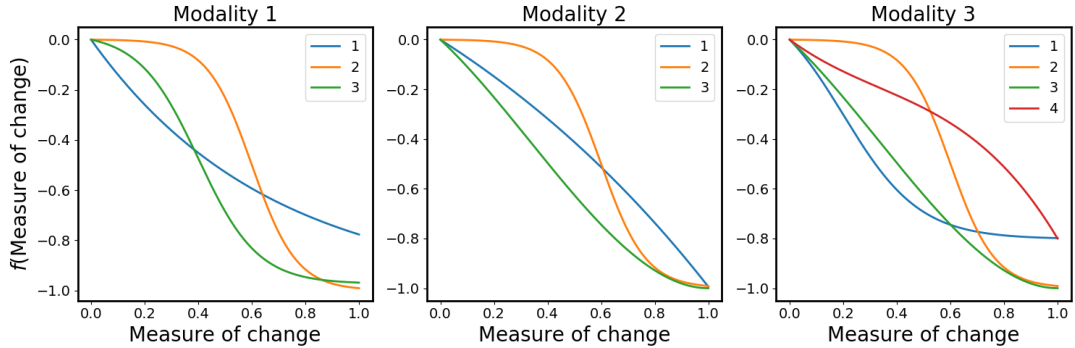


Figure 4.2: The simulated trajectories that each modality contained. The trajectory with a label of 2 is the same for all three modalities. For modalities two and three the trajectory with a label of 3 is the same. Hence, the total number of unique trajectories in the figure is seven.

A 2D spatial map was created for each modality and each of its voxels were assigned to one of the trajectories that belong to that modality. Although this is a somewhat idealistic simulation where there is no mixing between components in any voxel, it still provides a base case to test our method on. Additional voxels were also added to the spatial maps, defining regions that do not undergo any change with

respect to a_n . These spatial maps are displayed in figure 4.3, where the label numbers for a given modality correspond to its trajectory numbers in figure 4.2. Voxels that do not undergo any change have a label number of 0. In real data, different modalities can have different numbers of voxels and components, and each component can have different numbers of voxels relative to other components. We chose to reflect this in the simulations. The base number of voxels that each trajectory had for $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$ were $[500, 200, 150, 200]$, $[900, 300, 400, 100]$, $[700, 100, 300, 200, 200]$ respectively. The first element in each vector defines the number of non-changing voxels, while elements 2 onwards define the number of voxels for the trajectories, in order of their label number. These base numbers characterise the ratios of the number of voxels that different trajectories and modalities have. When actually simulating data, these vectors were multiplied by a constant factor, C , to determine the number of voxels. The spatial maps in figure 4.3 were created with $C = 9$.

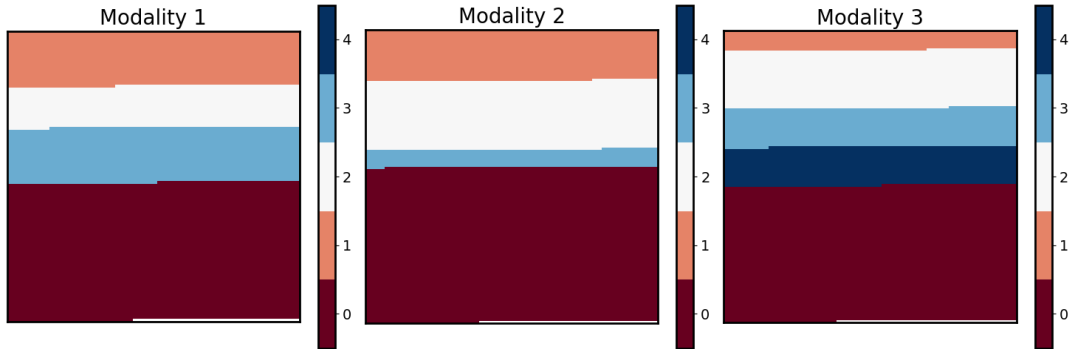


Figure 4.3: The spatial maps associated with the simulated trajectories. The label numbers correspond to the trajectory numbers in figure 4.2. Voxels that do not undergo any change have a label number of 0.

A total of N subjects were simulated, having a_n values uniformly spaced between 0 and 1. Spatial maps were created for them based on the label images in figure 4.3 and the trajectory values for their a_n . To reflect the fact that different modalities can have different SNRs, noise was added such that the mean SNR (defined by $\text{variance}(\text{signal})/\text{variance}(\text{noise})$), of $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$ were approximately $2S$, $0.5S$ and $1S$ respectively, where S is the global SNR factor that we vary in the simulations. In our case, the mean signal variance was calculated as the mean variance of each voxel's values across subjects. Noise sampled from a Gaussian distribution with

mean 0 and variance σ_t^2 was then added to the data for modality t , where the value of σ_t was set so as to achieve the desired SNR. Therefore, the SNR values for the simulations indicate the mean SNR of the voxels' trajectories. The non-changing voxels for a modality had values sampled from the same distribution that was used to sample the noise for that modality.

4.2.2.2 Influence of simulation and estimation parameters

We first measured how the accuracy of the estimated trajectories and component maps are affected by the simulation parameters and the choices made during estimation. The simulation parameters were the number of subjects N , SNR factor S and voxel factor C . The estimation parameters were the p value threshold p_{thresh} and the number of bins B . We ran our method for all combinations of these parameters, using k-means as the clustering algorithm. We skipped the concatenation step, meaning that a set of trajectories was estimated for each modality separately. The number of target clusters for k-means, L , was set to 7, which is higher than the ground truth value for all simulated modalities.

To compare the estimated trajectories and spatial maps to that of the ground truth ones, we binned, averaged (within the bins) and normalised the ground truth trajectories in the same way as steps 1, 3 and 5 during the estimation. The Hungarian algorithm [162] was subsequently used to match each of these bin-averaged and normalised trajectories to one of the estimated centroids from k-means. If we define a cost matrix, $\mathbf{K}$, where K_{ij} is the cost of matching the ground truth trajectory i to one of the estimated clusters j , then the Hungarian algorithm attempts to minimise the following:

$$\sum_i \sum_j K_{i,j} D_{i,j} \quad (4.6)$$

$\mathbf{D}$ is a boolean matrix where $D_{i,j}$ is equal to 1 if trajectory i has been matched to centroid j , and each row of $\mathbf{D}$ must be matched to at most one column. We set the cost of matching two components to be the Euclidean distance between them.

After matching, the similarity between the binary spatial map associated with

each ground truth trajectory and the binary spatial map associated with the ground truth’s matched trajectory was calculated. The statistic we used as the measure of similarity was the dice index [163]. The dice index is defined as:

$$\frac{2TP}{2TP + FN + FP} \quad (4.7)$$

TP , FN and FP are the total number of true positive, false negative and false positive voxels in the estimated spatial map respectively. We also measured the similarity between the ground truth and estimated trajectories. A smoothing splines regression model with five interior knots and a degree of three was fitted to each cluster’s subject weightings. Both the fitted splines function and the matched trajectory function were then evaluated for 1000 equally spaced values between 0 and 1 and stored in two separate vectors. The vectors were demeaned and variance normalised, and the Euclidean distance between them was calculated.

For each combination of parameters, we averaged the two similarity measures across all components and modalities, resulting in two final values. Moreover, each simulation run was repeated 10 times, and these two values were averaged again across the 10 runs.

4.2.2.3 Determining the correct number of components

We investigated the use of hierarchical clustering [164] to determine the correct number of trajectories that we should decompose a given dataset into. In the first step of the hierarchical clustering algorithms we consider, each voxel starts as its own cluster. In subsequent steps, two clusters that are closest to each other based on some linkage criteria are merged to create a new cluster. The clusters that were merged are removed from the list of clusters, and the newly merged cluster is added. This is repeated until just one cluster is left. This creates a hierarchy of clusters that can be viewed on a hierarchical tree called a dendrogram. The length of the ‘branches’ in a dendrogram are representative of the distance, which is defined by the linkage method, between the clusters that were merged. If a dataset contains clearly

separable clusters, then the distance required to merge these real clusters should dramatically increase relative to each cluster’s children. This would become visible on a dendrogram, hence a visual inspection of the dendrogram will aid in determining the correct number of components. Various automated methods for determining the correct number of clusters exist [165]. However, real datasets are unlikely to contain clearly separable clusters and the visual inspection of a dendrogram will allow the user to choose the most appropriate number of clusters for their data and experiment.

We simulated data with $N = 600$, $S = 1$, $C = 9$ and applied our method with $p = 0.005$ and $B = 10$, replacing k-means with a hierarchical clustering algorithm in step 6, as implemented in the `scipy.cluster.hierarchy` package in `scipy v1.1.0` [140]. We compared the dendrograms that resulted from the ‘single’, ‘complete’, ‘average’, ‘weighted’, ‘centroid’, ‘median’ and ‘ward’ linkage methods. We later focus on the complete and ward methods. At each step, the complete method merges the two clusters that have the smallest maximum Euclidean distance between their elements. The ward method meanwhile merges the two clusters that result in the smallest increase in the within cluster variance post merging. We considered the cases where the non-trajectory voxels are and are not included in the clustering, and further looked at both unimodal and multimodal trajectories.

4.2.2.4 Deriving continuous component maps

For reasons described in section 4.2.1, the final step in our method is to obtain continuous component spatial maps. The aim is to derive a value between 0 and 1 for each voxel, with this value being representative of the likeliness for that voxel to belong to a given component, and constrain the values for a voxel to sum to 1 across all components. Therefore, the resulting spatial maps can be thought of as ‘probability’ maps. We investigated two separate approaches for achieving this aim.

In the first approach, a multivariate Gaussian is first fitted, using the frequentist formulation, to each estimated cluster after step 6. Next, a bin-averaged dataset is obtained in the same way as step 3, except that the voxels that did not pass the

thresholding in step 2 are kept. This results in $\mathbf{X}_t$ of size $B \times M_t$, which is then concatenated as in step 4 yielding $\mathbf{X}^c$ of size $B \times M^c$ (the concatenation step is not carried out if only unimodal clustering has been implemented). Afterwards, the probability density functions corresponding to the L fitted multivariate Gaussians are evaluated for each of the column vectors in $\mathbf{X}^c$, producing L vectors of length M^c . These are stored in $\mathbf{S}^c$ of shape $L \times M^c$. Finally, the columns of $\mathbf{S}^c$ are normalised so that they sum to 1 and the rows can subsequently be reshaped into spatial maps.

In the second approach, $\mathbf{X}^c$ is again created. For each of the column vectors in $\mathbf{X}^c$, the squared Euclidean distance to the L estimated centroids after step 6 is calculated, resulting in L vectors of length M^c , which we denote by $\mathbf{d}_l$. Next, new vectors, $\mathbf{s}_l$, are calculated using $\mathbf{s}_l = (1/\mathbf{d}_l)^g$, which are stored in $\mathbf{S}^c$ of shape $L \times M^c$. Finally, $\mathbf{S}^c$ is normalised as before. The role of g is to control how much to accentuate the differences between time-courses. For example, if two components exist in a dataset and they both have similar time-courses, a voxel in $\mathbf{S}^c$ with $g = 1$ could have values of 0.45 and 0.55. However, if $g = 10$, the same voxel would have values of 0.12 and 0.88. We include this parameter in order to potentially give the user more control in their exploration of the components.

We applied these two approaches to the estimated clusters from the previous subsection. We initially arbitrarily set $g = 3$ in this study.

4.2.2.5 Comparisons to other methods

We applied two existing methods of multimodal decomposition to data simulated with $N = 600$, $S = 1$, $C = 9$. We chose joint ICA and informed multimodal PLS since we believe that they can produce outputs closest to what we aim to produce with our method. The independence assumption behind the former has potential to identify biologically distinct ageing components, while the latter can be explicitly informed on age. In addition, they can both be implemented in a unimodal setting, although we only explore their multimodal application in this study. Linked ICA would also have been a suitable method to compare to. However,

its computational demands for decomposing the UK biobank data would have been unfeasible. Furthermore, in initial testing on the simulated data, linked ICA was not able to produce reasonable components, possibly due to its priors not being compatible with the simulated data. Therefore, we do not report linked ICA results in this study.

When implementing joint ICA, the simulated data was normalised in the same way as described in Calhoun et al.’s study [152]. Each modality was divided by a constant factor so that the average sum of squares for a modality across all subjects and voxels is 1. This is to ensure that the units of the different modalities are the same. In addition, for any modality that only has positive values, they flipped the signs for alternate voxels. This is so that the modality would have approximately zero mean and voxels with large negative and positive values would correspond to voxels that originally had a large magnitude. Since all of our simulated data were positive valued, we implemented this step for all the modalities. The data from the different modalities were concatenated and FastICA, as implemented in scikit-learn v0.20.0 [142], was used to carry out spatial ICA on this concatenated data.

For the informed multimodal PLS, we concatenated the data and demeaned and variance normalised the columns, i.e. across subjects. Scikit-learn was then used to implement PLS on this data, with age set as the response variable.

4.2.3 UK biobank data

We applied our method to data from the UK biobank, utilising VBM and TBSS images. The TBSS data consisted of FA, MD and axial diffusivity (L1) measurements, resulting in a total of four modalities that we analysed.

Out of the 14521 subjects that had their data preprocessed, 2751 of them had missing data and we removed these subjects. Sampling subjects from the remaining set was carried out as in section 3.2.4. A total of 3036 subjects were sampled, selecting approximately equal numbers from each one-year age group to ensure uniformity. As before, subjects that had ages of 45, 46, 78, 79 and 80 were removed since there

were very few subjects that had these ages. This time, the 3036 subjects were split into a training and test set, with the training set consisting of 2134 subjects (approximately 70%). We applied our method to the training set and afterwards used the estimated components to calculate subject weightings for the test set.

To create a mask for the VBM data, we thresholded the UK biobank VBM template at 0.1 and binarised it. The VBM images for all the subjects were subsequently masked by this. The TBSS images were masked by a binarised version of the FA skeleton included in the FSL v5.0.11 [78] standard directory.

Both unimodal and multimodal versions of our method, using hierarchical clustering, were implemented, with $p_{thresh} = 1 \times 10^{-10}$ and $B = 10$. This p threshold was chosen such that approximately 13% of the original number of voxels were kept after step 2. This is due to the fact that hierarchical clustering initially creates a voxel by voxel distance matrix, and reducing the number of voxels to this would enable the distance matrix to fit in memory. In addition, for reasons described in step 6, the data for VBM and TBSS FA were multiplied by -1.

Finally, for comparison, joint ICA and informed multimodal PLS were applied to the same data. The sign flipping step in joint ICA was carried out for all the modalities, since they are all positive valued.

4.3 Results

4.3.1 Simulated data

4.3.1.1 Influence of simulation and estimation parameters

Figure 4.4 shows the dice index when comparing the estimated spatial maps to the ground truth ones for various simulation and estimation parameters. The values at the top of the figure show what the parameters not included in a plot were set to. For example, the top left plot shows the dice index for different combinations of SNR factor and numbers of subjects while the voxel factor, p threshold and number

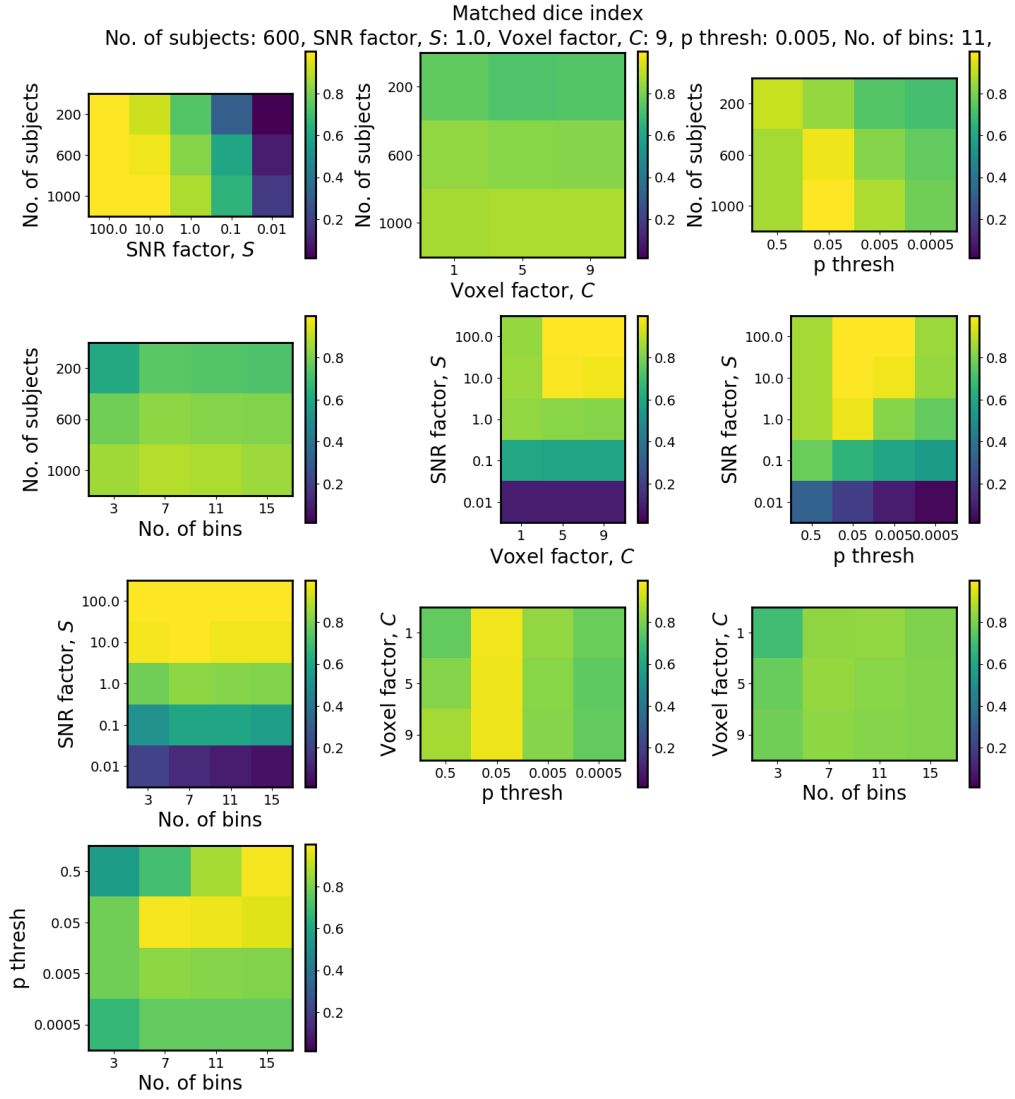


Figure 4.4: This figure shows how the dice index, when comparing the estimated binary spatial maps to the ground truth ones, is affected as various simulation and estimation parameters are varied.

of bins were 9, 0.005 and 11 respectively. Similarly, figure 4.5 shows the normalised distance between the estimated and ground truth trajectories.

As expected, increasing the number of subjects and SNR typically improved both the dice index and the distance measure. Increasing the number of voxels did not have much of an effect on the dice index, but did improve the distance measure. This is likely due to the subject weight calculation in step 7 of our method effectively being an averaging over the voxels. Hence if there are more voxels then this averaging is likely to result in more accurate subject weightings and fitted regression models. In general, when the number of bins is 3, the metrics are worse compared to the higher

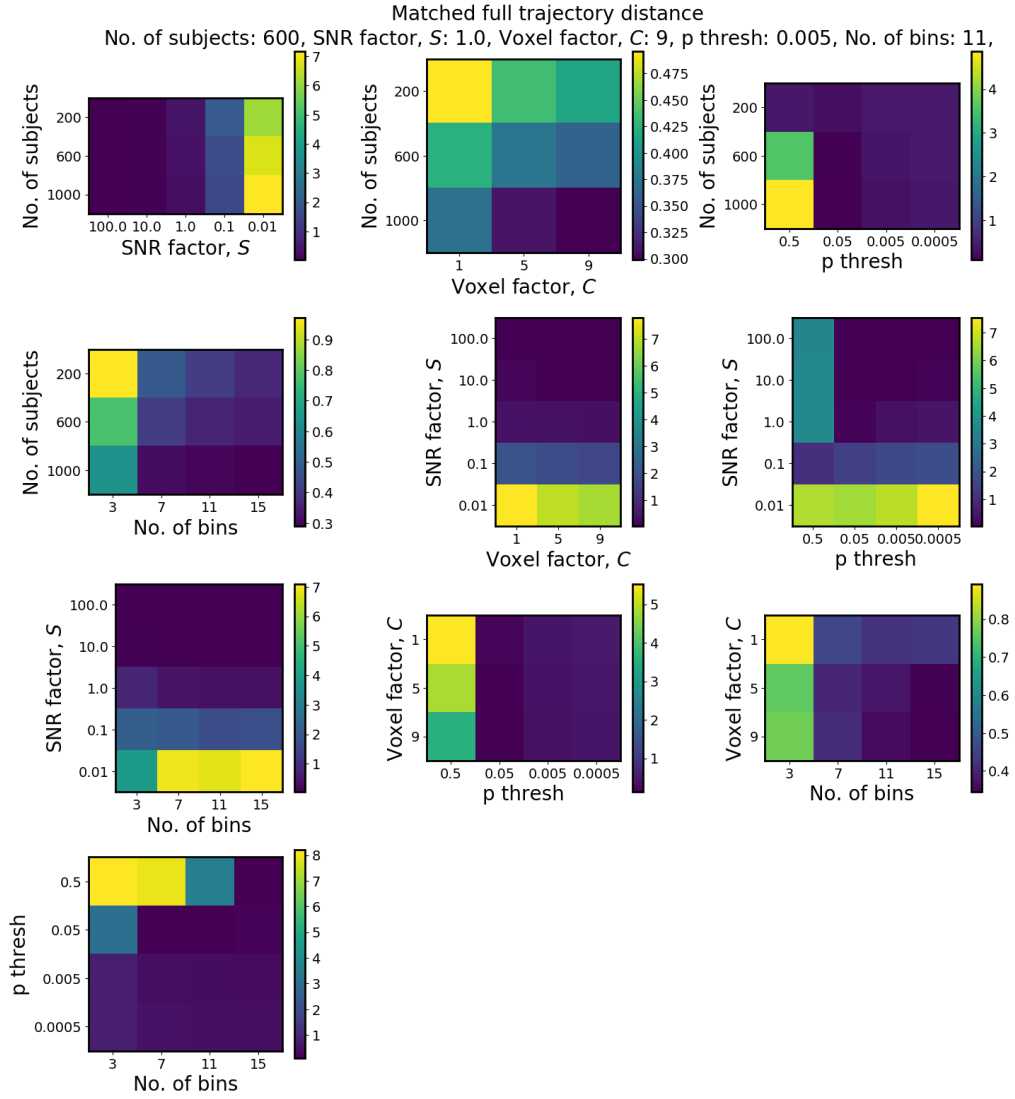


Figure 4.5: Same as figure 4.4, but showing the distance measure between the estimated ground truth trajectories instead.

numbers of bins. There are exceptions to this, such as when the SNR is at the lowest value. When the number of bins is this low the full shapes of the trajectories cannot be captured, meaning that the clustering algorithm cannot group the voxels into the correct components. When considering the distance measure, in general as the p threshold increases so does the performance. The exception to this is when the SNR is very low, since in this case a high p threshold will mean that not enough voxels will be left for clustering. When considering the dice index however, the relationship with the p threshold becomes more complicated. In the majority of the cases, the dice index improves when the p threshold is decreased from 0.5 to 0.05 but then starts to deteriorate after this.

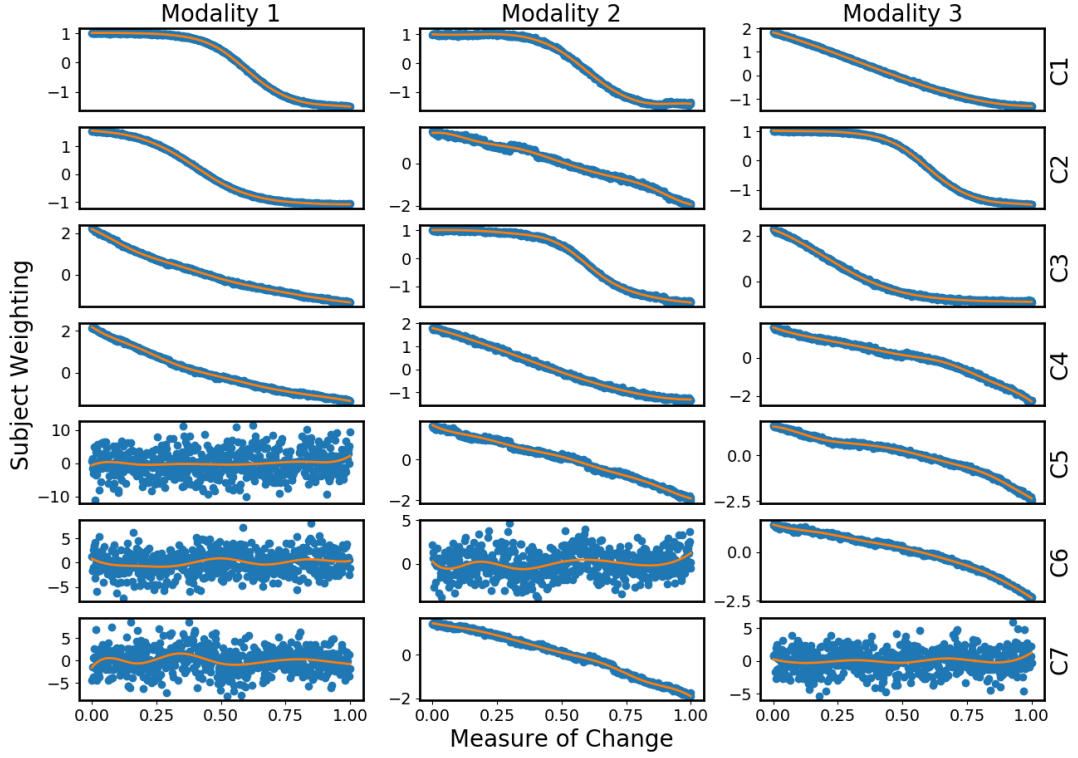


Figure 4.6: The estimated trajectories for $N = 600$, $S = 1$, $C = 9$, $p = 0.005$, $B = 10$ and $L = 7$.

To aid in understanding the relationship between the dice index and the p threshold, we present the estimated trajectories and associated binary spatial maps for $N = 600$, $S = 1$, $C = 9$, $p = 0.005$, $B = 10$ and $L = 7$ in figures 4.6 and 4.7. Since we requested more trajectories than actually existed in the data, for all modalities our method has split some trajectories into two separate ones. Consequently, the associated binary spatial maps have also been split into two separate maps. When calculating the distance measure, it does not matter which of the two trajectories get matched to the ground truth one since they are both still very close to it. However, each of the associated binary spatial maps only cover about half of the true region, meaning that the dice index suffers.

On the other hand, when $p = 0.05$, there are enough non-trajectory voxels left that the clustering algorithm does not need to split up one of the trajectories in order to achieve the requested number of clusters. Instead, it groups the non-trajectory voxels into a larger number of clusters. This result is displayed in figure 4.8.

These results indicate that our method may benefit from techniques which can

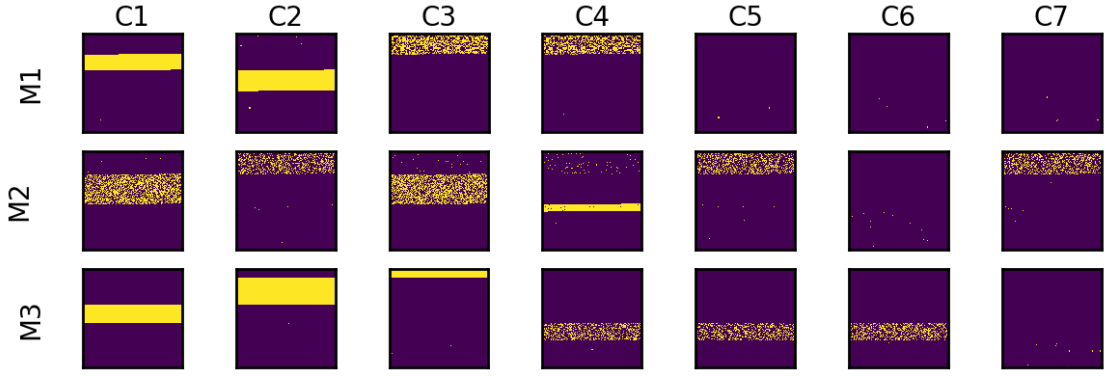


Figure 4.7: The estimated binary spatial maps associated with each of the trajectories in figure 4.6, where $N = 600$, $S = 1$, $C = 9$, $p = 0.005$, $B = 10$ and $L = 7$. Since the clustering was carried out for each modality separately, the relative ordering of the components does not have any significance.

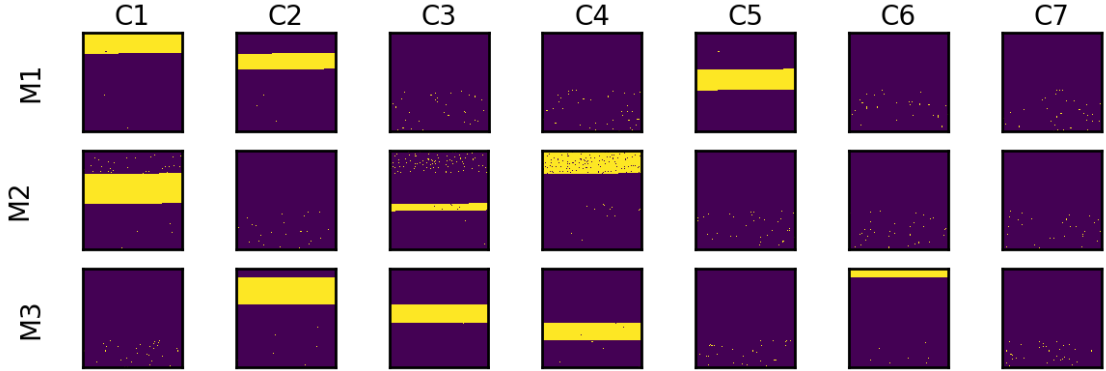


Figure 4.8: The estimated binary spatial maps when $N = 600$, $S = 1$, $C = 9$, $p = 0.05$, $B = 10$ and $L = 7$.

aid in estimating the correct number of clusters and reducing the dependence on the p threshold.

4.3.1.2 Determining the correct number of components

We explored the use of hierarchical clustering to determine the number of components to estimate. When no non-trajectory voxels were included in the simulated data, only the complete and ward methods resulted in interpretable dendrograms for unimodal clustering. These are shown in figures 4.9 and 4.10. We saw in figure 4.2 that modalities 1, 2 and 3 had three, three and four unique trajectories respectively. This is clearly represented in the structure of the ward dendrogram. It is represented in the complete dendrogram as well, although less clearly.

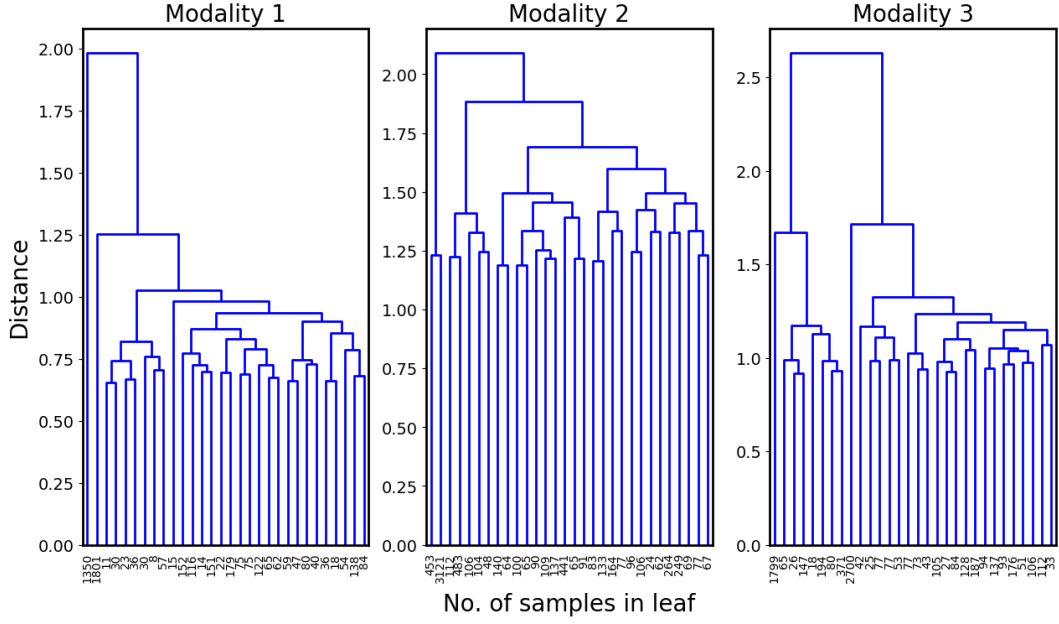


Figure 4.9: The resulting dendrogram when the ‘complete’ hierarchical clustering method was applied to simulated data that contains no non-trajectory voxels. The ‘leaves’ are the clusters at the lowest point in the hierarchy.

The complete and ward methods were next applied to simulated data that did contain the non-trajectory voxels. The dendrograms are shown in figures 4.11 and 4.12. There is still a clear structure in the ward dendrogram that reflects the ground truth number of trajectories. The non-trajectory voxels now appear as an extra branch, meaning that if the hierarchy was cut at e.g. 15 for all modalities, then for each modality there would be $L - 1$ components corresponding to the actual trajectories, and 1 component corresponding to noise. This time the structure of the complete dendrogram, however, does not reflect the ground truth. The clusters corresponding to non-trajectory voxels have been merged much higher up the hierarchy relative to the trajectory clusters, meaning that a suitable distance at which to cut the dendrogram does not exist. Based on these results the ward method is the superior approach, which is why we chose to use it for later experiments.

The ward dendrogram for multimodal clustering, i.e. when the data from different simulated modalities were concatenated, is shown in figure 4.13. The simulated data contained the non-trajectory voxels. There were seven unique trajectories in the concatenated simulated data and the dendrogram shows eight clear branches.

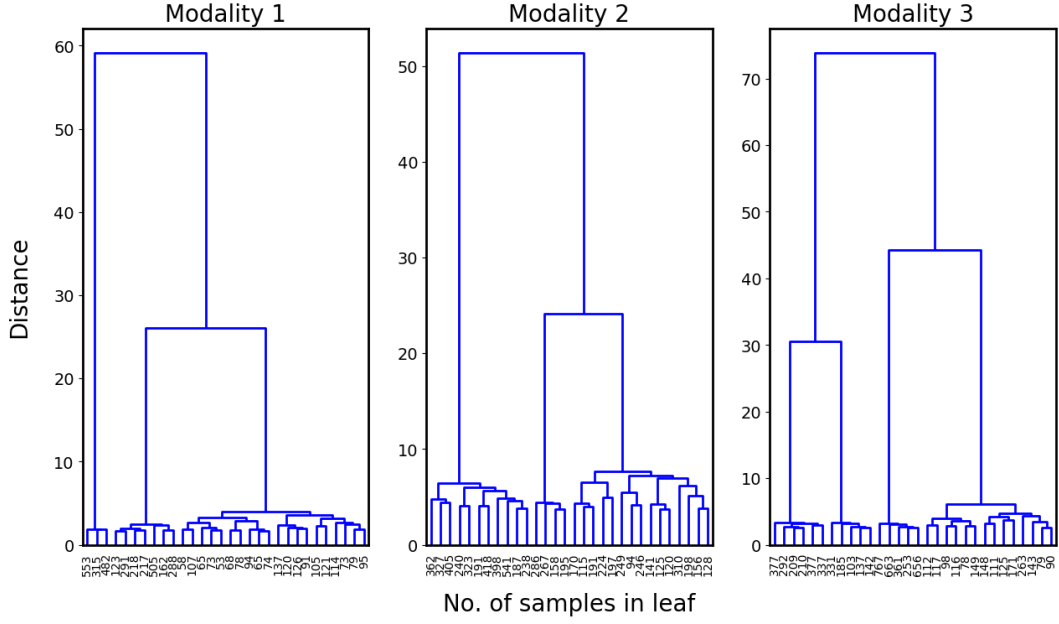


Figure 4.10: The resulting dendrogram when the ‘ward’ hierarchical clustering method was applied to simulated data that contains no non-trajectory voxels.

As before, one of the branches corresponds to noise and the rest to trajectories. This indicates that Ward has again been successful in determining the ground truth number of trajectories. The dendrogram was cut so that eight clusters are left and the consequent binary spatial maps are displayed in figure 4.14. By comparing these to the ground truth simulated trajectories and spatial maps in figures 4.2 and 4.3, we can see that each modality has been correctly decomposed, with component 6 being the noise component. In addition, voxels that had the same trajectory across the different modalities have been accurately grouped together.

There are also some cases where voxels have been attributed to the incorrect component. For example, component 7 corresponds to a trajectory that was only in modality three, and so the spatial maps for modalities one and two should not have any nonzero voxels. However, some of the voxels which should have been clustered with component eight for modality two appear in component seven and vice versa. This is due to the demeaned and variance normalised trajectories for these two components looking very similar. Nevertheless, the vast majority of the voxels for these two components have still been clustered correctly.

Instead of deriving clusters by directly cutting the dendrogram, it is possible

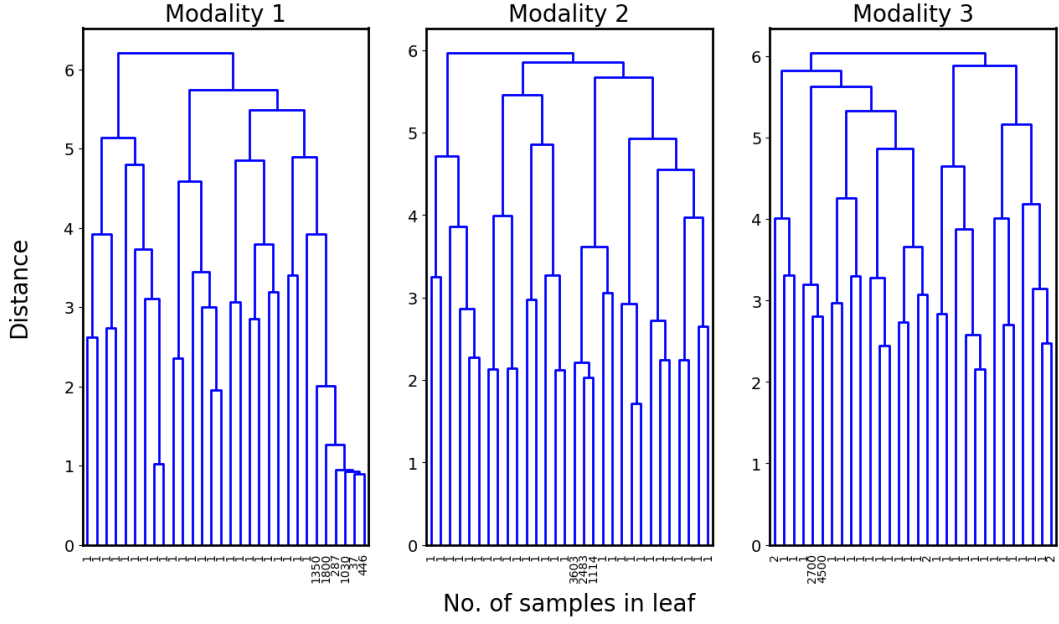


Figure 4.11: The resulting dendrogram when the complete hierarchical clustering method was applied to simulated data that also contains non-trajectory voxels.

to rerun our method using a different clustering algorithm and set the number of clusters to that learned from the dendrogram. We implemented this for the same simulated data with k-means and, as the spatial maps in figure 4.15 shows, the results are similar. However, since this cannot be guaranteed for every set of parameters and data, deriving clusters by directly cutting the dendrogram may be the better approach.

As we now illustrate, it is possible for hierarchical clustering to fail if the p threshold is too high. We simulated data using the same simulation parameters and ran our method using hierarchical clustering as before, except that we increased the p threshold to 0.1. The dendrogram for this is shown in figure 4.16. Since many more non-trajectory voxels are left for clustering, it becomes more likely that there will be clusters of non-trajectory voxels that are more different to each other than the trajectory clusters are to each other. These non-trajectory voxels get merged higher up the hierarchy and it becomes difficult to determine how many trajectory clusters exist, or at what distance to cut the dendrogram. However, in this case it is still obvious that the one large branch on the left corresponds to non-trajectory voxels since the sample counts for the leaves belonging to it are very small compared

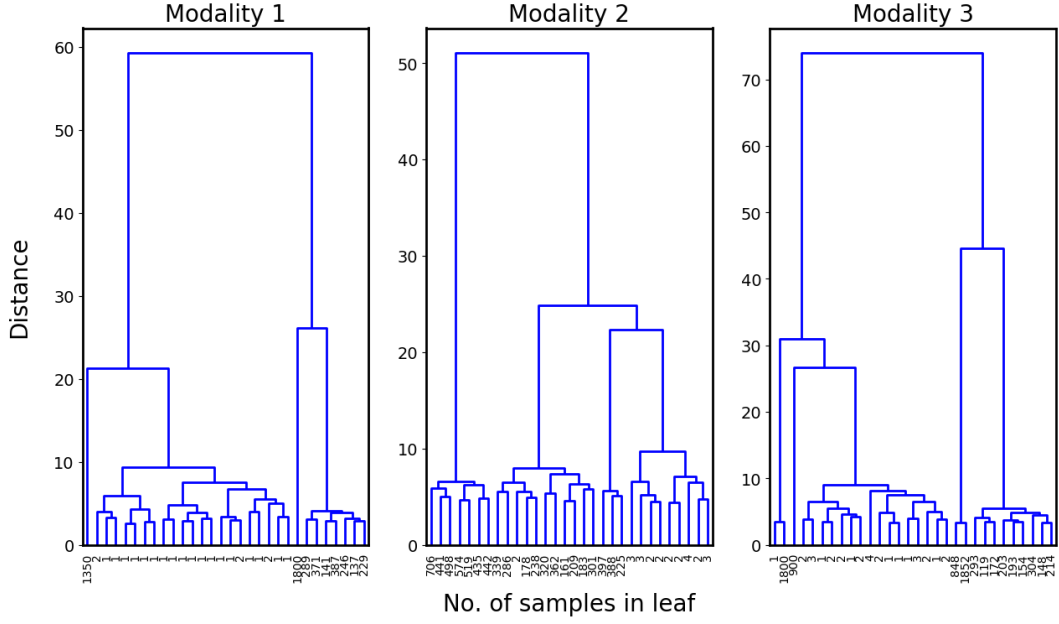


Figure 4.12: The resulting dendrogram when the ward hierarchical clustering method was applied to simulated data that also contains non-trajectory voxels.

to sample counts for the leaves belonging to the branches on the far right. This suggests that it may be possible to ‘prune’ the data by removing voxels that belong to leaves with very small relative sample counts, and then rerunning the hierarchical clustering on this pruned data.

We implemented this approach by removing voxels that belonged to leaves with a sample count of less than 30. The new dendrogram after rerunning hierarchical clustering on this pruned data is shown in figure 4.17. Since we effectively removed the noise branch by pruning, there are now seven clear branches in the dendrogram, each corresponding to a separate trajectory.

We cut the dendrogram such that seven clusters are created. The resulting spatial maps are in figure 4.18. These are again representative of the ground truth. The estimated trajectories corresponding to these component maps are shown in figure 4.19, which also reflects the ground truth.

4.3.1.3 Deriving continuous component maps

We applied the two approaches for obtaining continuous component spatial maps to the estimated clusters corresponding to figure 4.18. The spatial maps derived using

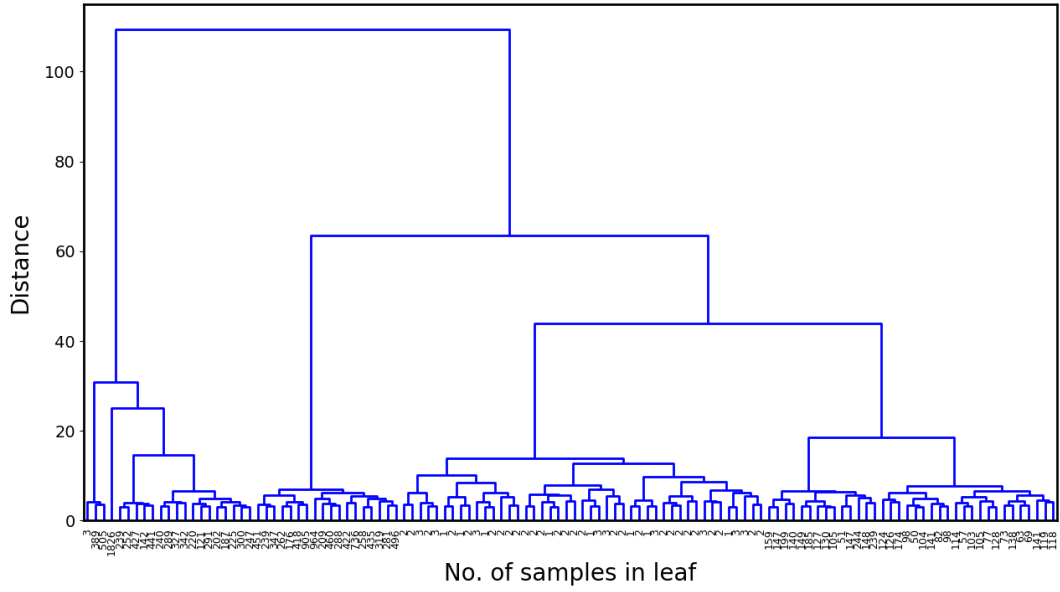


Figure 4.13: The dendrogram for multimodal hierarchical clustering using the ward method.

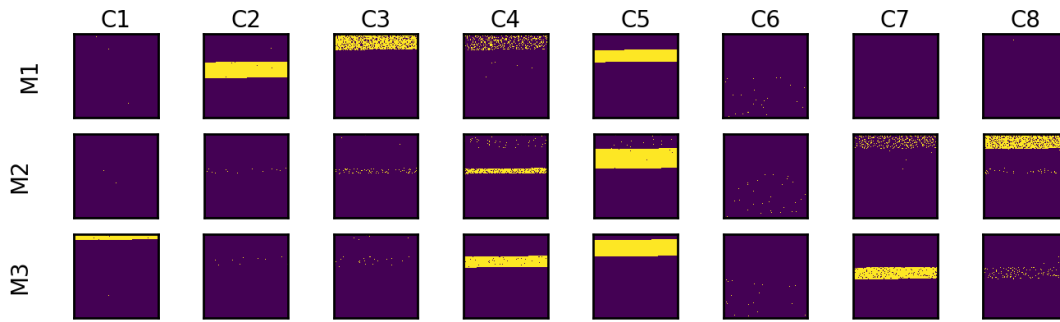


Figure 4.14: The resulting spatial maps when the dendrogram in figure 4.13 was cut such that eight clusters were left.

the multivariate Gaussian approach are shown in figure 4.20, while those derived from the squared Euclidean distance approach are shown in figure 4.21. Many of the voxels that were misclassified in the binary maps in figure 4.18, for example those that were assigned to component seven but should have been assigned to component six, now have more mid-range values in the spatial maps for both of these approaches. This suggests that there may be some useful information to be gained by considering these continuous maps in addition to the binary ones. Those from the multivariate Gaussian approach are closer to being binary, but the same could be obtained for the squared Euclidean distance approach by setting a larger g parameter. Due to the ability to set a g parameter based on what the user desires,

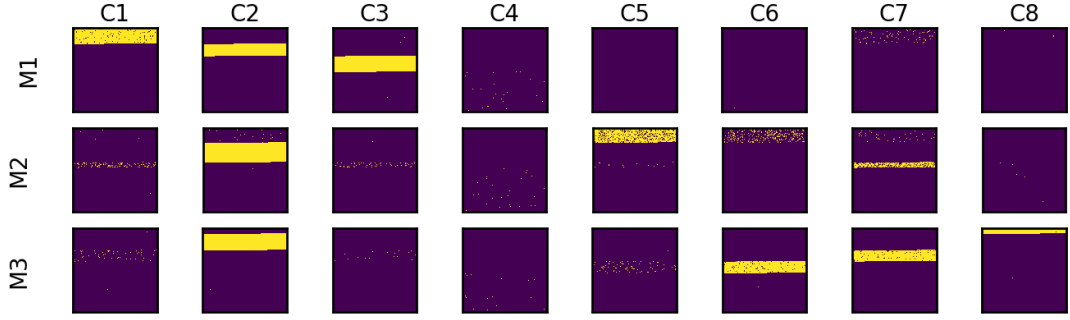


Figure 4.15: The resulting spatial maps when our method was rerun using k-means with eight target clusters as the clustering algorithm.

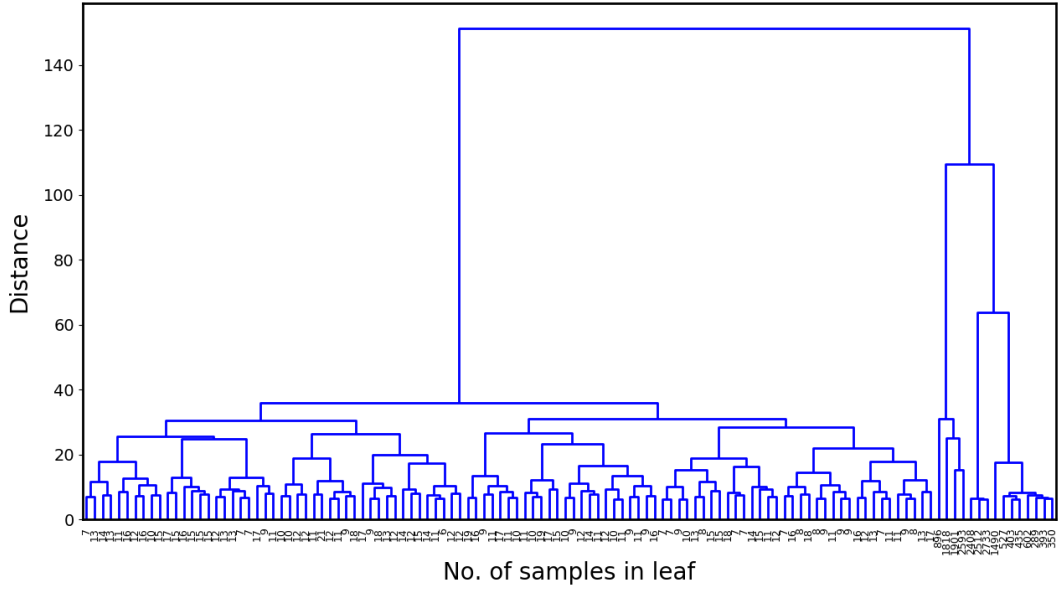


Figure 4.16: The dendrogram from multimodal hierarchical clustering when the p threshold was lowered to 0.1.

the squared Euclidean distance approach may be more useful.

4.3.1.4 Comparisons to other methods

To compare to some existing methods, we applied joint ICA and informed multimodal PLS to the simulated data. For both of these we made some use of the ground truth and chose to decompose the dataset into eight components, since we would ideally want to obtain seven trajectory components and one noise component. The estimated subject weightings for each component are plotted against age for both informed multimodal PLS and joint ICA in figure 4.22. The associated spatial maps for informed multimodal PLS and joint ICA are in figures 4.23 and 4.24 respectively.

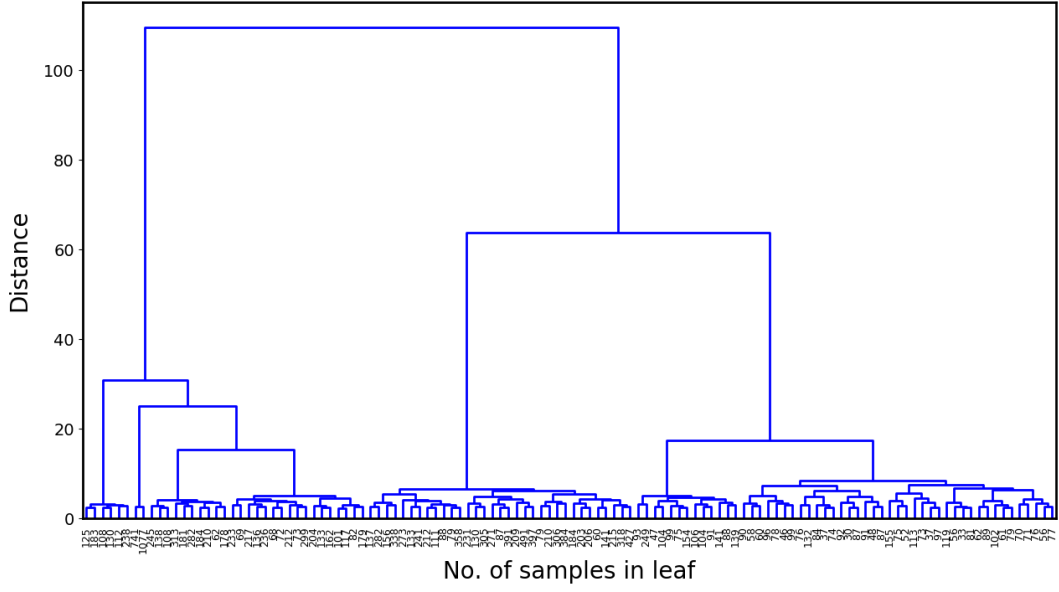


Figure 4.17: The resulting dendrogram after the voxels that belonged to leaves with a sample count of less than 30 in the dendrogram in figure 4.16 were removed from the data.

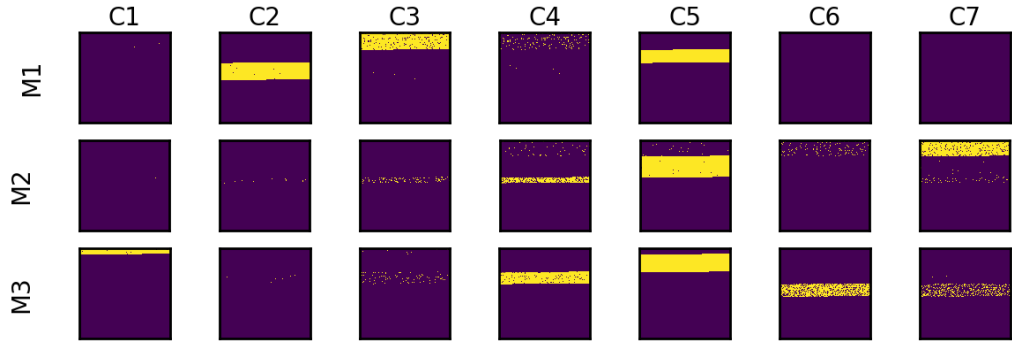


Figure 4.18: The spatial maps resulting from cutting the dendrogram in figure 4.17 such that seven clusters are created.

It is clear that neither method was successful at disentangling the data into the correct trajectories. The first three components from the PLS method are mixtures of the different trajectories, with the rest being mainly composed of noise. The ICA method was slightly more successful in that the trajectories are spread over more components, meaning that there were less noise components. In addition, component two actually corresponds to one of the ground truth components. However, the rest of the non-noise components are again mixtures of the ground truth trajectories.

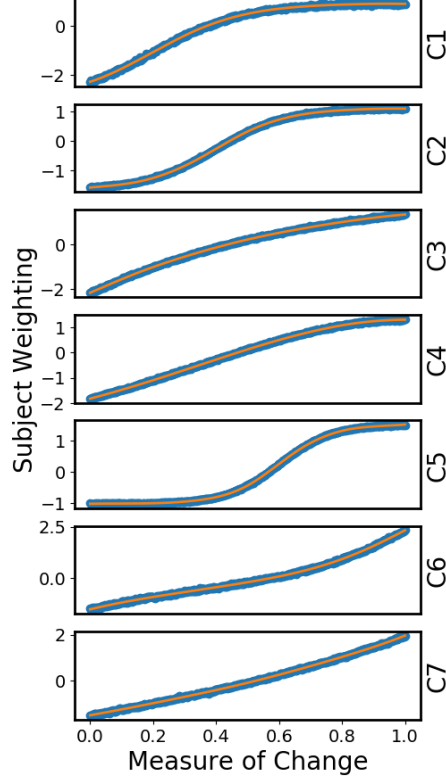


Figure 4.19: The trajectories corresponding to the spatial maps in figure 4.19.

4.3.2 UK biobank data

4.3.2.1 Single modality

We first applied our method to each UK biobank modality separately to obtain unimodal components. The resulting dendrograms are in figure 4.25. There were no branches that had very low relative leaf sample counts and so we did not prune any of the branches. Based on these dendrograms, we decided to cluster the VBM, FA, MD and L1 data into 3, 3, 4 and 4 components respectively. The corresponding component subject weights for the training set have been plotted against age in figure 4.26, while those for the test set have been plotted in figure 4.27. We correlated the test subject weights with age and the Pearson correlation coefficients are shown in table 4.1. The correlation values were all significant after Bonferroni correction. Smoothing spline regression models with degree two and one interior knot were also fitted to the subject weights from the training set. These are displayed in figure 4.28.

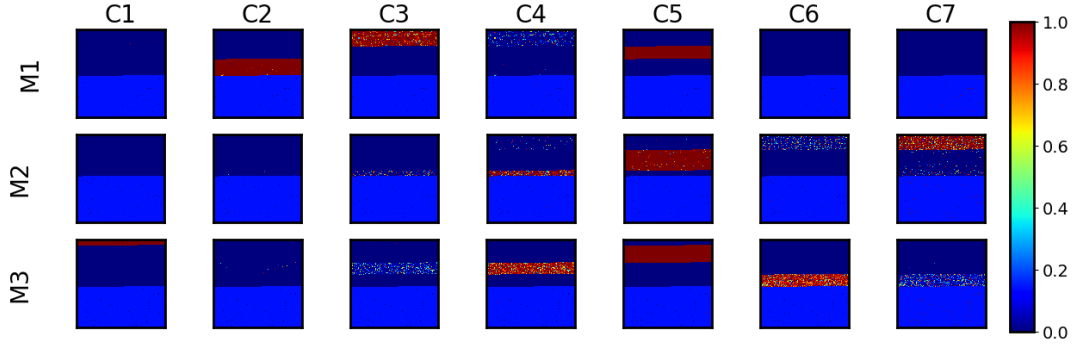


Figure 4.20: Continuous component spatial maps obtained using the multivariate Gaussian approach.

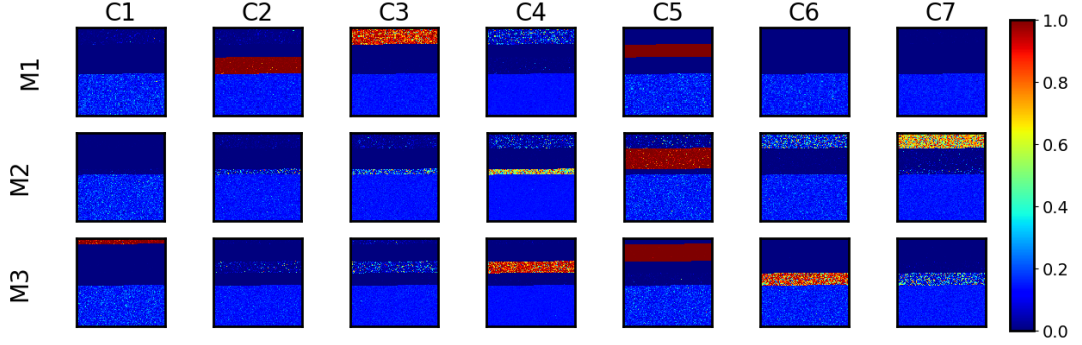


Figure 4.21: Continuous component spatial maps obtained using the squared Euclidean distance approach.

As described in section 4.2.2.4, we derive continuous maps with the Euclidean distance approach by calculating $(1/\mathbf{d}_l)^g$ and normalising so that the values for a given voxel across components sum to 1. However, when the number of components is small, as is the case for all the modalities here, it is not unlikely for a voxel's trajectory to be very far from all the components but, in relative terms, still be much closer to one component compared to the rest. Therefore, in the final step, when we normalise the voxel's distances to the components to sum to 1, it may appear as if the voxel is very strongly associated with one component, whereas in reality it was weakly associated with all of them. For this reason, we display the pure unnormalised inverse distances, i.e. $1/\mathbf{d}_l$, instead. For each component that we display, we selected the slices that were most representative of the features seen in that component. We show both the binary map and inverse distance map for the set of slices. The full extent of the spatial maps for all the components and methods can

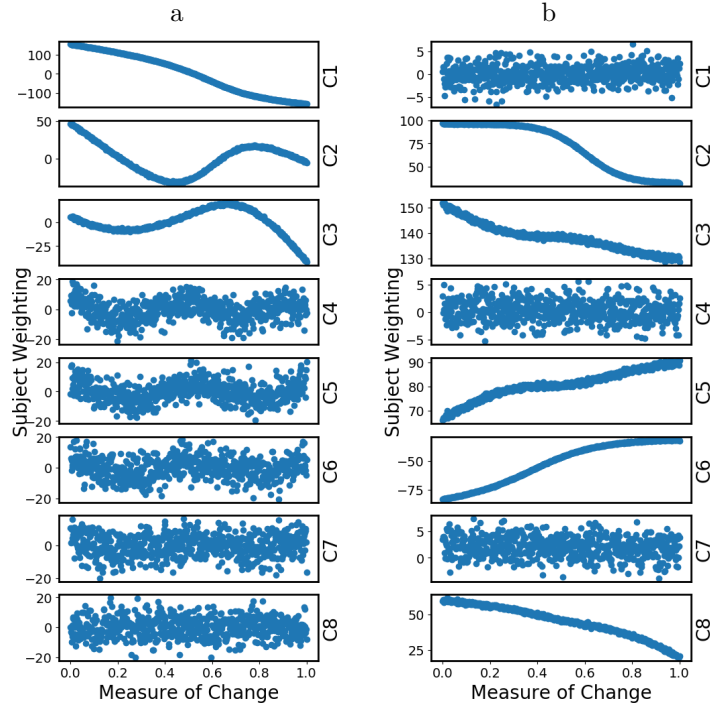


Figure 4.22: The estimated subject component weightings plotted against age for: a - informed multimodal PLS, b - joint ICA.

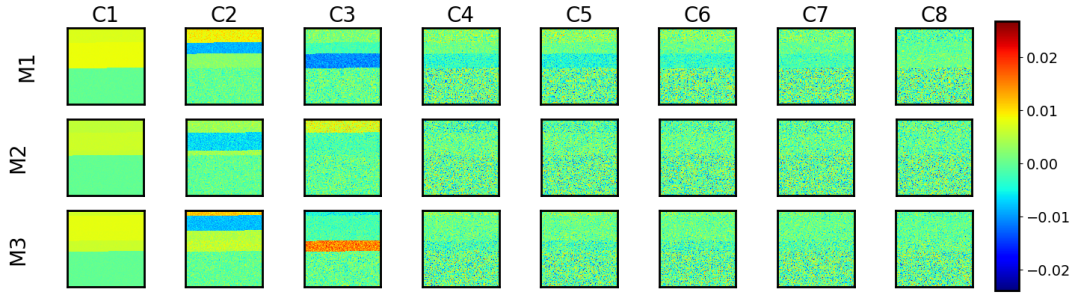


Figure 4.23: The estimated informed multimodal PLS component spatial maps associated with the trajectories in figure 4.22a.

be viewed in appendix C. For the VBM components, we used the Harvard-Oxford cortical and subcortical atlases as distributed in FSL to determine what regions the prominent features corresponded to. Although the Harvard-Oxford atlases are in MNI space and the UK biobank VBM images are in a cohort specific space, the two spaces were still closely aligned. Therefore, we were easily able to determine what region a voxel in a VBM component belonged to by visually checking the Harvard-Oxford atlases. For the TBSS components, we used the John Hopkins University (JHU) DTI white matter atlas, again distributed in FSL. Table 4.2 shows the total

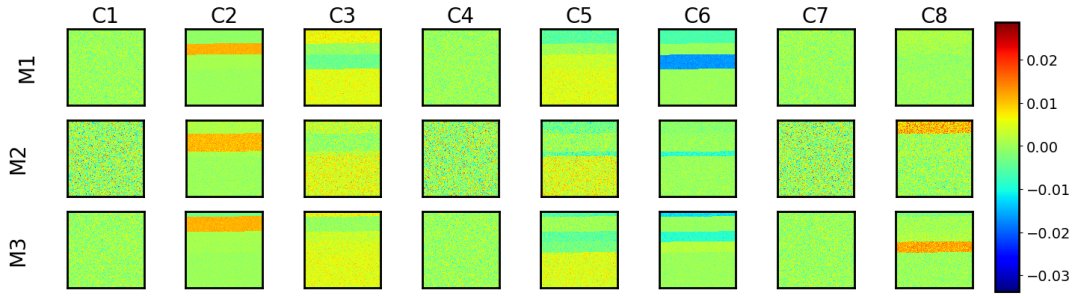


Figure 4.24: The estimated joint ICA component spatial maps associated with the trajectories in figure 4.22b.

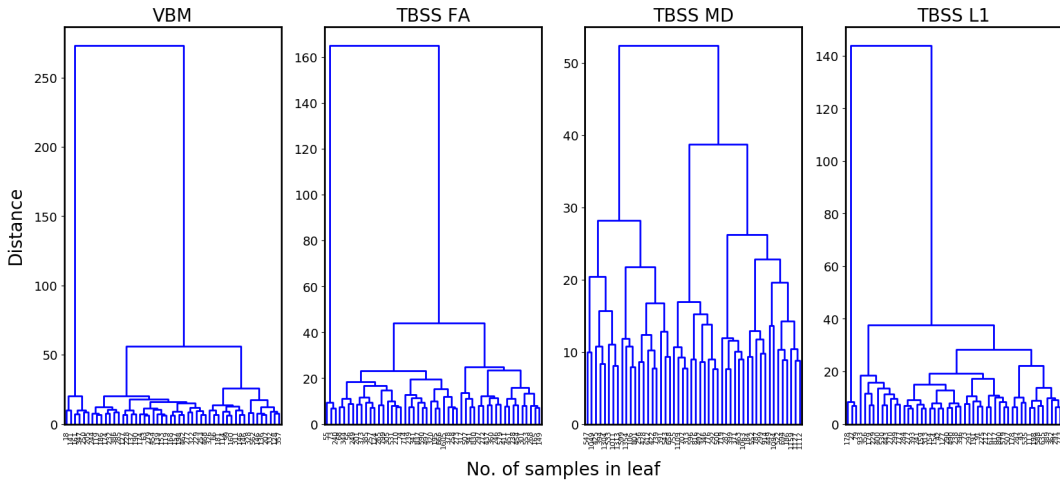


Figure 4.25: The unimodal dendrograms for the four UK biobank modalities that we analysed. Based on these, we cut the dendrograms such that 3, 3, 4 and 4 components were produced for VBM, TBSS FA, TBSS MD and TBSS L1 respectively.

number of voxels that each component contained.

The unimodal fitted trajectories for VBM in figure 4.28 show that component two has a near linear inverse relationship with age, while component three has a more quadratic one. The associated binary and inverse distance maps for component two in figure 4.29 show a clear pattern, highlighting most of the insular cortex. In addition, there are prominent sets of voxels in the opercular cortex, planum polare and the pre- and postcentral gyrus. The spatial maps for component three in figure 4.30 meanwhile, show prominent voxels in the accumbens, putamen, caudate and hippocampus.

The trajectory for the first VBM component shows an increasing trend with age, meaning that the values for voxels that belonged to this component increase with age. Figure 4.31 shows that most of the voxels belonging to this component lie on

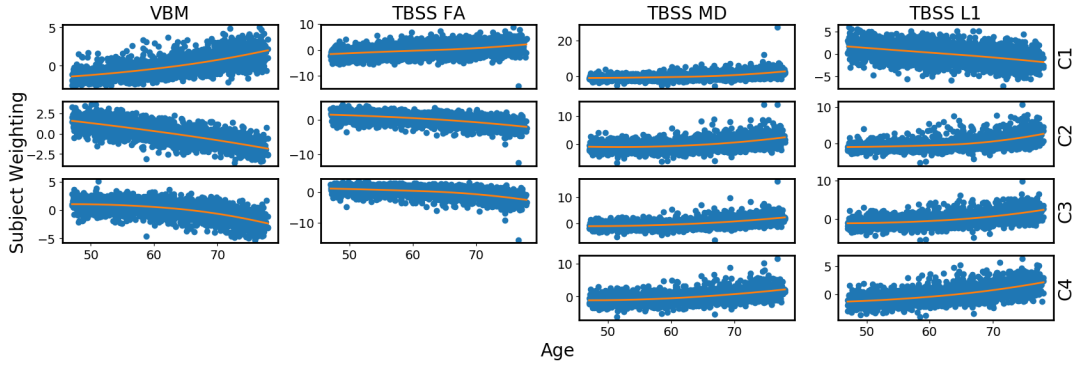


Figure 4.26: The unimodal subject weightings from the training set, shown for each modality and component. A smoothing spline regression model of degree two and one interior knot was fitted to each plot.

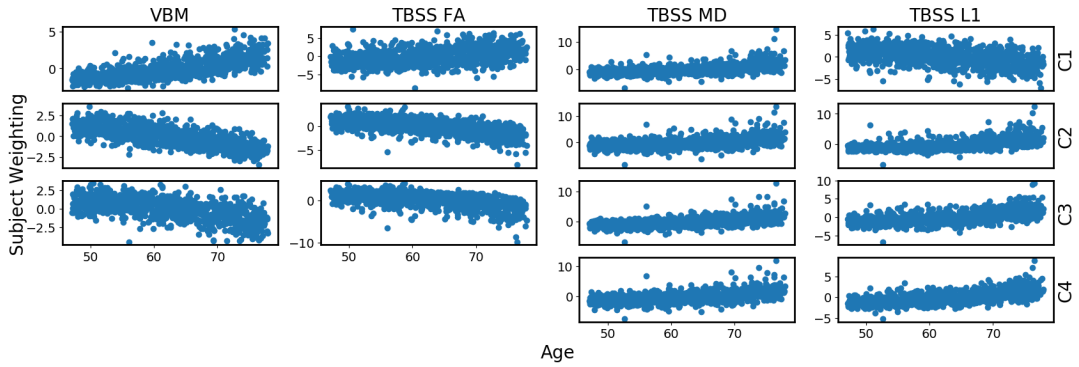


Figure 4.27: The unimodal subject weightings from the test set, shown for each modality and component.

grey matter-white matter or grey matter-CSF boundaries, hence this component could partly be driven by imperfections in the registration. For example, if the alignment in the presence of large ventricles has not gone far enough, then some grey matter structures will be further out than they should be, leading to higher grey matter with age in those regions. However, the coronal slice does show some prominent voxels in the more inferior parts of the insular cortex.

The unimodal fitted trajectories for FA show that it too had one component that showed an increase in voxel values with respect to age. Table 4.2 shows that binary map for this component contained relatively very few voxels and so it is unlikely to be an important component. Furthermore, visually the binary and inverse distance maps did not show any interesting features and so we do not consider it in any detail here, but they can be viewed in appendix C. The fitted trajectories for components two and three are visually not as different to each other when compared to those

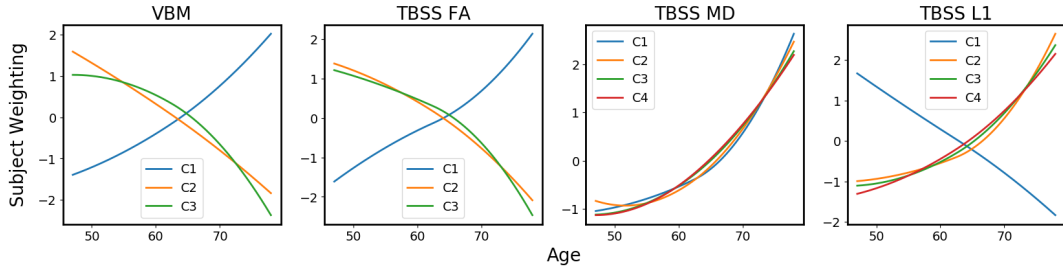


Figure 4.28: The fitted trajectories from figure 4.26 have been plotted on the same axes for each modality here.

	C1	C2	C3	C4
VBM	0.72	-0.73	-0.58	
TBSS FA	0.37	-0.62	-0.56	
TBSS MD	0.53	0.47	0.57	0.53
TBSS L1	-0.46	0.54	0.58	0.65

Table 4.1: The Pearson correlation coefficients corresponding to the plots for the test subject weights in figure 4.27. They were all significant under an unpaired two-tailed t-test after Bonferroni correction.

belonging to the unimodal VBM components. However, it can still be seen that component three has a slower rate of change in the earlier age range and a faster rate of change in the later age range. The spatial map for the second unimodal FA component in figure 4.32 shows prominent voxels mainly in the anterior and superior corona radiata.

In the axial slice of the binary spatial map for the third FA component in figure 4.33, a very prominent bilateral tract in the thalamic region can be seen. This set of voxels are in a region that is unclassified in the JHU DTI atlas but lie on the anterior thalamic radiation tract in the JHU tractography atlas. In the corresponding inverse distance map, this tract stays highly weighted and, in addition, the more medial tract next to it near the fornix, (which is unclassified in both of the JHU atlases) also becomes highly weighted. This shows that even though this tract was not clustered with this component in the binary spatial map, its trajectory was still very similar. The sagittal slice for this component also shows prominent voxels on the genu of the corpus callosum.

The fitted unimodal trajectories for MD show that, unlike FA and VBM, it did not contain any voxels that had a trajectory with a direction that is opposite to expectations. They also suggest that the MD data may have been broken down

	C1	C2	C3	C4	Total
VBM	1122 (10%)	6218 (56%)	3745 (34%)		11085
TBSS FA	368 (2%)	11154 (60%)	7176 (38%)		18698
TBSS MD	6576 (20%)	6909 (21%)	6702 (21%)	12381 (38%)	32568
TBSS L1	284 (2%)	4416 (27%)	8046 (48%)	3890 (23%)	16636

Table 4.2: The number of voxels in the binary maps of the unimodal components. The value in brackets is the number as a percentage of the total number of voxels that were clustered for that modality.

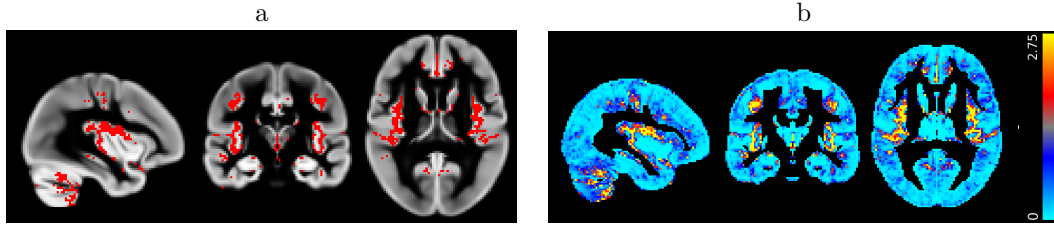


Figure 4.29: Two versions of the spatial map associated with the second unimodal VBM component: a - the binary map, b - the inverse distance map. Note that large distances, and hence small inverse distances, are not important. Only areas in yellow and red should be considered strongly associated with this component. The maximum of the colourmap range for the unimodal inverse distance maps that we display in this chapter was set to approximately the 99th percentile of the voxel values across all components for the modality.

into too many components since components three and four seem to have almost identical trajectories. The spatial map for component one in figure 4.34 shows that it corresponds to the same bilateral tract near the thalamus that was seen in the third FA component.

Table 4.2 shows that the fourth MD component contained many more voxels compared to the other components. The binary spatial map for this component in figure 4.35 shows that it was partly focused on the genu of the corpus callosum. The medial tract near the fornix is also highlighted in the binary spatial map. In the inverse distance map however, the voxels in the corpus callosum are generally weighted more heavily, suggesting that this component is more strongly associated with the corpus callosum as opposed to this medial tract. This component also contains many prominent voxels in the cortical regions which are unclassified in the JHU DTI atlases. Far fewer voxels in these regions passed the thresholding stage for the other TBSS modalities, and this is the main reason for the near double number of voxels that the MD components contained in table 4.2. The spatial maps for

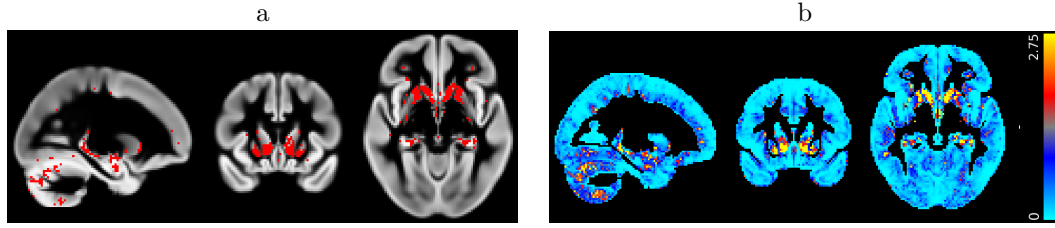


Figure 4.30: Two versions of the spatial map associated with the third unimodal VBM component: a - the binary map, b - the inverse distance map.

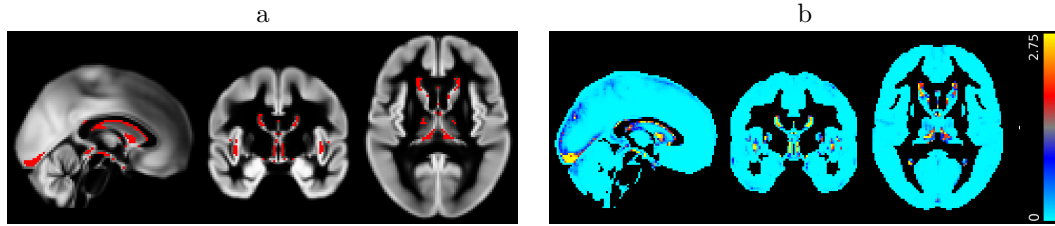


Figure 4.31: Two versions of the spatial map associated with the first unimodal VBM component: a - the binary map, b - the inverse distance map.

components two and three, which we do not show here, were less spatially coherent but contained voxels in similar regions as component four.

The first L1 component is similar to the first component from FA in that the direction of its trajectory is opposite to all the other components and it contains very few voxels. For the same reasons, we do not consider it in any detail. The spatial map for the second component in figure 4.36 again captures the bilateral tract near the thalamus. The third component in figure 4.37 meanwhile, contains prominent voxels on the internal capsule, anterior corona radiata and posterior corona radiata. The fourth component, which we do not show, contained prominent voxels in the same regions as the third component, but less coherently. This suggests that components three and four may in fact be one bigger component.

4.3.2.2 Multimodal

After applying our method to each modality separately, we applied it in a multimodal setting. The resulting dendrogram is shown in figure 4.38. Again, there were no branches that had very low relative leaf sample counts. Based on its structure, we decided to cut the dendrogram so that five components are produced.

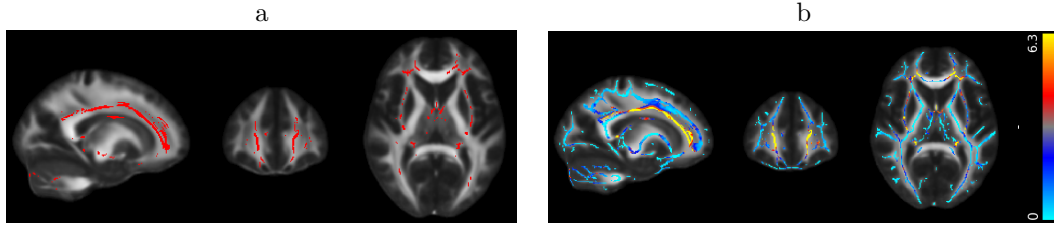


Figure 4.32: Two versions of the spatial map associated with the second unimodal FA component: a - the binary map, b - the inverse distance map.

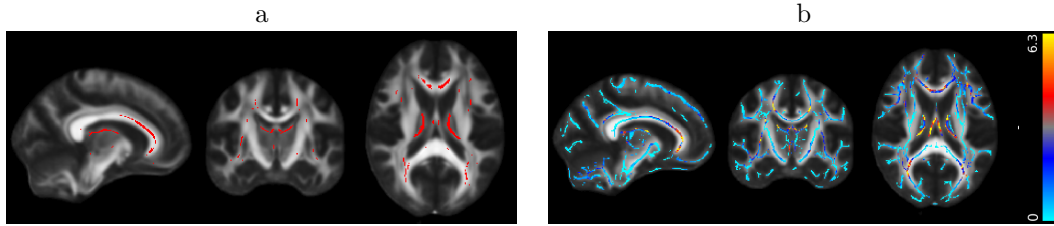


Figure 4.33: Two versions of the spatial map associated with the third unimodal FA component: a - the binary map, b - the inverse distance map.

Figure 4.39a shows the associated subject weights for subjects from the training set and 4.39b shows the subject weights for subjects from the test set. As with the unimodal case, we fitted smoothing spline regression models of degree two and one interior knot to the training subject weights. These are plotted on the same axes in figure 4.40. As explained in section 4.2.3, we multiplied the VBM and TBSS FA data by -1 before the clustering step. Therefore, a component that exhibits an increase in subject weighting with age corresponds to a decrease in voxel value for these two modalities. For TBSS MD and TBSS L1, an increase in subject weighting corresponds to an increase in their voxel values.

The same atlases as in the unimodal case were used to determine which brain region a given voxel belonged to. Table 4.3 shows the number of voxels from each modality that the components contained.

In the unimodal setting, we saw that the trajectory for the first component of VBM, FA and L1 changed in the opposite direction to all the other components. The multimodal fitted trajectory in figure 4.40 for the first component also shows this trend. This component captures voxels that are showing a trend that is opposite to expectation. The spatial maps was exactly the same as in the unimodal case and

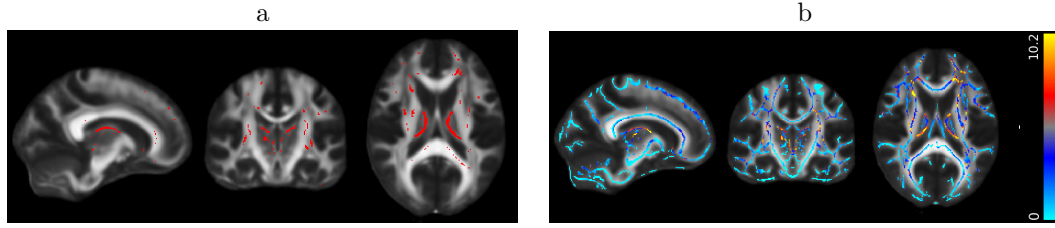


Figure 4.34: Two versions of the spatial map associated with the first unimodal MD component: a - the binary map, b - the inverse distance map.

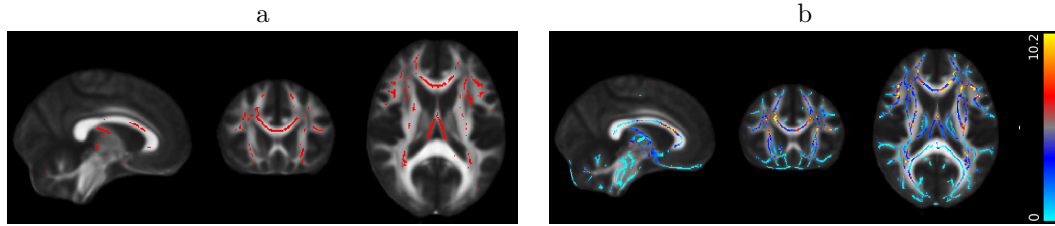


Figure 4.35: Two versions of the spatial map associated with the fourth unimodal MD component: a - the binary map, b - the inverse distance map.

we do not consider them in any detail here.

Out of the four remaining components, the fitted trajectory for the second multimodal component is the one that deviates most from linearity. The spatial maps for this component are shown in figure 4.41. All the TBSS modalities have the bilateral tract near the thalamus as a prominent feature. The inverse distance maps for the MD and L1 modalities also show prominent voxels in the anterior corona radiata. In the binary spatial map for VBM, the putamen, in addition to some voxels in the hippocampus, is a prominent region.

The fourth multimodal component has a fitted trajectory that is most linear compared to the other components. The spatial maps for this component are shown in figure 4.42. The binary and inverse distance maps for VBM clearly highlight voxels in the insular cortex and planum polare. The FA binary map highlight some voxels in the left anterior corona radiata, but this feature becomes more bilateral in the inverse distance map. The maps for both the MD and L1 modalities meanwhile, highlight the medial tract near the fornix.

The third and fifth multimodal components have very similar fitted trajectories. This is somewhat reflected in their binary and inverse distance maps in figures

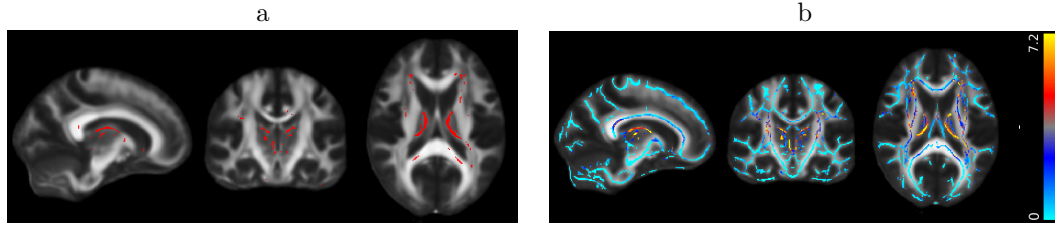


Figure 4.36: Two versions of the spatial map associated with the second unimodal L1 component: a - the binary map, b - the inverse distance map.

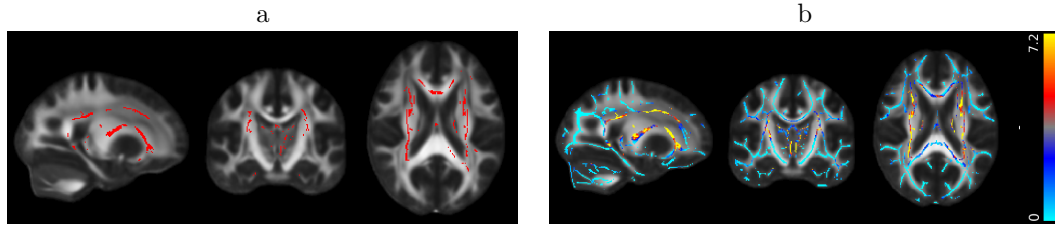


Figure 4.37: Two versions of the spatial map associated with the third unimodal L1 component: a - the binary map, b - the inverse distance map.

4.43 and 4.44. The MD binary and inverse distance maps for component five show highlighted voxels in the body of the corpus callosum, genu of the corpus callosum internal capsule, anterior corona radiata and superior corona radiator. The MD maps for component three also shows these regions, with the exception of the genu of the corpus callosum. The FA maps for both components highlight the superior and anterior corona radiata, but the genu of the corpus callosum is again only highlighted in component five. The L1 maps for both components also show the superior and anterior corona radiata, with the addition of the internal capsule. Although there are a few voxels from the genu of the corpus callosum in the binary map for the third L1 component, this region is much more prominent in component five. The VBM maps on the other hand, are very different across the two components. For component five the prominent regions are the accumbens and the caudate, in addition to some prominent voxels in the hippocampus. Component three meanwhile, contains prominent voxels in the insular cortex.

For comparison, we also decomposed the dataset using joint ICA and informed multimodal PLS. For both of these methods, we chose to decompose the data into five components. We correlated the test subject weights for all the components

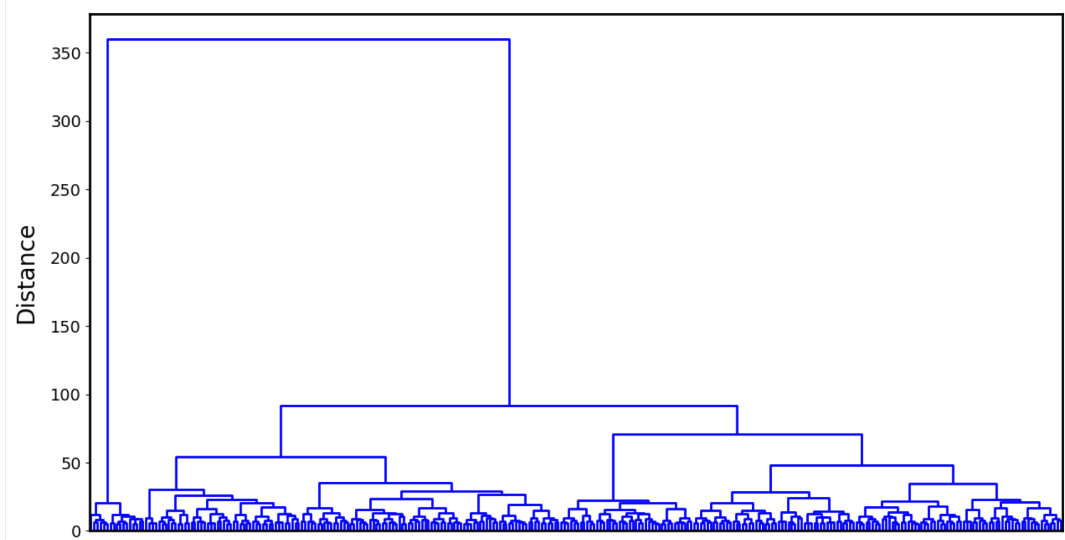


Figure 4.38: The multimodal dendrogram resulting from our method. We cut it such that five components are produced.

	C1	C2	C3	C4	C5	Total
VBM	1122 (10%)	1009 (9%)	2182 (20%)	3659 (33%)	3113 (28%)	11085
TBSS FA	368 (2%)	1136 (6%)	4904 (26%)	1860 (10%)	10430 (56%)	18698
TBSS MD	0 (0%)	6637 (20%)	9094 (28%)	984 (3%)	15853 (49%)	32568
TBSS L1	284 (2%)	3649 (22%)	6645 (40%)	930 (6%)	5128 (31%)	16636

Table 4.3: The number of voxels in the binary maps of the multimodal components. The value in brackets is the number as a percentage of the total number of voxels that were clustered for that modality.

and methods with age. Table 4.4 contains the Pearson correlation coefficients. On average, the correlation values for the components from our method was greater than that from joint ICA and informed multimodal PLS.

The subject weights from informed multimodal PLS are plotted in figure 4.45a for the training subjects and figure 4.45b for the test subjects. Table 4.4 shows that the test subject weights from the first component correlated most strongly with age. The spatial maps associated with this component are shown in figure 4.46. All the different regions that showed up in different components in our method have been grouped together into one component by PLS. For example, in the VBM maps the insular cortex and planum polare, accumbens, putamen and hippocampus are all highly weighted. Similarly, the bilateral tract near the thalamus, the corona radiata, corpus callosum and internal capsule are all highly weighted in the TBSS maps. The spatial maps for the second component in figure 4.47a show similar

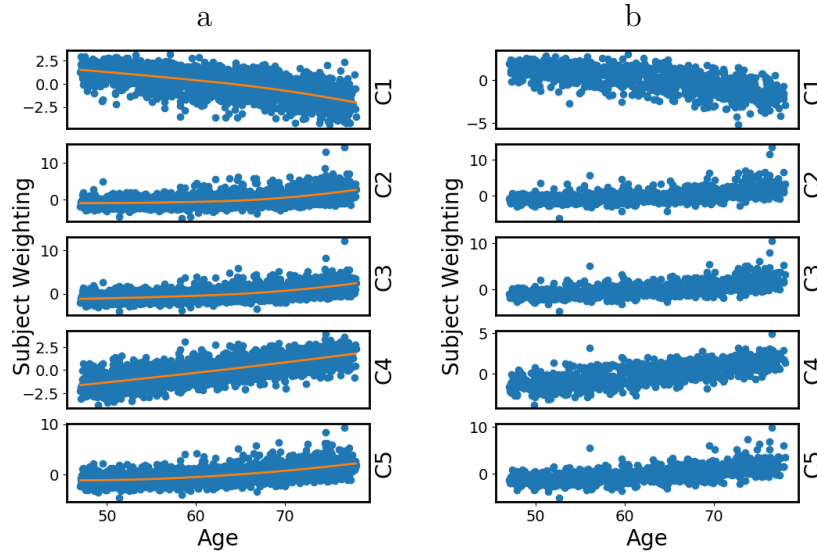


Figure 4.39: The associated subjects weights from our method for subjects from: a - the training set, b - the test set. A smoothing spline regression model was fitted to each component's training subject weights.

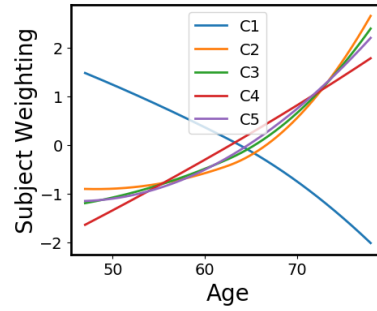


Figure 4.40: The fitted trajectories from figure 4.39a are plotted on the same axes here.

features in the VBM images and is in general weighted more weakly in the TBSS images. The VBM images in the spatial maps for the third component in figure 4.47b meanwhile, show a relatively large positive weighting in the thalamus. The TBSS images for this component show high weightings in the bilateral tract near the thalamus, but relatively the weighing is weak compared to the thalamus in the VBM maps. The fourth and fifth components showed a weak correlation with age, hence we do not consider them in any detail.

Figures 4.48a and 4.48b show the joint ICA components' subject weights for the training and test subjects. Table 4.4 shows that four out of five components were strongly correlated with age. The spatial maps for the first component in figure

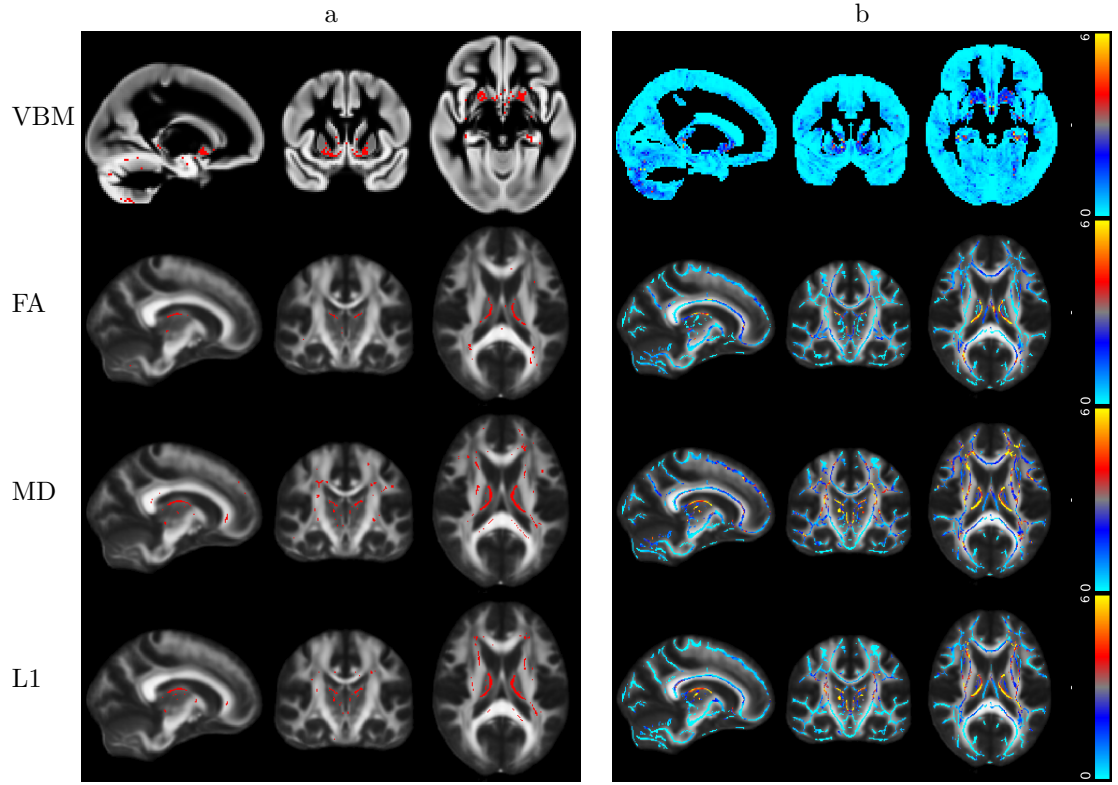


Figure 4.41: The (a) binary and (b) inverse distance maps corresponding to the second multimodal component from our method. The maximum of the colourmap range for the multimodal inverse distance maps that we display in this chapter was set to approximately the 99th percentile of the voxel values across all components and modalities.

4.49a shows that it is effectively an average of all the data, since the voxels are all weighted nearly uniformly. There are still some grey matter regions in the VBM maps that are weighted slightly higher, and the corpus callosum in the FA maps are also weighted slightly higher compared to the rest of the voxels. The MD and L1 maps however, do not highlight any specific regions. The VBM maps for the second component in figure 4.49b are relatively more localised, but various different regions are highlighted in the posterior of the brain and it is difficult to pinpoint specific regions. The MD and L1 maps on the other hand clearly highlight the medial tract near the fornix. The weightings in this tract is also stronger compared to the weightings in the VBM and FA maps, suggesting it is this region in these two modalities that is driving this component. Similarly, as figure 4.50a shows, the MD and L1 modalities seem to be driving the fourth component. The tracts near the thalamus and fornix are both weighted strongly relative to all the voxels in the

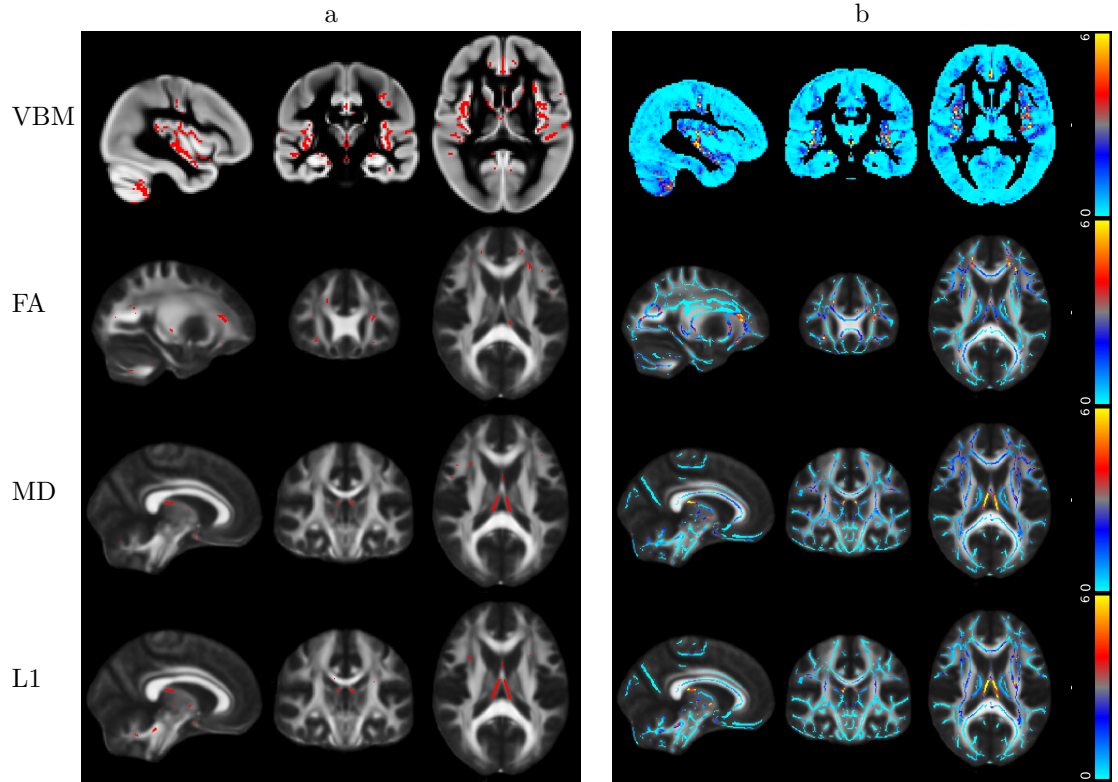


Figure 4.42: The (a) binary and (b) inverse distance maps corresponding to the fourth multimodal component from our method.

	C1	C2	C3	C4	C5
Our method	-0.67	0.53	0.63	0.74	0.61
Joint ICA	-0.44	-0.62	0.17	0.53	0.61
Informed multimodal PLS	-0.63	-0.42	0.34	0.19	-0.06

Table 4.4: The Pearson correlation coefficients corresponding to the plots for the test subject weights in figures 4.39b, 4.45b and 4.48b. With the exception of informed multimodal PLS component 5, they were all significant under an unpaired two-tailed t-test after Bonferroni correction.

VBM and FA modalities. Figure 4.50b shows that the MD and L1 tract near the fornix is again weighted heavily for the fifth component. This time however, small clusters of voxels in the insular cortex and the precuneous cortex-cingulate gyrus boundary are also heavily weighted. Since the third component exhibited a weak correlation with age, we do not consider it in any detail.

4.4 Discussion

In this study, we have proposed a novel approach to estimating multimodal ageing trajectories and associated spatial maps from multimodal neuroimaging data. Our

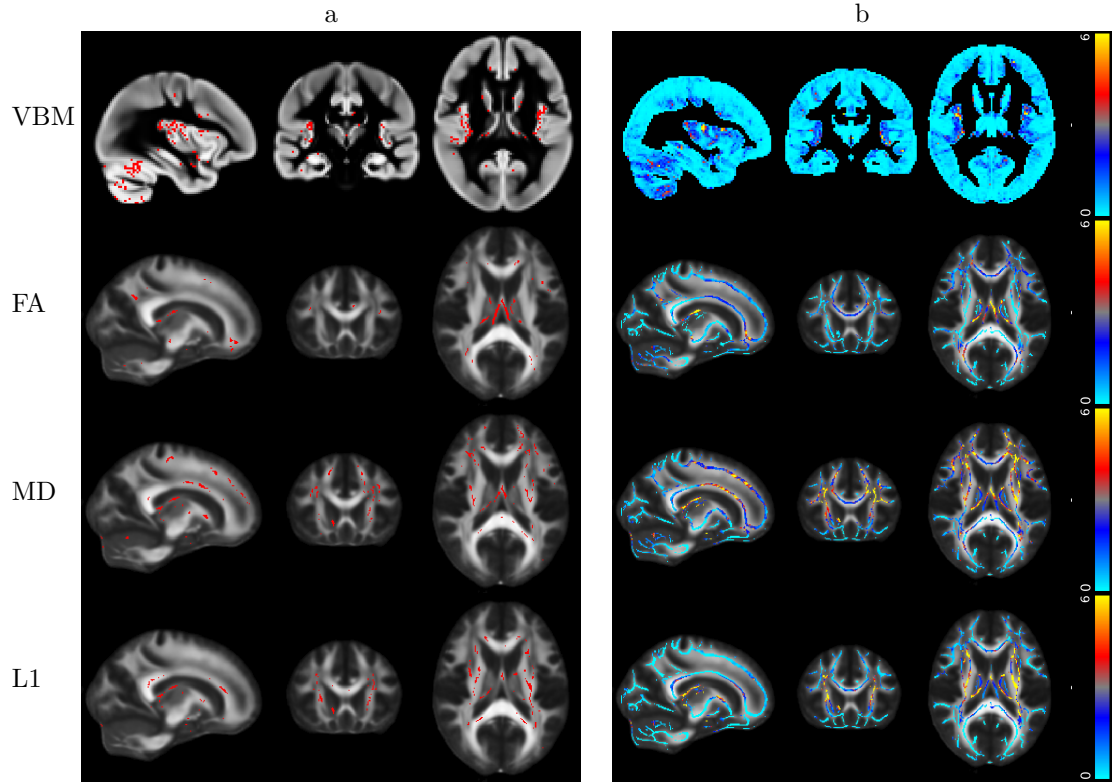


Figure 4.43: The (a) binary and (b) inverse distance maps corresponding to the third multimodal component from our method.

motivation for developing it was to define regions that age differently for use with later studies. It can also be used as an exploratory tool in its own right to investigate ageing and applied to variables other than age. For example, it can be used with a population of Alzheimer’s disease patients to estimate trajectories with respect to mini mental state examination (MMSE) scores [166].

When applied to simulated data, our method was able to accurately estimate trajectories and associated spatial maps. Additionally, we were able to show that through hierarchical clustering it is possible to estimate the correct number of components. Other more automated approaches for determining the number of clusters exist, such as the silhouette method. Their main disadvantage is that, when clustering real datasets, it is common for the optimal number of clusters to still be ambiguous [165]. Hierarchical clustering on the other hand, allows the user to make a more informed decision based on the structure seen in the dendrograms and their desired cluster resolution. It should be noted that automated techniques for determining the number of components for current decomposition methods also exist,

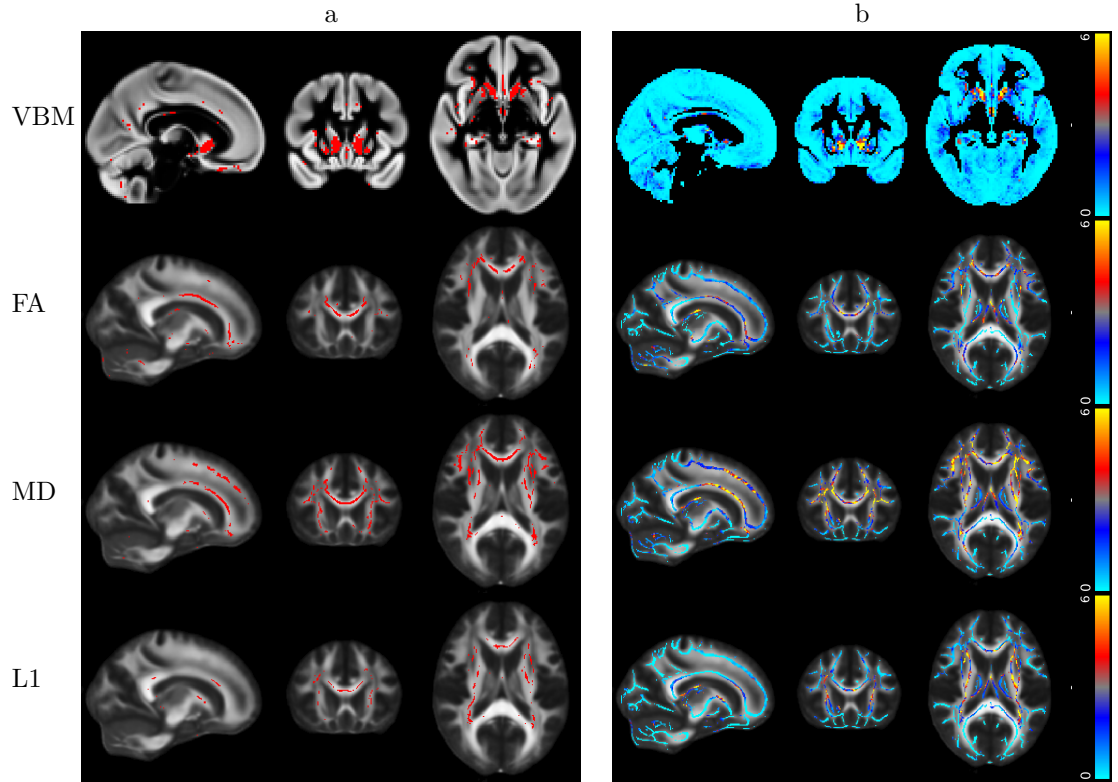


Figure 4.44: The (a) binary and (b) inverse distance maps corresponding to the fifth multimodal component from our method.

such as the minimum description length (MDL) approach that Calhoun et al. used with joint ICA [152]. These automated approaches can still behave poorly in some cases [167], however. In addition, it is reasonable to expect that some datasets will simply not have a natural optimum number of components. Hierarchical clustering is still informative in this scenario, whereas automated methods may not be.

The major downside with hierarchical clustering is its memory requirements. We experienced this when applying our method to the UK biobank data, as we had to set a very strict p threshold in order to reduce the data to a manageable size. This runs the risk of removing trajectories that are weakly varying but still significant. Using k-means as the clustering algorithm avoids this issue, with the obvious drawback being that no dendrogram will be created. Another limitation is related to our choice of ward as the hierarchical clustering algorithm, which merges clusters based on variance. We showed in simulations that it was the most suitable for our objectives. However, if there exists a trajectory that has much higher variance in voxel space compared to the other trajectories, then the sub-clusters associated with it will be

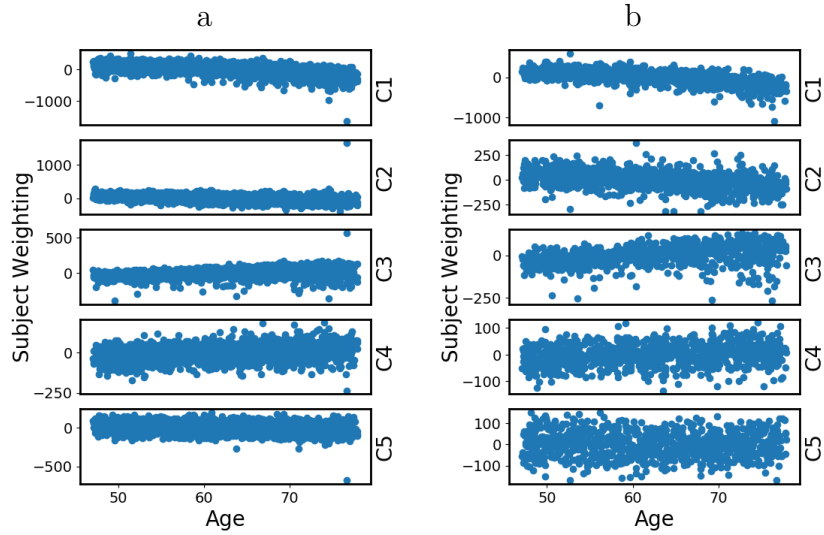


Figure 4.45: The subjects weights from informed multimodal PLS for subjects from: a - the training set, b - the test set.

merged much higher up the hierarchy compared to other trajectories. This may make it seem as if there are more clusters in the data than there actually exist. Another limitation of using hierarchical clustering dendrograms for determining the number of clusters is that the structure of the dendrograms are very heavily dependent on the linkage criteria and the distance metric. We empirically showed on simulated data that dendrograms from the Ward method most correctly reflected the ground truth. However, this may not generalise to all datasets. In addition, although we have argued that using hierarchical clustering dendrograms is potentially more beneficial compared to fully automated methods of determining cluster numbers, these other methods can still be used in a semi-automated manner in the same way that we have used dendrograms. For example, as opposed to automatically using the number of clusters that corresponds to the largest average silhouette, it is possible to look at the average silhouette plot and manually choose the number of clusters that is most appropriate for the data. Comparing the use of dendrograms to other semi-automated methods could be a useful further investigation.

The unimodal VBM and FA dendrograms in the biobank data clearly indicated that three clusters is a suitable number to choose. This is less than we were hoping for but we could not increase the cluster resolution, since a suitable lower level at

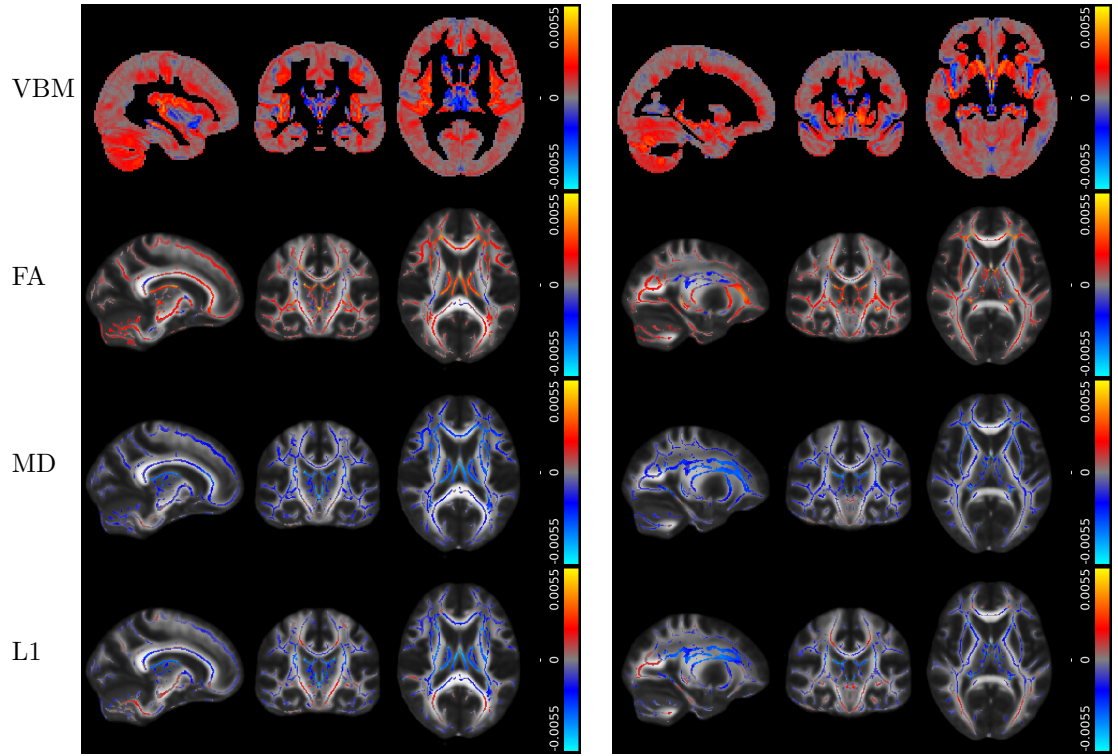


Figure 4.46: The spatial maps corresponding to the first multimodal component from informed multimodal PLS. For this component, we have displayed two sets of slices for each modality.

which to cut the dendrogram did not exist. The unimodal L1 dendrogram was less clear, as it can be argued that both three and four are suitable numbers of clusters. We decided to choose four since we preferred to decompose the data into more clusters. The unimodal MD dendrogram was the least clear, but again four seemed like a reasonable number and was consistent with the other three modalities, hence we chose four. As a whole, the unimodal dendrograms showcased the ability to make informed decisions on the numbers of clusters in our method.

In the unimodal decomposition of the VBM data, our method highlighted the anterior putamen, hippocampus, accumbens, caudate, insular cortex, opercular cortex, planum polare and precentral gyrus regions. The grey matter volumes of all of these structures have previously been shown to exhibit an inverse relationship with age [168, 169, 170, 171]. However, our method went beyond the methods used in these studies since it automatically grouped regions that showed similar trajectories together. The fact that cortical and subcortical regions were grouped together into

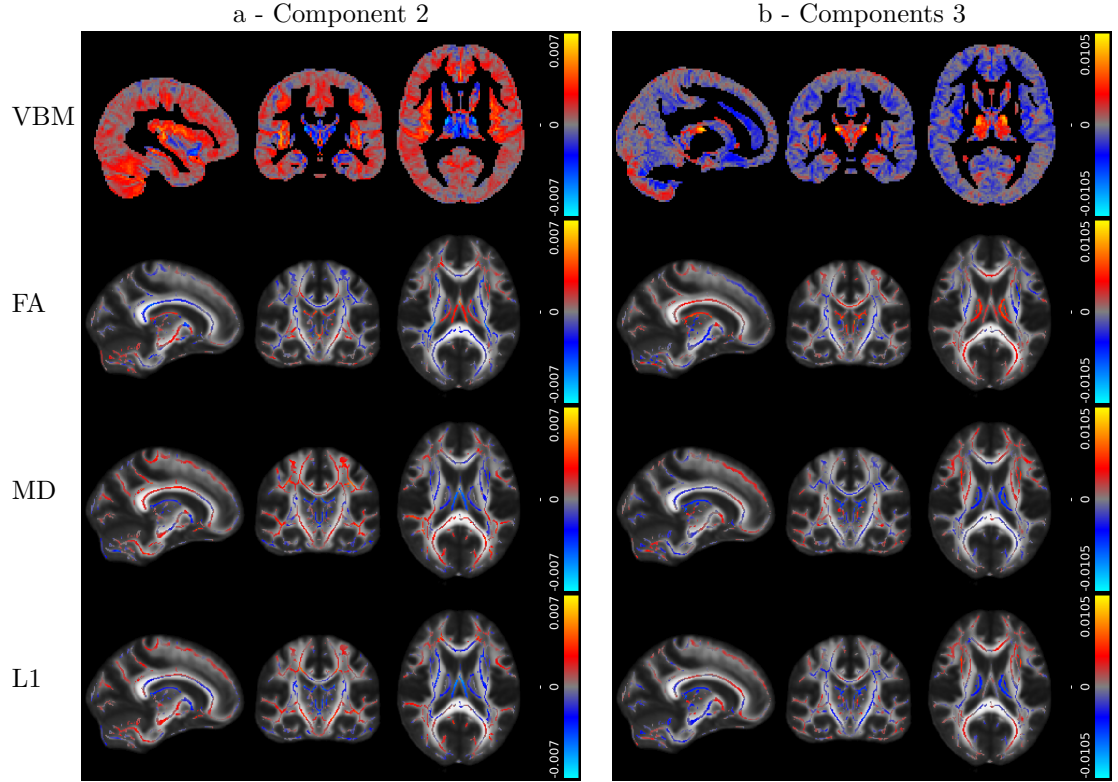


Figure 4.47: The spatial maps corresponding to (a) the second and (b) the third multimodal component from informed multimodal PLS.

their own respective components suggests that this grouping may be a reasonable one. The shapes of the trajectories for the two components are also generally in agreement with what has previously been reported. Walhovd et al. showed that the cortical grey matter volume exhibited a linear decrease with age, while out of the subcortical regions only the hippocampus exhibited the same quadratic trajectory that we found [168]. However, a later study by Goodro et al. that analysed nearly three times the number of subjects compared to Walhovd et al., found that in addition to the hippocampus, the accumbens also exhibited the same quadratic trajectory [172]. The caudate and the putamen on the other hand, did not. One reason for this may be that they segmented the whole of the putamen and caudate, whereas our method only highlighted the anterior parts of the putamen and the inferior parts of the caudate.

In the first multimodal and unimodal VBM components, some inferior insular cortex voxels showed an increasing trend with age. Although this is in disagreement with other reported studies, these voxels were vastly outnumbered by the number of

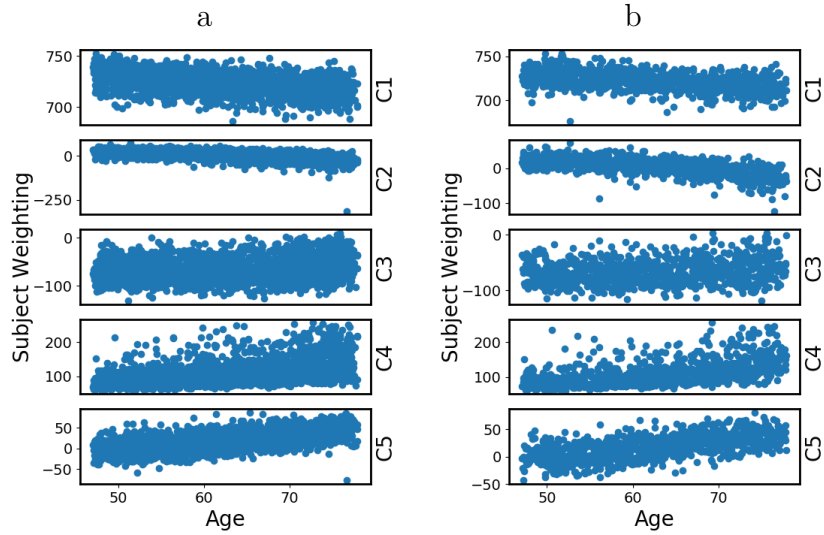


Figure 4.48: The subjects weights from joint ICA for subjects from: a - the training set, b - the test set.

insular cortex voxels that showed a decreasing trend with age in other components. This means that a segmentation of the whole insular cortex will likely still show an overall decreasing trend. This points to a possible advantage of our method over segmentation approaches, since it can automatically separate out voxels from the same region that are showing different trends.

Even though the unimodal trajectories for the TBSS modalities were not as distinctive when compared to the VBM trajectories, our method was still able to pick out some components that were spatially coherent. All three contained one component that mainly highlighted the bilateral tracks near the thalamus, which was part of the anterior thalamic radiation. Other prominent regions for all three TBSS modalities were the corona radiata and the corpus callosum. Both MD and FA have previously been shown to increase and decrease with age, respectively, in the corona radiata, corpus callosum and anterior thalamic radiation [173, 174, 20]. Moreover, we found the components associated with the anterior thalamic radiation to be the most non-linear. Cox et al. similarly showed a quadratic relationship between MD in the anterior thalamic radiation and age [20]. They found the relationship between FA in this region and age to be linear however, which could again be due to the fact that they considered the whole of the region whereas our method only highlighted

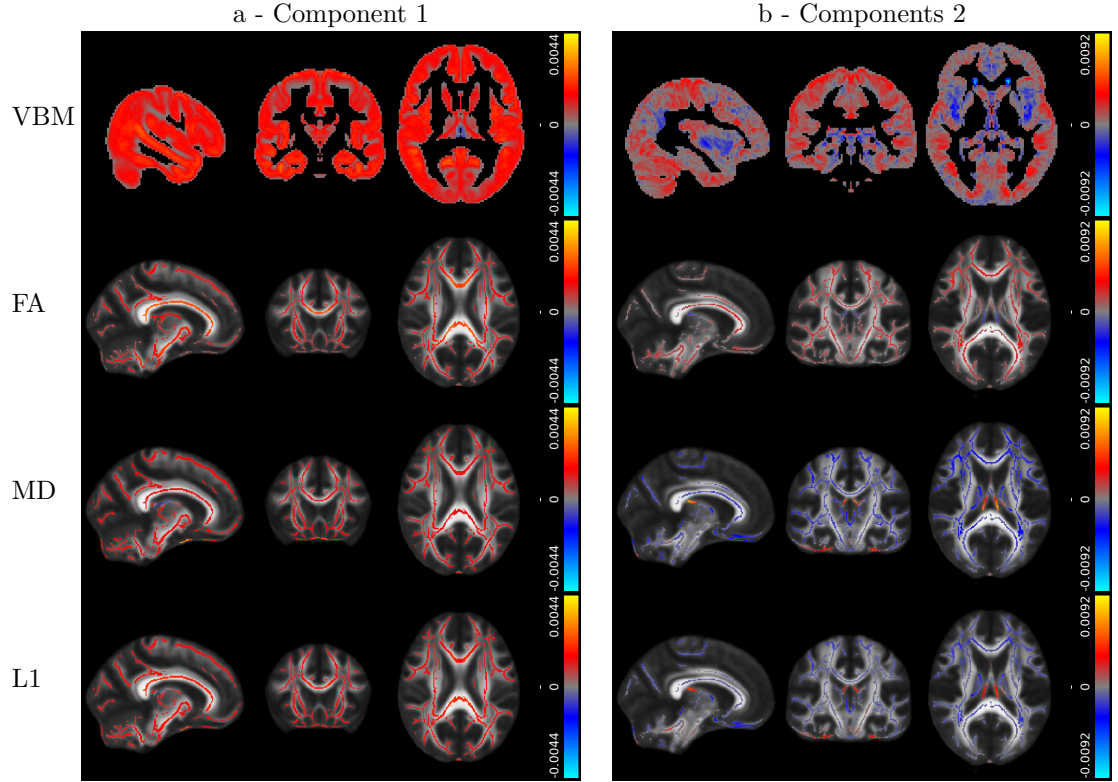


Figure 4.49: The spatial maps corresponding to (a) the first and (b) the second multimodal component from joint ICA.

a small part of it.

In the multimodal setting, we chose to decompose the data into five components based on the multimodal dendrogram. Again, this number was not clear-cut and it could be argued that three or six were also suitable numbers of clusters. The spatial maps for the second multimodal component, which had the most non-linear trajectory, mainly highlighted voxels in the anterior thalamic radiation in all of the TBSS modalities and grouped them with the putamen from VBM. This result indicates that these regions undergo age related changes at the same time, and could point to some common biological process affecting them. When considering the corresponding inverse distance map for this component however, the putamen is not as heavily weighted as the anterior thalamic radiation voxels in the TBSS maps. This suggests that this component may largely be driven by the TBSS modalities.

The fourth multimodal component, which was the most linear, mainly highlighted VBM voxels in the insular cortex and FA voxels in the anterior corona radiata. The corona radiata has previously been implicated in the structural con-

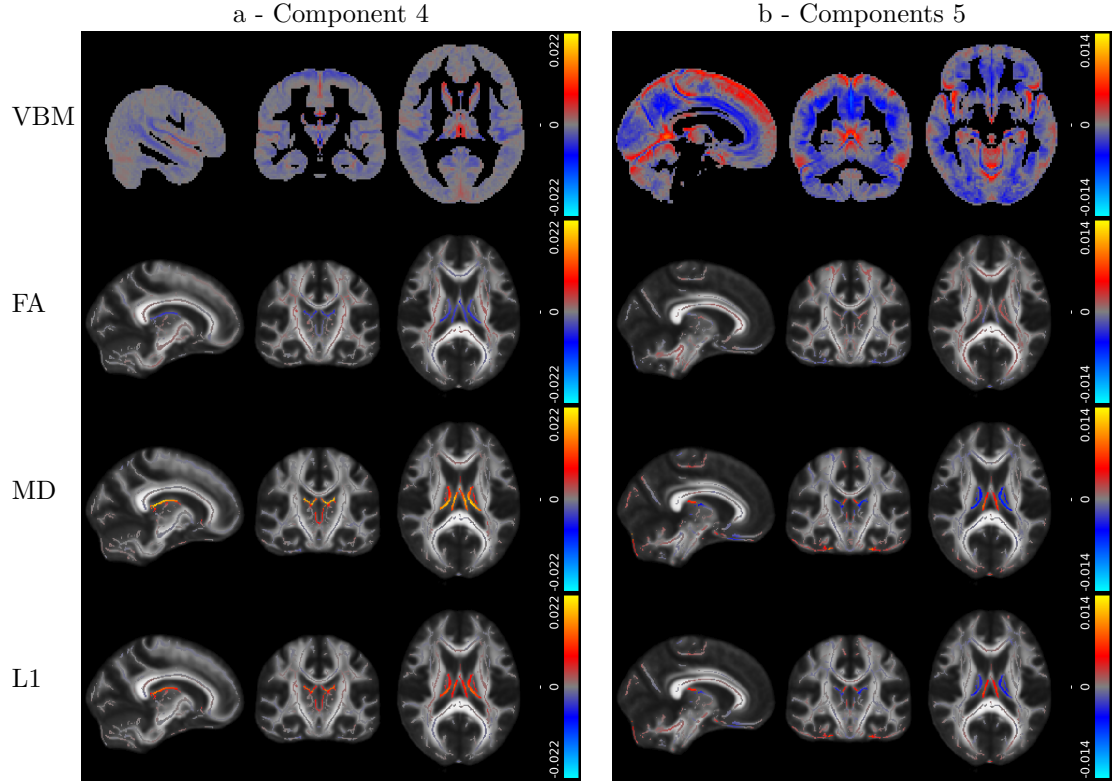


Figure 4.50: The spatial maps corresponding to (a) the fourth and (b) the fifth multimodal component from joint ICA.

nectivity of the insular cortex [175]. This means that there is a chance component four could be a biologically driven result and further investigation is warranted. This component also highlighted the medial tract near the fornix in the MD and L1 modalities. Interpreting this however, is made difficult since this tract is not defined in common atlases.

Our method is based on concatenation, meaning that it is still exposed to limitations that concatenation based methods suffer from. The third and fifth multimodal components reflect one of these limitations. In the unimodal decomposition of VBM, we saw that the trajectories associated with cortical and subcortical regions were very distinct. In the multimodal decomposition on the other hand, the third component contained cortical VBM regions and the fifth component contained subcortical VBM regions, yet their trajectories were very similar. Table 4.3 shows that this is due to the TBSS modalities, especially MD, contained relatively many more voxels in the binary spatial maps for these two components. Therefore, the subject weightings were overpowered by the TBSS modalities and, due to the TBSS modalities

having similar spatial maps, the trajectories ended up looking much the same.

Many of the same regions that our method highlighted, were also highlighted by informed multimodal PLS. However, the regions that our method assigned distinct trajectories to were instead grouped into one component by informed multimodal PLS. This effect was also seen in the simulated data. This is as a result of informed multimodal PLS only caring about maximising the covariance of the subject weights with the variable of interest, and it can potentially do this by combining distinct trajectories. These results showcase a major advantage that our method has over informed multimodal PLS when used as an exploratory tool. Informed multimodal PLS did pick out the thalamus in its third component however, a region that our method did not highlight. This was a consequence of the very strict p threshold that we set in our method, meaning that voxels in the thalamus were removed prior to the clustering. This issue can be solved by using a less strict p threshold, either through more computational resources or by replacing hierarchical clustering with an algorithm such as k-means. As with both our method and informed multimodal PLS, joint ICA also highlighted voxels on the anterior thalamic radiation tract and the tract near the fornix. It was difficult to see localised features in addition to this in the joint ICA spatial maps. It should be noted that typically a higher dimensional decomposition is carried out with ICA and it is possible that more localised regions may become visible in this case. Nonetheless, in simulated data joint ICA was not able to reproduce the correct trajectories. In addition, in the UK biobank, only four out of the five joint ICA components resulted in strong correlations with age. With informed multimodal PLS, it was only three. With our method, all the components had strong correlations with age, and these correlations were on average higher than the correlations for the strong age related joint ICA and informed multimodal PLS components.

The current implementation of our method requires for modalities where a decrease in a voxel's value represents pathology to be multiplied by -1. This is feasible in the context of ageing since the modalities that we have investigated here typically

show either an increasing or decreasing trend. However, if our method is applied to more complicated phenotypes or diseases, this implementation is limited in that it may not be clear whether to sign flip a given modality. In these cases, a different option could be to remove sign information completely, and have all voxels exhibiting the same direction of change.

Many of the trajectories that we found in this study, especially the TBSS trajectories, were not that distinct from each other. This could potentially be due to the UK biobank dataset having a small age range compared to typical ageing studies. Therefore, a focus of future work should be to apply our method to a healthy dataset with a wider age range. Applying it to a disease dataset would also be an interesting application and can serve as a further method of validation. In addition, in this study we only focused on trajectories that result after demeaning and variance normalising voxel time-courses. Probing other normalisation options may provide further insight into the data. For example, one option could be to divide each voxel's binned time-course by the value of the voxel's first bin. In this case, our method will cluster together voxels whose trajectories are showing the same fractional change with respect to the first time point.

4.5 Conclusion

We have presented a novel approach for decomposing neuroimaging data into distinct trajectories and associated spatial maps. Although our focus in this study was ageing, our method can be used to obtain trajectories with respect to any variable of interest. Our method also aids interpretation by assigning a brain region to only one trajectory. This is in contrast to existing decomposition methods, where each brain region is modelled as a combination of multiple trajectories. In simulations, we showed that unlike two existing decomposition methods, joint ICA and informed multimodal PLS, our method was able to accurately reproduce simulated multimodal trajectories and associated spatial maps. Through simulations, we also showed that we can use hierarchical clustering in our method to accurately estimate

the number of trajectories that exist in the data. When we applied it to the UK biobank dataset, our method was able to estimate trajectories and associated regions that were in agreement with previously published results. Moreover, it was able to produce interesting and interpretable components. In a unimodal setting, the most interesting of these were a subcortical VBM component that had a quadratic trajectory and a cortical VBM component that had a linear trajectory. In a multimodal setting meanwhile, it found a component that showed that VBM voxels in the insular cortex undergo the same rate of ageing as some TBSS FA voxels in the anterior corona radiata. However, further validation on a dataset with a wider age range is needed.

Chapter 5

Final discussion and future work

5.1 Summary of main thesis contributions

In this thesis, we have presented three novel methods, all with the aim of aiding neuroimaging biomarker development for neurodegenerative diseases.

In chapter 2, we outlined a novel algorithm for use with analysing rfMRI data in a discriminatory framework. The main advantage over current approaches is that the whole voxel by time data can be fed in without any prior parcellation or decomposition. This makes it ideal for use in discriminating between groups since no prior decision has to be made regarding the regions or brain networks to feed into the classification. In addition, the algorithm outputs two spatial maps that show regions of altered functional connectivity. Therefore, in addition to a classification tool, it can be used as a method for exploring differences between groups.

In chapter 3, we presented, to our knowledge, the first application of an EIV-based model to brain ageing. It explicitly models neuroimaging data as a function of brain age. As a consequence, compared to current approaches, it has the advantages of being able to model inverted U-shaped trajectories and has a framework for modelling interactions with disease. By calculating the difference between the chronological age and the estimated brain age of a subject, it can potentially be used to identify subjects at risk of age-related diseases and conditions. Moreover, it provides easy to visualise trajectories of neuroimaging data with respect to brain

age.

Finally, in chapter 4, we outlined a novel data-driven multimodal decomposition method that produces a set of components from neuroimaging data, with each component having its own associated subject weights and spatial map. The difference compared to most common current approaches is that the decomposition is driven by a variable of interest. This makes it suited for exploratory ageing studies, since age can be set as the variable of interest in order to obtain age-related components and trajectories. This is in contrast to current approaches, where the subject weights have to typically be correlated with age post-hoc. In addition, our method aids interpretability by providing a single trajectory to describe brain regions as opposed to a complicated mixture of many components.

5.2 Future work

5.2.1 Modelling errors associated with variables other than age

We have already suggested an approach for extending the brain age model in chapter 3 to investigate interactions between brain ageing and disease. A further extension would be to say that there is error associated with the measure of disease severity as well, i.e. that the measure of disease severity is a noisy measure of the ‘biological’ disease severity. If d_n is a measure of disease severity for subject n , then:

$$d_n = d'_n + \rho_n \quad (5.1)$$

where ρ_n is the error associated with d_n and d'_n is the biological disease severity. By assuming that $d_n \sim \mathcal{N}(0, \omega^2)$, equation 3.4 changes to:

$$J = \sum_{n=1}^N \left(\frac{(a_n - a'_n)^2}{2\nu^2} + \frac{(d_n - d'_n)^2}{2\omega^2} + \sum_{m=1}^M \frac{(y_{nm} - f_m(a'_n, d'_n, \boldsymbol{\theta}_m))^2}{2\sigma_m^2} \right) \quad (5.2)$$

The limitations of measures of disease severity, such as MMSE, are widely acknowledged [176]. This suggests that there is scope for this type of model to potentially result in more accurate trajectories and interactions.

Furthermore, it may aid in obtaining more accurate brain age estimates. Consider a function that has a planar surface when plotted against a'_n and d'_n . Now consider the data point of a subject that has been shifted by δ_n and ρ_n in the direction of the age and disease axes, respectively. When trying to recover the subject's original position, if shifts along both axes are allowed, then there is freedom to shift the subject's data point by the correct amount in both directions and get close to the function's surface. However, if only a shift along the age axis is allowed, then the subject's data point has to potentially be shifted by more than δ_n to get to the same distance away from the function's surface. This is of course just a motivating example and our formulation is not restricted to just linear or planar cases. In reality the shifts would be more complicated since the estimation of the surface itself would interact with the shifting.

5.2.2 Regional and multimodal brain ages

One issue with a whole-brain brain age is that its clinical applications are limited. For example, mean brain age deltas (chronological age minus brain age) of -2.6 to -5.5 , -4.6 to -5.9 and -5.5 to -6.2 years have been reported for schizophrenia, diabetes and MCI converters, respectively [93]. Consequently, if a subject has a brain age delta of -5.5 years, it will be unclear which of these conditions they may be susceptible to.

Different neurodegenerative diseases affect different parts of the brain. For example, one study comparing frontotemporal dementia patients (FTD) to AD patients showed that AD patients had more atrophy in the posterior parietal and occipital cortex, whereas FTLD patients had more atrophy in the anterior cingulate, frontal insula, subcallosal gyrus, and striatum [177]. Therefore, one possible solution is to consider regional brain ages instead. In this way, patterns of regional brain age

deltas can potentially be used to classify patients.

As we explained in the introduction chapter, different MRI modalities are sensitive to different properties of the brain. Hence, they may be sensitive to different aspects of brain ageing as well. For this reason, as well as considering region-specific brain ages, it may potentially be beneficial to consider modality-specific brain ages.

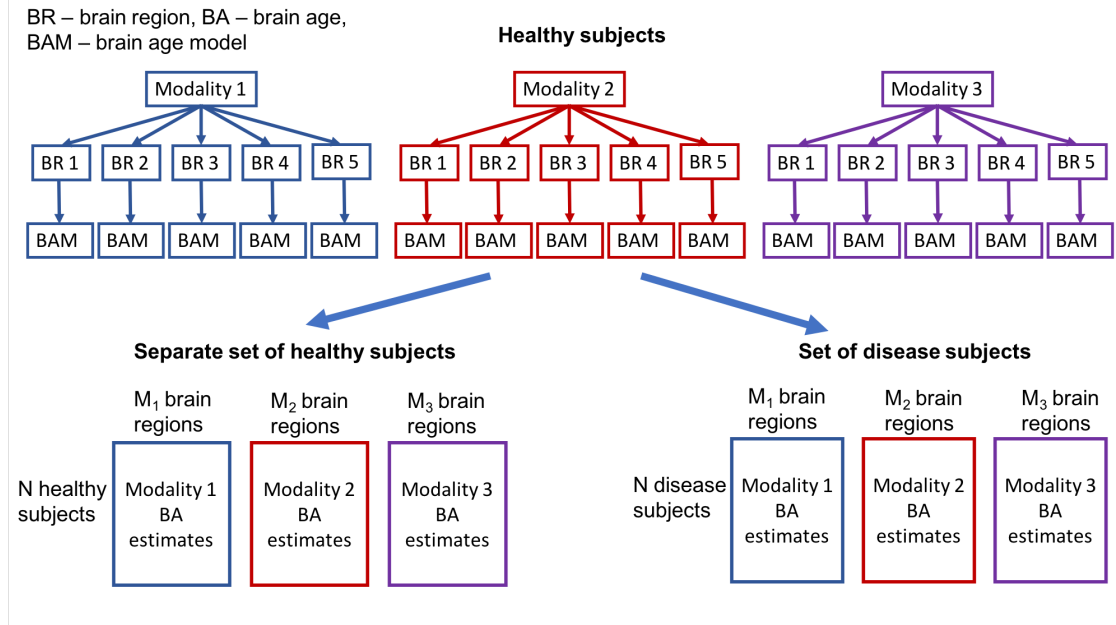


Figure 5.1: A diagram of a proposed study investigating the usefulness of region- and modality-specific brain ages.

A possible option for investigating the usefulness of region- and modality-specific brain ages is shown in figure 5.1. In this study, using data belonging to healthy subjects, separate brain age models would first be estimated for a set of regions specific to each modality. Note that in the diagram we have chosen five brain regions for each modality as an example, but in reality each modality can have different numbers of regions. Afterwards, the brain age models would be applied to a separate set of healthy subjects and a set of patients all having the same disease, e.g. AD. This would create a multimodal feature set (in the example the total number of features would be 15). Finally, a multimodal classification algorithm could be used to see whether this feature set adds any discriminatory power when discriminating between healthy and diseased subjects.

The work that we have presented in this thesis provides the basis for implement-

ing this type of study. The multimodal decomposition method in chapter 4 could be used for defining brain regions, while the model in chapter 3 could be used for building the brain age models. The algorithm in chapter 2 also provides a way to include rfMRI data, since it can be implemented to discriminate between young and old subjects. The region definition and brain age modelling can then be carried out with respect to the subject specific maps associated with one of the $\mathbf{v}$ or $\mathbf{w}$ vectors. For example, subject specific maps associated with $\mathbf{v}$ can be calculated using $\mathbf{x}_s = \mathbf{X}_s \mathbf{X}_s^T \mathbf{v}$, where $\mathbf{x}_s$ and $\mathbf{X}_s$ are the spatial map and the voxel by time rfMRI data for subject s , respectively.

5.2.3 Investigating the clinical relevance of biomarkers

Studies investigating machine learning-based biomarkers for neurodegeneration typically only consider binary class classification problems, i.e. discriminating between a disease and a healthy cohort. To increase the clinical relevance of such methods, their performance in a multiclass problem should also be evaluated. The ability to, for example, predict whether a patient with mild cognitive impairment will likely go on to develop AD, FTD or dementia with Lewy bodies will be a powerful attribute for a biomarker to have. Many different multiclass classification algorithms exist. Exploring strategies for adapting the IDLDA algorithm and the proposed study in section 5.2.2 into a multiclass framework based on one of these multiclass algorithms could comprise some of the future work.

There are sizeable financial costs associated with obtaining neuroimaging data. Due to this, another important potential study could be to investigate the extent of extra diagnostic power that methods employing neuroimaging data provide compared to relatively cheap methods, e.g. cognitive and behavioural assessments. Somewhat related to this is that in the clinic it is the subjects with an atypical presentation of a disease that are the most difficult to diagnose [178]. Investigating the performance on these atypical subjects would provide a further important validation.

Appendix A

Model derivatives

The derivatives of equation 3.4 with respect to a'_n and θ_m are:

$$\frac{\partial J}{\partial a'_n} = \frac{a'_n - a_n}{\nu^2} + \sum_{m=1}^M \frac{1}{\sigma_m^2} \left(f_m(a'_n, \theta_m) \frac{\partial f_m(a'_n, \theta_m)}{\partial a'_n} - y_{nm} \frac{\partial f_m(a'_n, \theta_m)}{\partial a'_n} \right) \quad (\text{A.1})$$

$$\frac{\partial J}{\partial \theta_m} = \sum_{n=1}^N \frac{1}{\sigma_m^2} \left(f_m(a'_n, \theta_m) \frac{\partial f_m(a'_n, \theta_m)}{\partial \theta_m} - y_{nm} \frac{\partial f_m(a'_n, \theta_m)}{\partial \theta_m} \right) \quad (\text{A.2})$$

If f is a polynomial:

$$\frac{\partial f}{\partial a'_n} = \sum_{p=1}^P p h_p a_n'^{p-1} \quad (\text{A.3})$$

$$\frac{\partial f}{\partial h_p} = a_n'^p \quad (\text{A.4})$$

If f is a basis spline:

$$\frac{\partial f}{\partial a'_n} = \sum_{i=0}^{T+q-1} \frac{(\beta_{i+1} - \beta_i) q B_{i+1,q-1}(a'_n)}{t_{i+q+1} - t_{i+1}} \quad (\text{A.5})$$

$$\frac{\partial f}{\partial \beta_i} = B_{i,q}(a'_n) \quad (\text{A.6})$$

If f is a sigmoid/logistic function:

$$\frac{\partial f}{\partial a'_n} = \frac{h k e^{-k(a'_n - a_0)}}{(e^{-k(a'_n - a_0)})^2} \quad (\text{A.7})$$

$$\frac{\partial f}{\partial h} = \frac{1}{1 + e^{-k(a'_n - a_0)}} \quad (\text{A.8})$$

$$\frac{\partial f}{\partial k} = \frac{-he^{-k(a'_n - a_0)}(-a'_n + a_0)}{(1 + e^{-k(a'_n - a_0)})^2} \quad (\text{A.9})$$

$$\frac{\partial f}{\partial a_0} = \frac{-hke^{-k(a'_n - a_0)}}{(1 + e^{-k(a'_n - a_0)})^2} \quad (\text{A.10})$$

$$\frac{\partial f}{\partial c} = 1 \quad (\text{A.11})$$

Appendix B

The chosen UK biobank IDPs

T1 SIENAX grey normalised volume; T1 SIENAX grey unnormalised volume; T1 SIENAX peripheral grey normalised volume; T1 SIENAX peripheral grey unnormalised volume; T1 SIENAX brain-normalised volume; T1 SIENAX brain-unnormalised volume; dMRI TBSS ISOVF Fornix; dMRI ProbtrackX L1 atr r; dMRI ProbtrackX MD atr r; dMRI ProbtrackX MD atr l; dMRI TBSS MD Fornix; dMRI TBSS L1 Fornix; dMRI ProbtrackX L1 atr l; dMRI TBSS L3 Fornix; dMRI TBSS L2 Fornix; T2 FLAIR BIANCA WMH volume; dMRI TBSS L2 Fornix cres+Stria terminalis R; dMRI TBSS FA Fornix; dMRI ProbtrackX L3 atr l; T1 SIENAX CSF normalised volume; T1 SIENAX CSF unnormalised volume; dMRI ProbtrackX L2 atr r; dMRI ProbtrackX L2 atr l; dMRI ProbtrackX L3 atr r; dMRI TBSS FA Fornix cres+Stria terminalis R; dMRI TBSS L2 Fornix cres+Stria terminalis L; dMRI TBSS L3 Fornix cres+Stria terminalis R; dMRI ProbtrackX L1 str r; T1 FIRST right thalamus volume; dMRI TBSS FA Fornix cres+Stria terminalis L; dMRI TBSS MD Superior fronto-occipital fasciculus L; dMRI ProbtrackX L1 str l; dMRI TBSS MD Superior fronto-occipital fasciculus R; dMRI ProbtrackX MD str r; dMRI TBSS ICVF Superior fronto-occipital fasciculus L; T1 FIRST left thalamus volume; dMRI TBSS L3 Anterior corona radiata L; dMRI TBSS L3 Superior fronto-occipital fasciculus L; dMRI TBSS L2 External capsule R; dMRI ProbtrackX MD str l; dMRI TBSS ICVF Superior fronto-occipital fasciculus R; dMRI TBSS L3 Anterior limb of internal capsule R; dMRI TBSS MD Fornix cres+Stria terminalis R; dMRI TBSS L3 Superior fronto-occipital fasciculus R; dMRI TBSS L3 Anterior limb of internal capsule L; dMRI TBSS L2 Anterior corona radiata L; dMRI TBSS L2 External capsule L; dMRI TBSS L3 Anterior corona radiata R; dMRI TBSS FA Anterior corona radiata L; T1 FAST ROIs L frontal pole; dMRI TBSS L3 Fornix cres+Stria terminalis L; dMRI TBSS MD Anterior corona radiata L; dMRI ProbtrackX ICVF atr l; dMRI TBSS MD Anterior limb of internal capsule R; dMRI TBSS MD Anterior limb of internal capsule L; dMRI ProbtrackX ICVF atr r; dMRI TBSS L3 Superior corona radiata L; dMRI TBSS L3

Superior corona radiata R; T1 FIRST right accumbens volume; dMRI TBSS L1 Posterior corona radiata R; dMRI ProbtrackX L2 ptr l; T1 FAST ROIs R frontal pole; T1 FAST ROIs L ventral striatum; dMRI ProbtrackX L2 ptr r; dMRI ProbtrackX MD ptr r; dMRI TBSS L1 Anterior limb of internal capsule L; T1 FAST ROIs L front orb cortex; dMRI TBSS L2 Anterior corona radiata R; T1 FAST ROIs R planum polare; dMRI TBSS MD Anterior corona radiata R; T1 FAST ROIs R heschl gyrus; dMRI TBSS MD Superior corona radiata R; dMRI TBSS L2 Superior fronto-occipital fasciculus L; rfMRI amplitudes (ICA100 node 38); dMRI TBSS ISOVF Body of corpus callosum; dMRI TBSS L1 Anterior limb of internal capsule R; dMRI TBSS MD External capsule L; dMRI ProbtrackX MD ptr l; dMRI TBSS MD Superior corona radiata L; dMRI TBSS MD External capsule R; dMRI TBSS L3 Genu of corpus callosum; T1 FAST ROIs R front orb cortex; T1 FAST ROIs R paracing gyrus; T1 FAST ROIs L cent operc cortex; dMRI TBSS L2 Superior fronto-occipital fasciculus R; T1 FAST ROIs brain stem; T1 FIRST left putamen volume; dMRI TBSS MD Genu of corpus callosum; dMRI TBSS ICVF External capsule L; dMRI TBSS ICVF Anterior corona radiata L; T1 FAST ROIs R amygdala; dMRI TBSS FA Anterior corona radiata R; dMRI ProbtrackX L3 ptr r; dMRI TBSS ICVF External capsule R; dMRI ProbtrackX MD unc l; dMRI ProbtrackX MD ifo r; dMRI TBSS ICVF Anterior corona radiata R; dMRI ProbtrackX MD unc r; dMRI ProbtrackX L3 str r; T1 FAST ROIs L paracing gyrus; dMRI TBSS MO Body of corpus callosum; T1 FAST ROIs L cerebellum VI; dMRI TBSS ISOVF Fornix cres+Stria terminalis R; T1 FAST ROIs R cerebellum crus I; T1 FAST ROIs L planum polare; T1 FAST ROIs R cent operc cortex; dMRI ProbtrackX L3 ptr l; T1 FAST ROIs L heschl gyrus; dMRI ProbtrackX L2 ilf r; dMRI ProbtrackX L3 str l; T1 FAST ROIs R cerebellum VI; dMRI ProbtrackX MD ilf r; T1 FIRST left accumbens volume; dMRI ProbtrackX L2 ifo r; T1 FIRST right putamen volume; dMRI TBSS L3 External capsule R; dMRI ProbtrackX L2 unc r; dMRI ProbtrackX MD ifo l; dMRI ProbtrackX L2 ifo l; dMRI ProbtrackX L2 unc l; dMRI ProbtrackX ISOVF str r; dMRI TBSS L1 Superior fronto-occipital fasciculus L; T1 FAST ROIs R ventral striatum; dMRI TBSS L3 External capsule L; dMRI TBSS MD Fornix cres+Stria terminalis L; dMRI TBSS MD Posterior corona radiata R; dMRI ProbtrackX L3 ilf r; dMRI ProbtrackX L1 ptr r; rfMRI amplitudes (ICA100 node 15); dMRI TBSS L3 Posterior corona radiata R; dMRI TBSS ISOVF Superior corona radiata R; dMRI ProbtrackX L2 ilf l; dMRI ProbtrackX L3 unc l; dMRI ProbtrackX MD slf r; T1 FAST ROIs L amygdala; dMRI ProbtrackX L3 ifo r; T1 FAST ROIs R precun cortex; dMRI ProbtrackX ISOVF str l; rfMRI amplitudes (ICA100 node 51); dMRI ProbtrackX L3 unc r; dMRI ProbtrackX L3 ifo l; dMRI TBSS L1 Posterior corona radiata L; dMRI TBSS ICVF Anterior limb of internal capsule R; dMRI TBSS L2 Posterior thalamic radiation R; dMRI TBSS MO Anterior limb of internal capsule R; dMRI ProbtrackX MD ilf l; dMRI TBSS MD Tapetum R; dMRI TBSS FA External capsule

R; dMRI TBSS OD Fornix; dMRI TBSS MO Fornix cres+Stria terminalis L; rfMRI amplitudes (ICA100 node 52);
 dMRI TBSS FA Cerebral peduncle L; dMRI TBSS L3 Tapetum R; rfMRI amplitudes (ICA25 node 18); T1 SIENAX
 white unnormalised volume; dMRI ProbtrackX L1 unc l; dMRI ProbtrackX L2 slf r; T1 SIENAX white normalised
 volume; dMRI ProbtrackX L1 slf r; rfMRI amplitudes (ICA100 node 2); dMRI ProbtrackX MD slf l; T1 FAST
 ROIs L cerebellum crus I; dMRI ProbtrackX L3 ilf l; dMRI ProbtrackX ISOVF ptr r; dMRI TBSS MD Body of
 corpus callosum; dMRI TBSS MD Posterior limb of internal capsule R; dMRI TBSS L2 Posterior thalamic radiation
 L; dMRI ProbtrackX ISOVF atr l; dMRI TBSS OD Fornix cres+Stria terminalis L; dMRI TBSS L1 Tapetum R;
 dMRI TBSS MD Posterior corona radiata L; dMRI TBSS L3 Posterior corona radiata L; dMRI ProbtrackX L1
 ifo r; dMRI TBSS MO Anterior limb of internal capsule L; dMRI ProbtrackX L1 ptr l; dMRI TBSS L3 Posterior
 thalamic radiation L; rfMRI amplitudes (ICA100 node 54); dMRI ProbtrackX L1 unc r; rfMRI amplitudes (ICA100
 node 55); dMRI TBSS ICVF Anterior limb of internal capsule L; T1 FAST ROIs R precentral gyrus; T1 FAST
 ROIs L front operc cortex; T1 FAST ROIs L precentral gyrus; dMRI TBSS FA Genu of corpus callosum; T1 FAST
 ROIs R latocc cortex sup; dMRI ProbtrackX L3 slf r; dMRI ProbtrackX L2 str r; dMRI ProbtrackX ISOVF ptr l;
 dMRI TBSS MD Superior longitudinal fasciculus R; T1 FAST ROIs L precun cortex; rfMRI amplitudes (ICA100
 node 18); dMRI TBSS L3 Posterior thalamic radiation R; dMRI ProbtrackX ICVF fmi; dMRI TBSS L2 Genu
 of corpus callosum; dMRI ProbtrackX ISOVF slf r; dMRI ProbtrackX ICVF unc l; rfMRI amplitudes (ICA100
 node 44); dMRI TBSS L1 Superior fronto-occipital fasciculus R; T1 FAST ROIs L temp fusif cortex post; dMRI
 TBSS FA Superior fronto-occipital fasciculus L; dMRI ProbtrackX L2 slf l; T1 FAST ROIs R occ fusif gyrus; dMRI
 ProbtrackX L3 slf l; dMRI ProbtrackX ISOVF atr r; T1 FAST ROIs L postcent gyrus; dMRI ProbtrackX ICVF unc
 r; dMRI ProbtrackX L1 slf l; dMRI TBSS FA External capsule L; T1 FAST ROIs R postcent gyrus; dMRI TBSS
 FA Posterior thalamic radiation R; rfMRI amplitudes (ICA100 node 35); dMRI TBSS ICVF Cingulum cingulate
 gyrus R; dMRI TBSS ISOVF Genu of corpus callosum; dMRI ProbtrackX L1 ilf r; T1 FAST ROIs R front operc
 cortex; dMRI ProbtrackX ICVF str r; rfMRI amplitudes (ICA100 node 19); dMRI ProbtrackX L2 str l; T1 FAST
 ROIs R cerebellum X; dMRI TBSS ICVF Cingulum cingulate gyrus L; rfMRI amplitudes (ICA100 node 21); T1
 FAST ROIs R cerebellum crus II; dMRI TBSS ISOVF Superior corona radiata L; rfMRI amplitudes (ICA100 node
 17); dMRI TBSS MO Fornix cres+Stria terminalis R; dMRI ProbtrackX MO cgc l; dMRI TBSS FA Posterior
 thalamic radiation L; rfMRI amplitudes (ICA100 node 47); T1 FAST ROIs L occ pole; dMRI ProbtrackX OD str
 r; dMRI ProbtrackX ICVF str l; dMRI ProbtrackX FA atr l; dMRI TBSS L1 Superior corona radiata L; dMRI
 TBSS L3 Posterior limb of internal capsule R; dMRI TBSS L3 Body of corpus callosum; T1 FAST ROIs L latocc

cortex sup; dMRI TBSS L1 Superior corona radiata R; dMRI TBSS L2 Cerebral peduncle L; dMRI TBSS MD Posterior thalamic radiation L; T1 FAST ROIs L subcallosal cortex; dMRI TBSS ICVF Superior corona radiata L; dMRI TBSS ISOVF Fornix cres+Stria terminalis L; dMRI ProbtrackX FA fmi; dMRI TBSS OD Fornix cres+Stria terminalis R; dMRI ProbtrackX L2 cgc l; T1 FAST ROIs R sup temp gyrus post; rfMRI amplitudes (ICA100 node 37); dMRI TBSS MD Superior longitudinal fasciculus L; rfMRI amplitudes (ICA100 node 3); T1 FAST ROIs L occ fusif gyrus; dMRI ProbtrackX ISOVF ilf r; T1 FAST ROIs R mid temp gyrus post; dMRI TBSS ISOVF Posterior corona radiata R; rfMRI amplitudes (ICA100 node 22); T1 FAST ROIs L temporal pole; dMRI ProbtrackX ICVF ifo l; dMRI TBSS L1 Genu of corpus callosum; T1 FAST ROIs R temporal pole; dMRI TBSS MD Posterior thalamic radiation R; dMRI TBSS ISOVF Superior longitudinal fasciculus R; rfMRI amplitudes (ICA100 node 8); dMRI ProbtrackX L1 ifo l; dMRI TBSS ICVF Superior corona radiata R; rfMRI amplitudes (ICA100 node 31); T1 FAST ROIs R planum temporale; dMRI TBSS L2 Body of corpus callosum; T1 FAST ROIs L hippocampus; rfMRI amplitudes (ICA25 node 16); dMRI TBSS FA Superior fronto-occipital fasciculus R; dMRI ProbtrackX ICVF ifo r; dMRI TBSS L3 Cerebral peduncle L; dMRI TBSS L3 Superior longitudinal fasciculus R; rfMRI amplitudes (ICA100 node 45); dMRI ProbtrackX FA ilf r; rfMRI amplitudes (ICA100 node 43); dMRI TBSS L1 Splenium of corpus callosum; dMRI TBSS MO External capsule R; rfMRI amplitudes (ICA100 node 23); T1 FAST ROIs L cerebellum X; dMRI ProbtrackX FA ilf l; dMRI ProbtrackX OD atr r; rfMRI amplitudes (ICA100 node 40); dMRI TBSS L2 Inferior cerebellar peduncle R; T1 FAST ROIs R hippocampus; dMRI TBSS L1 Tapetum L; T1 FAST ROIs L mid temp gyrus post; rfMRI amplitudes (ICA100 node 7); dMRI ProbtrackX OD str l; T1 FAST ROIs R temp fusif cortex post; dMRI TBSS L3 Sagittal stratum R; dMRI ProbtrackX ICVF cgc l; rfMRI amplitudes (ICA100 node 27); T1 FAST ROIs L cerebellum VIIa; rfMRI amplitudes (ICA25 node 15); dMRI TBSS L3 Superior longitudinal fasciculus L; dMRI TBSS MD Splenium of corpus callosum; dMRI TBSS ICVF Uncinate fasciculus L; T1 FAST ROIs L cerebellum crus II; dMRI TBSS MO Tapetum L; dMRI TBSS FA Cerebral peduncle R; T1 FAST ROIs L insular cortex; dMRI TBSS ISOVF Tapetum R; dMRI ProbtrackX L3 fmi; rfMRI amplitudes (ICA25 node 14); dMRI TBSS ICVF Uncinate fasciculus R; dMRI TBSS L3 Tapetum L; T1 FIRST right hippocampus volume; dMRI TBSS FA Body of corpus callosum; rfMRI amplitudes (ICA100 node 30); rfMRI amplitudes (ICA25 node 17); rfMRI amplitudes (ICA100 node 13); T1 FAST ROIs R cerebellum VIIa; dMRI ProbtrackX L2 fmi; dMRI ProbtrackX FA ptr l; dMRI TBSS L1 External capsule L; dMRI ProbtrackX L1 ilf l; dMRI TBSS L1 Superior longitudinal fasciculus R; dMRI TBSS MD Posterior limb of internal capsule L; dMRI TBSS MD Tapetum L; dMRI ProbtrackX ICVF cgc r; dMRI TBSS L3 Sagittal stratum L; dMRI TBSS ICVF Genu of corpus callosum; dMRI TBSS L3

Splenium of corpus callosum; dMRI TBSS ISOVF Posterior corona radiata L; dMRI TBSS MO Posterior corona radiata R; T1 FAST ROIs L inf front gyrus parsop; T1 FAST ROIs R cuneal cortex; T1 FIRST left hippocampus volume; dMRI TBSS L2 Inferior cerebellar peduncle L; dMRI ProbtrackX ISOVF slf l; rfMRI amplitudes (ICA100 node 33); dMRI TBSS L2 Cingulum cingulate gyrus R; rfMRI amplitudes (ICA25 node 20); dMRI TBSS MO External capsule L; dMRI TBSS MD Uncinate fasciculus R; T1 FAST ROIs L front med cortex; T1 FAST ROIs R occ pole; dMRI ProbtrackX ICVF slf l; dMRI TBSS OD Posterior corona radiata R; dMRI ProbtrackX L2 cgc r; dMRI TBSS L1 Superior cerebellar peduncle L; dMRI ProbtrackX FA ifo l; T1 FAST ROIs L cerebellum VIIb; dMRI TBSS L2 Tapetum R; T1 FAST ROIs R cerebellum VIIb; dMRI TBSS L1 Superior cerebellar peduncle R; dMRI ProbtrackX FA atr r; dMRI ProbtrackX ICVF ptr l; dMRI TBSS MD Retrolenticular part of internal capsule R; rfMRI amplitudes (ICA100 node 42); T1 FAST ROIs R front med cortex; dMRI TBSS L2 Cingulum cingulate gyrus L; T1 FAST ROIs L temp fusif cortex ant; dMRI TBSS MD Uncinate fasciculus L; rfMRI amplitudes (ICA100 node 25); dMRI TBSS L1 External capsule R; T1 FAST ROIs L sup temp gyrus post; dMRI ProbtrackX MD fmi; dMRI TBSS L3 Posterior limb of internal capsule L; rfMRI amplitudes (ICA100 node 12); dMRI TBSS L2 Superior longitudinal fasciculus R; dMRI TBSS MD Sagittal stratum R; dMRI TBSS L2 Cerebral peduncle R; dMRI TBSS L3 Cerebral peduncle R; rfMRI amplitudes (ICA100 node 41); T1 FAST ROIs L cerebellum V; dMRI ProbtrackX ISOVF ilf l; rfMRI amplitudes (ICA100 node 11); dMRI TBSS L1 Body of corpus callosum; dMRI TBSS ICVF Posterior corona radiata R; dMRI ProbtrackX ICVF slf r; rfMRI amplitudes (ICA100 node 14); dMRI ProbtrackX ICVF ptr r; dMRI ProbtrackX FA unc l; rfMRI amplitudes (ICA100 node 4); dMRI ProbtrackX FA slf l; dMRI TBSS L2 Superior corona radiata R; dMRI ProbtrackX FA slf r; T1 FAST ROIs R latocc cortex inf; dMRI TBSS MD Sagittal stratum L; dMRI ProbtrackX ISOVF unc l; dMRI ProbtrackX FA ifo r; dMRI TBSS L1 Anterior corona radiata L; T1 FAST ROIs R cerebellum V; rfMRI amplitudes (ICA25 node 4); rfMRI amplitudes (ICA25 node 21); dMRI TBSS ICVF Posterior thalamic radiation L; dMRI TBSS FA Cingulum cingulate gyrus L; dMRI TBSS ICVF Tapetum R; dMRI TBSS ICVF Superior longitudinal fasciculus L; dMRI ProbtrackX ICVF ilf l; dMRI TBSS L1 Retrolenticular part of internal capsule R; T1 FAST ROIs R lingual gyrus; dMRI TBSS L3 Cingulum cingulate gyrus L; T1 FAST ROIs R temp occ fusif cortex; BrainCoGX; rfMRI amplitudes (ICA100 node 53); T1 FAST ROIs R parietal operc cortex; dMRI TBSS OD Tapetum L; T1 FAST ROIs V cerebellum VIIa; dMRI TBSS ICVF Sagittal stratum L; T1 FAST ROIs R insular cortex; dMRI ProbtrackX L1 cgh r; dMRI ProbtrackX FA unc r; dMRI TBSS L1 Superior longitudinal fasciculus L; dMRI TBSS FA Inferior cerebellar peduncle R; dMRI TBSS ICVF Posterior corona radiata L; dMRI ProbtrackX MO cgc r; rfMRI amplitudes (ICA100 node 26); T1 FAST

ROIs L sup front gyrus; dMRI TBSS FA Superior corona radiata L; T1 FAST ROIs R supramarg gyrus post; rfMRI amplitudes (ICA25 node 3); dMRI TBSS FA Superior corona radiata R; dMRI ProbtrackX ICVF ilf r; dMRI TBSS MD Retrolenticular part of internal capsule L; dMRI ProbtrackX FA ptr r; dMRI TBSS L3 Retrolenticular part of internal capsule R; dMRI ProbtrackX MO str r; T1 FAST ROIs R inf front gyrus parstri; T1 FAST ROIs R cerebellum I-IV; dMRI TBSS ISOVF Anterior limb of internal capsule L; T1 FAST ROIs L cuneal cortex; T1 FAST ROIs R pallidum; rfMRI amplitudes (ICA100 node 1); dMRI TBSS FA Cingulum cingulate gyrus R; rfMRI amplitudes (ICA25 node 12); T1 FAST ROIs R temp fusif cortex ant; T1 FAST ROIs R inf front gyrus parsop; dMRI TBSS L1 Posterior limb of internal capsule R; dMRI TBSS ICVF Posterior thalamic radiation R; dMRI TBSS L3 Inferior cerebellar peduncle R; rfMRI amplitudes (ICA25 node 5); T1 FAST ROIs L inf front gyrus parstri; dMRI TBSS L2 Uncinate fasciculus L; dMRI TBSS MO Splenium of corpus callosum; dMRI TBSS L1 Anterior corona radiata R; dMRI TBSS ISOVF Anterior limb of internal capsule R; T1 FAST ROIs L cerebellum I-IV; dMRI ProbtrackX MO str l; rfMRI amplitudes (ICA100 node 9); T1 FAST ROIs L lingual gyrus; dMRI TBSS L2 Superior corona radiata L; dMRI TBSS L2 Superior longitudinal fasciculus L; dMRI TBSS FA Middle cerebellar peduncle; dMRI TBSS L3 Cingulum cingulate gyrus R; dMRI TBSS ISOVF Superior fronto-occipital fasciculus L; rfMRI amplitudes (ICA100 node 50); dMRI TBSS L2 Sagittal stratum R; rfMRI amplitudes (ICA100 node 29); dMRI TBSS ISOVF Posterior thalamic radiation L; dMRI TBSS ICVF Tapetum L; dMRI TBSS OD Cerebral peduncle L; T1 FAST ROIs R sup temp gyrus ant; T1 FAST ROIs L parahipp gyrus ant; dMRI ProbtrackX ISOVF ifo r; dMRI TBSS ICVF Sagittal stratum R; dMRI TBSS L3 Inferior cerebellar peduncle L; rfMRI amplitudes (ICA100 node 39); rfMRI amplitudes (ICA25 node 6); T1 FAST ROIs R juxta pos lobule cortex; dMRI ProbtrackX OD atr l; dMRI TBSS FA Tapetum R; dMRI TBSS ICVF Posterior limb of internal capsule R; dMRI TBSS MO Fornix; T1 FAST ROIs L mid front gyrus; dMRI TBSS ISOVF Posterior limb of internal capsule R; T1 FAST ROIs L sup temp gyrus ant; dMRI TBSS ICVF Cingulum hippocampus L; T1 FAST ROIs L planum temporale; dMRI TBSS ICVF Superior longitudinal fasciculus R; dMRI ProbtrackX MD cgh r; dMRI TBSS MO Inferior cerebellar peduncle R; dMRI TBSS ICVF Cingulum hippocampus R; T1 FAST ROIs L parahipp gyrus post; dMRI TBSS OD Splenium of corpus callosum; dMRI TBSS L3 Retrolenticular part of internal capsule L; rfMRI amplitudes (ICA100 node 32); T1 FAST ROIs V cerebellum VI; T1 FAST ROIs R mid front gyrus; dMRI TBSS ISOVF External capsule R; rfMRI amplitudes (ICA100 node 34); dMRI TBSS L2 Uncinate fasciculus R; T1 FAST ROIs R angular gyrus; rfMRI amplitudes (ICA25 node 8); dMRI TBSS MD Inferior cerebellar peduncle L; rfMRI amplitudes (ICA100 node 28); T1 FAST ROIs R cerebellum VIIIb; dMRI TBSS FA Sagittal stratum R; dMRI ProbtrackX MO fmi; T1 FAST ROIs R parahipp gyrus ant; dMRI TBSS

MD Inferior cerebellar peduncle R; dMRI TBSS OD Inferior cerebellar peduncle R; rfMRI amplitudes (ICA100 node 10); dMRI TBSS ICVF Fornix cres+Stria terminalis L; T1 FAST ROIs L sup parietal lobule; dMRI TBSS ISOVF External capsule L; rfMRI amplitudes (ICA25 node 7); dMRI ProbtrackX ICVF ar r; T1 FAST ROIs L temp occ fusif cortex; dMRI TBSS L2 Posterior corona radiata L; dMRI TBSS ISOVF Superior cerebellar peduncle R; dMRI TBSS FA Anterior limb of internal capsule R; dMRI TBSS ISOVF Superior cerebellar peduncle L; T1 FAST ROIs L parietal operc cortex; T1 FAST ROIs R sup front gyrus; dMRI TBSS FA Inferior cerebellar peduncle L; T1 FAST ROIs L latocc cortex inf; dMRI TBSS ICVF Posterior limb of internal capsule L; T1 FAST ROIs R parahipp gyrus post; dMRI ProbtrackX L2 mcp; T1 FAST ROIs L pallidum; dMRI TBSS ISOVF Posterior thalamic radiation R; dMRI ProbtrackX ISOVF cgh r; T1 FAST ROIs L mid temp gyrus ant; dMRI TBSS MO Cingulum cingulate gyrus R; dMRI ProbtrackX ICVF ar l; dMRI TBSS ICVF Retrolenticular part of internal capsule R; dMRI ProbtrackX ISOVF ifo l; dMRI TBSS L2 Retrolenticular part of internal capsule L; dMRI TBSS OD Body of corpus callosum; dMRI TBSS ISOVF Tapetum L; dMRI ProbtrackX FA cgc l; T1 FIRST right pallidum volume; dMRI TBSS L3 Uncinate fasciculus L; dMRI ProbtrackX L3 mcp; dMRI TBSS FA Anterior limb of internal capsule L; T1 FAST ROIs L juxtapos lobule cortex; dMRI ProbtrackX L2 cgh r; rfMRI amplitudes (ICA100 node 5); rfMRI amplitudes (ICA25 node 9); dMRI ProbtrackX ISOVF unc r; dMRI TBSS MD Superior cerebellar peduncle L; dMRI TBSS MD Cingulum cingulate gyrus R; dMRI TBSS MO Superior corona radiata L; T1 FAST ROIs R inf temp gyrus post; dMRI TBSS ISOVF Superior longitudinal fasciculus L; dMRI ProbtrackX L3 cst l; dMRI TBSS L2 Sagittal stratum L; dMRI TBSS ISOVF Splenium of corpus callosum; dMRI TBSS ICVF Fornix; dMRI ProbtrackX MD cgc r; T1 FAST ROIs L cerebellum VIIIb; dMRI TBSS MD Cingulum cingulate gyrus L; dMRI TBSS L3 Middle cerebellar peduncle; dMRI ProbtrackX L3 cst r; T1 FAST ROIs L cing gyrus post; dMRI TBSS FA Superior longitudinal fasciculus R; dMRI ProbtrackX MD cgh l; T1 FAST ROIs R mid temp gyrus ant; T1 FAST ROIs V cerebellum VIIb; dMRI TBSS MO Cerebral peduncle L; dMRI TBSS OD Cingulum cingulate gyrus R; T1 FIRST left caudate volume; dMRI TBSS MO Inferior cerebellar peduncle L; dMRI ProbtrackX MD cst r; rfMRI amplitudes (ICA25 node 13); dMRI TBSS ICVF Retrolenticular part of internal capsule L; rfMRI amplitudes (ICA25 node 11); dMRI ProbtrackX MD cgc l; dMRI ProbtrackX L2 cgh l; dMRI TBSS FA Superior longitudinal fasciculus L; dMRI TBSS OD Cerebral peduncle R; dMRI TBSS L2 Cingulum hippocampus L; dMRI ProbtrackX L3 cgh r; T1 FAST ROIs R sup parietal lobule; dMRI TBSS ICVF Splenium of corpus callosum; dMRI TBSS MO Tapetum R; dMRI ProbtrackX L3 cgc l; dMRI TBSS ISOVF Anterior corona radiata L; dMRI TBSS L3 Uncinate fasciculus R; rfMRI amplitudes (ICA100 node 36); dMRI TBSS ISOVF Cerebral peduncle L; dMRI TBSS MD Superior cerebellar pe-

duncle R; dMRI ProbtrackX L3 cgh l; dMRI TBSS L2 Anterior limb of internal capsule L; dMRI TBSS L3 Cingulum
 hippocampus R; rfMRI connectivity (ICA25 edge 151); rfMRI connectivity (ICA100 edge 579); rfMRI connectivity
 (ICA25 edge 58); dMRI TBSS OD Middle cerebellar peduncle; T1 FAST ROIs R intracalc cortex; rfMRI ampli-
 tudes (ICA100 node 6); dMRI TBSS FA Sagittal stratum L; dMRI TBSS ISOVF Retrolenticular part of internal
 capsule R; T1 FAST ROIs R supracalc cortex; tfMRI 90th-percentile zstat faces-shapes; dMRI TBSS MO Posterior
 thalamic radiation R; T1 FAST ROIs V cerebellum X; dMRI TBSS ICVF Fornix cres+Stria terminalis R; dMRI
 TBSS L3 Cingulum hippocampus L; dMRI TBSS MO Cingulum cingulate gyrus L; rfMRI amplitudes (ICA100 node
 24); dMRI TBSS ISOVF Cerebral peduncle R; dMRI ProbtrackX MD mcp; dMRI ProbtrackX MD cst l; dMRI
 ProbtrackX L1 cgh l; T1 FAST ROIs R subcallosal cortex; rfMRI amplitudes (ICA100 node 49); rfMRI amplitudes
 (ICA25 node 10); dMRI TBSS ISOVF Superior fronto-occipital fasciculus R; rfMRI amplitudes (ICA100 node 48);
 dMRI TBSS FA Posterior corona radiata L; T1 FAST ROIs R mid temp gyrus tempocc; dMRI ProbtrackX L3 cgc
 r; dMRI TBSS ISOVF Inferior cerebellar peduncle R; rfMRI connectivity (ICA100 edge 212); T1 FAST ROIs L
 inf temp gyrus ant; T1 FAST ROIs L intracalc cortex; dMRI TBSS ISOVF Retrolenticular part of internal capsule
 L; rfMRI connectivity (ICA100 edge 899); T1 FAST ROIs V cerebellum VIIIb; T1 FAST ROIs L inf temp gyrus
 post; dMRI TBSS L1 Posterior limb of internal capsule L; dMRI TBSS MO Posterior limb of internal capsule R;
 dMRI TBSS MO Superior corona radiata R; dMRI ProbtrackX L1 fmi; SWI T2star left putamen; dMRI ProbtrackX
 ISOVF cst r; dMRI TBSS L1 Retrolenticular part of internal capsule L; dMRI TBSS FA Corticospinal tract L; dMRI
 ProbtrackX OD ptr r; dMRI TBSS L1 Fornix cres+Stria terminalis R; dMRI TBSS L2 Anterior limb of internal
 capsule R; dMRI ProbtrackX FA cgc r; dMRI TBSS MO Posterior corona radiata L; dMRI ProbtrackX ISOVF cgh
 l; tfMRI median zstat faces-shapes; T1 FAST ROIs R cing gyrus post; dMRI TBSS OD Posterior corona radiata L;
 dMRI TBSS OD Superior cerebellar peduncle R; dMRI TBSS L2 Tapetum L; dMRI TBSS OD Inferior cerebellar
 peduncle L; rfMRI amplitudes (ICA25 node 2); dMRI ProbtrackX MO ptr l; dMRI TBSS ISOVF Cingulum hip-
 pocampus L; dMRI TBSS L1 Uncinate fasciculus R; SWI T2star right thalamus; dMRI TBSS MO Anterior corona
 radiata R; dMRI TBSS MO Anterior corona radiata L; dMRI ProbtrackX L2 cst r; tfMRI 90th-percentile BOLD
 faces-shapes; rfMRI connectivity (ICA100 edge 551); dMRI TBSS L2 Middle cerebellar peduncle; dMRI ProbtrackX
 ICVF cst r; rfMRI amplitudes (ICA25 node 19); tfMRI median zstat faces-shapes amygdala; dMRI TBSS MO Pos-
 terior thalamic radiation L; dMRI TBSS OD Cingulum cingulate gyrus L; T1 FAST ROIs L supracalc cortex; T1
 FAST ROIs R inf temp gyrus ant; dMRI TBSS OD Anterior corona radiata L; dMRI TBSS FA Corticospinal tract
 R; T1 FIRST right caudate volume; rfMRI amplitudes (ICA100 node 20); dMRI TBSS FA Medial lemniscus L;

dMRI TBSS FA Cingulum hippocampus L; dMRI ProbtrackX ICVF cst l; SWI T2star left thalamus; dMRI TBSS ISOVF Inferior cerebellar peduncle L; dMRI TBSS L1 Sagittal stratum R; dMRI TBSS L2 Posterior corona radiata R; dMRI TBSS MO Cerebral peduncle R; rfMRI connectivity (ICA100 edge 1079); dMRI TBSS ISOVF Sagittal stratum R; dMRI TBSS L2 Retrolenticular part of internal capsule R; dMRI ProbtrackX MO ilf l; T1 FIRST left pallidum volume; dMRI ProbtrackX OD ar r; dMRI TBSS FA Retrolenticular part of internal capsule L; dMRI TBSS OD Superior cerebellar peduncle L; dMRI TBSS L1 Sagittal stratum L; rfMRI connectivity (ICA100 edge 244); dMRI TBSS OD Posterior thalamic radiation R; rfMRI connectivity (ICA100 edge 82); rfMRI connectivity (ICA25 edge 55); dMRI TBSS L1 Uncinate fasciculus L; SWI T2star right putamen; dMRI TBSS FA Medial lemniscus R; T1 FAST ROIs L supramarg gyrus post; dMRI ProbtrackX L2 cst l; dMRI ProbtrackX ISOVF cst l; dMRI ProbtrackX ISOVF mcp; rfMRI connectivity (ICA25 edge 185); rfMRI amplitudes (ICA100 node 46); dMRI TBSS OD Medial lemniscus R; dMRI TBSS ISOVF Posterior limb of internal capsule L; dMRI TBSS OD Tapetum R; rfMRI connectivity (ICA100 edge 683); rfMRI connectivity (ICA100 edge 990); dMRI ProbtrackX MD ar r; tfMRI median BOLD faces-shapes; dMRI ProbtrackX L1 ar r; dMRI TBSS L1 Cerebral peduncle L; rfMRI connectivity (ICA25 edge 120); rfMRI connectivity (ICA25 edge 172); rfMRI connectivity (ICA25 edge 101); dMRI TBSS OD Corticospinal tract R; rfMRI connectivity (ICA100 edge 488); rfMRI connectivity (ICA100 edge 467); dMRI TBSS ICVF Superior cerebellar peduncle R; tfMRI median BOLD faces-shapes amygdala; rfMRI connectivity (ICA100 edge 595); dMRI TBSS MO Posterior limb of internal capsule L; T1 FAST ROIs L angular gyrus; rfMRI connectivity (ICA100 edge 6); rfMRI connectivity (ICA100 edge 98); dMRI TBSS MD Cerebral peduncle R; dMRI TBSS MD Cerebral peduncle L; dMRI ProbtrackX ICVF fma; dMRI TBSS MO Genu of corpus callosum; dMRI TBSS ICVF Middle cerebellar peduncle; dMRI TBSS FA Tapetum L; dMRI TBSS L1 Posterior thalamic radiation L; dMRI ProbtrackX L2 ar r; dMRI TBSS MD Cingulum hippocampus L; dMRI ProbtrackX OD ifo r; rfMRI connectivity (ICA100 edge 408); rfMRI connectivity (ICA100 edge 251); rfMRI connectivity (ICA100 edge 560); dMRI TBSS FA Uncinate fasciculus L; rfMRI connectivity (ICA25 edge 202); rfMRI connectivity (ICA100 edge 1052); dMRI TBSS FA Cingulum hippocampus R; rfMRI connectivity (ICA100 edge 135); rfMRI connectivity (ICA100 edge 992); dMRI TBSS OD Medial lemniscus L; dMRI TBSS OD Cingulum hippocampus L; rfMRI connectivity (ICA100 edge 129); dMRI ProbtrackX L1 mcp; dMRI ProbtrackX MO ifo l; T1 FAST ROIs L mid temp gyrus tempocc; tfMRI 90th-percentile zstat faces-shapes amygdala; rfMRI connectivity (ICA100 edge 674); rfMRI connectivity (ICA100 edge 1080); dMRI ProbtrackX ICVF mcp; dMRI TBSS OD Corticospinal tract L; rfMRI connectivity (ICA25 edge 161); dMRI TBSS MD Cingulum hippocampus R; dMRI TBSS L2 Cingulum hippocampus R; T1 FAST ROIs R

cerebellum IX; dMRI ProbtrackX MO ilf r; rfMRI connectivity (ICA25 edge 130); rfMRI connectivity (ICA25 edge 155); dMRI ProbtrackX L3 ar r; dMRI TBSS ISOVF Cingulum cingulate gyrus R; rfMRI connectivity (ICA100 edge 517); rfMRI connectivity (ICA100 edge 831); dMRI ProbtrackX ICVF cgh l; dMRI TBSS ISOVF Anterior corona radiata R; dMRI TBSS MO Uncinate fasciculus L; rfMRI connectivity (ICA100 edge 999); dMRI TBSS MD Middle cerebellar peduncle; dMRI TBSS OD Posterior thalamic radiation L; rfMRI amplitudes (ICA100 node 16); T1 FAST ROIs L cerebellum IX; rfMRI connectivity (ICA100 edge 94); T1 FAST ROIs R supramarg gyrus ant; dMRI ProbtrackX L1 ml l; rfMRI connectivity (ICA100 edge 327); rfMRI connectivity (ICA100 edge 707); T1 FAST ROIs V cerebellum crus I; rfMRI connectivity (ICA100 edge 395); rfMRI connectivity (ICA100 edge 471); dMRI TBSS FA Posterior corona radiata R; rfMRI connectivity (ICA100 edge 12); rfMRI connectivity (ICA100 edge 485); rfMRI connectivity (ICA25 edge 203); dMRI ProbtrackX L1 cst r; rfMRI connectivity (ICA25 edge 182); dMRI ProbtrackX ISOVF cgc l; dMRI TBSS ISOVF Uncinate fasciculus L; rfMRI connectivity (ICA100 edge 928); dMRI TBSS L3 Medial lemniscus L; rfMRI connectivity (ICA25 edge 29); dMRI ProbtrackX FA cst r; rfMRI connectivity (ICA100 edge 318); dMRI TBSS OD Anterior corona radiata R; T1 FAST ROIs L supramarg gyrus ant; rfMRI connectivity (ICA100 edge 651); rfMRI connectivity (ICA100 edge 342); dMRI ProbtrackX ISOVF cgc r; rfMRI connectivity (ICA25 edge 23); rfMRI connectivity (ICA100 edge 694); dMRI ProbtrackX FA cst l; dMRI TBSS FA Uncinate fasciculus R; rfMRI connectivity (ICA25 edge 183); rfMRI connectivity (ICA100 edge 675); rfMRI connectivity (ICA100 edge 192); rfMRI connectivity (ICA100 edge 878); dMRI TBSS FA Pontine crossing tract; rfMRI connectivity (ICA100 edge 88); rfMRI connectivity (ICA25 edge 89); rfMRI connectivity (ICA100 edge 97); rfMRI connectivity (ICA25 edge 171); T1 FAST ROIs V cerebellum IX; rfMRI connectivity (ICA25 edge 131); dMRI ProbtrackX FA mcp; rfMRI connectivity (ICA100 edge 647); rfMRI connectivity (ICA25 edge 146); T1 FAST ROIs L thalamus; dMRI ProbtrackX MO fma; rfMRI connectivity (ICA100 edge 751); T1 FAST ROIs V cerebellum crus II; rfMRI connectivity (ICA100 edge 703); rfMRI connectivity (ICA100 edge 764); dMRI TBSS MO Cingulum hippocampus L; dMRI ProbtrackX OD cgc l; rfMRI connectivity (ICA100 edge 876); rfMRI connectivity (ICA100 edge 963); rfMRI connectivity (ICA25 edge 117); dMRI ProbtrackX MO ifo r; rfMRI connectivity (ICA25 edge 115); rfMRI connectivity (ICA100 edge 237); rfMRI connectivity (ICA100 edge 177); rfMRI connectivity (ICA100 edge 1239); dMRI TBSS MO Corticospinal tract R; dMRI ProbtrackX OD ptr l; dMRI TBSS FA Retrolenticular part of internal capsule R; rfMRI connectivity (ICA25 edge 102); rfMRI connectivity (ICA100 edge 657); rfMRI connectivity (ICA100 edge 693); dMRI TBSS ISOVF Sagittal stratum L; dMRI ProbtrackX MO ptr r; rfMRI connectivity (ICA100 edge 558); dMRI TBSS L2 Posterior limb of internal capsule R; rfMRI connectivity

(ICA100 edge 290); rfMRI connectivity (ICA100 edge 95); rfMRI connectivity (ICA100 edge 824); rfMRI connectivity (ICA25 edge 7); dMRI ProbtrackX MD fma; rfMRI connectivity (ICA100 edge 896); rfMRI connectivity (ICA25 edge 108); dMRI TBSS ISOVF Cingulum cingulate gyrus L; rfMRI connectivity (ICA100 edge 588); rfMRI connectivity (ICA100 edge 345); rfMRI connectivity (ICA100 edge 495); rfMRI connectivity (ICA25 edge 176); rfMRI connectivity (ICA25 edge 38); rfMRI connectivity (ICA100 edge 399); dMRI TBSS MO Retrolenticular part of internal capsule L; rfMRI connectivity (ICA100 edge 895); dMRI TBSS L1 Posterior thalamic radiation R; rfMRI connectivity (ICA100 edge 943); rfMRI connectivity (ICA100 edge 953); dMRI TBSS OD Superior corona radiata L; dMRI ProbtrackX MD ml l; rfMRI connectivity (ICA100 edge 769); rfMRI connectivity (ICA100 edge 131); rfMRI connectivity (ICA100 edge 155); rfMRI connectivity (ICA100 edge 164); dMRI ProbtrackX L1 cst l; dMRI TBSS L3 Medial lemniscus R; rfMRI connectivity (ICA25 edge 87); dMRI TBSS OD Posterior limb of internal capsule R; dMRI TBSS FA Posterior limb of internal capsule R; rfMRI connectivity (ICA100 edge 1430); rfMRI connectivity (ICA100 edge 1284); dMRI TBSS MO Uncinate fasciculus R; rfMRI connectivity (ICA100 edge 20); rfMRI connectivity (ICA100 edge 788); rfMRI connectivity (ICA25 edge 50); rfMRI connectivity (ICA100 edge 557); dMRI ProbtrackX L1 fma; dMRI ProbtrackX L1 ml r; T1 FAST ROIs L inf temp gyrus tempocc; rfMRI connectivity (ICA100 edge 633); dMRI TBSS OD Superior corona radiata R; rfMRI connectivity (ICA25 edge 75); rfMRI connectivity (ICA100 edge 1289); SWI T2star left caudate; dMRI TBSS OD Pontine crossing tract; rfMRI connectivity (ICA100 edge 378); rfMRI connectivity (ICA100 edge 783); rfMRI connectivity (ICA100 edge 593); rfMRI connectivity (ICA100 edge 930); rfMRI connectivity (ICA100 edge 546); rfMRI connectivity (ICA100 edge 847); rfMRI connectivity (ICA100 edge 1259); dMRI TBSS ICVF Superior cerebellar peduncle L; rfMRI connectivity (ICA100 edge 548); rfMRI connectivity (ICA100 edge 629); T1 FAST ROIs R caudate; rfMRI connectivity (ICA100 edge 861); rfMRI connectivity (ICA100 edge 205); dMRI ProbtrackX OD cgc r; dMRI ProbtrackX L3 fma; dMRI ProbtrackX FA cgh l; rfMRI connectivity (ICA100 edge 285); rfMRI connectivity (ICA100 edge 125); rfMRI connectivity (ICA100 edge 1067); rfMRI connectivity (ICA100 edge 188); rfMRI connectivity (ICA100 edge 944); dMRI TBSS MO Corticospinal tract L; rfMRI connectivity (ICA25 edge 178); dMRI TBSS ICVF Inferior cerebellar peduncle L; T1 FAST ROIs R inf temp gyrus tempocc; rfMRI connectivity (ICA25 edge 90); rfMRI connectivity (ICA25 edge 100); rfMRI connectivity (ICA100 edge 668); rfMRI connectivity (ICA100 edge 838); SWI T2star right caudate; dMRI TBSS ISOVF Uncinate fasciculus R; rfMRI connectivity (ICA100 edge 1395); rfMRI connectivity (ICA100 edge 22); rfMRI connectivity (ICA25 edge 105); rfMRI connectivity (ICA100 edge 93); rfMRI connectivity (ICA100 edge 802); rfMRI connectivity (ICA100 edge 142); rfMRI connectivity (ICA100 edge 428); tfMRI 90th-

percentile BOLD faces-shapes amygdala; rfMRI connectivity (ICA25 edge 207); rfMRI connectivity (ICA100 edge 605); dMRI ProbtrackX OD slf r; rfMRI connectivity (ICA100 edge 849); rfMRI connectivity (ICA100 edge 286); dMRI TBSS L2 Medial lemniscus R; dMRI TBSS MO Middle cerebellar peduncle; rfMRI connectivity (ICA100 edge 924); rfMRI connectivity (ICA100 edge 1242); dMRI TBSS L2 Medial lemniscus L; dMRI ProbtrackX L2 fma; rfMRI connectivity (ICA100 edge 1182); rfMRI connectivity (ICA100 edge 388); dMRI ProbtrackX MO unc r; dMRI ProbtrackX FA ar l; rfMRI connectivity (ICA100 edge 434); dMRI ProbtrackX FA fma; dMRI TBSS L1 Cerebral peduncle R; rfMRI connectivity (ICA25 edge 169); dMRI TBSS L1 Corticospinal tract R; dMRI TBSS OD Retrolenticular part of internal capsule R; rfMRI connectivity (ICA25 edge 181); rfMRI connectivity (ICA100 edge 969); rfMRI connectivity (ICA100 edge 25); rfMRI connectivity (ICA100 edge 811); rfMRI connectivity (ICA100 edge 59); rfMRI connectivity (ICA100 edge 581); dMRI ProbtrackX L2 ar l; rfMRI connectivity (ICA100 edge 486); rfMRI connectivity (ICA25 edge 32); rfMRI connectivity (ICA100 edge 641); rfMRI connectivity (ICA100 edge 73); rfMRI connectivity (ICA100 edge 425); dMRI TBSS OD External capsule R; dMRI TBSS FA Posterior limb of internal capsule L; rfMRI connectivity (ICA100 edge 159); rfMRI connectivity (ICA100 edge 687); rfMRI connectivity (ICA100 edge 92); rfMRI connectivity (ICA100 edge 421); rfMRI connectivity (ICA25 edge 177); dMRI TBSS MO Superior fronto-occipital fasciculus L; rfMRI connectivity (ICA100 edge 662); rfMRI connectivity (ICA100 edge 75); rfMRI connectivity (ICA100 edge 652); rfMRI connectivity (ICA100 edge 219); rfMRI connectivity (ICA100 edge 113); rfMRI connectivity (ICA100 edge 715); rfMRI connectivity (ICA100 edge 452); rfMRI connectivity (ICA100 edge 1136); rfMRI connectivity (ICA100 edge 100); rfMRI connectivity (ICA100 edge 949); rfMRI connectivity (ICA100 edge 1045); rfMRI connectivity (ICA100 edge 1269); rfMRI connectivity (ICA25 edge 46); rfMRI connectivity (ICA100 edge 91); rfMRI connectivity (ICA25 edge 41); rfMRI connectivity (ICA100 edge 479); rfMRI connectivity (ICA100 edge 741); dMRI ProbtrackX OD slf l; rfMRI connectivity (ICA25 edge 125); dMRI TBSS OD Cingulum hippocampus R; rfMRI connectivity (ICA100 edge 1222); rfMRI connectivity (ICA25 edge 193); rfMRI connectivity (ICA100 edge 86); rfMRI connectivity (ICA100 edge 1474); rfMRI connectivity (ICA100 edge 678); rfMRI connectivity (ICA100 edge 1055); dMRI TBSS L3 Corticospinal tract L; rfMRI connectivity (ICA25 edge 44); rfMRI connectivity (ICA100 edge 956); rfMRI connectivity (ICA100 edge 312); rfMRI connectivity (ICA100 edge 1184); rfMRI connectivity (ICA100 edge 948); dMRI ProbtrackX L3 ml l; rfMRI connectivity (ICA25 edge 199); rfMRI connectivity (ICA100 edge 536); rfMRI connectivity (ICA100 edge 1033); dMRI TBSS L2 Superior cerebellar peduncle L; rfMRI connectivity (ICA100 edge 563); rfMRI connectivity (ICA100 edge 660); rfMRI connectivity (ICA25 edge 122); rfMRI connectivity (ICA100 edge 28); dMRI TBSS L3 Superior cerebellar peduncle L;

rfMRI connectivity (ICA100 edge 1339); rfMRI connectivity (ICA100 edge 591); dMRI ProbtrackX ICVF cgh r;
rfMRI connectivity (ICA25 edge 4); dMRI ProbtrackX L3 ar l; rfMRI connectivity (ICA100 edge 777).

Appendix C

Full spatial maps

The full UK biobank unimodal binary maps from our method are shown in figures C.1, C.3, C.5 and C.7 for VBM, TBSS FA, TBSS MD and TBSS L1 respectively. The corresponding unimodal inverse distance maps are shown in figures C.2, C.4, C.6 and C.8. Meanwhile, the full multimodal UK biobank binary maps from our method are shown in figures C.9, C.11, C.13, C.15 and C.17 for components one to five respectively. The corresponding multimodal inverse distance maps are shown in figures C.10, C.12, C.14, C.16 and C.18. The informed multimodal PLS UK biobank spatial maps are displayed in figures C.19-C.23, while those from joint ICA are displayed in figures C.24-C.28. All the images that we show in this section are axial slices ranging from $z = -8$ mm to $z = 40$ mm in 8 mm steps.

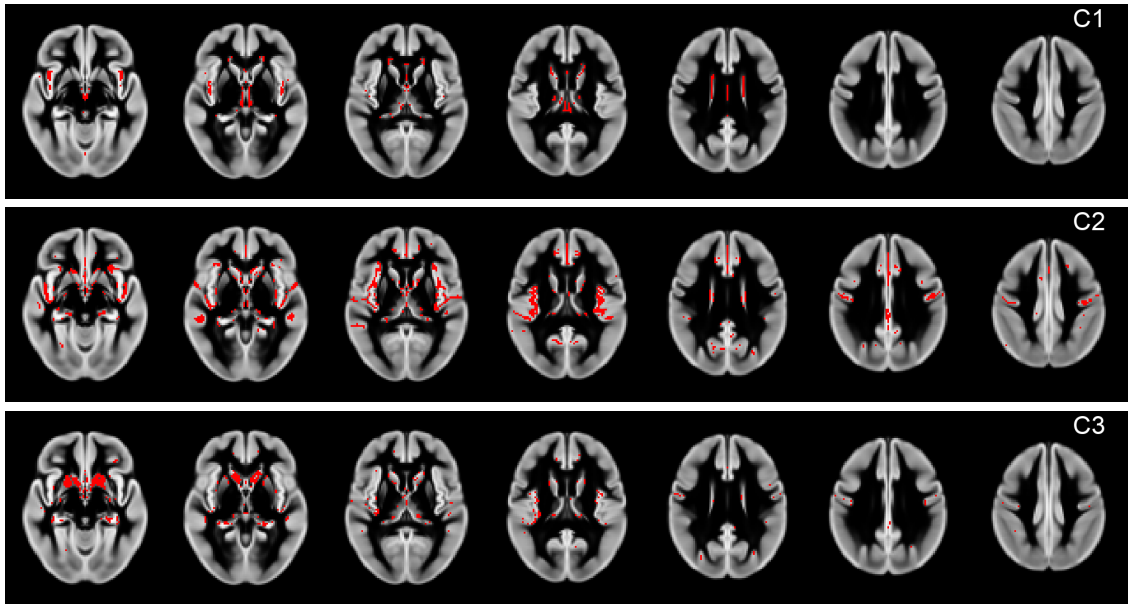


Figure C.1: The unimodal binary maps associated with the estimated VBM components.

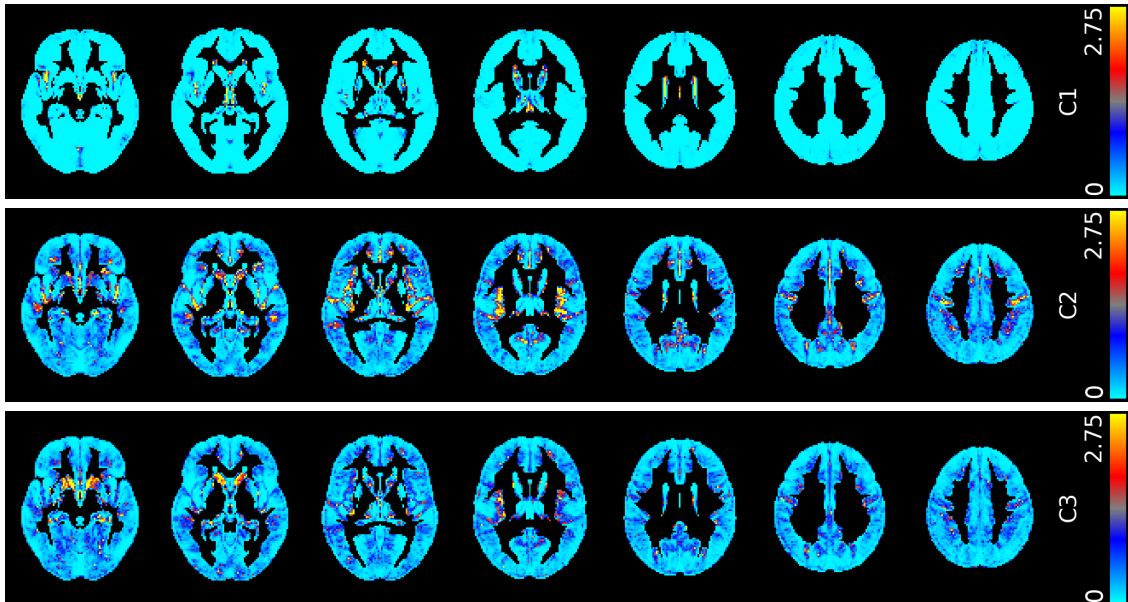


Figure C.2: The unimodal inverse distance spatial maps associated with the estimated VBM components.

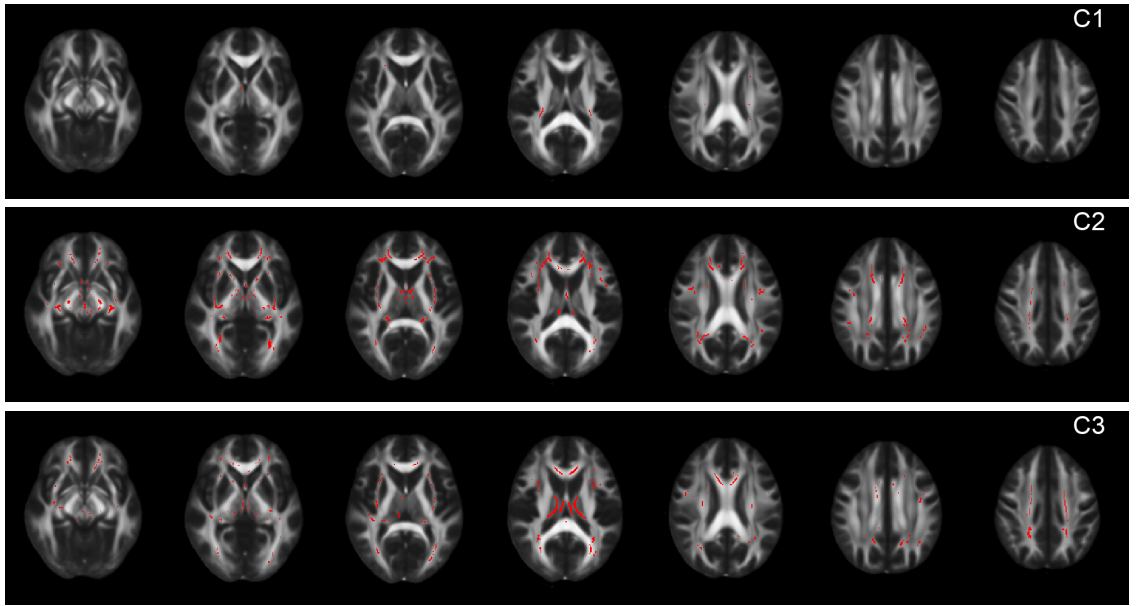


Figure C.3: The unimodal binary maps associated with the estimated TBSS FA components.

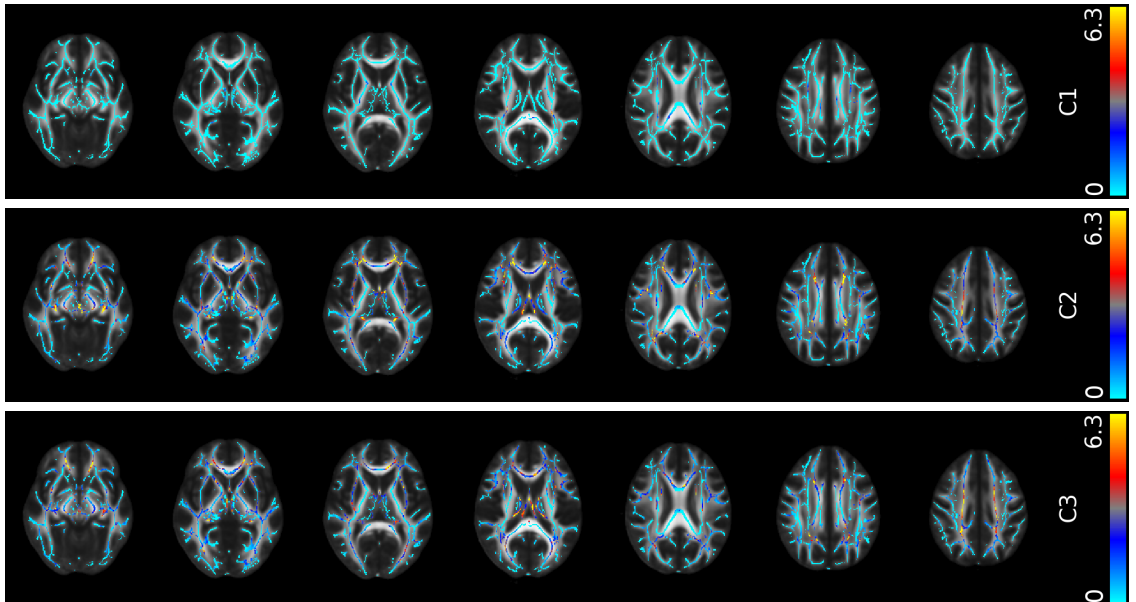


Figure C.4: The unimodal inverse distance spatial maps associated with the estimated TBSS FA components.

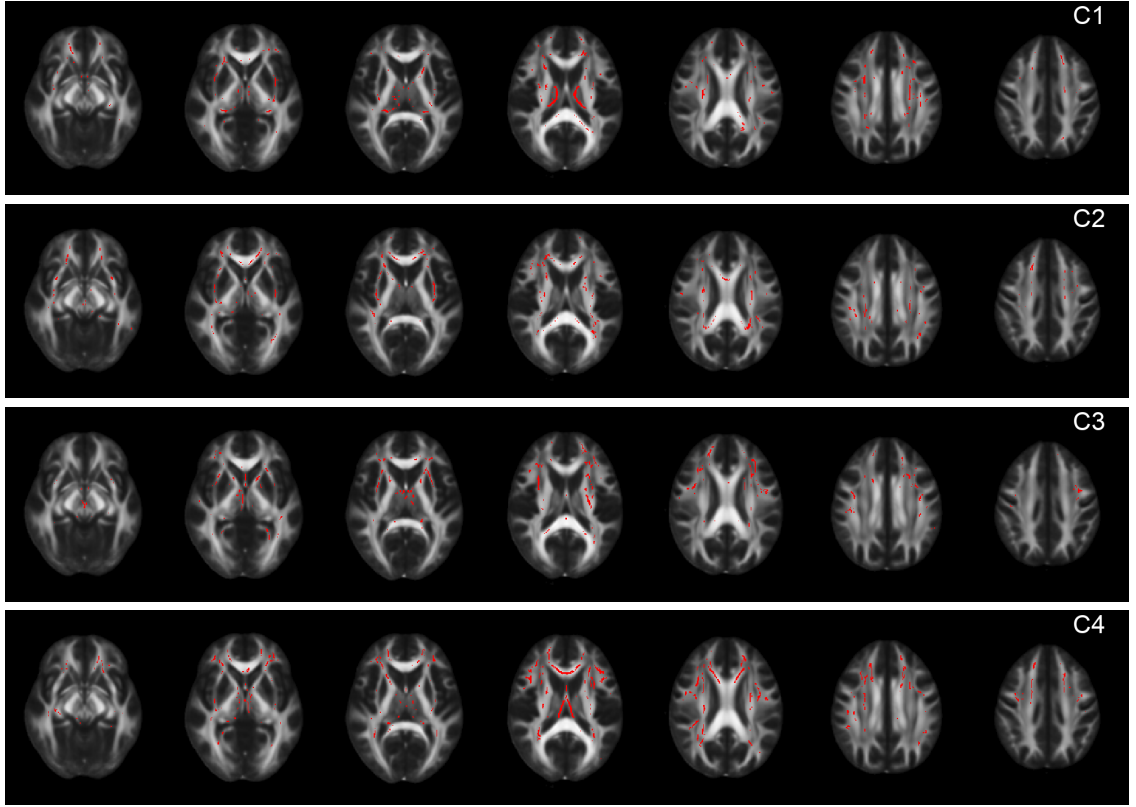


Figure C.5: The unimodal binary maps associated with the estimated TBSS MD components.

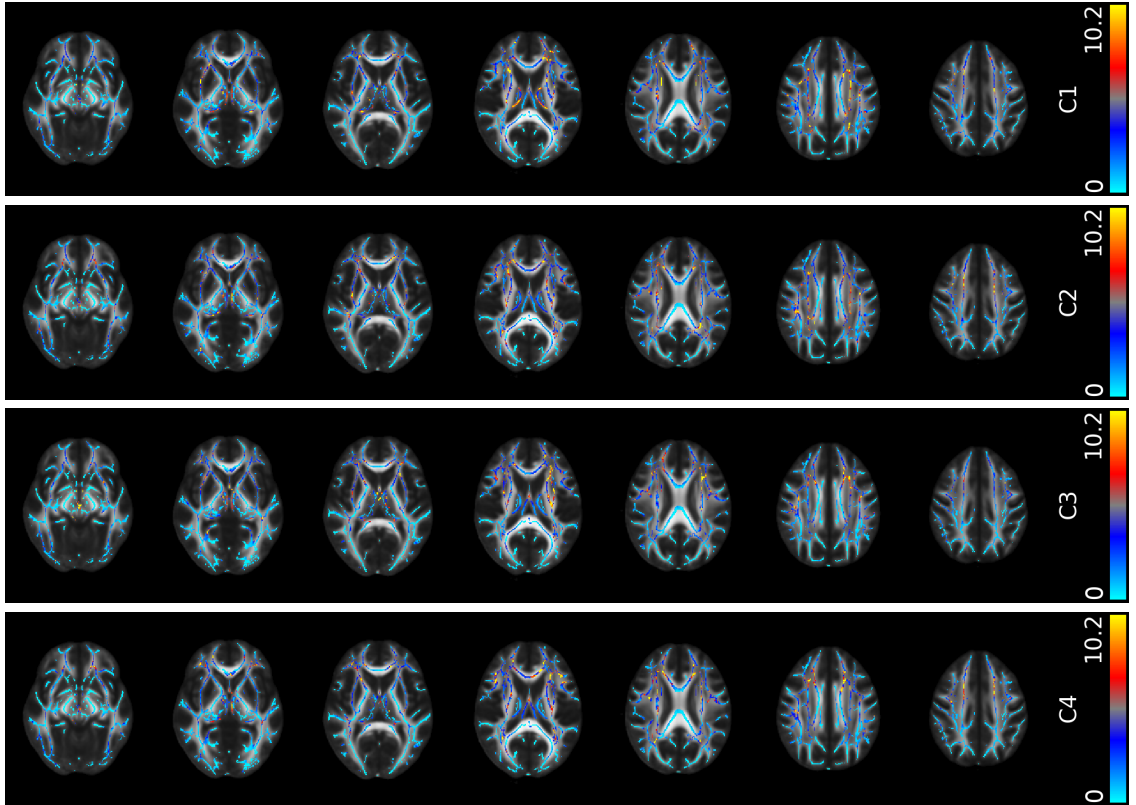


Figure C.6: The unimodal inverse distance spatial maps associated with the estimated TBSS MD components.

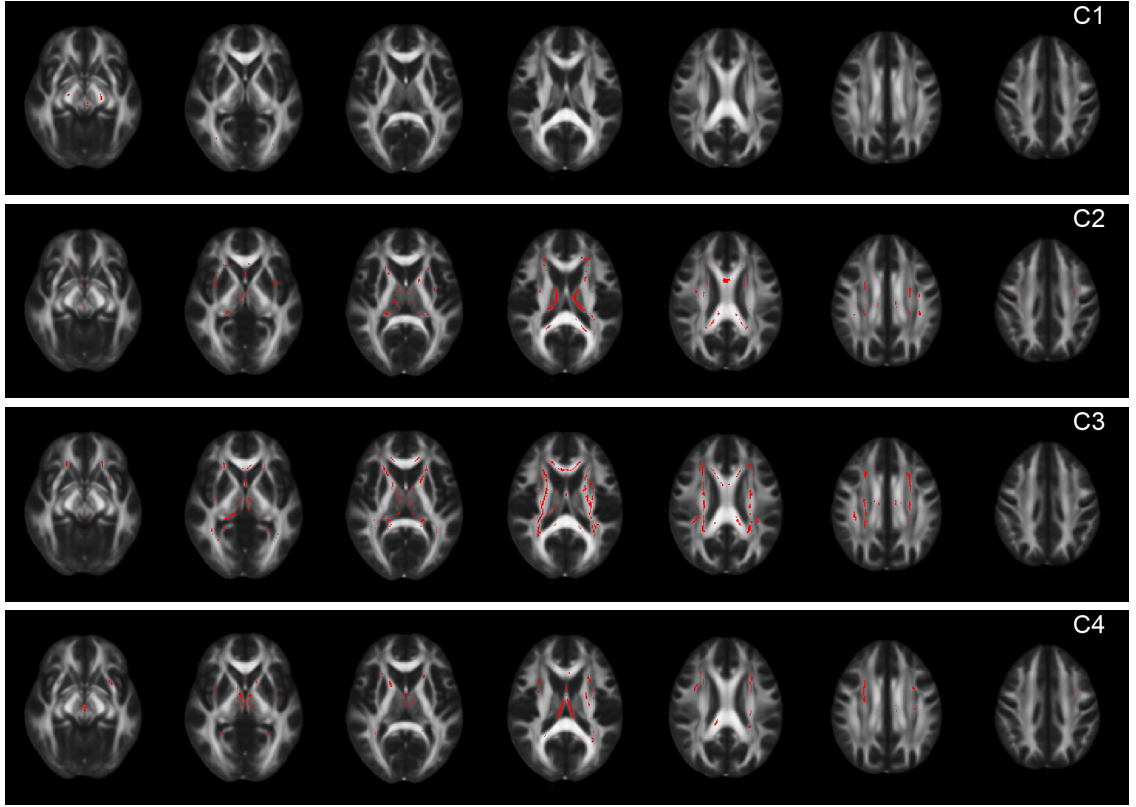


Figure C.7: The unimodal binary maps associated with the estimated TBSS L1 components.

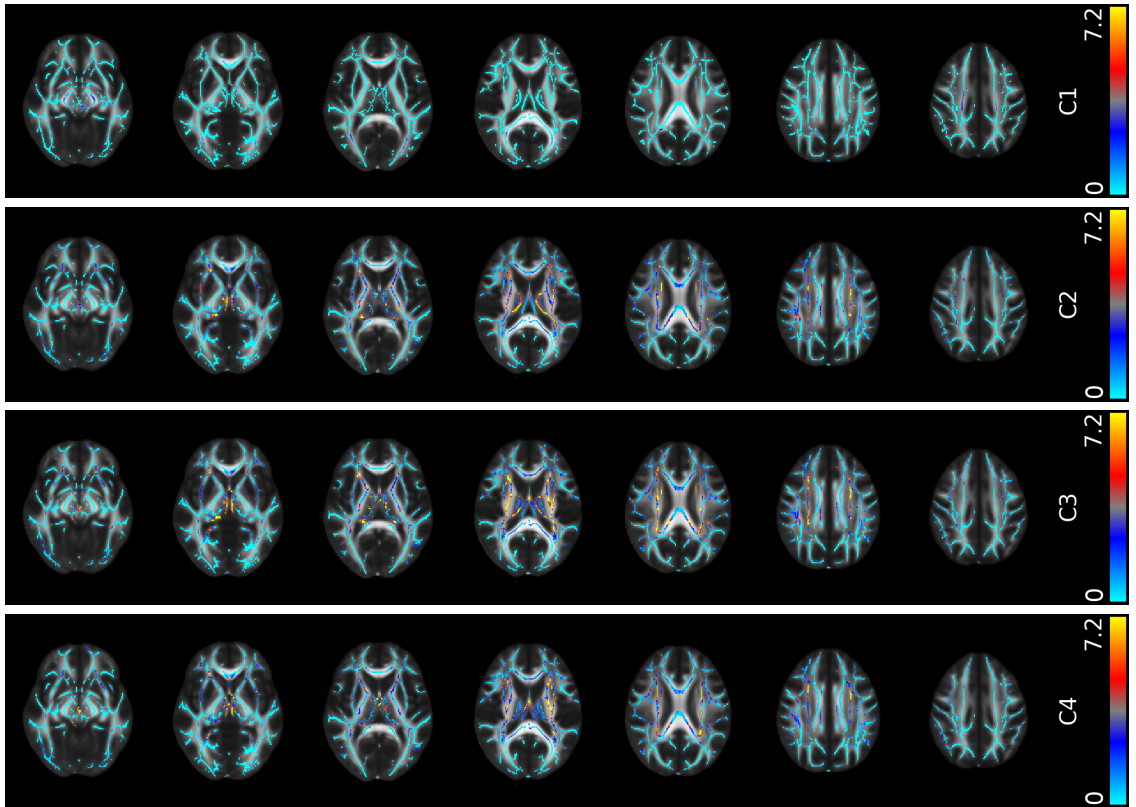


Figure C.8: The unimodal inverse distance spatial maps associated with the estimated TBSS L1 components.

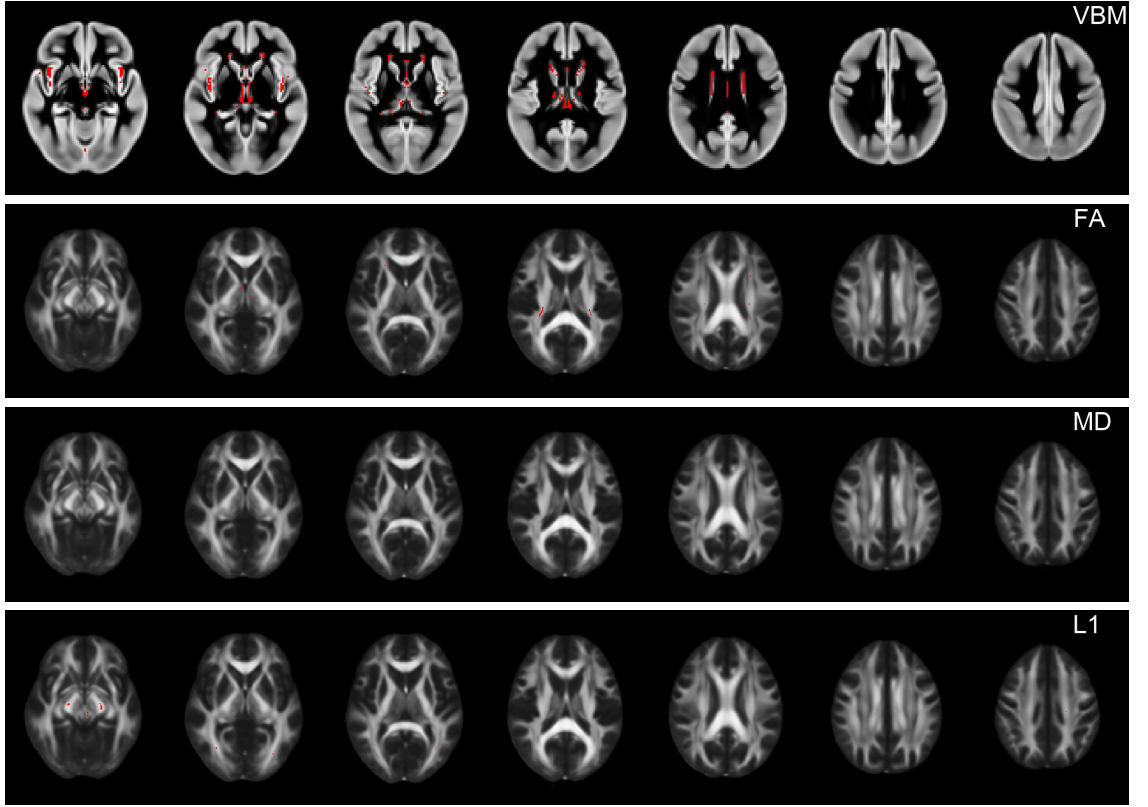


Figure C.9: The multimodal binary maps corresponding to component one from our method.

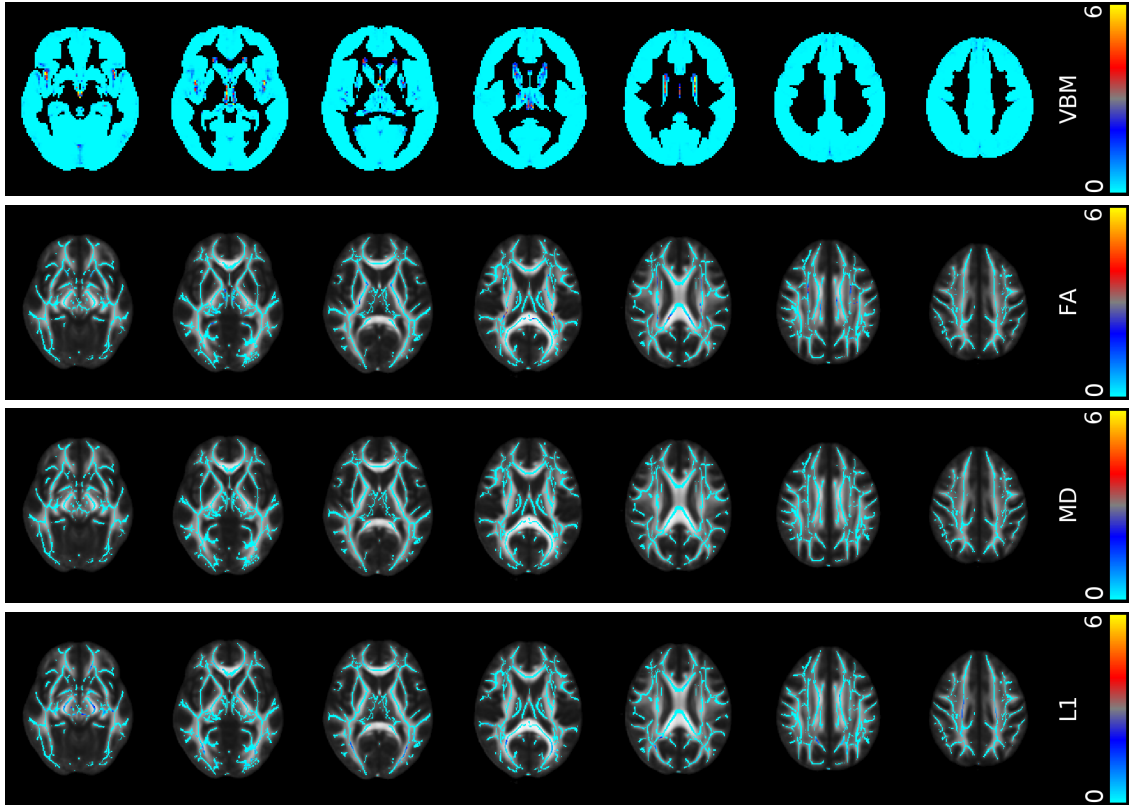


Figure C.10: The multimodal inverse distance maps corresponding to component one from our method.

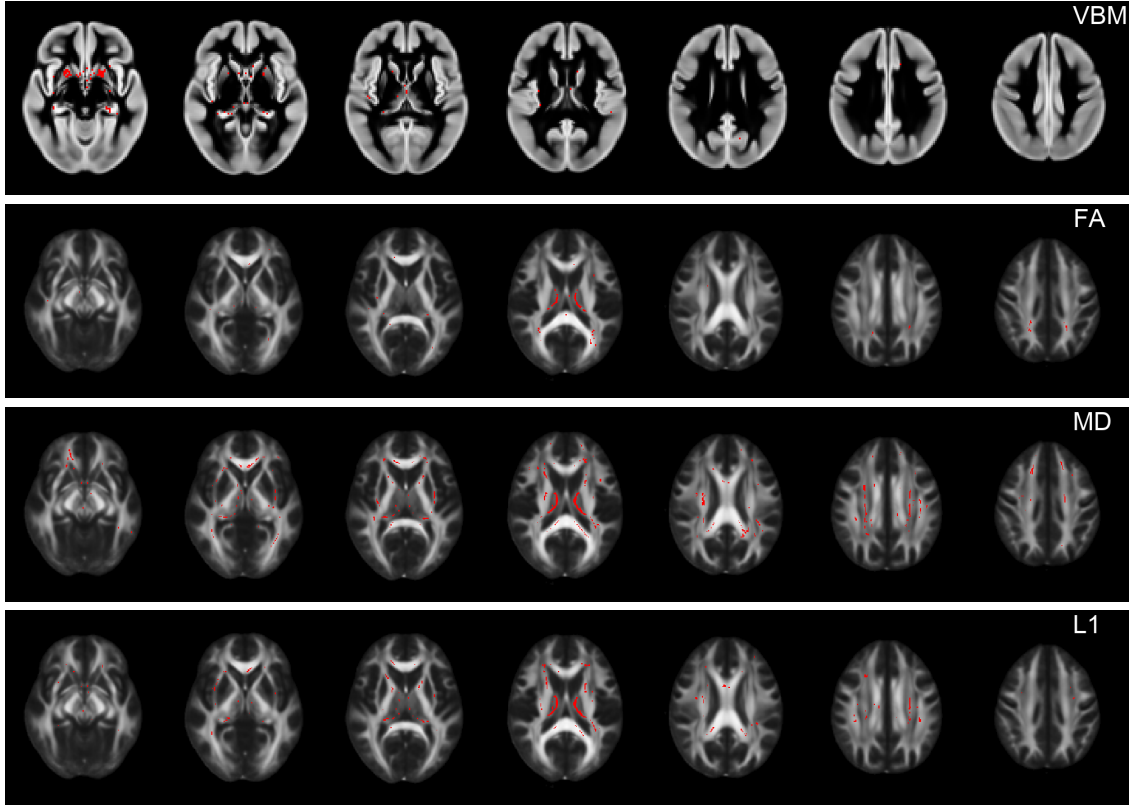


Figure C.11: The multimodal binary maps corresponding to component two from our method.

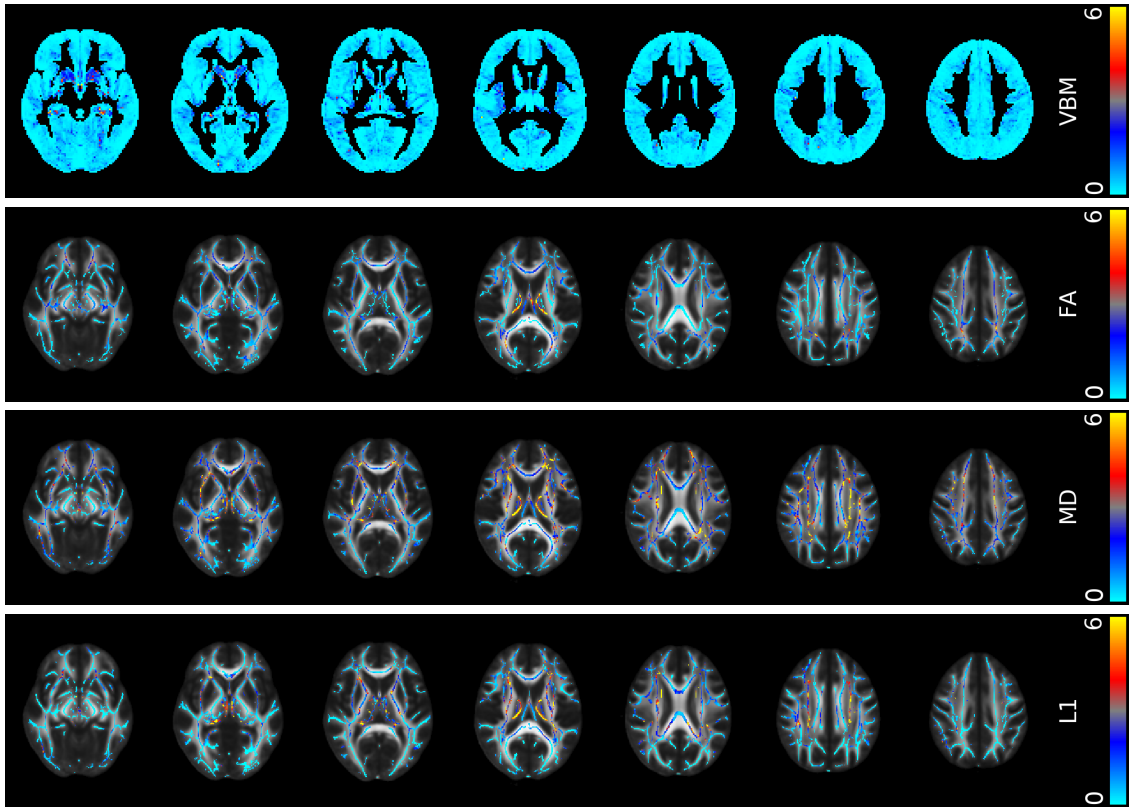


Figure C.12: The multimodal inverse distance maps corresponding to component two from our method.

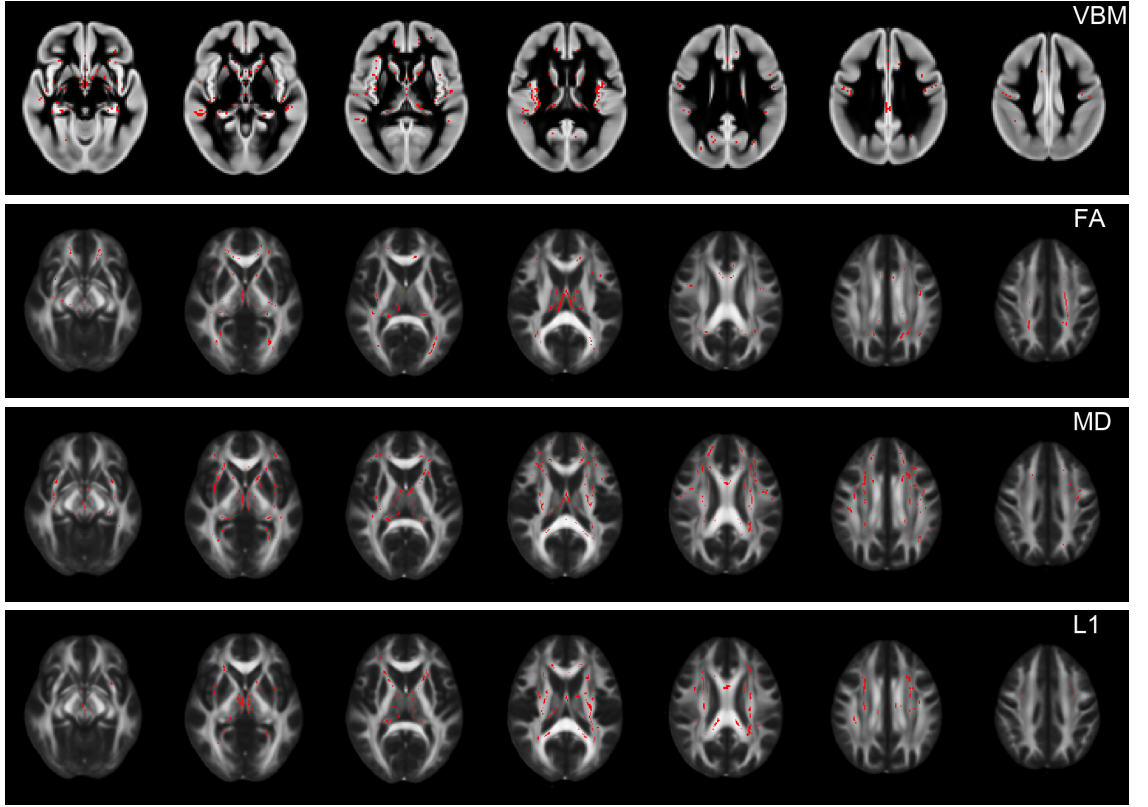


Figure C.13: The multimodal binary maps corresponding to component three from our method.

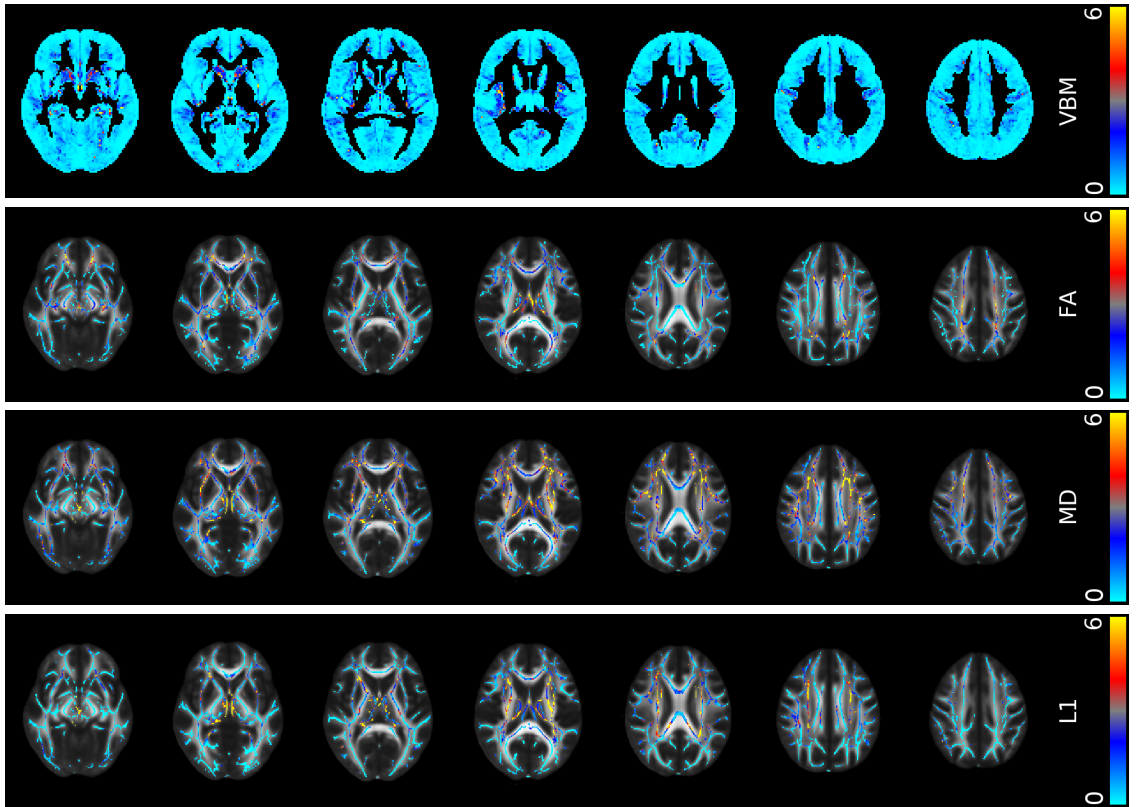


Figure C.14: The multimodal inverse distance maps corresponding to component three from our method.

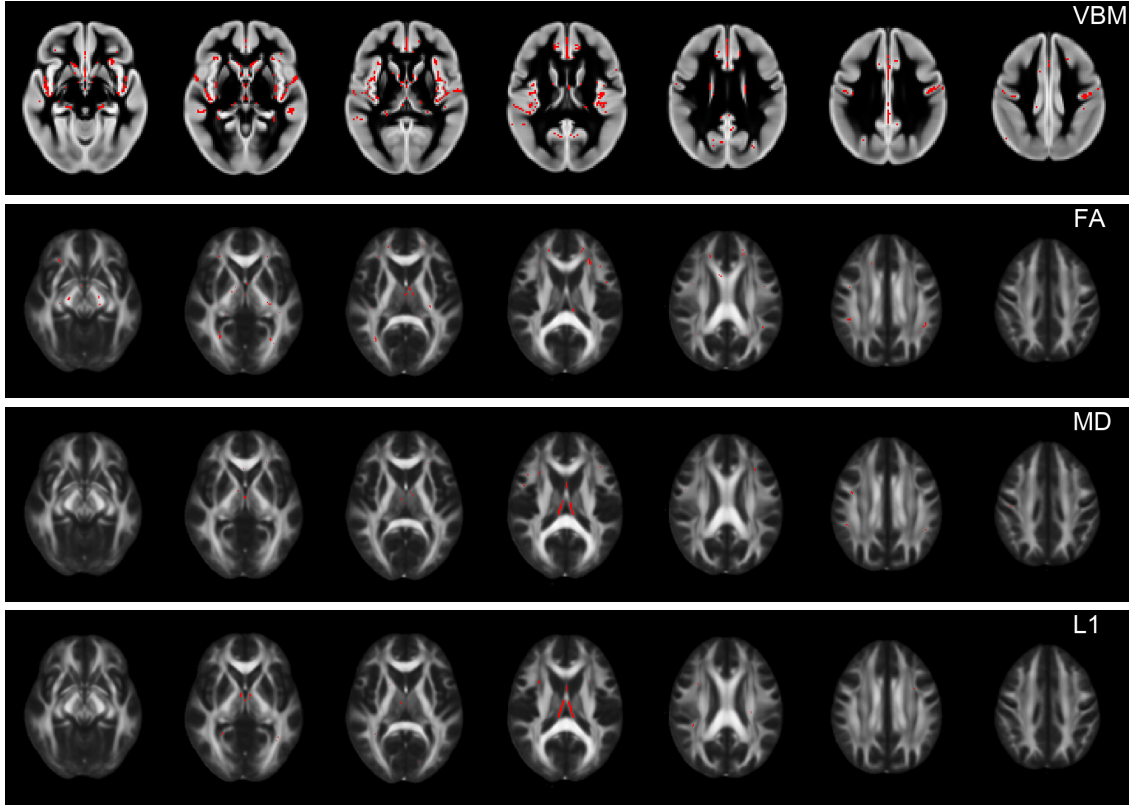


Figure C.15: The multimodal binary maps corresponding to component four from our method.

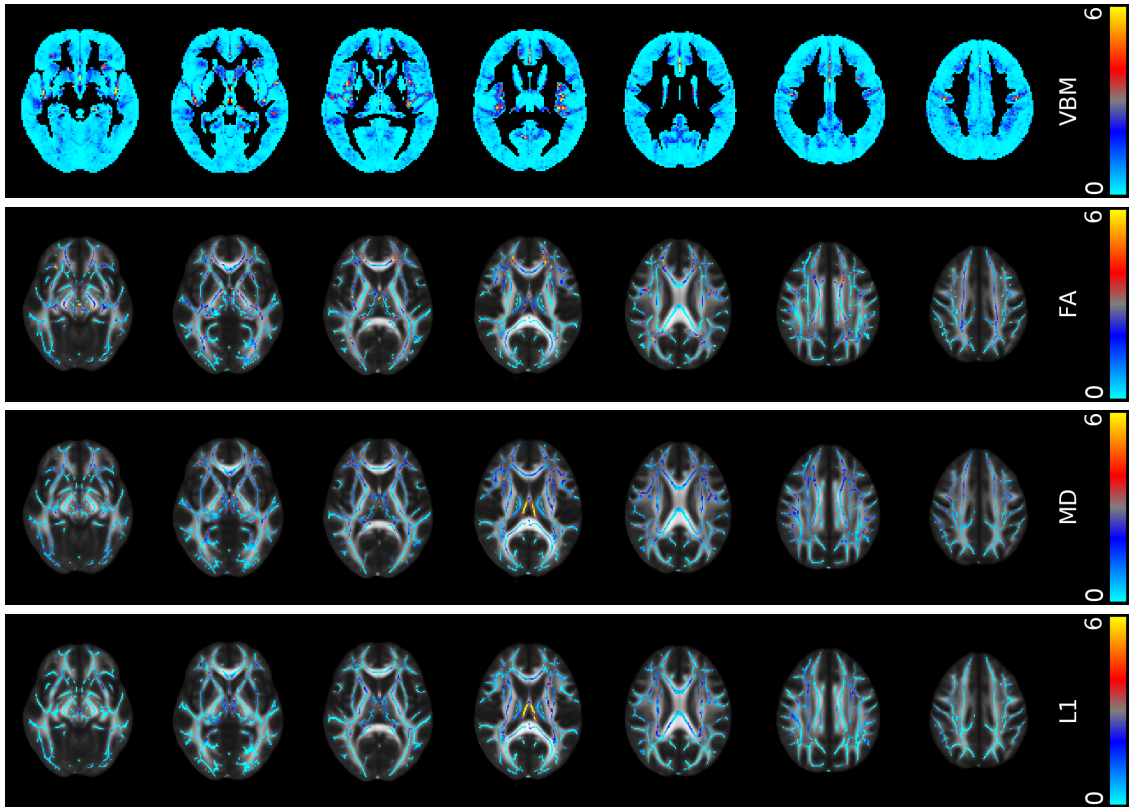


Figure C.16: The multimodal inverse distance maps corresponding to component four from our method.

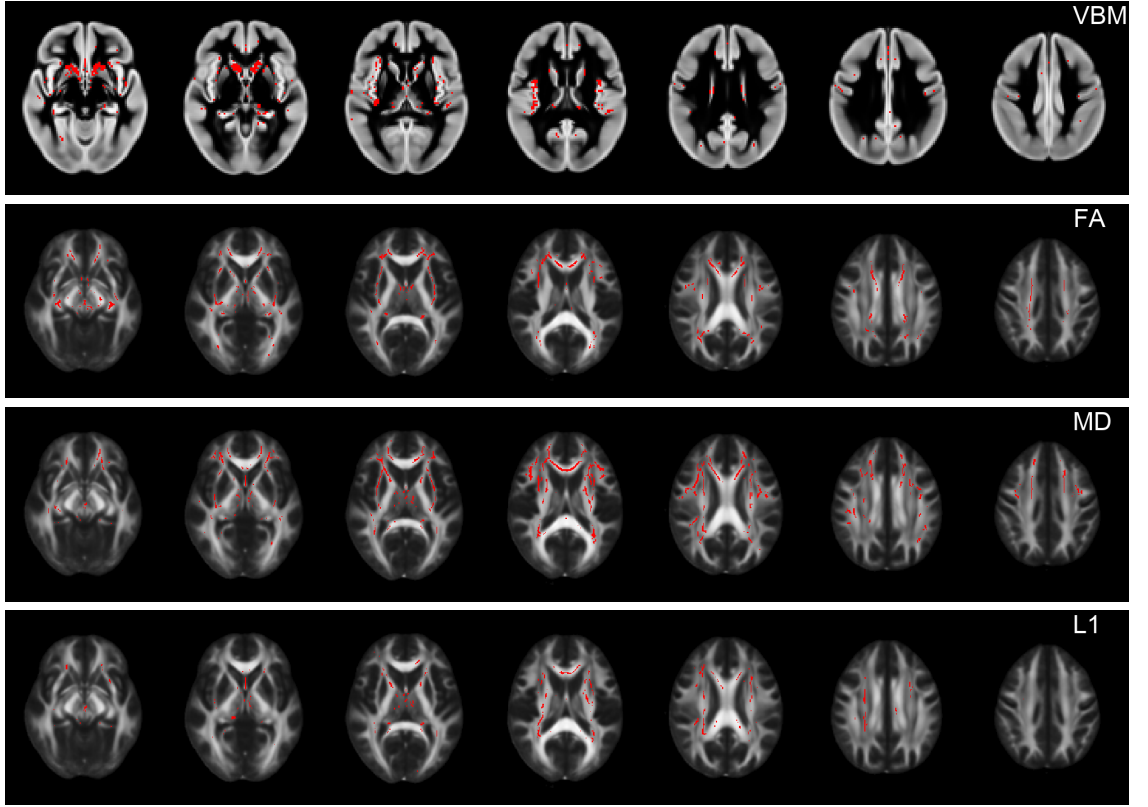


Figure C.17: The multimodal binary maps corresponding to component five from our method.

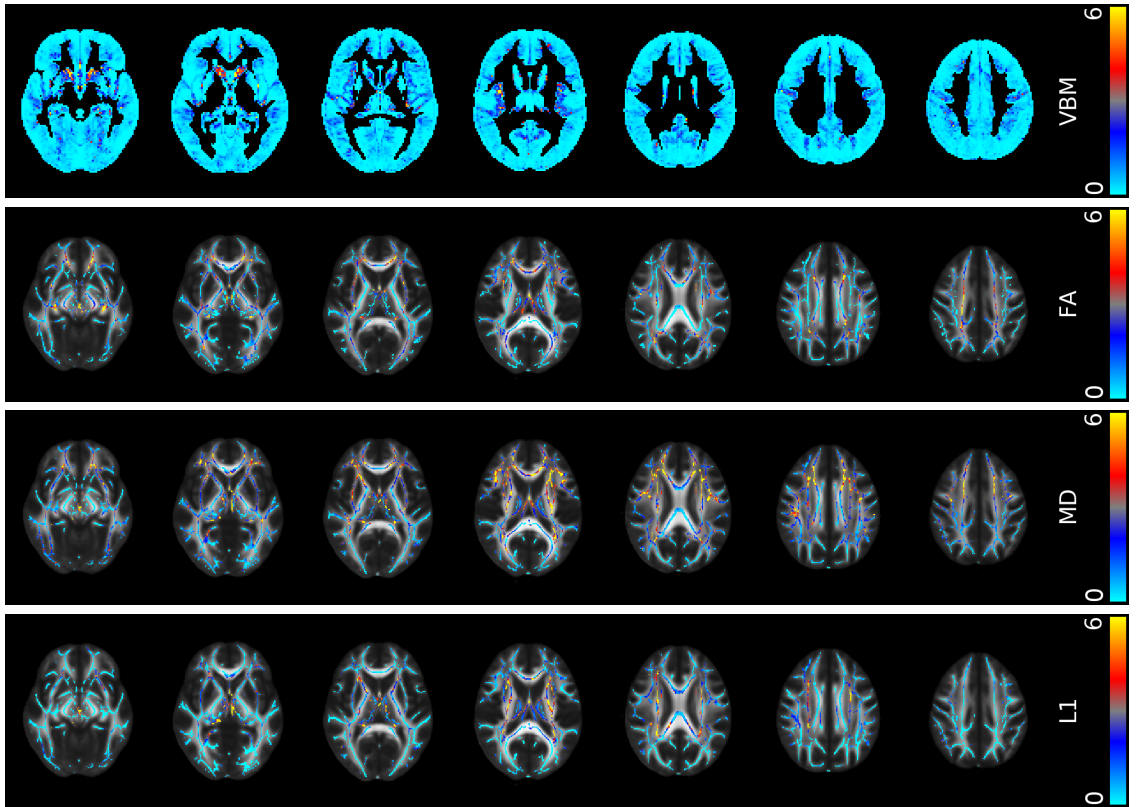


Figure C.18: The multimodal inverse distance maps corresponding to component five from our method.

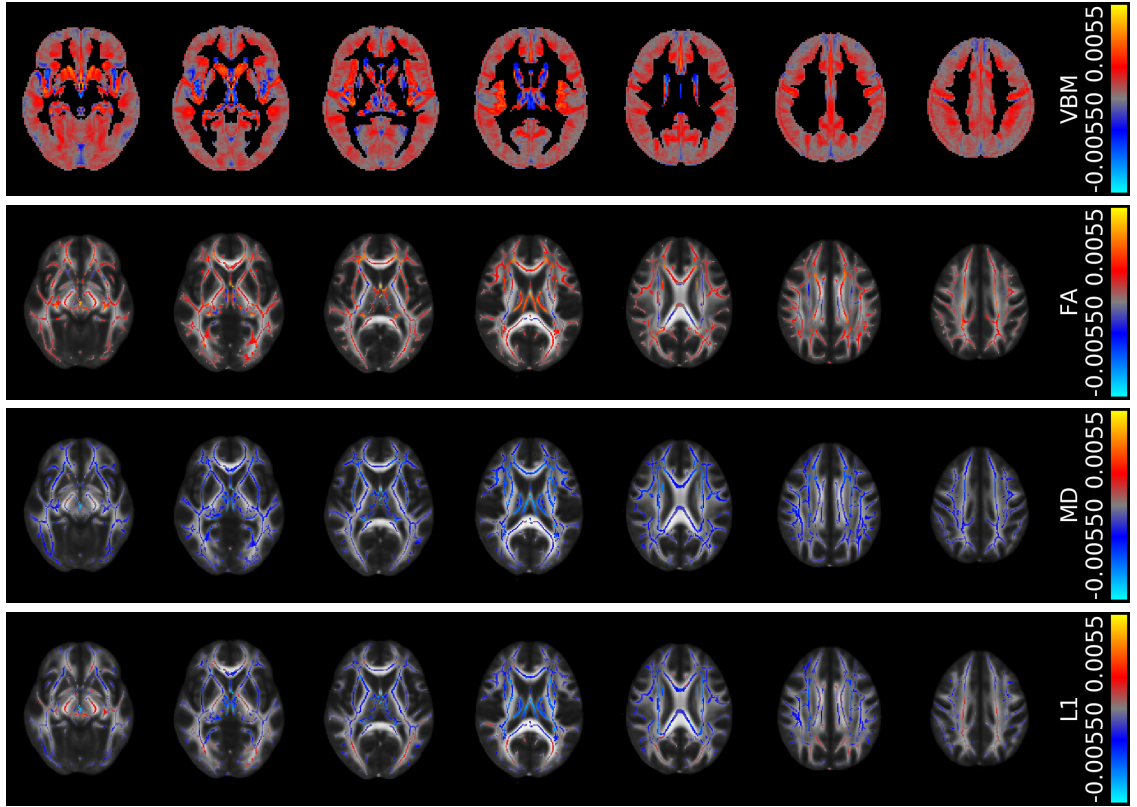


Figure C.19: The spatial maps for informed multimodal PLS component one.

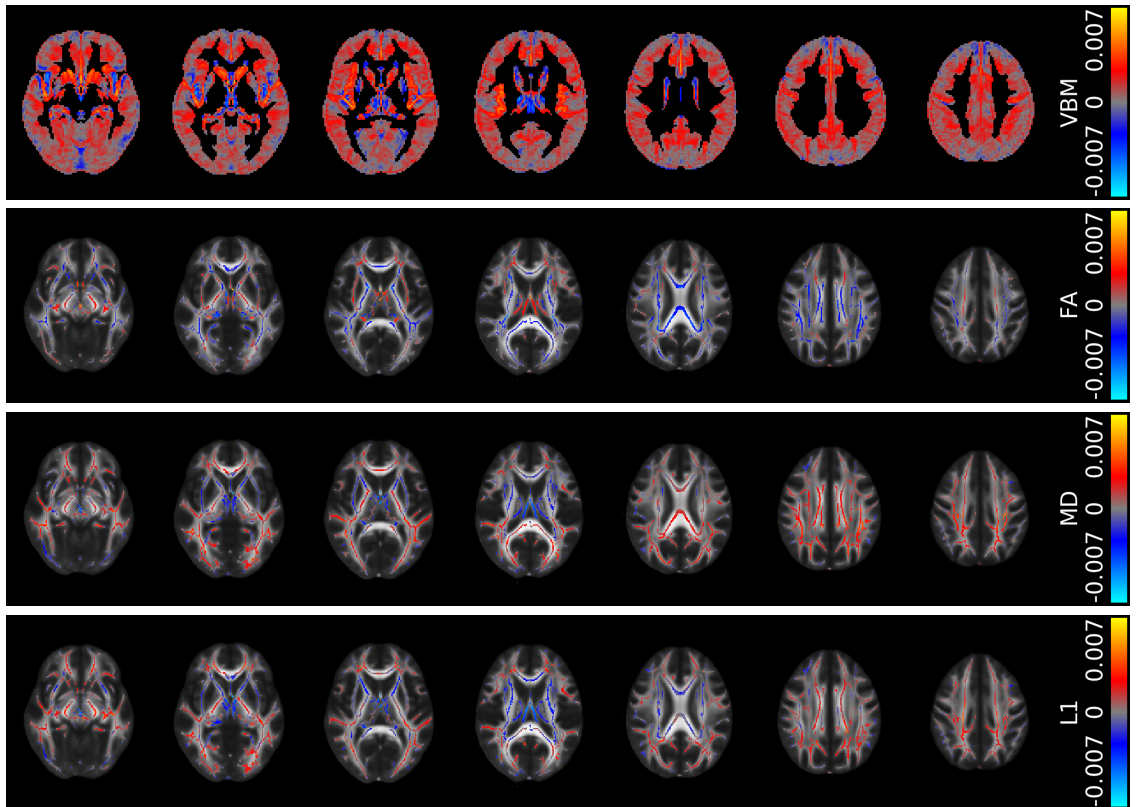


Figure C.20: The spatial maps for informed multimodal PLS component two.

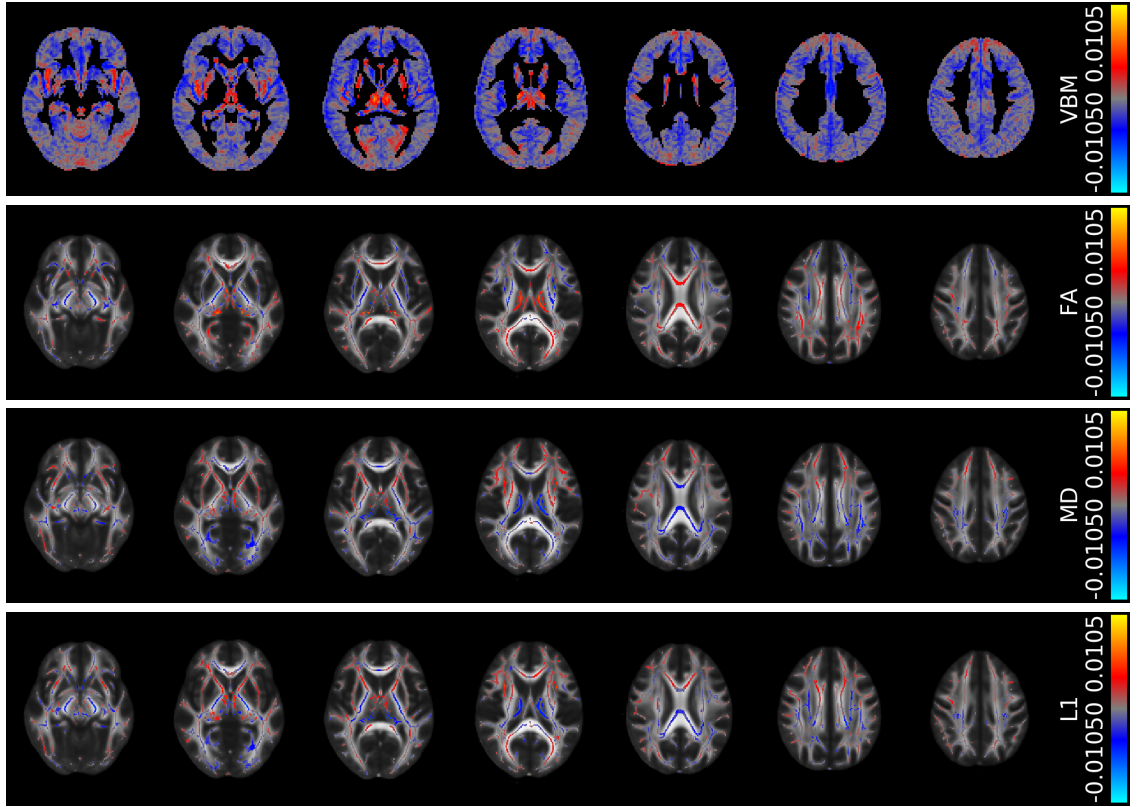


Figure C.21: The spatial maps for informed multimodal PLS component three.

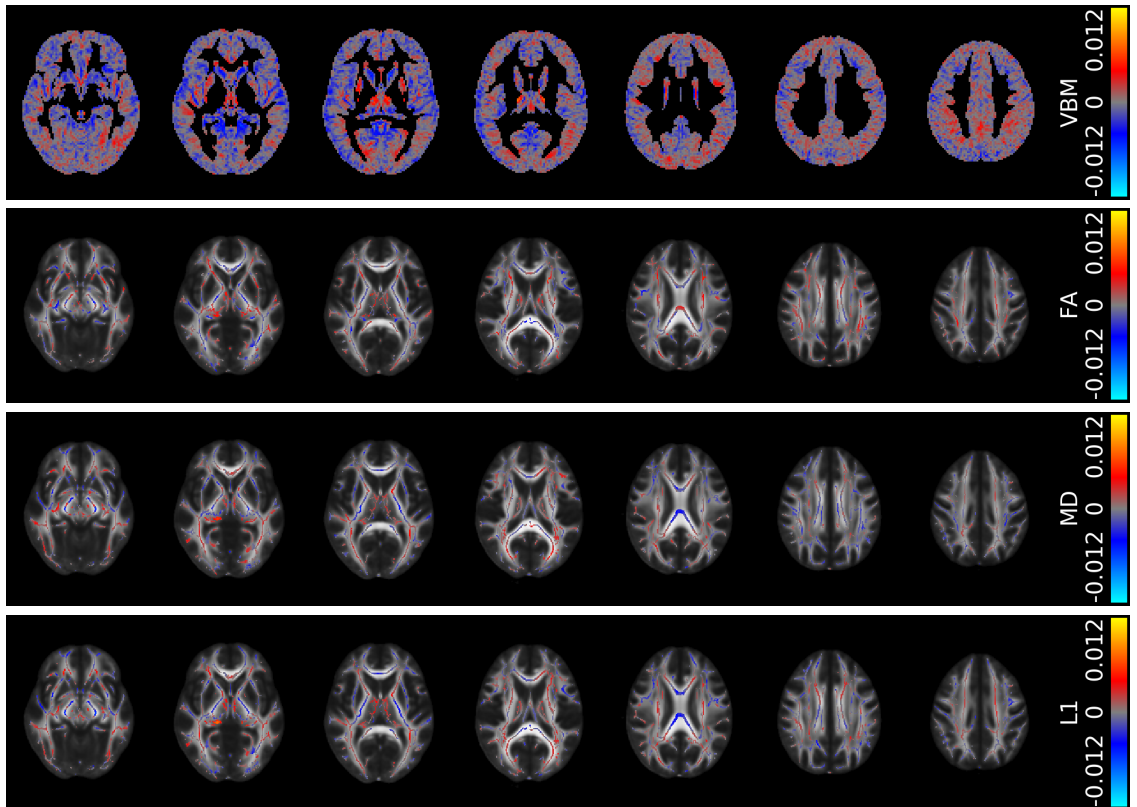


Figure C.22: The spatial maps for informed multimodal PLS component four.

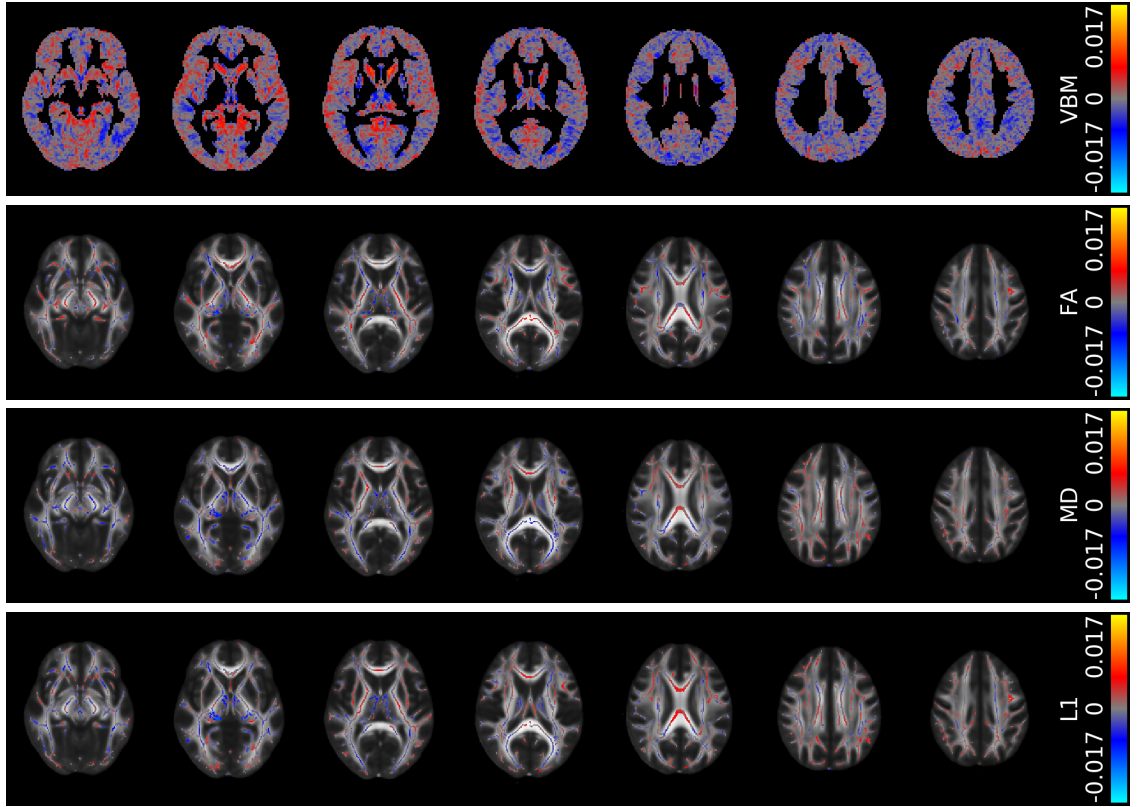


Figure C.23: The spatial maps for informed multimodal PLS component five.

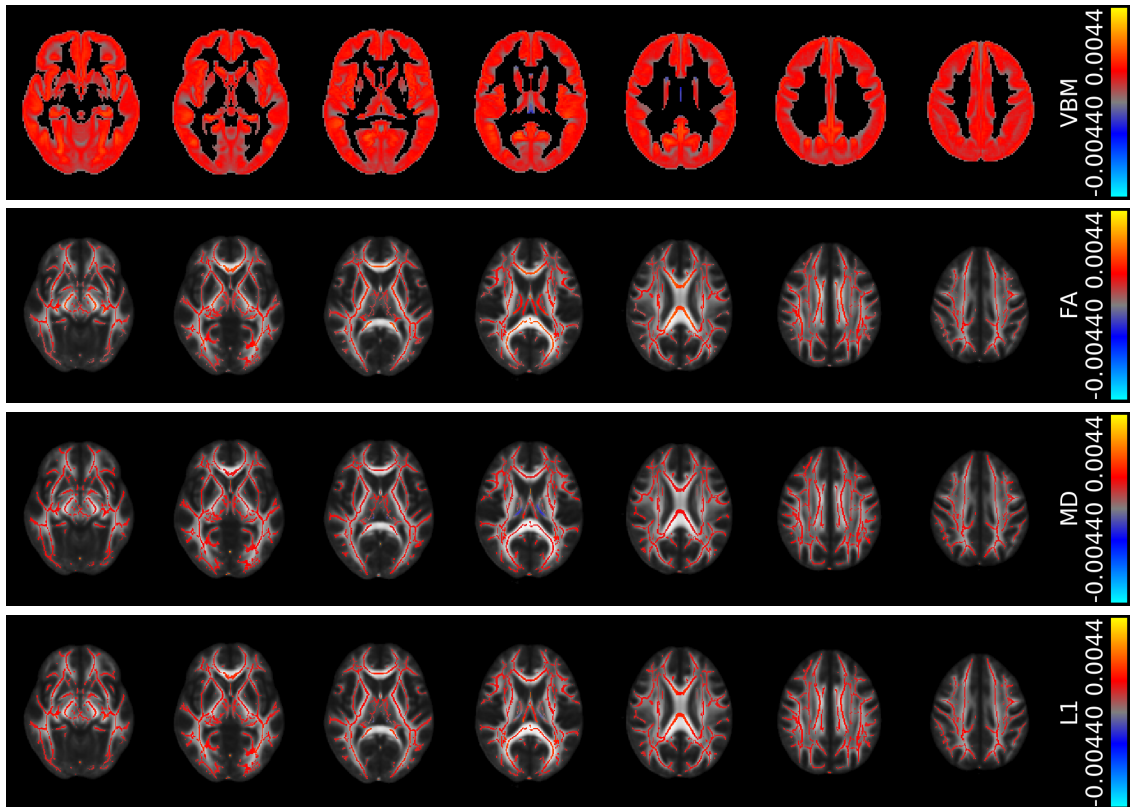


Figure C.24: The spatial maps for joint ICA component one.

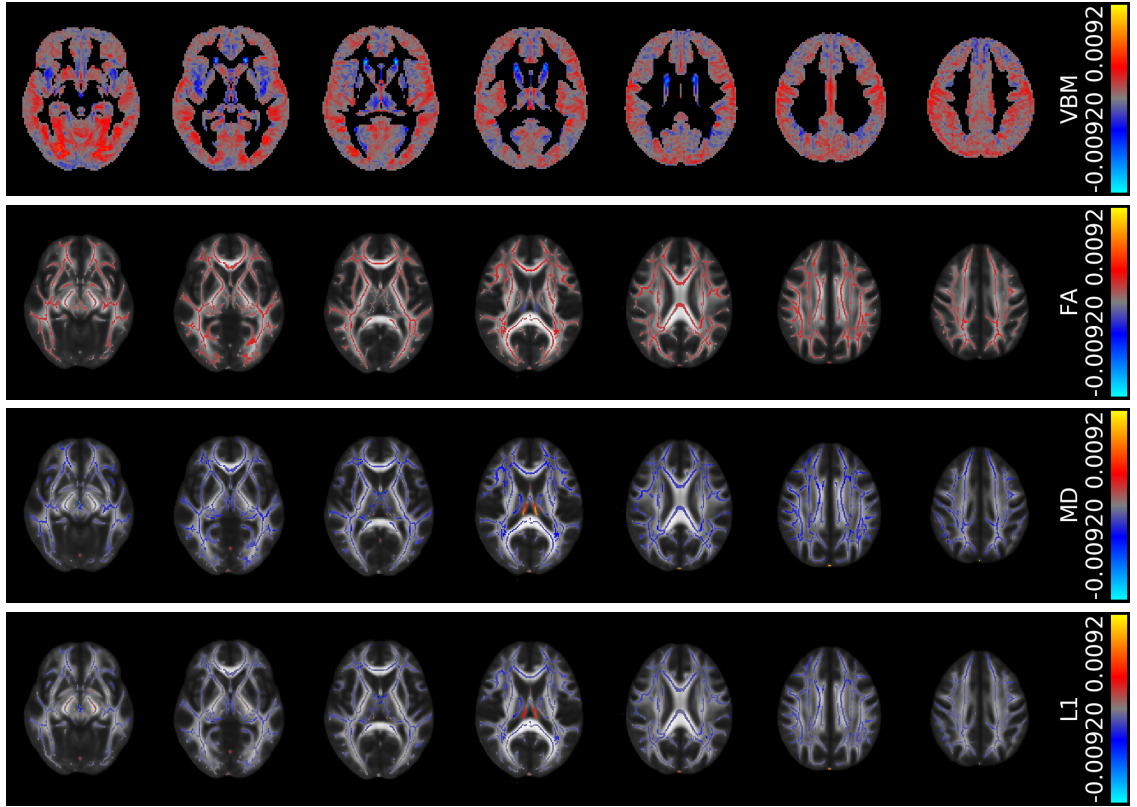


Figure C.25: The spatial maps for joint ICA component two.

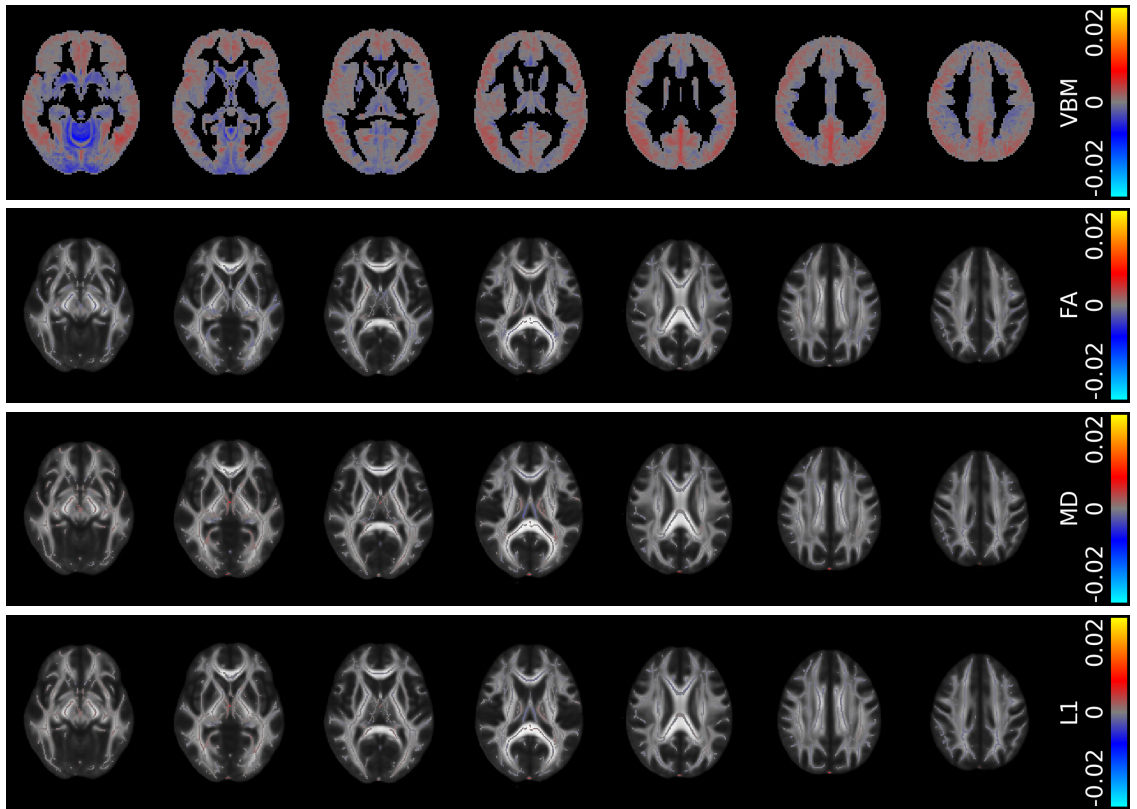


Figure C.26: The spatial maps for joint ICA component three.

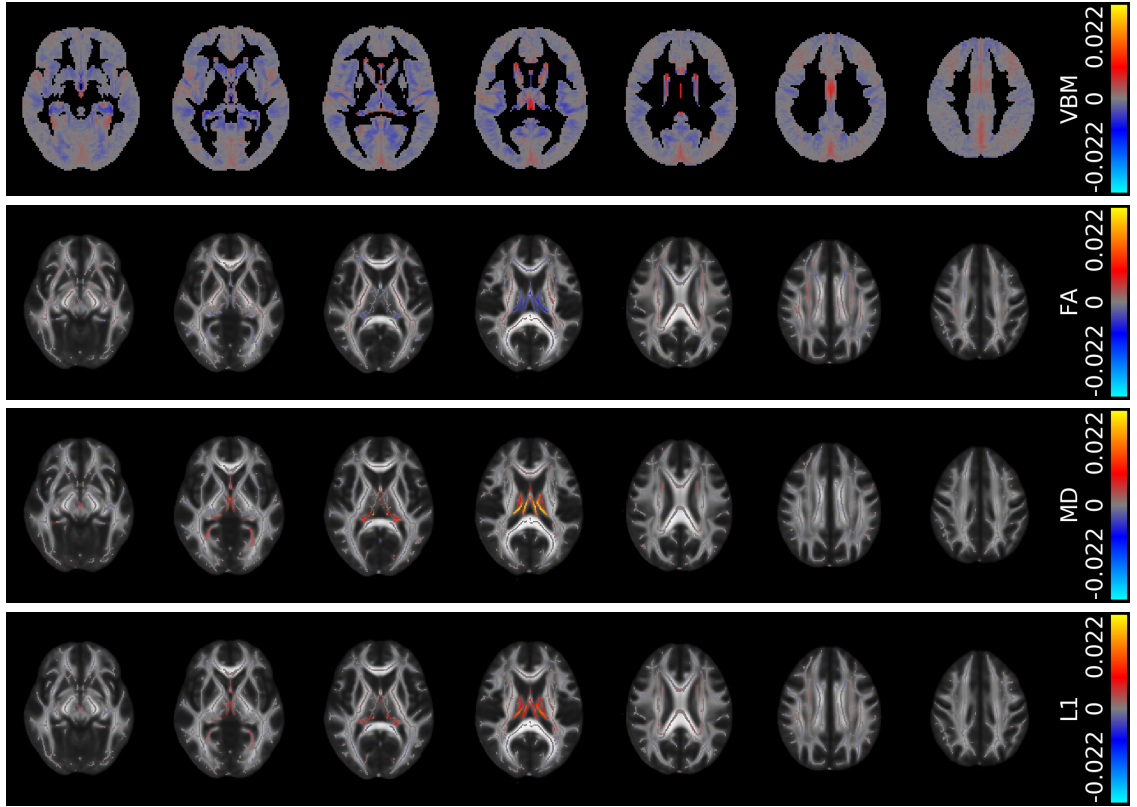


Figure C.27: The spatial maps for joint ICA component four.

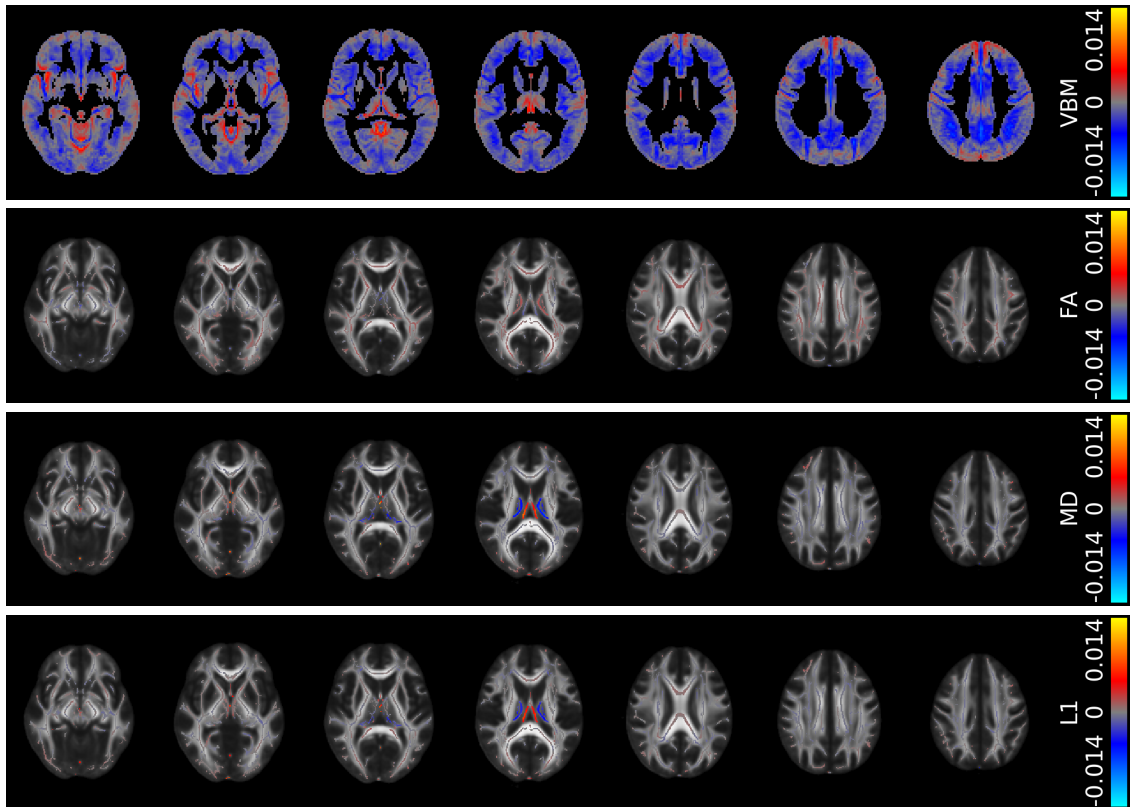


Figure C.28: The spatial maps for joint ICA component five.

Bibliography

- [1] V. Rachakonda, T. H. Pan, and W. D. Le. “Biomarkers of neurodegenerative disorders: How good are they?” In: *Cell Research* 14.5 (2004), pp. 349–358.
- [2] S. Karantzoulis and J. E. Galvin. “Distinguishing Alzheimer’s disease from other major forms of dementia”. In: *Expert Review of Neurotherapeutics* 11.11 (2011), pp. 1579–1591.
- [3] J.-C. Lambert, N. Coyle, and C. Lendon. “The allelic modulation of apolipoprotein E expression by oestrogen: potential relevance for Alzheimer’s disease.” In: *Journal of medical genetics* 41.2 (2004), pp. 104–12.
- [4] S. Lesage and A. Brice. “Parkinson’s disease: from monogenic forms to genetic susceptibility factors”. In: *Human Molecular Genetics* 18.R1 (2009), R48–R59.
- [5] T. Sunderland, G. Linker, N. Mirza, K. T. Putnam, D. L. Friedman, L. H. Kimmel, J. Bergeson, G. J. Manetti, M. Zimmermann, B. Tang, J. J. Bartko, and R. M. Cohen. “Decreased β -Amyloid_{1–42} and Increased Tau Levels in Cerebrospinal Fluid of Patients With Alzheimer Disease”. In: *JAMA* 289.16 (2003), pp. 2094–2103.
- [6] T. Wyss-Coray. “Ageing, neurodegeneration and brain rejuvenation”. In: *Nature* 539.7628 (2016), pp. 180–186.
- [7] A. Burns and S. Iliffe. “Alzheimer’s disease.” In: *BMJ (Clinical research ed.)* 338 (2009), b158.
- [8] N. Okamura, R. Harada, S. Furumoto, H. Arai, K. Yanai, and Y. Kudo. “Tau PET imaging in Alzheimer’s disease”. In: *Current neurology and neuroscience reports* 14.11 (2014), p. 500.
- [9] S. T. Schwarz, M. Afzal, P. S. Morgan, N. Bajaj, P. A. Gowland, and D. P. Auer. “The ‘swallow tail’ appearance of the healthy nigrosome - a new accurate test of Parkinson’s disease: a case-control and retrospective cross-sectional MRI study at 3T.” In: *PloS one* 9.4 (2014), e93814.

- [10] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, P. M. Thompson, and P. M. Thompson. “The clinical use of structural MRI in Alzheimer disease.” In: *Nature reviews. Neurology* 6.2 (2010), pp. 67–77.
- [11] A. M. Fjell and K. B. Walhovd. “Structural brain changes in aging: courses, causes and cognitive consequences.” In: *Reviews in the neurosciences* 21.3 (2010), pp. 187–221.
- [12] M. Jenkinson and M. Chappell. *Introduction to Neuroimaging Analysis*. New York, USA: Oxford University Press, 2018.
- [13] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. Henson, K. J. Friston, and R. S. Frackowiak. “A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains”. In: *NeuroImage* 14.1 (2001), pp. 21–36.
- [14] G. F. Busatto, B. S. Diniz, and M. V. Zanetti. “Voxel-based morphometry in Alzheimer’s disease”. In: *Expert Review of Neurotherapeutics* 8.11 (2008), pp. 1691–1702.
- [15] A. Mechelli, C. Price, K. Friston, and J. Ashburner. “Voxel-Based Morphometry of the Human Brain: Methods and Applications”. In: *Current Medical Imaging Reviews* 1.2 (2005), pp. 105–113.
- [16] A. L. Alexander, J. E. Lee, M. Lazar, and A. S. Field. “Diffusion tensor imaging of the brain.” In: *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics* 4.3 (2007), pp. 316–29.
- [17] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, and T. E. Behrens. “Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data”. In: *NeuroImage* 31.4 (2006), pp. 1487–1505.
- [18] T. M. Nir, N. Jahanshad, J. E. Villalon-Reina, A. W. Toga, C. R. Jack, M. W. Weiner, and P. M. Thompson. “Effectiveness of regional DTI measures in distinguishing Alzheimer’s disease, MCI, and normal aging”. In: *NeuroImage: Clinical* 3 (2013), pp. 180–195.
- [19] J. Acosta-Cabronero and P. J. Nestor. “Diffusion tensor imaging in Alzheimer’s disease: insights into the limbic-diencephalic network and methodological considerations.” In: *Frontiers in aging neuroscience* 6 (2014), p. 266.
- [20] S. R. Cox, S. J. Ritchie, E. M. Tucker-Drob, D. C. Liewald, S. P. Hagenaars, G. Davies, J. M. Wardlaw, C. R. Gale, M. E. Bastin, and I. J. Deary. “Ageing and brain white matter structure in 3,513 UK Biobank participants”. In: *Nature Communications* 7.1 (2016), p. 13629.

- [21] P. M. Matthews and P. Jezzard. “Functional magnetic resonance imaging”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 75.1 (2004), pp. 6–12.
- [22] C. Rosazza and L. Minati. “Resting-state brain networks: literature review and clinical applications”. In: *Neurological Sciences* 32.5 (2011), pp. 773–785.
- [23] A. Mahmoudi, S. Takerkart, F. Regragui, D. Boussaoud, and A. Brovelli. “Multivoxel pattern analysis for fMRI data: a review.” In: *Computational and mathematical methods in medicine* 2012 (2012), p. 961257.
- [24] C. Beckmann and S. Smith. “Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging”. In: *IEEE Transactions on Medical Imaging* 23 (2004), pp. 137–152.
- [25] M. J. Mckeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. “Analysis of fMRI data by blind separation into independent spatial components”. In: *Human Brain Mapping* 6.3 (1998), pp. 160–188.
- [26] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon. “Default-mode network activity distinguishes Alzheimer’s disease from healthy aging: Evidence from functional MRI”. In: *Proceedings of the National Academy of Sciences* 101.13 (2004), pp. 4637–4642.
- [27] C. Beckmann, C. Mackay, N. Filippini, and S. Smith. “Group comparison of resting-state fMRI data using multi-subject ICA and dual regression”. In: *NeuroImage* 47 (2009), S148.
- [28] A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith, and T. E. Nichols. “Permutation inference for the general linear model”. In: *NeuroImage* 92 (2014), pp. 381–397.
- [29] S. M. Smith and T. E. Nichols. “Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference”. In: *NeuroImage* 44.1 (2009), pp. 83–98.
- [30] A. Tessitore, F. Esposito, C. Vitale, G. Santangelo, M. Amboni, A. Russo, D. Corbo, G. Cirillo, P. Barone, and G. Tedeschi. “Default-mode network connectivity in cognitively unimpaired patients with Parkinson disease.” In: *Neurology* 79.23 (2012), pp. 2226–32.
- [31] K. Szewczyk-Krolikowski, R. A. Menke, M. Rolinski, E. Duff, G. Salimi-Khorshidi, N. Filippini, G. Zamboni, M. T. Hu, and C. E. Mackay. “Functional connectivity in the basal ganglia network differentiates PD patients from controls”. In: *Neurology* 83.3 (2014), pp. 208–214.
- [32] D. M. Cole, C. F. Beckmann, C. J. Long, P. M. Matthews, M. J. Durcan, and J. D. Beaver. “Nicotine replacement in abstinent smokers improves cognitive withdrawal symptoms with modulation of resting brain network dynamics”. In: *NeuroImage* 52.2 (2010), pp. 590–599.

- [33] L. Griffanti, M. Rolinski, K. Szewczyk-Krolikowski, R. A. Menke, N. Filippini, G. Zamboni, M. Jenkinson, M. T. Hu, and C. E. Mackay. “Challenges in the reproducibility of clinical studies with resting state fMRI: An example in early Parkinson’s disease”. In: *NeuroImage* 124 (2016), pp. 704–713.
- [34] J. Bijsterbosch, S. Smith, and C. Beckmann. *An Introduction to Resting State fMRI Functional Connectivity*. New York, USA: Oxford University Press, 2017.
- [35] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li. “Classification of Alzheimer Disease, Mild Cognitive Impairment, and Normal Cognitive Status with Large-Scale Network Analysis Based on Resting-State Functional MR Imaging”. In: *Radiology* 259.1 (2011), pp. 213–221.
- [36] J. W. Bohland, H. Bokil, C. B. Allen, and P. P. Mitra. “The Brain Atlas Concordance Problem: Quantitative Comparison of Anatomical Parcellations”. In: *PLoS ONE* 4.9 (2009). Ed. by O. Sporns, e7200.
- [37] B. Thirion, G. Varoquaux, E. Dohmatob, and J.-B. Poline. “Which fMRI clustering gives good brain parcellations?” In: *Frontiers in Neuroscience* 8 (2014), p. 167.
- [38] D. Meunier, R. Lambiotte, A. Fornito, K. Ersche, and E. T. Bullmore. “Hierarchical modularity in human brain functional networks”. In: *Frontiers in Neuroinformatics* 3 (2009), p. 37.
- [39] R. C. Craddock, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. “Disease state prediction from resting state functional connectivity”. In: *Magnetic Resonance in Medicine* 62.6 (2009), pp. 1619–1628.
- [40] M. R. Arbabshirani, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun. “Classification of schizophrenia patients based on resting-state functional network connectivity”. In: *Frontiers in Neuroscience* 7.7 JUL (2013), pp. 1–16.
- [41] S. Dey, a. R. Rao, and M. Shah. “Exploiting the brain’s network structure in identifying ADHD subjects.” In: *Frontiers in systems neuroscience* 6.November (2012), p. 75.
- [42] F. de Vos, M. Koini, T. M. Schouten, S. Seiler, J. van der Grond, A. Lechner, R. Schmidt, M. de Rooij, and S. A. Rombouts. “A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer’s disease”. In: *NeuroImage* 167 (2018), pp. 62–72.

- [43] W. Koch, S. Teipel, S. Mueller, J. Benninghoff, M. Wagner, A. L. W. Bokde, H. Hampel, U. Coates, M. Reiser, and T. Meindl. “Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer’s disease”. In: *Neurobiology of Aging* 33.3 (2012), pp. 466–478.
- [44] J. Zhou, M. D. Greicius, E. D. Gennatas, M. E. Growdon, J. Y. Jang, G. D. Rabinovici, J. H. Kramer, M. Weiner, B. L. Miller, and W. W. Seeley. “Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer’s disease”. In: *Brain* 133.5 (2010), pp. 1352–1367.
- [45] Y. Fan, Y. Liu, H. Wu, Y. Hao, H. Liu, Z. Liu, and T. Jiang. “Discriminant analysis of functional connectivity patterns on Grassmann manifold”. In: *NeuroImage* 56.4 (2011), pp. 2058–2067.
- [46] W. Du, V. D. Calhoun, H. Li, S. Ma, T. Eichele, K. a. Kiehl, G. D. Pearlson, and T. Adali. “High Classification Accuracy for Schizophrenia with Rest and Task fMRI Data”. In: *Frontiers in Human Neuroscience* 6 (2012), p. 145.
- [47] M. Rahim, B. Thirion, A. Abraham, M. Eikenberg, E. Fohmatob, C. Comtat, and G. Varoquaux. “Integrating Multimodal Priors in Predictive Models for the Functional Characterization of Alzheimer’s Disease”. In: *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 18.Pt 1 (2015), pp. 207–214.
- [48] C. Z. Zhu, Y. F. Zang, Q. J. Cao, C. G. Yan, Y. He, T. Z. Jiang, M. Q. Sui, and Y. F. Wang. “Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder”. In: *NeuroImage* 40.1 (2008), pp. 110–120.
- [49] G. Deshpande, Z. Li, P. Santhanam, C. D. Coles, M. E. Lynch, S. Hamann, and X. Hu. “Recursive cluster elimination based support vector machine for disease state prediction using resting state functional and effective brain connectivity”. In: *PLoS ONE* 5.12 (2010), pp. 1–10.
- [50] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. Dubray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher, A. L. Alexander, E. D. Bigler, N. Lange, and J. E. Lainhart. “Functional connectivity magnetic resonance imaging classification of autism”. In: *Brain* 134.12 (2011), pp. 3739–3751.
- [51] D. a. Fair, J. T. Nigg, S. Iyer, D. Bathula, K. L. Mills, N. U. F. Dosenbach, B. L. Schlaggar, M. Mennes, D. Gutman, S. Bangaru, J. K. Buitelaar, D. P. Dickstein, A. Di Martino, D. N. Kennedy, C. Kelly, B. Luna, J. B. Schweitzer, K. Velanova, Y.-F. Wang, S. Mostofsky, F. X. Castellanos, and M. P. Milham. “Distinct neural signatures detected for ADHD

- subtypes after controlling for micro-movements in resting state functional connectivity MRI data.” In: *Frontiers in systems neuroscience* 6:February (2012), p. 80.
- [52] D. L. Murdaugh, S. V. Shinkareva, H. R. Deshpande, J. Wang, M. R. Pennick, and R. K. Kana. “Differential Deactivation during Mentalizing and Classification of Autism Based on Default Mode Network Connectivity”. In: *PLoS ONE* 7.11 (2012).
 - [53] W. Cheng, X. Ji, J. Zhang, and J. Feng. “Individual classification of ADHD patients by integrating multiscale neuroimaging markers and advanced pattern recognition techniques.” In: *Frontiers in systems neuroscience* 6:AUG 2012 (2012), p. 58.
 - [54] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius. “Network analysis of intrinsic functional brain connectivity in Alzheimer’s disease”. In: *PLoS Computational Biology* 4.6 (2008).
 - [55] H. Shen, L. Wang, Y. Liu, and D. Hu. “Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI”. In: *NeuroImage* 49.4 (2010), pp. 3110–3121.
 - [56] D. S. Bassett, B. G. Nelson, B. A. Mueller, J. Camchong, and K. O. Lim. “Altered resting state complexity in schizophrenia”. In: *NeuroImage* 59.3 (2012), pp. 2196–2207. arXiv: NIHMS150003.
 - [57] M. Liu, L.-L. Zeng, H. Shen, Z. Liu, and D. Hu. “Potential risk for healthy siblings to develop schizophrenia”. In: *NeuroReport* 23.5 (2012), pp. 265–269.
 - [58] A. Lord, D. Horn, M. Breakspear, and M. Walter. “Changes in community structure of resting state functional connectivity in unipolar depression”. In: *PLoS ONE* 7.8 (2012).
 - [59] J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, B. Greco, P. Hagmann, M. Schluep, P. Vuilleumier, and D. Van De Ville. “Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity.” In: *NeuroImage* 62.3 (2012), pp. 2021–33.
 - [60] Y. Tang, L. Wang, F. Cao, and L. Tan. “Identify schizophrenia using resting-state functional connectivity: an exploratory research and analysis”. In: *BioMedical Engineering OnLine* 11.1 (2012), p. 50.
 - [61] A. Venkataraman, T. J. Whitford, C.-F. Westin, P. Golland, and M. Kubicki. “Whole brain resting state functional connectivity abnormalities in schizophrenia”. In: *Schizophrenia Research* 139.1-3 (2012), pp. 7–12.

- [62] L. L. Zeng, H. Shen, L. Liu, L. Wang, B. Li, P. Fang, Z. Zhou, Y. Li, and D. Hu. “Identifying major depression using whole-brain functional connectivity: A multivariate pattern analysis”. In: *Brain* 135.5 (2012), pp. 1498–1507.
- [63] J. Zhang, W. Cheng, Z. Wang, Z. Zhang, W. Lu, G. Lu, and J. Feng. “Pattern classification of large-scale functional brain networks: Identification of informative neuroimaging markers for epilepsy”. In: *PLoS ONE* 7.5 (2012).
- [64] C.-Y. Wee, Y. Li, B. Jie, Z.-W. Peng, and D. Shen. “Identification of MCI using optimal sparse MAR modeled effective connectivity networks.” In: *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 16.Pt 2 (2013), pp. 319–27.
- [65] Y. Yu, H. Shen, H. Zhang, L.-L. Zeng, Z. Xue, and D. Hu. “Functional connectivity-based signatures of schizophrenia revealed by multiclass pattern analysis of resting-state fMRI from schizophrenic patients and their healthy siblings.” In: *Biomedical engineering online* 12 (2013), p. 10.
- [66] B. Jie, D. Zhang, W. Gao, Q. Wang, C. Y. Wee, and D. Shen. “Integration of network topological and connectivity properties for neuroimaging classification”. In: *IEEE Transactions on Biomedical Engineering* 61.2 (2014), pp. 576–589.
- [67] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani. “Gaussian process classification of Alzheimer’s disease and mild cognitive impairment from resting-state fMRI”. In: *NeuroImage* 112 (2015), pp. 232–243.
- [68] L. Dodero, F. Sambataro, V. Murino, and D. Sona. “Kernel-Based Analysis of Functional Brain Connectivity on Grassmann Manifold”. In: *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 18.Pt 3 (2015), pp. 604–611.
- [69] H.-I. Suk, S.-W. Lee, and D. Shen. “A Hybrid of Deep Network and Hidden Markov Model for MCI Identification with Resting-State fMRI.” In: *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 18.Pt 1 (2015), pp. 573–580.
- [70] X. Zhang, B. Hu, X. Ma, and L. Xu. “Resting-state whole-brain functional connectivity networks for MCI classification using L2-regularized logistic regression”. In: *IEEE Transactions on Nanobioscience* 14.2 (2015), pp. 237–247.

- [71] H. Chen, X. Duan, F. Liu, F. Lu, X. Ma, Y. Zhang, L. Q. Uddin, and H. Chen. “Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity-A multi-center study”. In: *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 64 (2016), pp. 1–9.
- [72] J. Kim, V. D. Calhoun, E. Shim, and J. H. Lee. “Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia”. In: *NeuroImage* 124 (2016), pp. 127–146.
- [73] R. J. Meszlényi, K. Buza, and Z. Vidnyánszky. “Resting State fMRI Functional Connectivity-Based Classification Using a Convolutional Neural Network Architecture”. In: *Frontiers in Neuroinformatics* 11 (2017), p. 61.
- [74] R. A. Fisher. “The use of multiple measurements in taxonomic problems”. In: *Annals of Eugenics* 7.2 (1936), pp. 179–188. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [75] G. Zamboni, G. K. Wilcock, G. Douaud, E. Drazich, E. McCulloch, N. Filippini, I. Tracey, J. C. W. Brooks, S. M. Smith, M. Jenkinson, and C. E. MacKay. “Resting functional connectivity reveals residual functional activity in Alzheimer’s disease”. In: *Biological Psychiatry* 74.5 (2013), pp. 375–383.
- [76] N. Filippini, L. D. Nickerson, C. F. Beckmann, K. P. Ebmeier, G. B. Frisoni, P. M. Matthews, S. M. Smith, and C. E. Mackay. “Age-related adaptations of brain function during a memory task are also present at rest”. In: *NeuroImage* 59.4 (2012), pp. 3821–3828.
- [77] N. Filippini, K. P. Ebmeier, B. J. MacIntosh, A. J. Trachtenberg, G. B. Frisoni, G. K. Wilcock, C. F. Beckmann, S. M. Smith, P. M. Matthews, and C. E. Mackay. “Differential effects of the APOE genotype on brain function across the lifespan”. In: *NeuroImage* 54.1 (2011), pp. 602–610.
- [78] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith. “FSL”. In: *NeuroImage* 62.2 (2012), pp. 782–790.
- [79] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images”. In: *NeuroImage* 17.2 (2002), pp. 825–841.
- [80] L. Griffanti, G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippini, C. E. Mackay, S. Moeller, J. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K. L. Miller, and S. M. Smith. “ICA-based artefact removal and

- accelerated fMRI acquisition for improved resting state network imaging”. In: *NeuroImage* 95 (2014), pp. 232–247.
- [81] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith. “Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers”. In: *NeuroImage* 90 (2014), pp. 449–468.
 - [82] R. Ní Mhuircheartaigh, C. Warnaby, R. Rogers, S. Jbabdi, and I. Tracey. “Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans.” In: *Science translational medicine* 5.208 (2013), 208ra148.
 - [83] B. Mohammadi, K. Kollewé, A. Samii, K. Krampfl, R. Dengler, and T. F. Münte. “Changes of resting state brain networks in amyotrophic lateral sclerosis”. In: *Experimental Neurology* 217.1 (2009), pp. 147–153.
 - [84] A. G. Hudetz. “General anesthesia and human brain connectivity.” In: *Brain connectivity* 2.6 (2012), pp. 291–302.
 - [85] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.” In: *NeuroImage* 15.1 (2002), pp. 273–289.
 - [86] R. Martuzzi, R. Ramani, M. Qiu, N. Rajeevan, and R. T. Constable. “Functional connectivity and alterations in baseline brain state in humans”. In: *NeuroImage* 49.1 (2010), pp. 823–834.
 - [87] P. Boveroux, A. Vanhaudenhuyse, M.-A. Bruno, Q. Noirhomme, S. Lauwick, A. Luxen, C. Degueldre, A. Plenevaux, C. Schnakers, C. Phillips, J.-F. Brichant, V. Bonhomme, P. Maquet, M. D. Greicius, S. Laureys, and M. Boly. “Breakdown of within- and between-network Resting State Functional Magnetic Resonance Imaging Connectivity during Propofol-induced Loss of Consciousness”. In: *Anesthesiology* 113.5 (2010), pp. 1038–1053.
 - [88] B. J. A. Palanca, A. Mitra, L. Larson-Prior, A. Z. Snyder, M. S. Avidan, and M. E. Raichle. “Resting-state Functional Magnetic Resonance Imaging Correlates of Sevoflurane-induced Unconsciousness.” In: *Anesthesiology* 123.2 (2015), pp. 346–56.
 - [89] M. D. Greicius, V. Kiviniemi, O. Tervonen, V. Vainionpää, S. Alahuhta, A. L. Reiss, and V. Menon. “Persistent default-mode network connectivity during light sedation.” In: *Human brain mapping* 29.7 (2008), pp. 839–47.
 - [90] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer. “The Hallmarks of Aging”. In: *Cell* 153.6 (2013), pp. 1194–1217.

- [91] D. Harman. “Aging: overview.” In: *Annals of the New York Academy of Sciences* 928 (2001), pp. 1–21.
- [92] A. Abbott. “Dementia: a problem for our age.” In: *Nature* 475.7355 (2011), S2–4.
- [93] J. H. Cole, R. E. Marioni, S. E. Harris, and I. J. Deary. “Brain age and other bodily ‘ages’: implications for neuropsychiatry”. In: *Molecular Psychiatry* (2018).
- [94] J. Jylhävä, N. L. Pedersen, and S. Hägg. “Biological Age Predictors.” In: *EBioMedicine* 21 (2017), pp. 29–36.
- [95] E. H. Blackburn, E. S. Epel, and J. Lin. “Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection.” In: *Science (New York, N.Y.)* 350.6265 (2015), pp. 1193–8.
- [96] S. L. Bakaysa, L. A. Mucci, P. E. Slagboom, D. I. Boomsma, G. E. McClearn, B. Johansson, and N. L. Pedersen. “Telomere length predicts survival independent of genetic influences”. In: *Aging Cell* 6.6 (2007), pp. 769–774.
- [97] X. Zhu, W. Han, W. Xue, Y. Zou, C. Xie, J. Du, and G. Jin. “The association between telomere length and cancer risk in population studies”. In: *Scientific Reports* 6.1 (2016), p. 22243.
- [98] P. C. Haycock, E. E. Heydon, S. Kaptoge, A. S. Butterworth, A. Thompson, and P. Willeit. “Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis.” In: *BMJ (Clinical research ed.)* 349 (2014), g4227.
- [99] S. Hägg, Y. Zhan, R. Karlsson, L. Gerritsen, A. Ploner, S. J. van der Lee, L. Broer, J. Deelen, R. E. Marioni, A. Wong, A. Lundquist, G. Zhu, N. K. Hansell, E. Sillanpää, I. O. Fedko, N. A. Amin, M. Beekman, A. J. M. de Craen, S. Degerman, S. E. Harris, K.-J. Kan, C. M. Martin-Ruiz, G. W. Montgomery, A. N. Adolfsson, C. A. Reynolds, N. J. Samani, H. E. D. Suchiman, A. Viljanen, T. von Zglinicki, M. J. Wright, J.-J. Hottenga, D. I. Boomsma, T. Rantanen, J. A. Kaprio, D. R. Nyholt, N. G. Martin, L. Nyberg, R. Adolfsson, D. Kuh, J. M. Starr, I. J. Deary, P. E. Slagboom, C. M. van Duijn, V. Codd, N. L. Pedersen, and N. L. Pedersen. “Short telomere length is associated with impaired cognitive performance in European ancestry cohorts”. In: *Translational Psychiatry* 7.4 (2017), e1100.
- [100] Y. Zhan, C. Song, R. Karlsson, A. Tillander, C. A. Reynolds, N. L. Pedersen, and S. Hägg. “Telomere Length Shortening and Alzheimer Disease—A Mendelian Randomization Study”. In: *JAMA Neurology* 72.10 (2015), p. 1202.

- [101] A. Müezzlinler, A. K. Zaineddin, and H. Brenner. “A systematic review of leukocyte telomere length and age in adults”. In: *Ageing Research Reviews* 12.2 (2013), pp. 509–519.
- [102] B. Jin, Y. Li, and K. D. Robertson. “DNA methylation: superior or subordinate in the epigenetic hierarchy?” In: *Genes & cancer* 2.6 (2011), pp. 607–17.
- [103] S. Horvath. “DNA methylation age of human tissues and cell types”. In: *Genome Biology* 14.10 (2013), R115.
- [104] G. Hannum, J. Guinney, L. Zhao, L. Zhang, G. Hughes, S. Sadda, B. Klotzle, M. Bibikova, J.-B. Fan, Y. Gao, R. Deconde, M. Chen, I. Rajapakse, S. Friend, T. Ideker, and K. Zhang. “Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates”. In: *Molecular Cell* 49.2 (2013), pp. 359–367.
- [105] R. E. Marioni, S. Shah, A. F. McRae, S. J. Ritchie, G. Muniz-Terrera, S. E. Harris, J. Gibson, P. Redmond, S. R. Cox, A. Pattie, J. Corley, A. Taylor, L. Murphy, J. M. Starr, S. Horvath, P. M. Visscher, N. R. Wray, and I. J. Deary. “The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936”. In: *International Journal of Epidemiology* 44.4 (2015), pp. 1388–1396.
- [106] D. W. Belsky, T. E. Moffitt, A. A. Cohen, D. L. Corcoran, M. E. Levine, J. A. Prinz, J. Schaefer, K. Sugden, B. Williams, R. Poulton, and A. Caspi. “Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing?” In: *American Journal of Epidemiology* 187.6 (2017), pp. 1220–1230.
- [107] A. Quach, M. E. Levine, T. Tanaka, A. T. Lu, B. H. Chen, L. Ferrucci, B. Ritz, S. Bandinelli, M. L. Neuhouser, J. M. Beasley, L. Snetselaar, R. B. Wallace, P. S. Tsao, D. Absher, T. L. Assimes, J. D. Stewart, Y. Li, L. Hou, A. A. Baccarelli, E. A. Whit-sel, and S. Horvath. “Epigenetic clock analysis of diet, exercise, education, and lifestyle factors”. In: *Aging* (2017).
- [108] L. P. Breitling, K.-U. Saum, L. Perna, B. Schöttker, B. Holleczek, and H. Brenner. “Frailty is associated with the epigenetic clock but not with telomere length in a German cohort”. In: *Clinical Epigenetics* 8.1 (2016), p. 21.
- [109] M. E. Levine, A. T. Lu, D. A. Bennett, and S. Horvath. “Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer’s disease related cognitive functioning”. In: *Aging* 7.12 (2015), pp. 1198–1211.
- [110] S. Horvath and B. R. Ritz. “Increased epigenetic age and granulocyte counts in the blood of Parkinson’s disease patients”. In: *Aging* 7.12 (2015), pp. 1130–1142.

- [111] M. J. Peters et al. “The transcriptional landscape of age in human peripheral blood”. In: *Nature Communications* 6.1 (2015), p. 8570.
- [112] J. Krištić, F. Vučković, C. Menni, L. Klarić, T. Keser, I. Beceheli, M. Pučić-Baković, M. Novokmet, M. Mangino, K. Thaqi, P. Rudan, N. Novokmet, J. Šarac, S. Missoni, I. Kolčić, O. Polašek, I. Rudan, H. Campbell, C. Hayward, Y. Aulchenko, A. Valdes, J. F. Wilson, O. Gornik, D. Primorac, V. Zoldoš, T. Spector, and G. Lauc. “Glycans Are a Novel Biomarker of Chronological and Biological Ages”. In: *The Journals of Gerontology: Series A* 69.7 (2014), pp. 779–789.
- [113] J. Hertel, N. Friedrich, K. Wittfeld, M. Pietzner, K. Budde, S. Van der Auwera, T. Lohmann, A. Teumer, H. Völzke, M. Nauck, and H. J. Grabe. “Measuring Biological Age via Metabonomics: The Metabolic Age Score”. In: *Journal of Proteome Research* 15.2 (2016), pp. 400–410.
- [114] M. E. Levine. “Modeling the Rate of Senescence: Can Estimated Biological Age Predict Mortality More Accurately Than Chronological Age?” In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 68.6 (2013), pp. 667–674.
- [115] K. Franke, G. Ziegler, S. Klöppel, and C. Gaser. “Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters”. In: *NeuroImage* 50.3 (2010), pp. 883–892.
- [116] M. E. Tipping. “Sparse Bayesian Learning and the Relevance Vector Machine”. In: *Journal of Machine Learning Research* 1.Jun (2001), pp. 211–244.
- [117] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, M. W. Weiner, and M. W. Weiner. “Alzheimer’s Disease Neuroimaging Initiative (ADNI): clinical characterization.” In: *Neurology* 74.3 (2010), pp. 201–9.
- [118] J. H. Cole, S. J. Ritchie, M. E. Bastin, M. C. Valdés Hernández, S. Muñoz Maniega, N. Royle, J. Corley, A. Pattie, S. E. Harris, Q. Zhang, N. R. Wray, P. Redmond, R. E. Marioni, J. M. Starr, S. R. Cox, J. M. Wardlaw, D. J. Sharp, and I. J. Deary. “Brain age predicts mortality.” In: *Molecular psychiatry* 23.5 (2018), pp. 1385–1392.
- [119] J. Steffener, C. Habeck, D. O’Shea, Q. Razlighi, L. Bherer, and Y. Stern. “Differences between chronological and brain age are related to education and self-reported physical activity”. In: *Neurobiology of Aging* 40 (2016), pp. 138–144.

- [120] E. Luders, N. Cherbuin, and C. Gaser. “Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners”. In: *NeuroImage* 134 (2016), pp. 508–513.
- [121] K. Franke and C. Gaser. “Longitudinal Changes in Individual BrainAGE in Healthy Aging, Mild Cognitive Impairment, and Alzheimer’s Disease.” In: *GeroPsych* 25.4 (2012), pp. 235–245.
- [122] H. G. Schnack, N. E. van Haren, M. Nieuwenhuis, H. E. Hulshoff Pol, W. Cahn, and R. S. Kahn. “Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study”. In: *American Journal of Psychiatry* 173.6 (2016), pp. 607–616.
- [123] N. Koutsouleris, C. Davatzikos, S. Borgwardt, C. Gaser, R. Bottlender, T. Frodl, P. Falkai, A. Riecher-Rossler, H.-J. Moller, M. Reiser, C. Pantelis, and E. Meisenzahl. “Accelerated Brain Aging in Schizophrenia and Beyond: A Neuroanatomical Marker of Psychiatric Disorders”. In: *Schizophrenia Bulletin* 40.5 (2014), pp. 1140–1153.
- [124] M. Guggenmos, K. Schmack, M. Sekutowicz, M. Garbusow, M. Sebold, C. Sommer, M. N. Smolka, H.-U. Wittchen, U. S. Zimmermann, A. Heinz, and P. Sterzer. “Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence”. In: *Translational Psychiatry* 7.12 (2017), p. 1279.
- [125] J. H. Cole, R. Leech, and D. J. Sharp. “Prediction of brain age suggests accelerated atrophy after traumatic brain injury”. In: *Annals of Neurology* 77.4 (2015), pp. 571–581.
- [126] J. H. Cole, T. Annus, L. R. Wilson, R. Remtulla, Y. T. Hong, T. D. Fryer, J. Acosta-Cabronero, A. Cardenas-Blanco, R. Smith, D. K. Menon, S. H. Zaman, P. J. Nestor, and A. J. Holland. “Brain-predicted age in Down syndrome is associated with beta amyloid deposition and cognitive decline”. In: *Neurobiology of Aging* 56 (2017), pp. 41–49.
- [127] H. R. Pardoe, J. H. Cole, K. Blackmon, T. Thesen, and R. Kuzniecky. “Structural brain changes in medically refractory focal epilepsy resemble premature brain aging”. In: *Epilepsy Research* 133 (2017), pp. 28–32.
- [128] F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. Kharabian Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock, S. Riedel-Heller, T. Luck, M. Loeffler, M. L. Schroeter, A. V. Witte, A. Villringer, and D. S. Margulies. “Predicting brain-age from multimodal imaging data captures cognitive impairment”. In: *NeuroImage* 148 (2017), pp. 179–188.
- [129] K. Franke, C. Gaser, B. Manor, and V. Novak. “Advanced BrainAGE in older adults with type 2 diabetes mellitus”. In: *Frontiers in Aging Neuroscience* 5 (2013), p. 90.

- [130] J. H. Cole, J. Underwood, M. W. A. Caan, D. De Francesco, R. A. van Zoest, R. Leech, F. W. N. M. Wit, P. Portegies, G. J. Geurtsen, B. A. Schmand, M. F. Schim van der Loeff, C. Franceschi, C. A. Sabin, C. B. L. M. Majoie, A. Winston, P. Reiss, D. J. Sharp, and C. collaboration. “Increased brain-predicted aging in treated HIV disease.” In: *Neurology* 88.14 (2017), pp. 1349–1357.
- [131] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector, and G. Montana. “Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker”. In: *NeuroImage* 163 (2017), pp. 115–124.
- [132] L. Ronan, A. F. Alexander-Bloch, K. Wagstyl, S. Farooqi, C. Brayne, L. K. Tyler, and P. C. Fletcher. “Obesity associated with increased brain age from midlife”. In: *Neurobiology of Aging* 47 (2016), pp. 63–70.
- [133] G. Douaud, A. R. Groves, C. K. Tamnes, L. T. Westlye, E. P. Duff, A. Engvig, K. B. Walhovd, A. James, A. Gass, A. U. Monsch, P. M. Matthews, A. M. Fjell, S. M. Smith, and H. Johansen-Berg. “A common brain network links development, aging, and vulnerability to disease.” In: *Proceedings of the National Academy of Sciences of the United States of America* 111.49 (2014), pp. 17648–53.
- [134] M. Toepper. “Dissociating Normal Aging from Alzheimer’s Disease: A View from Cognitive Neuroscience.” In: *Journal of Alzheimer’s disease : JAD* 57.2 (2017), pp. 331–352.
- [135] W. A. Fuller. *Measurement error models*. USA: Wiley, 1987, p. 440.
- [136] Z. Griliches and J. A. Hausman. “Errors in variables in panel data”. In: *Journal of Econometrics* 31.1 (1986), pp. 93–118.
- [137] P. Klemmera and S. Doubal. “A new approach to the concept and computation of biological age”. In: *Mechanisms of Ageing and Development* 127.3 (2006), pp. 240–248.
- [138] J. S. Racine. *A primer on regression splines*. Tech. rep. [Online https://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf; accessed 21-08-2018].
- [139] R. (Fletcher. *Practical methods of optimization*. New York: Wiley, 1987, p. 436.
- [140] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy: Open source scientific tools for Python*. [Online <http://www.scipy.org>; accessed 21-08-2018]. 2001–.
- [141] T. E. Oliphant. *Guide to NumPy*. USA: Continuum Press, 2015.

- [142] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12.Oct (2011), pp. 2825–2830.
- [143] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins. “UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age”. In: *PLOS Medicine* 12.3 (2015), e1001779.
- [144] K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson, L. Griffanti, G. Douaud, T. W. Okell, P. Weale, I. Dragonu, S. Garratt, S. Hudson, R. Collins, M. Jenkinson, P. M. Matthews, and S. M. Smith. “Multimodal population brain imaging in the UK Biobank prospective epidemiological study”. In: *Nature Neuroscience* 19.11 (2016), pp. 1523–1536.
- [145] F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. Andersson, L. Griffanti, G. Douaud, S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, E. Vallee, D. Vidaurre, M. Webster, P. McCarthy, C. Rorden, A. Daducci, D. C. Alexander, H. Zhang, I. Dragonu, P. M. Matthews, K. L. Miller, and S. M. Smith. “Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank”. In: *NeuroImage* 166 (2018), pp. 400–424.
- [146] E. J. Vinke, M. de Groot, V. Venkatraghavan, S. Klein, W. J. Niessen, M. A. Ikram, and M. W. Vernooij. “Trajectories of imaging markers in brain aging: the Rotterdam Study”. In: *Neurobiology of Aging* 71 (2018), pp. 32–40.
- [147] W. B. Kannel. “Blood Pressure as a Cardiovascular Risk Factor”. In: *JAMA* 275.20 (1996), p. 1571.
- [148] P. Sebastiani, B. Thyagarajan, F. Sun, N. Schupf, A. B. Newman, M. Montano, and T. T. Perls. “Biomarker signatures of aging”. In: *Aging Cell* 16.2 (2017), pp. 329–338.
- [149] A. M. Valdes, J. B. Richards, J. P. Gardner, R. Swaminathan, M. Kimura, L. Xiaobin, A. Aviv, and T. D. Spector. “Telomere length in leukocytes correlates with bone mineral density and is shorter in women with osteoporosis”. In: *Osteoporosis International* 18.9 (2007), pp. 1203–1210.

- [150] C. Menni, G. Kastenmüller, A. K. Petersen, J. T. Bell, M. Psatha, P.-C. Tsai, C. Gieger, H. Schulz, I. Erte, S. John, M. J. Brosnan, S. G. Wilson, L. Tsaprouni, E. M. Lim, B. Stuckey, P. Deloukas, R. Mohny, K. Suhre, T. D. Spector, and A. M. Valdes. “Metabolomic markers reveal novel pathways of ageing and early development in human populations”. In: *International Journal of Epidemiology* 42.4 (2013), pp. 1111–1119.
- [151] H. R. Pardoe and R. Kuzniecky. “NAPR: a Cloud-Based Framework for Neuroanatomical Age Prediction”. In: *Neuroinformatics* 16.1 (2018), pp. 43–49.
- [152] V. D. Calhoun, T. Adali, N. R. Giuliani, J. J. Pekar, K. A. Kiehl, and G. D. Pearlson. “Method for multimodal analysis of independent source differences in schizophrenia: Combining gray matter structural and auditory oddball functional data”. In: *Human Brain Mapping* 27.1 (2006), pp. 47–62.
- [153] K. Chen, E. M. Reiman, Z. Huan, R. J. Caselli, D. Bandy, N. Ayutyanont, and G. E. Alexander. “Linking functional and structural brain images with multivariate network analyses: A novel application of the partial least square method”. In: *NeuroImage* 47.2 (2009), pp. 602–610.
- [154] A. R. Groves, C. F. Beckmann, S. M. Smith, and M. W. Woolrich. “Linked independent component analysis for multimodal data fusion”. In: *NeuroImage* 54.3 (2011), pp. 2198–2217.
- [155] N. M. Correa, Y.-O. Li, T. Adali, and V. D. Calhoun. “Canonical Correlation Analysis for Feature-Based Fusion of Biomedical Imaging Modalities and Its Application to Detection of Associative Networks in Schizophrenia.” In: *IEEE journal of selected topics in signal processing* 2.6 (2008), pp. 998–1007.
- [156] N. M. Correa, T. Eichele, T. Adali, Y.-O. Li, and V. D. Calhoun. “Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI”. In: *NeuroImage* 50.4 (2010), pp. 1438–1445.
- [157] J. Liu, G. Pearlson, A. Windemuth, G. Ruano, N. I. Perrone-Bizzozero, and V. Calhoun. “Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA”. In: *Human Brain Mapping* 30.1 (2009), pp. 241–255.
- [158] J. Sui, T. Adali, G. D. Pearlson, V. P. Clark, and V. D. Calhoun. “A method for accurate group difference detection by constraining the mixing coefficients in an ICA framework”. In: *Human Brain Mapping* 30.9 (2009), pp. 2953–2970.

- [159] E. Martinez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen. “Concurrent EEG/fMRI analysis by multiway Partial Least Squares”. In: *NeuroImage* 22.3 (2004), pp. 1023–1034.
- [160] V. M. Vergara, A. Ulloa, V. D. Calhoun, D. Boutte, J. Chen, and J. Liu. “A three-way parallel ICA approach to analyze links among genetics, brain structure and brain function”. In: *NeuroImage* 98 (2014), pp. 386–394.
- [161] V. D. Calhoun and J. Sui. “Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing Link(s) in Complex Mental Illness”. In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1.3 (2016), pp. 230–244.
- [162] H. W. Kuhn. “The Hungarian method for the assignment problem”. In: *Naval Research Logistics Quarterly* 2.1-2 (1955), pp. 83–97.
- [163] L. R. Dice. “Measures of the Amount of Ecologic Association Between Species”. In: *Ecology* 26.3 (1945), pp. 297–302.
- [164] L. Rokach and O. Maimon. “Clustering Methods”. In: *Data Mining and Knowledge Discovery Handbook*. New York: Springer-Verlag, 2005, pp. 321–352.
- [165] D. J. Ketchen Jr. and C. L. Shook. “The application of cluster analysis in strategic management research: an analysis and critique”. In: *Strategic Management Journal* 17.6 (1996), pp. 441–458.
- [166] I. Arevalo-Rodriguez, N. Smailagic, M. Roqué i Figuls, A. Ciapponi, E. Sanchez-Perez, A. Giannakou, O. L. Pedraza, X. Bonfill Cosp, and S. Cullum. “Mini-Mental State Examination (MMSE) for the detection of Alzheimer’s disease and other dementias in people with mild cognitive impairment (MCI)”. In: *Cochrane Database of Systematic Reviews* 3 (2015).
- [167] P. Grunwald. *A tutorial introduction to the minimum description length principle*. 2004. eprint: [arXiv:math/0406077](https://arxiv.org/abs/math/0406077).
- [168] K. B. Walhovd, A. M. Fjell, I. Reinvang, A. Lundervold, A. M. Dale, D. E. Eilertsen, B. T. Quinn, D. Salat, N. Makris, and B. Fischl. “Effects of age on volumes of cortex, white matter and subcortical structures”. In: *Neurobiology of Aging* 26 (2005), pp. 1261–1270.
- [169] J. C. Churchwell and D. A. Yurgelun-Todd. “Age-related changes in insula cortical thickness and impulsivity: Significance for emotional development and decision-making”. In: *Developmental Cognitive Neuroscience* 6 (2013), pp. 80–86.

- [170] A. Hafkemeijer, I. Altmann-Schneider, A. J. M. de Craen, P. E. Slagboom, J. van der Grond, and S. A. R. B. Rombouts. “Associations between age and gray matter volume in anatomical brain networks in middle-aged to older adults”. In: *Aging Cell* 13.6 (2014), pp. 1068–1074.
- [171] H. Lemaitre, A. L. Goldman, F. Sambataro, B. A. Verchinski, A. Meyer-Lindenberg, D. R. Weinberger, and V. S. Mattay. “Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume?” In: *Neurobiology of Aging* 33.3 (2012), 617.e1–617.e9.
- [172] M. Goodro, M. Sameti, B. Patenaude, and G. Fein. “Age effect on subcortical structures in healthy adults”. In: *Psychiatry Research: Neuroimaging* 203.1 (2012), pp. 38–45.
- [173] A. Giorgio, L. Santelli, V. Tomassini, R. Bosnell, S. Smith, N. De Stefano, and H. Johansen-Berg. “Age-related changes in grey and white matter structure throughout adulthood”. In: *NeuroImage* 51.3 (2010), pp. 943–951.
- [174] E. V. Sullivan, T. Rohlfing, and A. Pfefferbaum. “Longitudinal Study of Callosal Microstructure in the Normal Adult Aging Brain Using Quantitative DTI Fiber Tracking”. In: *Developmental Neuropsychology* 35.3 (2010), pp. 233–256.
- [175] L. Q. Uddin, J. S. Nomi, B. Hébert-Seropian, J. Ghaziri, and O. Boucher. “Structure and Function of the Human Insula.” In: *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 34.4 (2017), pp. 300–306.
- [176] C. Carnero-Pardo. “Should the Mini-Mental State Examination be retired?” In: *Neurología (English Edition)* 29.8 (2014), pp. 473–481.
- [177] G. Rabinovici, W. Seeley, E. Kim, M. Gorno-Tempini, K. Rascovsky, T. Pagliaro, S. Allison, C. Halabi, J. Kramer, J. Johnson, M. Weiner, M. Forman, J. Trojanowski, S. DeArmond, B. Miller, and H. Rosen. “Distinct MRI Atrophy Patterns in Autopsy-Proven Alzheimer’s Disease and Frontotemporal Lobar Degeneration”. In: *American Journal of Alzheimer’s Disease & Other Dementias* 22.6 (2008), pp. 474–488.
- [178] J. D. Rohrer. “Structural brain imaging in frontotemporal dementia”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1822.3 (2012), pp. 325–332.