

NEW ADVANCES IN STATISTICAL NEUROIMAGE PROCESSING

FRANCISCO JESÚS MARTÍNEZ MURCIA



TESIS DOCTORAL

Doctor en Ingeniería (Dr.-Ing.)

Departamento de Teoría de la Señal, Telemática y Comunicaciones
Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Universidad de Granada

Junio 2017 – version 4.2

Ohana means family.
Family means nobody gets left behind, or forgotten.
— Lilo & Stitch

Dedicated to the loving memory of Rudolf Miede.

1939 – 2005

DECLARACIÓN

El doctorando Francisco Jesús Martínez Murcia y los directores de la tesis Juan Manuel Górriz Sáez y Javier Ramírez Pérez de Inestrosa garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, Junio 2017

Directores de la Tesis:

Juan Manuel Górriz Sáez

Javier Ramírez Pérez de
Inestrosa

Doctorando:

Francisco Jesús Martínez
Murcia

ABSTRACT

The rise of neuroimaging in the last years has provided physicians and radiologist with the ability to study the brain with unprecedented ease. This led to a new biological perspective in the study of neurodegenerative diseases, allowing the characterization of different anatomical and functional patterns associated with them. Computer Aided Diagnostic (CAD) systems use statistical techniques for preparing, processing and extracting information from neuroimaging data pursuing a major goal: optimize the process of analysis and diagnosis of neurodegenerative diseases and mental conditions.

With this thesis we focus on three different stages of the CAD pipeline: preprocessing, feature extraction and validation. For preprocessing, we have developed a method that target a relatively recent concern: the confounding effect of false positives due to differences in the acquisition at multiple sites. Our method can effectively merge datasets while reducing the acquisition site effects. Regarding feature extraction, we have studied decomposition algorithms (independent component analysis, factor analysis), texture features and a complete framework called Spherical Brain Mapping, that reduces the 3-dimensional brain images to two-dimensional statistical maps. This allowed us to improve the performance of automatic systems for detecting Alzheimer's and Parkinson's diseases. Finally, we developed a brain simulation technique that can be used to validate new functional datasets as well as for educational purposes.

Guide:

<https://plg.uwaterloo.ca/~migod/research/beck00PSLA.html>

RESUMEN

Resumen de la tesis en español.

PUBLICACIONES

Some ideas and figures have appeared previously in the following publications, that we divide here in articles and conference presentations.

*We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.*

— knuth:1974 [knuth:1974]

AGRADECIMIENTOS

Put your acknowledgments here.

Many thanks to everybody who already sent me a postcard!

Regarding the typography and other help, many thanks go to Marco Kuhlmann, Philipp Lehman, Lothar Schlesier, Jim Young, Lorenzo Pantieri and Enrico Gregorio¹, Jörg Sommer, Joachim Köstler, Daniel Gottschlag, Denis Aydin, Paride Legovini, Steffen Prochnow, Nicolas Repp, Hinrich Harms, Roland Winkler, Jörg Weber, Henri Menke, Claus Lahiri, Clemens Niederberger, Stefano Bragaglia, Jörn Hees, and the whole L^AT_EX-community for support, ideas and some great software.

Regarding LyX: The LyX port was intially done by Nicholas Mariette in March 2009 and continued by Ivo Pletikosić in 2011. Thank you very much for your work and for the contributions to the original style.

¹ Members of GuIT (Gruppo Italiano Utilizzatori di T_EX e L^AT_EX)

CONTENTS

I	INTRODUCTION	1
1	INTRODUCTION	3
1.1	Motivation	3
1.1.1	The Small Sample Size Problem	3
1.2	Neuroimaging Modalities	4
1.2.1	Magnetic Resonance Imaging	4
1.2.2	Single Photon Emission Computed Tomography	4
1.2.3	Positron Emission Tomography	5
1.2.4	Other Modalities	5
1.3	Machine Learning in Neuroimaging	5
1.4	Overview	5
2	MEDICAL BACKGROUND	7
2.1	Alzheimer's Disease	7
2.1.1	Diagnosis	7
2.2	Parkinsonism	7
2.2.1	Parkinson's Disease	7
2.2.2	Extrapyramidal Symptoms	7
2.2.3	Diagnosis	7
2.3	Autism Spectrum Disorder	7
2.3.1	Diagnosis	7
3	COMPUTER AIDED DIAGNOSIS	9
3.1	Statistical Parametric Mapping (SPM)	9
3.2	Voxel Based Morphometry (VBM)	9
3.3	Voxels As Features (VAF)	9
4	IMAGE PRE-PROCESSING	11
4.1	Spatial normalization or registration	11
4.1.1	The MNI Space	11
4.1.2	Registration algorithms	11
4.1.3	Co-registration	11
4.2	Smoothing	11
4.3	Segmentation	11
4.4	Intensity Normalization	11
4.5	Voxel-wise Analyses	11
4.5.1	Statistical Parametric Mapping	11
4.5.2	Voxel Based Morphometry	11
4.5.3	Voxels as Features	11
II	FEATURE SIZE REDUCTION	13
5	IMAGE DECOMPOSITION	15
5.1	Feature Selection	15
5.1.1	t-Test	15
5.1.2	Mann-Whitney-Wilcoxon	15

5.1.3	Relative Entropy	16
5.2	Factor Analysis	16
5.3	Independent Component Analysis	18
5.3.1	FastICA	19
5.4	Results	20
6	TEXTURE FEATURES	21
6.1	Introduction	21
6.2	Haralick Texture Features	21
6.3	Results	21
6.4	Discussion	21
7	SPHERICAL BRAIN MAPPING	23
7.1	Introduction	23
7.2	Spherical Brain Mapping	23
7.3	Volumetric Radial LBP	23
7.4	Path vía Hidden Markov Models	23
7.5	Results	23
III	SAMPLE SIZE INCREASING	25
8	SIGNIFICANCE WEIGHTED PRINCIPAL COMPONENT ANALYSIS	27
8.1	Significance Weighted Principal Component Analysis	28
8.1.1	Principal Component Analysis	28
8.1.2	One-Way Analysis of Variance	29
8.1.3	Weighting Function	30
8.2	Results for AIMS-MRI Dataset	31
8.2.1	Experiment 1: Effect of Acquisition Site	33
8.2.2	Experiment 2: Within-site Between-Group Differences	33
8.2.3	Experiment 3: Effect of SWPCA on Group Differences	35
8.2.4	Discussion	38
8.3	Results for DaTSCAN Datasets	41
9	SIMULATION OF FUNCTIONAL BRAIN IMAGES	47
9.1	Simulation Procedure	47
9.1.1	Decomposition via PCA	47
9.1.2	Probability Density Modelling using Kernel Density Estimation	47
9.1.3	Probability Density Modelling using Multivariate Gaussian	48
9.1.4	Random Number Generation	48
9.1.5	Brain Image Synthesis	48
9.2	Experimental Setup	48
9.3	Results for ADNI-PET Dataset	49
9.3.1	Experiment 1	49
9.4	Results for DaTSCAN Datasets	49

IV	GENERAL DISCUSSION AND CONCLUSIONS	51
V	APPENDIX	53
A	DATASETS	55
A.1	Magnetic Resonance Imaging	55
A.1.1	ADNI-MRI, Alzheimer's Disease Neuroimaging Initiative	55
A.1.2	AIMS-MRI, MRC-AIMS Consortium	55
A.2	Positron Emission Tomography	57
A.2.1	ADNI-PET, Alzheimer's Disease Neuroimaging Initiative	57
A.3	Single Photon Emission Computed Tomography	57
A.3.1	VDLN-HMPAO, Virgen de las Nieves	57
A.3.2	VDLN-DAT, Virgen de las Nieves	57
A.3.3	VDLV-DAT, Virgen de la Victoria Hospital	57
A.3.4	PPMI-DAT, Parkinson's Progression Markers Initiative	57
B	BACKGROUND ON SUPPORT VECTOR MACHINES	59
	BIBLIOGRAPHY	61

LIST OF FIGURES

Figure 1.1	Example of a T ₁ -weighted MRI image.	4
Figure 4.1	Typical pre-processing pipeline in MRI	11
Figure 5.1	Several selected slices of the significance map obtained for each brain coordinate by applying the Relative Entropy criterium to the ADNI database.	16
Figure 5.2	a) Original PET image composed by N = 7000 selected voxels and b) reconstruction using Factor Analysis with K = 13 factors extracted.	17
Figure 5.3	Specific variance of reconstruction error Ψ using Factor Analysis, in function of number of factors extracted (K) for ADNI database (the behaviour is similar in the SPECT database).	18
Figure 8.1	Summary of the Significance Weighted Principal Component Analysis (SWPCA) algorithm, along with its context in the pipeline used in this article. Circles represent the input data, both images (green shading) and class (group and acquisition site, purple shading). Rectangles represent the different procedures applied, comprising the DARTEL normalization and registration, the different steps contained in SWPCA , Analysis Of Variance (ANOVA) and obtaining the weighting function $\Lambda(c)$ - and the subsequent analysis.	28
Figure 8.2	Weighting function $\Lambda_c(p_c, p_{th})$ used in SWPCA .	31
Figure 8.3	Box-plot of the distribution of the component scores at each site of the AIMS-MRI dataset (see Sections 8.2 and A.1.2) in the four first components. We assume that bigger differences between distributions imply a bigger influence of the acquisition site on the portion of variance modelled by that component and therefore, to parse out those differences, the resulting weight will be smaller.	31

Figure 8.4	Brain t-map (voxel-based morphometry) of significant ($p < 0.01, t > 2.57$) grey matter (GM) and white matter (WM) between-group differences using quantitative T ₁ - weighted (qT ₁), quantitative T ₂ - weighted (qT ₂), simulated T ₁ - weighted Inversion Recovery (synT ₁), GM and WM modalities after applying SWPCA to remove site effects. 34
Figure 8.5	Brain t-map (voxel-based morphometry) of significant ($p < 0.01, t > 2.57$) GM and WM differences in Autism Spectrum Disorder (ASD) using qT ₁ , qT ₂ , synT ₁ , GM and WM maps before and after applying SWPCA to remove site effects. 36
Figure 8.6	Brain Z-map (component-based morphometry) of significant ($p < 0.01, t > 2.57$) GM and WM differences using qT ₁ , qT ₂ , synT ₁ , GM and WM maps before and after applying SWPCA to remove site effects. 37
Figure 9.1	Schema of the brain image synthesis algorithm. 47
Figure 9.2	Comparison between simulated and original images from Alzheimer's Disease (AD) and Control Subject (CTL) classes. 47

LIST OF TABLES

Table 8.1	Between-site classification accuracy (\pm standard deviation) for different modalities and masks without and with SWPCA correction. 42
Table 8.2	Classification accuracy (Acc), sensitivity (Sen) and specificity (Spec) \pm standard deviation for each modality and mask using the participants acquired at the LON and CAM sites. 43
Table 8.3	Classification accuracy (Acc), sensitivity (Sen), and specificity (Spec) \pm STD for the different modalities and masks using ALL, before and after applying SWPCA. 44
Table 8.4	Performance measures for the combined DaTSCAN dataset 45
Table 9.1	Baseline performance of the set, using the original dataset. 49

Table 9.2	Performance of Exp 1, demonstrating the predictive ability of the simulated images over the real dataset.	49
Table 9.3	Performance of the Exp 3 proves the independence of the simulated images with respect to the originals.	49
Table A.1	Summary of the datasets used in this thesis.	55
Table A.2	Demographics of the AIMS-MRI dataset.	56

LISTINGS

PCA	Principal Component Analysis
ICA	Independent Component Analysis
SPECT	Single Photon Emission Computed Tomography
PET	Positron Emission Tomography
AD	Alzheimer's Disease
PKS	Parkinsonism
ASD	Autism Spectrum Disorder
MRI	Magnetic Resonance Imaging
SWPCA	Significance Weighted Principal Component Analysis
SBM	Spherical Brain Mapping
VBM	Voxel Based Morphometry
CTL	Control Subject
CAD	Computer Aided Diagnosis
ADNI	Alzheimer's Disease Neuroimaging Initiative
PPMI	Parkinson's Progression Markers Initiative
VDLN	Virgen de las Nieves Hospital

VDLV	Virgen de la Victoria Hospital
MRC-AIMS	Medical Research Council Autism Imaging Multicentre Study
MNI	Montreal Neurological Institute
synT ₁	simulated T ₁ - weighted Inversion Recovery
qT ₁	quantitative T ₁ - weighted
qT ₂	quantitative T ₂ - weighted
GM	grey matter
WM	white matter
ANOVA	Analysis Of Variance
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVC	Support Vector Classifier
CBM	Component Based Morphometry
KDE	Kernel Density Estimation
MCI	Mild Cognitive Impairment

Part I

INTRODUCTION

In this part, we will focus on the motivation of this Thesis, examining the state-of-the-art methodology used in clinical practice. We will also provide a brief medical background on the diseases studied in this work, as well as an examination of the computational methodology in neuroimaging.

INTRODUCTION

1.1 MOTIVATION

In recent years, there has been a rise in the use of neuroimaging in the clinical practice. It has improved and speeded the procedure of diagnostic, providing unprecedented insight into the brain. Neuroimaging is a very extended tool in research as well. Different fields such as psychiatry, neurology, psychology, behavioural science or biology make extensive use of brain imaging in their studies.

The basis of these studies are common: a selection procedure by which a representative set of subjects is recruited, a performance of the experiment on (or by) each subject and a statistical analysis of the acquired data. Particularly, when studying a certain disease, it is common to recruit subjects affected by the disease and non-affected, healthy subjects, usually known as **CTLs**. Then, in this typical example, both affected and **CTLs** are scanned, and brain anatomy or function is analysed using statistical tools. The result of this analysis is a list of significant differences between structure or function that are linked to the disease.

Computer Aided Diagnosis (**CAD**) systems provide a set of tools to help setting up and performing these studies. It is currently a thriving area of research involving multidisciplinary teams, combining computer science, mathematics, medicine, artificial intelligence, statistics, machine learning, and many others. The main aim is to assist clinicians in the procedure of diagnosis and study of the diseases by providing software that can effectively recognize disease patterns, characterize differences and make predictions. [**Martinez-Murcia2016**]

One fundamental issue often found in this studies is the sample size. The number of subjects frequently ranges from tens to hundreds, whereas the number of features (namely voxels) to be analysed can add up to millions. This causes the so-called *Small Sample Size Problem* [**Duin2000**] which negatively affects the statistical power of any experiment performed using these datasets [**Button2013**].

1.1.1 *The Small Sample Size Problem*

The Small Sample Size Problem arises as a loss of statistical power when the number of samples is small compared to the number of features. This loss of statistical power translates to neuroimaging as false positives (the system detects signal where there is not) or false negatives (the system is unable to detect some signals). In differen-

tial diagnosis studies, it leads to places where differences are stated where there is not

In addition to some confounding effects, such as population bias or scanner differences,

Solucion tipica: multilple comparison correction.

Main idea: Neuroimaging studies need are subject to the small sample size problem. To overcome that problem we must either decrease the number of features or increase the sample size.

From Chapter: Signal decomposition techniques are widely used in many applications, ranging from one-dimensional signals such as audio or electroencephalography (EEG) to multidimensional arrays, and are frequently applied as feature reduction to overcome the small sample-size problem, that is, the loss of statistical power due to a larger number of features compared to the number of samples.

1.2 NEUROIMAGING MODALITIES

There exist a variety of imaging modalities used in neuroscience. By far, the most extended is Magnetic Resonance Imaging ([MRI](#)), which provides intensity maps that represent the internal structure of the brain. Other modalities are aimed at studying the function of the brain, by injecting radioactive ligands that, linked to a receptor, can measure its distribution. This is the case of Positron Emission Tomography ([PET](#)) and Single Photon Emission Computed Tomography ([SPECT](#)).

1.2.1 *Magnetic Resonance Imaging*

Magnetic Resonance Imaging ([MRI](#)) is perhaps the most widespread imaging modality, and it is used to analyse both structural and functional properties of the brain.

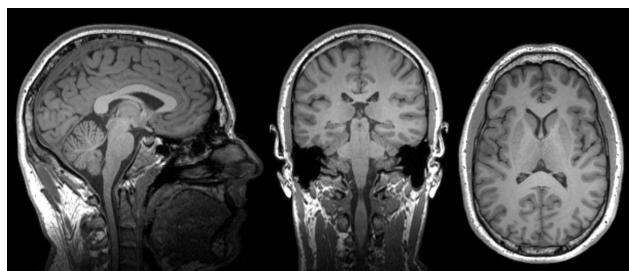


Figure 1.1: Example of a T1-weighted MRI image.

Technology. Physics fundamentals.

1.2.2 *Single Photon Emission Computed Tomography*

Single Photon Emission Computed Tomography ([SPECT](#)) is

Ref: has been the cornerstone of nuclear medicine and it is widely used to detect molecular changes in cardiovascular, oncological and neurological diseases.

Technology.

Radiotracers used. Here: DATSCAN. HMPAO.

1.2.3 *Positron Emission Tomography*

1.2.4 *Other Modalities*

1.3 MACHINE LEARNING IN NEUROIMAGING

In contrast to traditional visual inspection a semiquantitative analysis of neuroimaging, machine learning is nowadays a trend in the field. Machine learning is the subfield of computer science that provides computers with the ability to learn from data instead of being programmed for an explicit task. Applications range from automatic processing of images to Computer Aided Diagnosis ([CAD](#)),

Statistical techniques such as Principal Component Analysis ([PCA](#)), are used for feature extraction and hypothesis testing for feature selection in systems intended to

[CAD](#)

1.4 OVERVIEW

This thesis explores how to use signal processing and machine learning to overcome the small sample size problem in neuroimaging.

First strategy: decrease the number of features -> Feature extraction. This is the subject of Decomposition techniques, Texture Analysis and Spherical Brain Mapping ([SBM](#)).

Second strategy: increase the sample size. Most popular option: multi-site studies where subjects are acquired using similar techniques at different sites. This poses a major problem: inhomogeneities, etc. To overcome this we propose the [SWPCA](#). Other option: simulate new subjects from the existent database, in order to increase sample size.

2

MEDICAL BACKGROUND

First of all, it is interesting to provide some medical background about the diseases that we have applied or methodology to. This is the case of [AD](#), Parkinsonism ([PKS](#)) and [ASD](#). In this chapter, we will explore what we currently know about causes, symptoms and particularities of these diseases, and how these can be identified using neuroimaging.

2.1 ALZHEIMER'S DISEASE

[AD](#) is whatever..

Wikipedia: The cause of Alzheimer's disease is poorly understood [1]. About 70% of the risk is believed to be genetic with many genes usually involved.[6] Other risk factors include a history of head injuries, depression, or hypertension.[1] The disease process is associated with plaques and tangles in the brain.[6] A probable diagnosis is based on the history of the illness and cognitive testing with medical imaging and blood tests to rule out other possible causes.[7] Initial symptoms are often mistaken for normal ageing.[1] Examination of brain tissue is needed for a definite diagnosis.[6] Mental and physical exercise, and avoiding obesity may decrease the risk of AD.[6] There are no medications or supplements that decrease risk.[8]

2.1.1 *Diagnosis*

TEsts

2.2 PARKINSONISM

2.2.1 *Parkinson's Disease*

2.2.2 *Extrapyramidal Symptoms*

2.2.3 *Diagnosis*

2.3 AUTISM SPECTRUM DISORDER

2.3.1 *Diagnosis*

3

COMPUTER AIDED DIAGNOSIS

3.1 STATISTICAL PARAMETRIC MAPPING (SPM)

3.2 VOXEL BASED MORPHOMETRY (VBM)

3.3 VOXELS AS FEATURES (VAF)

asdñlkfj as laksdfa s dfa

4

IMAGE PRE-PROCESSING

Brief introduction to image pre-processing.

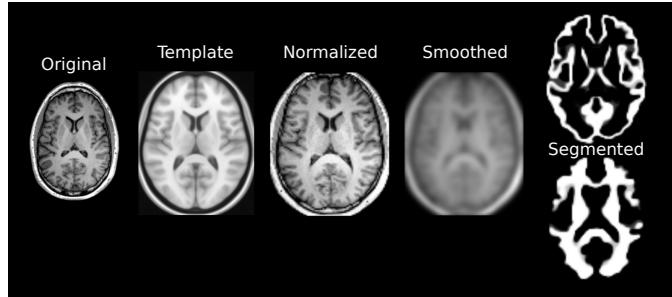


Figure 4.1: Typical pre-processing pipeline in [MRI](#).

4.1 SPATIAL NORMALIZATION OR REGISTRATION

4.1.1 *The MNI Space*

4.1.2 *Registration algorithms*

4.1.3 *Co-registration*

4.2 SMOOTHING

4.3 SEGMENTATION

4.4 INTENSITY NORMALIZATION

4.5 VOXEL-WISE ANALYSES

4.5.1 *Statistical Parametric Mapping*

4.5.2 *Voxel Based Morphometry*

Whole pipeline

4.5.3 *Voxels as Features*

Part II

FEATURE SIZE REDUCTION

The first approach to reduce the small sample size problem in neuroimaging studies is based on reducing the number of features .

5

IMAGE DECOMPOSITION

In this, we will focus on [**Martinez201141**, **Martinez-Murcia20129676**,
Martinez-Murcia2013255, **Martinez-Murcia201458**]

5.1 FEATURE SELECTION

[**Martinez201141**, **Martinez-Murcia20129676**]

5.1.1 t-Test

Student's t-Test (t-Test) is a widely used statistical test which quantifies the differences between two different classes. It uses a common estimation of variance for both classes. The value of statistical t can be computed [Fay10] as:

$$t = \frac{\bar{\Omega}_1 - \bar{\Omega}_2}{\sigma_{\Omega_1 \Omega_2}^2 \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (5.1)$$

where

$$\sigma_{\Omega_1 \Omega_2}^2 = \sqrt{\frac{(n_1 - 1)\sigma_{\Omega_1}^2 + (n_2 - 1)\sigma_{\Omega_2}^2}{n_1 + n_2 - 2}} \quad (5.2)$$

$\sigma_{\Omega_1 \Omega_2}^2$ is an estimator of common standard deviation of both samples, $\bar{\Omega}_1$ and $\bar{\Omega}_2$ are the mean of each class, n_1 is the number of samples in class ω_1 and n_2 is the number of samples in class ω_2 .

5.1.2 Mann-Whitney-Wilcoxon

Mann-Whitney-Wilcoxon U-test(MWW) uses the absolute value of the statistical U to rank voxels. Calculation of U value is done by the following expression [Fay10]:

$$U_i = R_i - \frac{n_i(n_i + 1)}{2} \quad (5.3)$$

where n_i is the sample size for sample i, and R_i is the sum of the ranks in sample i (where $i = 1, 2$). Smaller U_i value is taken as the final U value.

This statistical test measures the dissimilarity between two groups of values, and, although is similar to Student's t-Test, is less likely

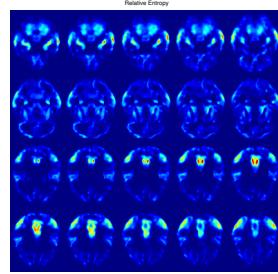


Figure 5.1: Several selected slices of the significance map obtained for each brain coordinate by applying the Relative Entropy criterium to the ADNI database.

than it to spuriously indicate significance because of the presence of outliers. As we make use of real data, and we have no knowledge about the images' statistical distribution, MW test could be a very good choice [Fay10].

5.1.3 *Relative Entropy*

Relative Entropy (or Kullback-Leibler divergence) is a non-symmetric measure of the difference between two probability distributions Ω_1 and Ω_2 . Because of its non-symmetric property, we can make use of this to evaluate the difference between CTRL and AD images for each voxel. Relative Entropy can be calculated with equation 5.4 [EntropyBishop].

Let Ω_1 and Ω_2 be two discrete random variables. Relative Entropy is defined as:

$$\text{KL}_{\omega_1 \omega_2} = \int_V \omega_1 \log \frac{\omega_1}{\omega_2} d\mu \quad (5.4)$$

where μ is any measure of V , the set of all voxels that are placed on a certain brain coordinate, in which $\omega_1 = \frac{d\Omega_1}{d\mu}$ and $\omega_2 = \frac{d\Omega_2}{d\mu}$ exist. Figure 5.1 depicts values of Relative Entropy for each voxel.

5.2 FACTOR ANALYSIS

[Martinez20141, Martinez-Murcia20129676]

To this purpose, a number of key features (K) are extracted using Factor Analysis technique [Harman73]. It is assumed that each image in the database is a different observation of the experiment. Factor analysis models each of the n observations as the expression of fewer unobserved variables, which are called factors. Each observation has

N variables, which are modelled as linear combinations of the K factors, plus errors, as described in Eq. 5.5.

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\lambda}\mathbf{F} + \boldsymbol{\epsilon} \quad (5.5)$$

where \mathbf{x} is the vector containing the observed variable (dimension n), $\boldsymbol{\mu}$ is the mean of the variable, $\boldsymbol{\lambda}$ is a vector of K factor loadings for this observation, \mathbf{F} is a matrix of dimension $N \times K$ which contains the common factors and $\boldsymbol{\epsilon}$ is the error of reconstruction. For all observations, Eq. 5.5 can be rewritten as:

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\epsilon} \quad (5.6)$$

where \mathbf{X} is a matrix of observed variables of dimension $n \times N$, $\boldsymbol{\mu}$ is a vector of means of length n , $\boldsymbol{\Lambda}$ is a matrix of dimension $n \times K$ which contains the maximum likelihood estimate of the factor loadings for each observation, \mathbf{F} is a matrix of dimension $N \times K$ which contains the common factors and $\boldsymbol{\epsilon}$ is a vector of length n containing reconstruction errors.

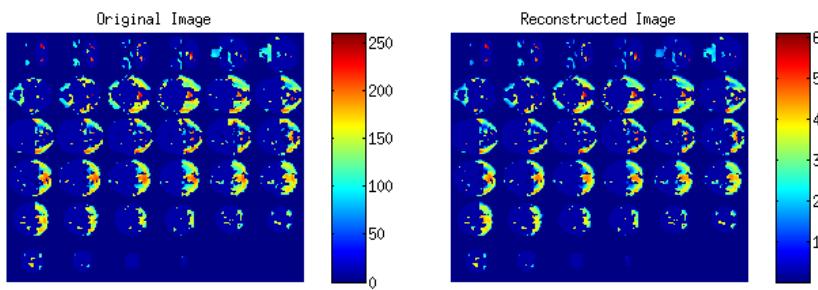


Figure 5.2: a) Original PET image composed by $N = 7000$ selected voxels and b) reconstruction using Factor Analysis with $K = 13$ factors extracted.

Original input image can be reconstructed with computed factors and factor loadings, as we have seen in Eq. 5.5. In Fig. 5.2 original image and Factor Analysis reconstruction are shown. The selected regions pinpoint the disease affected areas which are class-discriminative. Specifically, it highlights the posterior cingulate gyri and precunei, as well as the temporo-parietal region, both considered as typically affected by glucose hypometabolism in the AD [Claus1994]. A closer look shows that it also selects small thalamus regions, which has never been described as a relevant region for diagnosis. To compute representativeness of different factors, we can rewrite Eq. 5.6 as:

$$\text{Cov}(\mathbf{X} - \boldsymbol{\mu}) = \text{Cov}(\boldsymbol{\Lambda}\mathbf{F} - \boldsymbol{\epsilon}) \quad (5.7)$$

If Σ stands for $\text{Cov}(\mathbf{X} - \mu)$, Eq. 5.7 can be rewritten as:

$$\Sigma = \Lambda \text{Cov}(\mathbf{F}) \Lambda^T - \text{Cov}(\boldsymbol{\epsilon}) \quad (5.8)$$

$$\Sigma = \Lambda \Lambda^T - \Psi \quad (5.9)$$

and extract Ψ , the diagonal matrix containing the specific variances of the reconstruction error.

The choice of K requires a deeper analysis. If K is large, the image can be modelled very well, and therefore, the reconstruction error should be small. However, a large number of features can be counter-productive for the performance of the classifier due to the well known small sample size problem [Duinoo]. Thus, we should find a trade-off between the length of the feature vectors (K) and the ability of reconstruction. According to Fig. 5.3, the variance reconstruction error tends to stabilize as K increases, and the improvements are no longer significant. In the experimental Section ?? a detailed discussion about K parameter selection is shown in addition with other experimental findings.

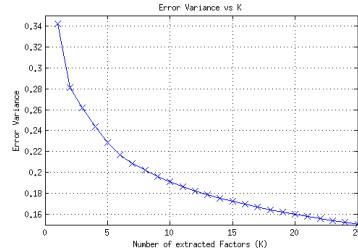


Figure 5.3: Specific variance of reconstruction error Ψ using Factor Analysis, in function of number of factors extracted (K) for ADNI database (the behaviour is similar in the SPECT database).

5.3 INDEPENDENT COMPONENT ANALYSIS

[Martinez-Murcia2013255, Martinez-Murcia201458]

Independent Component Analysis (ICA) [Hyvarinen2000], is a statistical technique that represents a multidimensional random vector as a linear combination of non-gaussian random variables (the so-called "independent components") to be as independent as possible, and has been used widely on segmentation and clustering of medical images [DeMartino2007, Alvarez2009, IllanTesis, Theis2005]. It can be considered as a non-gaussian version of Factor Analysis. Assume that we observe n linear mixtures $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ of length N that can be modelled as an expression of K independent components (IC). These independent components are defined as $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K)$,

where each \mathbf{s}_K vector has a length of N . So, each random vector \mathbf{x}_n can be described as a linear combination of K independent components:

$$\mathbf{x}_n = a_{1n}\mathbf{s}_1 + a_{2n}\mathbf{s}_2 + \dots + a_{Kn}\mathbf{s}_K \quad (5.10)$$

Without loss of generality we can assume that both the observed vectors and the independent components are zero mean. If the previous conditions are not met, the \mathbf{x} variables can be centered by subtracting the sample mean. To use a vector-matrix notation, more convenient in this case, we denote as matrix \mathbf{X} the random vector whose elements are $\mathbf{x}_1, \dots, \mathbf{x}_n$. We also denote as \mathbf{A} the matrix that contains all a_{Kn} elements, the "mixing matrix" that projects each image into the space defined by the IC. Using this notation, the mixing model above remains as follows:

$$\mathbf{X} = \mathbf{AS} \quad (5.11)$$

The starting point of ICA is the assumption that all components \mathbf{s}_K are statistically independent. To measure independence, we assume that all independent components have a non-gaussian statistical distribution. It is assumed that a sum of independent signal trends to gaussianity, so if non-gaussianity is maximized with any independence criteria F , for instance, the kurtosis or negentropy, we obtain signals that are more independent than the previous ones [Hyvarinen1999, Hyvarinen2000]. After estimating the matrix \mathbf{A} , we can compute its inverse, \mathbf{W} and obtain the projection \mathbf{S} of the images in the dataset into the IC space with:

$$\mathbf{S} = \mathbf{WX} \quad (5.12)$$

5.3.1 FastICA

Adaptive algorithms based on gradient descend can be problematic when they are used on an environment in which adaptation is not necessary, like this case. The convergence is often slow, and depends on the choice of convergence parameters. As a solution to this problem, block algorithms based on fixed-point iteration [Oja1997, FastICA99] can be used. In [Oja1997] a fixed-point algorithm based on kurtosis is introduced. In [FastICA99], this algorithm, known as FastICA, is generalized to general contrast functions. The single unit FastICA algorithm has the following form:

$$\mathbf{w}(k) = E\{\mathbf{x}g(\mathbf{w}(k-1)^T\mathbf{x})\} - E\{g'(\mathbf{w}(k-1)^T\mathbf{x})\}\mathbf{w}(k-1) \quad (5.13)$$

where the loadings vector \mathbf{w} is normalized to unit norm in each iteration, and the function $g(x)$ is a derivative of the contrast function

G defined in [Hyvarinen1999]. The expected values are estimated in practice by using the mean of a significantly high number of samples of the input data. The speed of convergence of the fixed-point algorithms is clearly superior to more neural algorithms. Improvements between 10 and 100 times the speed are observed frequently [Giannakopoulos1998].

5.4 RESULTS

Dataset used: ADNI-PET, SPECT

6

TEXTURE FEATURES

6.1 INTRODUCTION

[Martinez-Murcia2013²⁶⁶, Martinez-Murcia2014]

6.2 HARALICK TEXTURE FEATURES

6.3 RESULTS

6.4 DISCUSSION

7

SPHERICAL BRAIN MAPPING

7.1 INTRODUCTION

7.2 SPHERICAL BRAIN MAPPING

Original SBM [Martinez-Murcia2014²²⁵, Martinez-Murcia2015¹⁹, Martinez-Murcia2016, Martinez-Murcia2016hbm]

7.3 VOLUMETRIC RADIAL LBP

```
def foo():
    hola amigo
    print('amigo')

eh = foo("amigo")

string title = "This is a Unicode $\pi$ in the sky"
/*
Defined as $ \pi=\lim_{n\rightarrow\infty}\frac{P_n}{d} $ where $P$ 
    is the perimeter
of an $n$-sided regular polygon circumscribing a
circle of diameter $d$.
*/
const double pi = 3.1415926535
```

7.4 PATH VÍA HIDDEN MARKOV MODELS

7.5 RESULTS

Part III

SAMPLE SIZE INCREASING

The first approach to reduce the small sample size problem in neuroimaging studies is based on reducing the number of features .

8

SIGNIFICANCE WEIGHTED PRINCIPAL COMPONENT ANALYSIS

Multicentre studies with structural (sMRI) and functional Magnetic Resonance Imaging (fMRI) are increasingly common, allowing for recruitment of larger samples in shorter periods of time. However, the use of images acquired at different sites still poses a major challenge. In addition to logistical difficulties, such as regulatory approvals and data protection, a number of technical and methodological issues can potentially affect the resulting maps, introducing undesired intensity and geometric variance. This issue has been addressed in other neurological conditions, such as Alzheimer's Disease (Jovicich, et al., 2006; Stonnington, et al., 2008), where group differences are well known, and demonstrating that the impact of a correction for site on the resulting neurobiological differences is relatively small. However, these effects have a stronger impact in psychiatric conditions where the atypical radiological signs on MRI are often subtle and require large samples of patients to observe on-average differences relative to control samples. Recent meta-analyses point to differences being inconsistently reported in schizophrenia (Friedman and Glover, 2006; Turner, et al., 2013), psychosis (Clementz, et al., 2016; Wang, et al., 2015), and [ASD](#) (using the multi-centre ABIDE database) [\[haar2014anatomical\]](#)

These inconsistencies can arise from a variety of variance sources, ranging from the multi-level (phenotypic, neurobiological, and etiological) heterogeneities of the conditions to technical issues that include differences in scanner make, model, manufacturer, static field strength, field inhomogeneities, slew rates and image reconstruction (Van Horn and Toga, 2009), as well as acquisition problems such as within-acquisition participant head motion. Field inhomogeneities are a source of misinterpretation of the data even when the same MRI system manufacturer and model are used (Van Horn and Toga, 2009). Furthermore, results in (Pearlson, 2009) demonstrate that a single scanner can change with time, which makes some widely used strategies, for example collecting controls first and patients later, a flawed approach. Recent neuroimaging research on ASD (Haar, et al., 2014) has shown that, while analyses performed on a particular database (acquired on a single platform) could yield coherent regions, the atypical structures are often inconsistent across the wider literature using different databases. Therefore, new methodologies focused on reducing multi-site variance may be potentially helpful

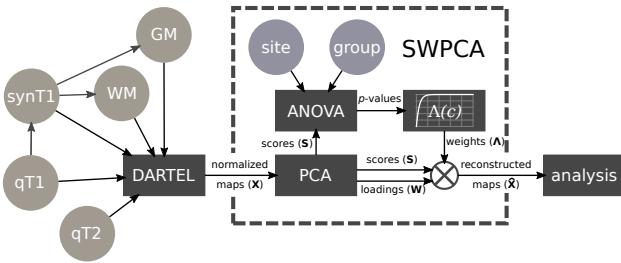


Figure 8.1: Summary of the [SWPCA](#) algorithm, along with its context in the pipeline used in this article. Circles represent the input data, both images (green shading) and class (group and acquisition site, purple shading). Rectangles represent the different procedures applied, comprising the DARTEL normalization and registration, the different steps contained in [SWPCA](#), [ANOVA](#) and obtaining the weighting function $\Lambda(c)$ - and the subsequent analysis.

in increasing the power to identify the characteristic neurobiological signature of autism, should there be one.

[[Martinez-Murcia2016a](#)]

8.1 SIGNIFICANCE WEIGHTED PRINCIPAL COMPONENT ANALYSIS

The Significance Weighted Principal Component Analysis ([SWPCA](#)) is an algorithm to reduce, in this case, undesired intensity variance introduced by multi-site image acquisition. [SWPCA](#) takes any dataset of pre-processed images, spatially normalized, and decomposes them into their variance components to then provide a corrected dataset where these undesired variance components have been reduced. To do so, [PCA](#) was applied to each modality in turn to obtain the component scores and component loadings. Since [PCA](#) is a data-driven approach, it was only used to decompose the source images, and after this procedure, a one-way [ANOVA](#) estimated the relation between each variance component and a given categorical variable, in our case, the acquisition site. The between-site variability in the variance component was then identified by its corresponding p -value. Finally, these p -values were transformed into a weighting matrix Λ that weighted the influence of each variance component in a final [PCA](#) reconstruction of the corrected maps. The procedure is summarized in Figure 8.1.

8.1.1 Principal Component Analysis

The first step in the [SWPCA](#) algorithm was to perform a [PCA](#) decomposition of the dataset into a set of orthogonal components that model the variance present in the images.

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations \mathbf{X} of possibly correlated variables, where \mathbf{X} is a $K \times N$ matrix, with K participants (in this case, with one image per participant) and N the number of voxels, into a set of N linearly uncorrelated variables called Principal Components (PC, also known as component loadings or the mixing matrix) \mathbf{W} of size $N \times N$ whose linear combination using a vector of component scores \mathbf{s}_K can perfectly recompose each image. The set of these component scores \mathbf{S} (size $K \times N$) was estimated as:

$$\mathbf{S} = \mathbf{X}\mathbf{W}^T \quad (8.1)$$

This transformation computes a sequence of PCs, maximally explaining the variability of the data while maintaining orthogonality between components. PCA was computed using Singular Value Decomposition (SVD):

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^* \quad (8.2)$$

where \mathbf{U} is an $K \times K$ orthogonal matrix, Σ is an $K \times N$ diagonal matrix with non-negative real numbers on the diagonal, and the $N \times N$ unitary matrix \mathbf{V}^* denotes the conjugate transpose of the $N \times N$ unitary matrix \mathbf{V} . With this decomposition both the component scores and estimates of the set of components loadings \mathbf{W} were obtained. In this work the truncated form of SVD was used such that only the first C components were considered, where most of the variability of the data was concentrated:

$$\mathbf{S}_C = \mathbf{U}_C\Sigma_C = \mathbf{X}\mathbf{W}_C \quad (8.3)$$

where \mathbf{S}_C is the set of component scores using the first C components (size $K \times C$). To achieve reasonable performance with minimal information loss, it was assumed that the number of components was the same as the number of images, $C = K$. Thus, a partial reconstruction of the original signal could be undertaken:

$$\hat{\mathbf{X}} = \mathbf{S}_C\mathbf{A}_C \quad (8.4)$$

where \mathbf{A}_C is the pseudoinverse of the truncated matrix of component loadings \mathbf{W}_C , and $\hat{\mathbf{X}}$ is the reconstructed set of images.

8.1.2 One-Way Analysis of Variance

The estimated PCs effectively model the variability of the image dataset. The next step was to assess each PC as a source of inter-site variance with one-way Analysis Of Variance (ANOVA). ANOVA

estimates the F-statistic, defined as the ratio between the estimated variance within groups and the variance between groups:

$$F = \frac{MS_{within}}{MS_{between}} = \frac{SS_{within}/(G - 1)}{SS_{between}/(K - G)} = \frac{\sum_i n_i (\bar{Y}_i - \bar{Y})^2 / (G - 1)}{\sum_{ij} (Y_{ij} - \bar{Y}_i)^2 / (K - G)} \quad (8.5)$$

Where MS_{within} and $MS_{between}$ are the mean squares within- and between-groups respectively, G is the number of separate groups (in our case, two), \bar{Y} is the sample mean of a certain feature (in our case, the sample mean of all K values of a given component score), \bar{Y}_i is the sample mean of the features belonging to group $i = 1 \dots G$, Y_{ij} is the j_{th} observation of a feature belonging to group i and n_i is the number of participants in the i_{th} group. The F-distribution allows an easy computation of p-values, given the number of groups and degrees of freedom. The F-statistic and p-values were computed independently for each component score and acquisition site, and then used in the [SWPCA](#) algorithm.

8.1.3 Weighting Function

To obtain a set of corrected maps, a new signal matrix of all maps of the same modality, $\hat{\mathbf{X}}$, was estimated with the influence of the PCs with variance related to acquisition site, assessed via the p-values, reduced. To do so, equation 8.4 was modified to include a square matrix Λ (dimension $C \times C$) whose diagonal contains a weight λ_c for each component that depends on its p-value; that is,

$$\hat{\mathbf{X}} = \mathbf{S} \Lambda \mathbf{A} \quad (8.6)$$

The computation of each λ_c , for each component, was performed using the Laplace distribution, modified so that the weights were on the interval $[0, 1]$:

$$\lambda_c(p_c, p_{th}) = 1 - e^{\frac{-p_c}{p_{th}}} \quad \forall p_c \in [0, 1] \quad (8.7)$$

where p_c is the statistical significance of the c_{th} component with respect to the acquisition site and p_{th} is the statistical threshold for significance; that is, $p_{th}=0.05$. A plot of the univariate weighting function $\lambda_c(p_c, p_{th})$ can be found in Figure 8.2. This weighting ensured that most of the components of variance that are not related to the acquisition site are kept unchanged, while at the same time it strongly reduces the influence of components with p-values less than the threshold.

This procedure is illustrated in Figure 8.3, where a boxplot of the distribution of the first four principal component scores is shown. Since we have assumed that substantial differences imply a bigger influence of the acquisition site on the portion of variance modelled

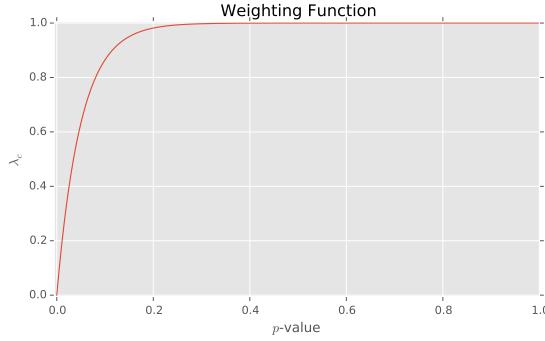


Figure 8.2: Weighting function $\Lambda_c(p_c, p_{th})$ used in [SWPCA](#).

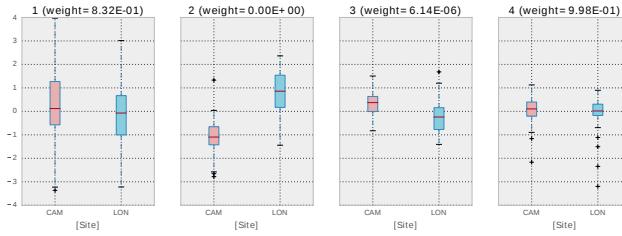


Figure 8.3: Box-plot of the distribution of the component scores at each site of the AIMS-MRI dataset (see Sections 8.2 and A.1.2) in the four first components. We assume that bigger differences between distributions imply a bigger influence of the acquisition site on the portion of variance modelled by that component and therefore, to parse out those differences, the resulting weight will be smaller.

by that component, the resulting weight is reduced, and the contribution of that component to the reconstructed signal will be smaller. After computing all weights, most of the sources that are related to the acquisition site (for example, the second and third components) have been parsed out while keeping all other sources of variance.

8.2 RESULTS FOR AIMS-MRI DATASET

To validate the effects of the [SWPCA](#) algorithm on the inter-site variance, experiments were undertaken to assess the reduction of the undesired site variance in the original datasets, and its impact on the between-group signal. Two kind of analysis were performed: a characterization of voxel-wise differences, and a classification analysis.

Voxel-wise differences between groups were characterized using Voxel Based Morphometry ([VBM](#)) [Ashburner2000], comprising pre-processing (registration, smoothing) and mass-univariate t-test on the smoothed maps from each modality. [SWPCA](#) is included (when needed) in this pipeline as a plug in, after the smoothing and before the computation of the test. Permutation testing assessed the significance of the relationship between the tested and target vari-

ables. A max-type procedure was used to obtain family-wise, whole-brain corrected p -values (Freedman and Lane, 1983). Additionally, a Component Based Morphometry (**CBM**), based on Source Based Morphometry (SBM) [xu2009source] was used. This procedure provided Z-maps for visual inspection comparable to those obtained in **VBM**, by selecting component loadings \mathbf{W} , scaling them to unit standard deviation and weighting their contribution to the final map with their statistical significance, computed using the same permutation inference as in **VBM**.

A classification analysis was undertaken using a common classification pipeline (Khedher, et al., 2015; López, et al., 2009) consisting of preprocessing, feature extraction and classification. **SWPCA** is used as a plug-in here as well, after the preprocessing and before the feature extraction step. We used **PCA** on the images for feature reduction and a Support Vector Classifier (**SVC**) with linear kernel, as implemented in LIBSVM [Chang2001], to classify the component scores in both corrected and uncorrected datasets (i.e. with and without **SWPCA**).

The classification was validated using stratified 10-fold cross-validation [Kohavi1995a]. In brief, 9 subsets of the dataset were used for extraction of the PCs and training of the classifier with the remaining subset used for testing. This procedure was repeated for each subset, repeated 10 times to avoid possible bias and random effects of the partitions. The average and standard deviation of the accuracy (acc), sensitivity (sens) and specificity (spec) values for each repetition were recorded.

For each modality independently, the following experiments were performed:

- **Experiment 1:** To demonstrate the ability of the **SWPCA** algorithm to reduce undesired effects due to acquisition site, the **PCA + SVC** pipeline was applied to the datasets labelled by acquisition site. Classification accuracy was compared to datasets with and without **SWPCA**. **VBM** was then applied to identify the spatial location of the between-site differences. This was undertaken on the whole database (ALL), and subgroups containing only **ASD** or **ASD** participants.
- **Experiment 2:** The discrimination ability of each modality, acquired at different sites was assessed by classification performance of individuals from London (LON) and Cambridge (CAM) was separately assessed, using group (**ASD** and **CTL**) as the labels.
- **Experiment 3:** To assess the impact of **SWPCA** on the datasets when characterizing the differences between **ASD** and **CTL** groups, the classification pipeline comprising **PCA + SVC**, as well as **VBM** and **CBM**, were applied to all participants with group as the labels.

8.2.1 Experiment 1: Effect of Acquisition Site

The first experiment was to demonstrate the ability of SWPCA to reduce the intensity variance related to acquisition site. To do so, we first performed a VBM analysis in all five modalities (*qT₁*, *qT₂*, *synT₁*, *GM* and *WM*) separately, with the uncorrected (without applying SWPCA) and the corrected (after applying SWPCA) maps, using the acquisition site as labels.

To illustrate where the sources of variance of the acquisition sites are located, Figure 8.4 shows a brain t-map of significant ($p < 0.01$, $|t| > 2.57$) *GM* and *WM* between-site differences. The biggest reductions in variance were found in *qT₁* and *synT₁* maps, where high variability between acquisition sites, especially in the right hemisphere, was substantially reduced after the application of SWPCA. The reduction in the *qT₂*, *GM* and *WM* maps was smaller, although noticeable.

To quantify the impact of this variance reduction on the between-groups effects, the classification analysis was undertaken. Higher accuracy values imply that the maps contain site-related patterns that were significant, whereas accuracy close to 0.5 indicates that the site-related variance was low. The test was applied to ALL, and also to the ASD and CTL subgroups. The classification results are presented in Table 8.1.

Performance results indicate clear advantages of using SWPCA, in particular in the case of *qT₁* and *synT₁* which were associated with strong site-dependent variance. These results are also consistent with the reduction of significant between-group areas observed in Figure 8.4.

The between-site differences were smaller for *GM* and *WM* maps, possibly due their reduced sensitivity. Since fractional occupancy values are abstract, unitless values derived from each image they are less influenced by the acquisition site effects. For *qT₂* maps, the site-related differences were greater for the CTL participants than ASD where, according to the classification accuracy, they were nearly indistinguishable. Acquisition site differences were therefore noticeably reduced in the CTL and ALL databases, but not in the ASD.

8.2.2 Experiment 2: Within-site Between-Group Differences

In this second experiment, accuracy, sensitivity and specificity in the between-group comparison were recorded for images acquired from each site. This is an estimation of the discrimination ability of the different modalities without the influence of the site effects; Table 8.2. For all modalities, most of the values are close to a random classifier (~50%), indicative of having either no significant differences between groups, or having spatially heterogeneous patterns of sMRI measures across individuals where mass-univariate approaches are

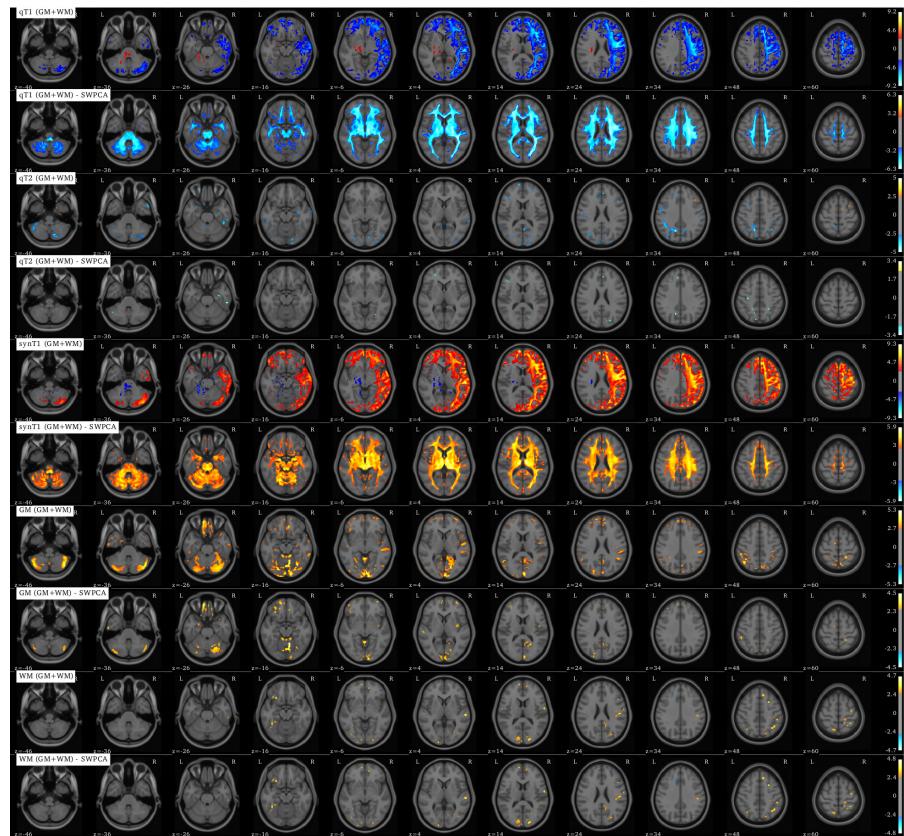


Figure 8.4: Brain t-map (voxel-based morphometry) of significant ($p < 0.01$, $|t| > 2.57$) GM and WM between-group differences using qT_1 , qT_2 , $synT_1$, GM and WM modalities after applying SWPCA to remove site effects.

sub-optimal in detecting group differences. It is interesting to note that the London sample contained more between-group differences than those acquired in Cambridge.

8.2.3 Experiment 3: Effect of SWPCA on Group Differences

Finally, group differences were characterised with and without applying site-effects reduction via SWPCA to the five modalities.

Whole-brain VBM analysis was performed on the corrected and uncorrected maps from each modality. Figure 4 depicts the brain t-maps of significant ($p < 0.01, |t| > 2.57$) **qT₁**, **qT₂**, **synT₁**, **GM** and **WM** between-group differences, using ALL, with the GM+WM mask, before and after applying **SWPCA**, so that the reduction of site-related variability can be observed. Some of the highlighted areas after applying **SWPCA** are inconsistent across modalities, with spurious peaks and noise, including a large area around the ventricles in the **qT₁** and **synT₁** modalities related to some abnormal participants that will be discussed later. However, there were some areas that were consistent across modalities. Significant areas found across at least 4 of the 5 modalities correspond to the Advanced Automated Labelling (AAL) (Tzourio-Mazoyer, et al., 2002) areas of: A) right superior frontal gyrus, Brodmann areas 6 ($z=60$); B) the pars opercularis of the left inferior frontal gyrus, Brodmann areas 44; C) the pars triangularis of the left inferior frontal gyrus, Brodmann areas 45; D) the posterior part of the left middle temporal gyrus ($z=24$); CSF filled spaces on the margins of the ventricles ($z=-6,4,14,24$); and the left crus I of cerebellar hemisphere ($z=-26$).

The complementary CBM (Section 2.4) analysis was performed on the most significant components. The resulting regions, statistically thresholded with $Z > 2.57$ (corresponding to $p < 0.01$), were superimposed on the Montreal Neurological Institute (**MNI**) template, and are depicted in Figure 8.6. A reduction of significant between-group areas after applying SWPCA is evident in most modalities, but particularly noticeable in the **qT₁** and **qT₂**. In **WM** no significant regions were observed, neither before nor after SWPCA. The significant regions identified in any modality corresponded to the AAL areas of the CSF filled areas around the ventricles (planes $z=-6, 4, 14, 24$), the right middle temporal gyrus (plane $z=14$) and the left crus I of cerebellar hemisphere (plane $z=-26$). However, none of these regions were repeated over more than two of the modalities, except for the large areas around ventricles that were caused by abnormalities in three participants, which will be discussed later.

Performance results for the classification analysis applied to ALL are shown in Table 8.3. Between-group results were quite similar before or after applying SWPCA, although reducing between-site variance generally reduced the performance towards a random classifier.

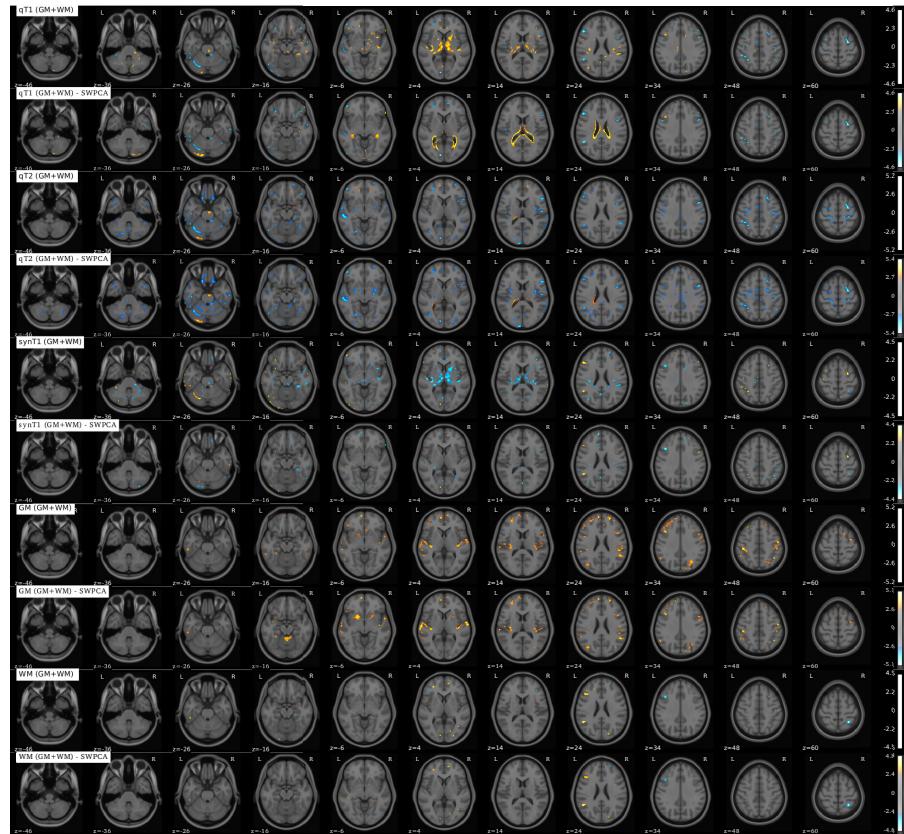


Figure 8.5: Brain t-map (voxel-based morphometry) of significant ($p < 0.01$, $|t| > 2.57$) GM and WM differences in ASD using qT₁, qT₂, synT₁, GM and WM maps before and after applying SWPCA to remove site effects.

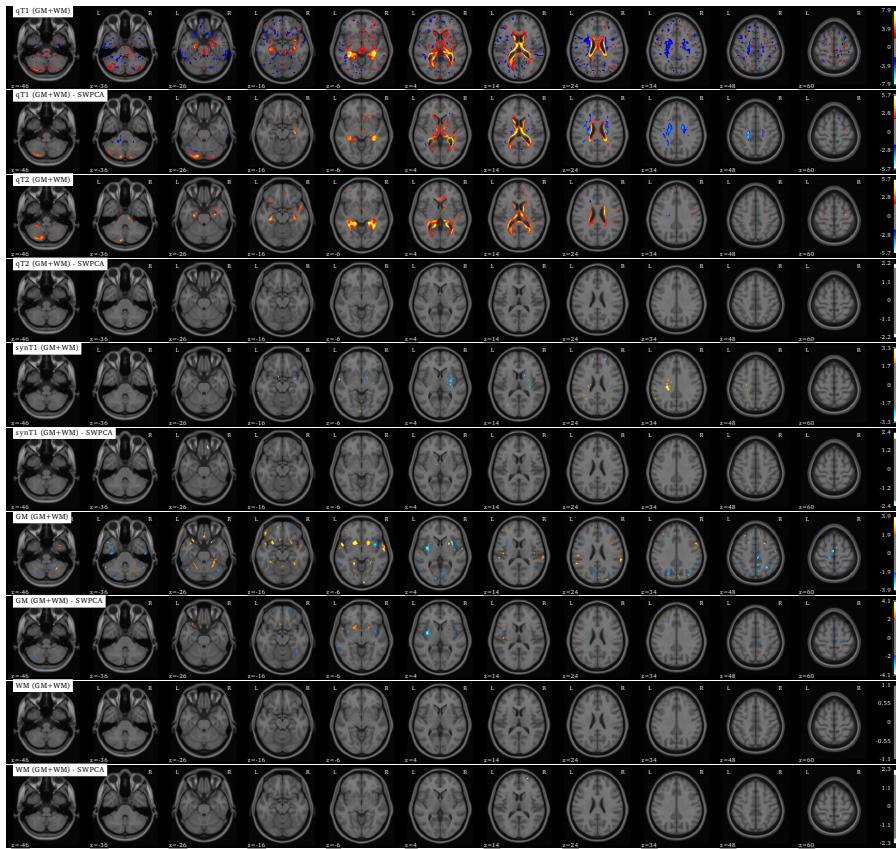


Figure 8.6: Brain Z-map (component-based morphometry) of significant ($p < 0.01, |t| > 2.57$) GM and WM differences using qT₁, qT₂, synT₁, GM and WM maps before and after applying SWPCA to remove site effects.

The results in this table match the overall effects that were found in Figure 4, where most spurious significance peaks disappeared after applying SWPCA, but some regions were highlighted. These regions, where SWPCA did not seem to eliminate the significant areas but enhanced them, could be responsible for the accuracy increment in the analysis of the qT₂ modality, and the GM with GM mask.

8.2.4 Discussion

Brain anatomical and functional differences between ASD participants and controls have been explored by a number of previous studies (Di Martino, et al., 2014; Ecker, et al., 2015; Hernandez, et al., 2015; Lenroot and Yeung, 2013; Zürcher, et al., 2015). Many affected structures have been proposed in each of these studies, however as a recent large-scale study points out (Haar, et al., 2014), these are frequently inconsistent throughout the literature. Researchers argue that most of these structures are database-dependent, and since many studies use multi-site acquisition procedures, the variance introduced by each acquisition site is a probable source of Type I errors.

The technical and logistical drawbacks of multicentre studies are widely documented, including participant recruitment procedures (Pearlson, 2009) and technical effects that range from the usage of different equipment or acquisition parameters (Van Horn and Toga, 2009) to physical changes that affect the performance of MRI scanners across time (Pearlson, 2009). There is general recognition that standardization is needed to ensure the uniformity of the acquired maps. Different approaches have been used in large-scale studies, such as Alzheimer’s Disease Neuroimaging Initiative (ADNI) where human “phantoms” were used to perform a preparatory optimisation of MRI scanning platforms (Friedman and Glover, 2006).

There are two major types of site effects, regardless of their source: geometric distortions and intensity inhomogeneities. In this work, we focused on the latter, since much of the geometric distortion has been eliminated during acquisition (see Section 2.1), and the DARTEL normalization and registration acts as a homogenizing step, reducing both between-site and between-subject geometric differences, substantially reducing the impact of the site-related geometric differences.

Regarding intensity correction, in the MRC AIMS database used in this study (Ecker, et al., 2013; Ecker, et al., 2012), a standardization procedure based on quantitative imaging (Deoni, et al., 2008) was used to minimize inter-site variance and improve the signal-to-noise contrast. However, as the between-site analysis in Section 8.2.1 suggests, this strategy still results in variance that makes it easier to distinguish scanning sites than diagnostic groups. For example, when using qT₁ the accuracy for LON vs. CAM classification was >80%,

whilst when classifying ASD vs. CTL it was 52%. This marks the substantial effect of site variance on the maps' intensity distribution, even when the multi-site study employs quantitative imaging protocol on the same model of scanner platform across sites. However, with the inclusion of GM and WM maps, we can observe that the inhomogeneities found on qT₁ or synT₁ barely affected the segmentation procedure.

In this work, the approach we have taken is to perform a multivariate decomposition of each dataset into a number of components that explain different portions of variance. The following step was to identify the components of variance that are due to multi-site acquisition and reduce them. Decomposition was completed using PCA and then, to identify which of the components were linked to acquisition site, we performed an ANOVA on the component scores. Finally, using the weighting function defined in Sec. 8.2.3, we reconstructed the original signal reducing the undesired variance, in what we called Significance Weighted PCA (SWPCA). The method has proven its ability in reducing undesired variance, quantifiable by means of the accuracy obtained in a site vs. site classification. In this case, SWPCA reduced the accuracy from >0.8 to approximately ~0.5, a random classifier, suggesting that most site-related variance was eliminated.

A simpler approach such as applying a voxel-by-voxel ANOVA would also be useful to reduce the acquisition site effects (Suckling, et al., 2012). However, SWPCA is a multivariate approach that still offers major advantages over this voxel-wise algorithm, and similar algorithms have found utility in text document searches (Kriegel, et al., 2008; Tavoli, et al., 2013; Zhang and Nguyen, 2005). First, PCA models the different sources of variance of the dataset, whereas a simple voxel-wise ANOVA only removes mean site differences, which might result in less statistical power. Secondly, SWPCA is multivariate in nature, where each component contains information that potentially affects all voxels. Together, these two features allow SWPCA to identify the components linked to the undesired effects, and reduce their impact with a weighted reconstruction approach, reducing the general variance related to the acquisition site. However, this increased power reveals a major drawback: SWPCA needs at least a moderate number of participants to work properly. That is the reason why we cannot apply SWPCA to databases such as ADNI (Friedman and Glover, 2006) or ABIDE (Di Martino, et al., 2014), where the number of participants acquired at each site is small, or to the six travelling phantoms used in the calibration of the MRC AIMS study.

There exist a number of similar multivariate methods that model the influence of categorical variables, such as the well-known Partial Least Squares (PLS) algorithm (Vinzi, et al., 2010) or Surrogate Variable Analysis (SVA) (Leek and Storey, 2007). In the first case, both PLS and SWPCA take categorical variables \mathbf{Y} along with the data \mathbf{X}

as inputs to partition the influence of these into components. However, the most significant difference is the underlying model. Whilst SWPCA estimates the principal components blindly using their variance, which is what we aim to reduce, and performs an ANOVA afterwards, PLS uses the categorical variable in the computation of the covariance matrix and then estimates the components.

On the other hand, SVA, used for gene expression studies (Leek and Storey, 2007), is more comparable to SWPCA. The SVA algorithm uses a number of decomposition and significance estimation steps to construct a set of surrogate variables; that is, variables that account for the unmodeled variance and expression heterogeneity. While similar to SWPCA in the steps used (i.e. SVD decomposition and significance estimation), their approaches are fundamentally different. SVA constructs a higher complexity model that starts by eliminating the contribution of primary variables to produce a number of unknown hidden (surrogate) variables, whereas SWPCA is intended to reduce complexity by producing variance-reduced maps to reduce the influence of previously known, but unconsidered, variables and facilitate a subsequent analysis focused only on the relevant variables.

Focusing on the [VBM](#) results, after performing the site-effects removal by SWPCA significant between-group differences were noted in five areas: A) the right superior frontal gyrus; B) the pars opercularis of the left inferior frontal gyrus; C) the pars triangularis of the left inferior frontal gyrus; D) the posterior part of the left middle temporal gyrus; and E) the left crus I of cerebellar hemisphere. The first three regions are within Brodmann areas 6, 44 and 45. However, when examining the projection of the region D onto the MNI template (see Figure 6), it is also located in the posterior part of the left superior temporal gyrus. Therefore, D corresponds closely with the region between Brodmann areas 22 and 39, the Temporo-Parietal Junction (TPJ), with negative t-value at the left side (containing Wernicke's area) and positive t-value at the right side.

The role of these regions in autism has received much attention. Brodmann areas 44 and 45, that together make the Broca's Area (of importance in speech production and a proposed part of the human mirror neuron system (Nishitani, et al., 2005)), is a region where mirror neuron dysfunction has been consistently reported in ASD-affected children (Dapretto, et al., 2006) and adults (Hadjikhani, et al., 2006; Lopez-Hurtado and Prieto, 2008; Verly, et al., 2014). Wernicke's area, contained in the left TPJ, is also linked to language, and has been associated with ASD in several works (Hadjikhani, et al., 2006; Kriegel, et al., 2008; Verly, et al., 2014). Additionally, the right TPJ has been proposed as related to mentalizing and has been repeatedly implicated in autism (Barnea-Goraly, et al., 2004), including a [fMRI](#) study of a subsample of this same AIMS dataset (Lombardo, et al., 2011). The right superior frontal gyrus (region A) is more equivocal,

with some studies (Ecker, et al., 2010; Ecker, et al., 2012) reporting abnormalities in this area, while others (Hadjikhani, et al., 2006; Segovia, et al., 2014) report no significant differences. Our analyses reveal no differences in the insula and amygdala, brain structures frequently linked to autism.

Some regions, particularly in *qT₂*, *synT₁* and segmented *GM* maps show potentially spurious significance peaks around the ventricles and especially in the left crus I of cerebellar hemisphere (region E). After examining the database, two individuals had appreciable structural abnormalities in the form of abnormal ventricle size and cerebellar atrophy, as can be seen in Figure 7. It is possible that these participants influenced the computation of the *t*-maps, and therefore are responsible for the significance in region E and areas surrounding the ventricles and, since they are part of the LON subdataset, could also be responsible for the increased classification accuracy of the quantitative T₁ and T₂, and the synthetic T₁ maps in this subdataset.

After observing the influence of these participants on the computation of the *t*-maps, we can assume that most of the structural differences in ASD are so subtle that the influence of just one or two images can impact on the final results. This, along with the poor performance of the classification pipeline presented in Section 3, dramatically reduces the significance of the aforementioned *t*-maps. Therefore, the existing evidence leads to the conclusion that ASD presents as either undetectable structural differences or, more likely, with such heterogeneous differences that are difficult to establish a common pattern even after reducing the variance introduced by acquisition site.

It may be the case that cohorts of individuals examined at different sites are somehow systematically biased towards a specific type of patient (in ways that we cannot see simply based on phenotypic information), then site-related intensity variability is also enriched with important variability about nested autism subgroups. So with any technique trying to remove the site-related inhomogeneity, the subgroup information could also be removed. Together, the evidence supports the claim that defining meaningful subgroups based on different measures, such as genetic profiling, clinical co-morbidities or sensory sensitivities, is the most urgent next step for ASD research (Haar, et al., 2014).

8.3 RESULTS FOR DATSCAN DATASETS

MODALITY	MASK	ALL		CTL		ASD	
		NO-SWPCA	SWPCA	NO-SWPCA	SWPCA	NO-SWPCA	SWPCA
qT₁	GM+WM	0.875 ± 0.083	0.530 ± 0.130	0.847 ± 0.141	0.543 ± 0.115	0.769 ± 0.145	0.553 ± 0.093
	GM	0.849 ± 0.085	0.535 ± 0.107	0.835 ± 0.154	0.501 ± 0.090	0.712 ± 0.161	0.575 ± 0.084
	WM	0.865 ± 0.082	0.447 ± 0.071	0.876 ± 0.128	0.441 ± 0.058	0.813 ± 0.127	0.575 ± 0.153
qT₂	GM+WM	0.596 ± 0.128	0.503 ± 0.093	0.615 ± 0.196	0.454 ± 0.075	0.506 ± 0.192	0.476 ± 0.103
	GM	0.596 ± 0.126	0.493 ± 0.097	0.549 ± 0.187	0.478 ± 0.108	0.497 ± 0.197	0.425 ± 0.091
	WM	0.612 ± 0.131	0.560 ± 0.128	0.576 ± 0.195	0.550 ± 0.146	0.541 ± 0.185	0.575 ± 0.172
synT₁	GM+WM	0.904 ± 0.073	0.563 ± 0.060	0.919 ± 0.100	0.440 ± 0.057	0.807 ± 0.151	0.631 ± 0.098
	GM	0.879 ± 0.090	0.576 ± 0.035	0.899 ± 0.108	0.526 ± 0.079	0.800 ± 0.145	0.587 ± 0.042
	WM	0.904 ± 0.076	0.582 ± 0.047	0.894 ± 0.111	0.574 ± 0.038	0.859 ± 0.112	0.468 ± 0.101
GM	GM+WM	0.595 ± 0.133	0.586 ± 0.141	0.582 ± 0.192	0.566 ± 0.093	0.481 ± 0.169	0.468 ± 0.152
	GM	0.620 ± 0.141	0.585 ± 0.078	0.604 ± 0.227	0.574 ± 0.038	0.499 ± 0.188	0.525 ± 0.114
WM	GM+WM	0.659 ± 0.139	0.448 ± 0.066	0.635 ± 0.180	0.507 ± 0.144	0.522 ± 0.206	0.525 ± 0.198
	WM	0.639 ± 0.124	0.549 ± 0.072	0.578 ± 0.194	0.516 ± 0.126	0.549 ± 0.160	0.526 ± 0.136

Table 8.1: Between-site classification accuracy (\pm standard deviation) for different modalities and masks without and with SWPCA correction..

MODALITY	MASK	LONDON			CAMBRIDGE		
		ACC.	SENS.	SPEC.	ACC.	SENS.	SPEC.
qT_1	GM+WM	0.603 ± 0.175	0.512 ± 0.260	0.692 ± 0.237	0.504 ± 0.193	0.492 ± 0.276	0.515 ± 0.307
	GM	0.501 ± 0.157	0.440 ± 0.244	0.565 ± 0.245	0.484 ± 0.201	0.488 ± 0.300	0.480 ± 0.327
	WM	0.505 ± 0.174	0.485 ± 0.248	0.526 ± 0.242	0.451 ± 0.197	0.465 ± 0.297	0.435 ± 0.296
qT_2	GM+WM	0.628 ± 0.168	0.535 ± 0.246	0.719 ± 0.237	0.467 ± 0.181	0.527 ± 0.307	0.417 ± 0.314
	GM	0.539 ± 0.149	0.425 ± 0.220	0.654 ± 0.222	0.491 ± 0.196	0.548 ± 0.316	0.430 ± 0.298
	WM	0.619 ± 0.194	0.585 ± 0.262	0.655 ± 0.250	0.472 ± 0.195	0.448 ± 0.283	0.492 ± 0.290
$synT_1$	GM+WM	0.665 ± 0.158	0.578 ± 0.224	0.755 ± 0.238	0.479 ± 0.201	0.478 ± 0.318	0.475 ± 0.316
	GM	0.547 ± 0.159	0.475 ± 0.237	0.622 ± 0.252	0.514 ± 0.218	0.477 ± 0.322	0.555 ± 0.342
	WM	0.515 ± 0.185	0.520 ± 0.288	0.506 ± 0.254	0.509 ± 0.209	0.472 ± 0.317	0.542 ± 0.316
GM	GM+WM	0.513 ± 0.171	0.507 ± 0.252	0.518 ± 0.245	0.488 ± 0.202	0.445 ± 0.318	0.528 ± 0.285
	GM	0.586 ± 0.174	0.610 ± 0.247	0.564 ± 0.270	0.521 ± 0.187	0.522 ± 0.303	0.535 ± 0.289
WM	GM+WM	0.471 ± 0.181	0.455 ± 0.245	0.488 ± 0.278	0.489 ± 0.206	0.502 ± 0.319	0.483 ± 0.314
	WM	0.465 ± 0.174	0.445 ± 0.243	0.484 ± 0.268	0.468 ± 0.210	0.488 ± 0.292	0.448 ± 0.305

Table 8.2: Classification accuracy (Acc), sensitivity (Sen) and specificity (Spec) ± standard deviation for each modality and mask using the participants acquired at the LON and CAM sites.

MODALITY	MASK	NO-SWPICA			SWPCA		
		ACC.	SENS.	SPEC.	ACC.	SENS.	SPEC.
qT_1	GM+WM	0.564 ± 0.123	0.503 ± 0.179	0.625 ± 0.177	0.435 ± 0.123	0.499 ± 0.181	0.371 ± 0.178
	GM	0.523 ± 0.112	0.468 ± 0.162	0.580 ± 0.192	0.458 ± 0.120	0.477 ± 0.187	0.441 ± 0.210
	WM	0.504 ± 0.131	0.475 ± 0.191	0.533 ± 0.194	0.484 ± 0.123	0.511 ± 0.179	0.456 ± 0.194
qT_2	GM+WM	0.578 ± 0.115	0.487 ± 0.208	0.669 ± 0.178	0.593 ± 0.136	0.546 ± 0.206	0.640 ± 0.194
	GM	0.554 ± 0.135	0.492 ± 0.194	0.614 ± 0.181	0.526 ± 0.144	0.512 ± 0.209	0.543 ± 0.222
	WM	0.516 ± 0.138	0.508 ± 0.198	0.522 ± 0.216	0.499 ± 0.137	0.477 ± 0.209	0.521 ± 0.196
SynT_1	GM+WM	0.596 ± 0.132	0.509 ± 0.194	0.680 ± 0.172	0.577 ± 0.130	0.479 ± 0.208	0.676 ± 0.183
	GM	0.587 ± 0.139	0.509 ± 0.210	0.665 ± 0.169	0.483 ± 0.136	0.489 ± 0.218	0.480 ± 0.200
	WM	0.496 ± 0.139	0.500 ± 0.189	0.492 ± 0.194	0.487 ± 0.134	0.513 ± 0.189	0.461 ± 0.211
GM	GM+WM	0.498 ± 0.120	0.486 ± 0.197	0.507 ± 0.203	0.490 ± 0.123	0.514 ± 0.197	0.465 ± 0.182
	GM	0.574 ± 0.121	0.571 ± 0.189	0.579 ± 0.163	0.593 ± 0.127	0.602 ± 0.172	0.587 ± 0.190
WM	GM+WM	0.499 ± 0.132	0.506 ± 0.189	0.487 ± 0.181	0.521 ± 0.129	0.510 ± 0.209	0.532 ± 0.180
	WM	0.506 ± 0.143	0.488 ± 0.219	0.526 ± 0.197	0.507 ± 0.122	0.521 ± 0.165	0.492 ± 0.193

Table 8.3: Classification accuracy (Acc), sensitivity (Sen), and specificity (Spec) \pm STD for the different modalities and masks using ALL, before and after applying SWPCA.

PERFORMANCE					
SWPCA	NORM.	ACC.	SENS.	SPEC.	
no	max	0.883 ± 0.030	0.855 ± 0.058	0.915 ± 0.058	
	int	0.877 ± 0.035	0.849 ± 0.073	0.908 ± 0.079	
	stable	0.898 ± 0.033	0.883 ± 0.057	0.915 ± 0.079	
yes	max	0.539 ± 0.100	0.527 ± 0.373	0.550 ± 0.337	
	int	±	±	±	
	stable	0.361 ± 0.102	0.394 ± 0.295	0.322 ± 0.270	

Table 8.4: Performance measures for the combined DaTSCAN dataset found before and after applying [SWPCA](#).

SIMULATION OF FUNCTIONAL BRAIN IMAGES

9.1 SIMULATION PROCEDURE

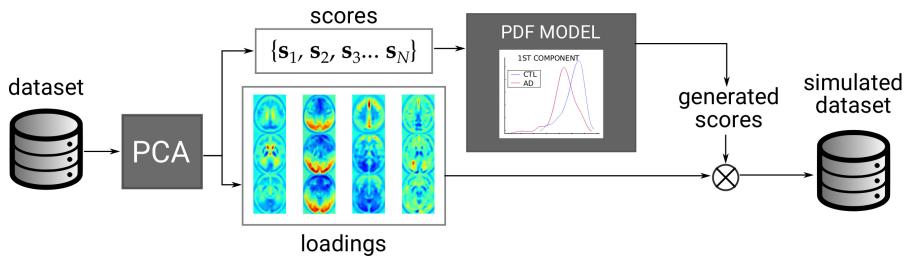


Figure 9.1: Schema of the brain image synthesis algorithm.

9.1.1 Decomposition via PCA

The first step in our simulation algorithm is to project the original dataset to a new space defined by the principal components of the set; that is, the eigenbrain space. In this space, each subject from the original dataset is projected to a point, and we can afterwards use the space basis (the principal components) to reconstruct that particular subject. In this work we will use the first N components for performance, where N is the number of subjects that are used in the computation of PCA. For more details about PCA, see Section 8.1.1.

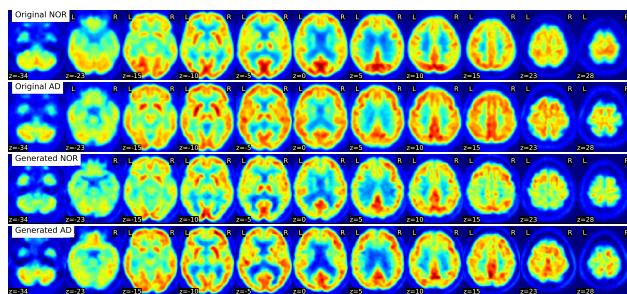


Figure 9.2: Comparison between simulated and original images from AD and CTL classes.

9.1.2 Probability Density Modelling using Kernel Density Estimation

Kernel Density Estimation (KDE) is used here to model the statistical distribution of the projected subjects in the eigenbrain space, and it is applied independently to each AD, Mild Cognitive Impair-

ment ([MCI](#)) and [CTL](#) class. The [KDE](#) estimates the probability density function f from a number of independent and identically distributed samples (x_1, x_2, \dots, x_n), in the following manner:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad (9.1)$$

where $h > 0$ is the bandwidth, a smoothing parameter. The [KDE](#) via diffusion [[Botev2010](#)] used in this article uses a data-driven automatic estimation of the bandwidth, which unlike most methods, does not rely on arbitrary normal reference rules.

9.1.3 Probability Density Modelling using Multivariate Gaussian

9.1.4 Random Number Generation

9.1.5 Brain Image Synthesis

9.2 EXPERIMENTAL SETUP

To validate the simulated dataset, we have performed two different experiments:

- **Exp. 1:** We have estimated the predictive power of the simulated images by generating new images from the original training set in each cross-validation iteration and using them to predict the original test set.
- **Exp. 2:** We tested that the simulated images are independent from the original ones, although preserving similar characteristics. To do so, following a Voxel as Features (VAF) approach [[Stoeckel04](#)], we extract a small subset (10 AD and 10 NOR) from the original dataset. Then, we trick the classifier, training it with the whole subset -instead of the training set only-, and testing it against the test set. Therefore, the performance of the tricked system must be close to 1. Then, we generate a new set of simulated images (100 AD and 100 NOR) from the reduced subset and proceed similarly. If our simulated images are independent from the originals, the performance of the system should decrease substantially.

Classification is performed using a Support Vector Machine (SVM) classifier with linear kernel. Estimation of parameter C is performed in an inner cross-validation loop within the training set. Values of accuracy (acc), sensitivity (sens) and specificity (spec) and their standard deviation (SD) are estimated.

9.3 RESULTS FOR ADNI-PET DATASET

9.3.1 Experiment 1

The performance results for the proposed experiments are shown in Table 9.1. Exp. 1 is applied to three different scenarios: only AD vs NOR (95 vs 101 subjects), and after incorporating MCI subjects, using them as NOR or AD.

SCENARIO	ACC (\pm SD)	SENS (\pm SD)	SPEC (\pm SD)
AD vs NOR	0.882 ± 0.012	0.865 ± 0.091	0.901 ± 0.118
MCI as NOR	0.727 ± 0.119	0.769 ± 0.155	0.789 ± 0.151
MCI as AD	0.739 ± 0.126	0.747 ± 0.147	0.845 ± 0.146

Table 9.1: Baseline performance of the set, using the original dataset.

SCENARIO	ACC (\pm SD)	SENS (\pm SD)	SPEC (\pm SD)
AD vs NOR	0.801 ± 0.095	0.782 ± 0.202	0.821 ± 0.191
MCI as NOR	0.751 ± 0.078	0.433 ± 0.201	0.851 ± 0.262
MCI as AD	0.712 ± 0.048	0.821 ± 0.062	0.382 ± 0.248

Table 9.2: Performance of Exp 1, demonstrating the predictive ability of the simulated images over the real dataset.

SCENARIO	ACC (\pm SD)	SENS (\pm SD)	SPEC (\pm SD)
Original	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000
Simulated	0.839 ± 0.094	0.830 ± 0.228	0.849 ± 0.206

Table 9.3: Performance of the Exp 3 proves the independence of the simulated images with respect to the originals.

9.4 RESULTS FOR DATSCAN DATASETS

Part IV

GENERAL DISCUSSION AND CONCLUSIONS

Part V
APPENDIX

A

DATASETS

Many dataset are used in this thesis, covering three imaging modalities and three disorders. A summary of these can be found on Table A.1, followed by a longer description of each one.

ACRONYM	ORIGIN	DISEASE	MODALITY	DRUG
ADNI-MRI	ADNI	AD	MRI	-
AIMS-MRI	MRC-AIMS	ASD	MRI	-
ADNI-PET	ADNI	AD	PET	HMPAO!
VDLN-HMPAO	VDLN	AD	SPECT	HMPAO!
VDLN-DAT	VDLN	PKS	SPECT	DaTSCAN
VDLV-DAT	VDLV	PKS	SPECT	DaTSCAN
PPMI-DAT	PPMI	PKS	SPECT	DaTSCAN

Table A.1: Summary of the datasets used in this thesis.

A.1 MAGNETIC RESONANCE IMAGING

A.1.1 *ADNI-MRI, Alzheimer's Disease Neuroimaging Initiative*

AD

A.1.2 *AIMS-MRI, MRC-AIMS Consortium*

Structural MRI were analysed from 136 adult, right-handed males (68 with ASD and 68 matched controls) with no significant mean differences in age and full-scale IQ, acquired from the centres contributing to the UK Medical Research Council Autism Imaging Multi-centre Study (MRC AIMS) (Ecker, et al., 2013; Ecker, et al., 2012) and recruited by advertisement. In this work, only participants recruited at the Institute of Psychiatry, King's College London (LON) and the Autism Research Centre, University of Cambridge (CAM) were included where an equivalent set of images were acquired from each participant.

Participants were excluded from the study if they had a history of major psychiatric disorder or medical illness affecting brain function (e.g. psychosis or epilepsy), or current drug misuse (including al-

cohol), or were taking antipsychotic medication, mood stabilizers or benzodiazepines.

All participants with ASD were diagnosed according to International Classification of Diseases, 10th Revision (ICD-10) research criteria, and confirmed using the Autism Diagnostic Interview-Revised (ADI-R) (Lord, et al., 1994). Autism Diagnostic Observation Schedule (ADOS) (Lord, et al., 2000) was performed, but the score was not considered as an inclusion criteria. ASD participants, to be included, must have scored above the ADI-R cut-off in the three domains of impaired reciprocal social interaction, communication and repetitive behaviours and stereotyped patterns, although failure to reach cut-off in one of the domains by one point was permitted. Intellectual ability was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999), ensuring the participants fell within the high-functioning range on the spectrum defined by a full-scale IQ > 70. The demographics of the participants are shown in detail in Table A.2.

	DATABASE	GROUP	N	AGE ($\mu \pm \sigma$ YEARS)	IQ ($\mu \pm \sigma$)
LON		ASD	39	28.74 ± 6.52	111.28 ± 13.13
		CTL	40	25.30 ± 6.62	104.67 ± 11.16
CAM		ASD	29	26.83 ± 4.64	115.83 ± 11.88
		CTL	28	26.75 ± 7.32	115.25 ± 13.67
ALL		ASD	68	25.90 ± 6.95	109.03 ± 13.31
		CTL	68	27.93 ± 5.87	113.22 ± 12.81

Table A.2: Demographics of the AIMS-MRI dataset.

Structural MRI were obtained using Driven Equilibrium Single Pulse Observation of T₁ and T₂ (DESPOT₁, DESPOT₂) (Deoni, et al., 2008) at King's College London and University of Cambridge, both with 3T GE Medical Systems HDx scanners. Using multiple Spoilt Gradient Recall (SPGR) acquisitions in the DESPOT₁ sequence and Steady State Free Procession (SSPF) acquisitions in the DESPOT₂ sequence, with different flip angles and repetition times, qT₁ and qT₂ maps were calculated with a custom ImageJ plug-in package. Correction of main and transmit magnetic field (B₀ and B₁) inhomogeneity effects was performed during the estimation of T₁ and T₂.

For accurate registration to the standard stereotatic space of the MNI, a synT₁ images were created based on the qT₁ maps (Ecker, et al., 2013; Ecker, et al., 2012; Lai, et al., 2012). The synT₁ images were then segmented using New Segment into GM and WM maps, and normalized to the MNI space using DARTEL in SPM8 (Friston, et al., 2007), with modulation (preserve volume) to retain information

of regional/local **GM** and **WM** volumes, and smoothed with a 3mm FWHM Gaussian Kernel to account for inter-subject mis-registration. The **synT₁**, **qT₁** and **qT₂** maps were also registered to the standard **MNI** space using the same DARTEL flow fields, but without modulation (preserve concentration) to retain information of regional/local T₁ contrast, T₁ relaxation time, and T₂ relaxation times, and smoothed with a 3mm FWHM Gaussian kernel. Therefore, there were five different modalities: **qT₁**, **qT₂**, **synT₁** map, **GM** and **WM** maps, for each every participant, which allows us to observe the impact of our **SWPCA** correction of site-related undesired variance on quantitative (**qT₁** and **qT₂**), simulated (**synT₁**) images and probability maps (**GM** and **WM**).

During the pre-processing of the images, several procedures targeted the reduction of inter-subject and inter-site geometric distortion, amongst them the correction of Bo and B₁ field inhomogeneity effects and the registration to **MNI** space. Many other algorithms have been proposed to help in this task. However, the study of their relative performance lies beyond the scope of this article. Following image registration, it was assumed that only the intensity of the maps was affected between sites.

A.2 POSITRON EMISSION TOMOGRAPHY

A.2.1 *ADNI-PET, Alzheimer's Disease Neuroimaging Initiative*

A.3 SINGLE PHOTON EMISSION COMPUTED TOMOGRAPHY

A.3.1 *VDLN-HMPAO, Virgen de las Nieves*

SPECT HMPAO

A.3.2 *VDLN-DAT, Virgen de las Nieves*

SPECT DATSCAN

A.3.3 *VDLV-DAT, Virgen de la Victoria Hospital*

A.3.4 *PPMI-DAT, Parkinson's Progression Markers Initiative*

B

BACKGROUND ON SUPPORT VECTOR MACHINES

Support Vector Machine ([SVM](#)), introduced in the late 70s [[Vapnik1982](#)], are a set of related supervised learning methods widely used in pattern recognition, voice activity detection (VAD), classification and regression analysis.

We suppose the data to be linearly separable. In this case, a data point is viewed as a p-dimensional vector. Our objective is to separate a set of binary labelled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyper-plane). To do so, we build a function $f : \Re^n \rightarrow \{\pm 1\}$ using training data that is, p-dimensional patterns x_i and class labels y_i :

$$(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \in \Re^n \times \{\pm 1\} \quad (\text{B.1})$$

so that f will correctly classify new examples (x, y) .

Linear discriminant functions define decision hypersurfaces or hyperplanes in a multidimensional feature space, that is:

$$g(x) = \mathbf{w}^T x + \omega_0 = 0, \quad (\text{B.2})$$

where \mathbf{w} is known as the weight vector and ω_0 as the threshold. The weight vector \mathbf{w} is orthogonal to the decision hyperplane and the optimization task consists of finding the unknown parameters ω_i , $i = 1, \dots, n$ defining the decision hyperplane.

Let $x_i, i = 1, 2, \dots, n$ be the feature vectors of the training set, X . These belong to either of the two classes, ω_1 or ω_2 . If the classes were linearly separable, the objective would be to design a hyperplane that classifies correctly all the training vectors. The hyperplane is not unique, and the selection process is focused on maximizing the generalization performance of the classifier, that is, the ability of the classifier, designed using the training set, to operate satisfactorily with new data. Among the different design criteria, the maximal margin hyperplane is usually selected since it leaves the maximum margin of separation between the two classes. Since the distance from a point x to the hyperplane is given by $z = |g(x)|/\|\mathbf{w}\|$, scaling \mathbf{w} and \mathbf{w}_0 so that the value of $g(x)$ is $+1$ for the nearest point in ω_1 and -1 for the nearest points in ω_2 , reduces the optimization problem to maximizing the margin: $2/\|\mathbf{w}\|$ with the constraints:

$$\mathbf{w}^T x + \omega_0 \geq 1, \forall x \in \omega_1 \quad (\text{B.3})$$

$$\mathbf{w}^T \mathbf{x} + w_0 \leq 1, \forall \mathbf{x} \in \omega_2 \quad (\text{B.4})$$

DECLARACIÓN

D. Juan Manuel Górriz Sáez, Doctor por la Universidad de Cádiz y la Universidad de Granada y Catedrático del Departamento de Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada y

D. Javier Ramírez Pérez de Inestrosa, Doctor por la Universidad de Granada y Catedrático del Departamento de Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada,

MANIFIESTAN:

Que la presente Memoria titulada "New Advances in Statistical Neuroimage Processing", presentada por Francisco Jesús Martínez Murcia para optar al grado de Doctor por la Universidad de Granada, ha sido realizada bajo nuestra dirección. Con esta fecha, autorizamos la presentación de la misma.

Granada, Junio 2017

Fdo: Juan Manuel Górriz Sáez

Fdo: Javier Ramírez Pérez de Inestrosa

Memoria presentada por Francisco Jesús Martínez Murcia para optar al Grado de Doctor por la Universidad de Granada.

Francisco Jesús Martínez Murcia

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both `LATEX` and `LYX`:

<https://bitbucket.org/amiede/classicthesis/>

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>