

# Automated morphometry of transgenic mouse brains in MR images

Alize Scheenstra



Alize Scheenstra

in MR images

transgenic mouse brains

Automated morphometry of

## Colophone

### *About the cover*

The displayed mouse is a C57BL/6J mouse, which is the most widely used inbred strain and the first to have its genome sequenced. The mirrored image of the mouse represents its genetically modified relative, which has the same genome except for one gene that codes for a certain disease.

Automated morphometry of transgenic mouse brains in MR images  
Scheenstra, Alize Elske Hiltje

Printed by Ipskamp Drukkers, Enschede, The Netherlands

ISBN-13: 978-90-9026004-4

©2011 A.E.H. Scheenstra, Leiden, The Netherlands

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

# Automated morphometry of transgenic mouse brains in MR images

Automatische morfometrie van transgene  
muizenhersenen in MR beelden

Proefschrift  
ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van de Rector Magnificus Prof. mr. P.F. van der Heijden ,  
volgens besluit van het College voor Promoties  
te verdedigen op donderdag 24 maart 2011  
klokke 16:15 uur

door

**Alize Elske Hiltje Scheenstra**  
geboren te Gouda  
in 1981

## Promotiecommissie

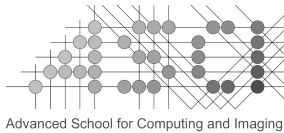
Promotor: Prof. dr. ir. J.H.C. Reiber

Co-promotor: Dr.ir. J. Dijkstra

Overige leden: Prof. dr. M. van Buchem

Prof. dr. D. Rueckert (Imperial College London )

dr. L. van der weerd



Advanced School for Computing and Imaging

This work was carried out in the ASCI graduate school.  
ASCI dissertation series number 229.

This work was supported by funds from CYTTRON within the BSIK program (Besluit subsidies investeringen kennisinfrastructuur).

*Voor Riëna van Rijn - Haasnoot*

*Ter nagedachtenis aan Hiltje Scheenstra - Betlehem*



---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Transgenic mouse models . . . . .	1
1.2	Mouse brain anatomy . . . . .	3
1.3	High resolution magnetic resonance imaging . . . . .	3
1.4	Aim of the thesis . . . . .	5
1.5	Thesis outline . . . . .	6
<b>2</b>	<b>Morphometry on rodent brains</b>	<b>7</b>
2.1	Introduction . . . . .	9
2.2	Volumetry . . . . .	9
2.3	Automated morphometry . . . . .	9
2.4	Method comparison . . . . .	11
2.5	Limitations to automated morphometry . . . . .	12
2.A	Multiple-test correction . . . . .	13
<b>3</b>	<b>Early detection of Alzheimer’s disease in MR images</b>	<b>15</b>
3.1	Introduction . . . . .	17
3.2	Alzheimer mouse models . . . . .	17
3.3	Relaxometry . . . . .	19
3.4	Analysis and models of plaque burden . . . . .	23
3.5	Cerebral amyloid angiopathy . . . . .	24
3.6	Volumetric methods . . . . .	26
3.7	Texture analysis . . . . .	27
3.8	Discussion and conclusion . . . . .	28
3.A	The most commonly used AD mouse models . . . . .	32

## Contents

---

<b>4</b>	<b>Automated segmentation of mouse brains</b>	<b>35</b>
4.1	Introduction . . . . .	37
4.2	Materials and methods . . . . .	38
4.3	Results . . . . .	42
4.4	Discussion . . . . .	47
4.5	Conclusion . . . . .	49
<b>5</b>	<b>The generalized Moore-Rayleigh test</b>	<b>51</b>
5.1	Introduction . . . . .	53
5.2	The Moore-Rayleigh test . . . . .	54
5.3	The two-sample test . . . . .	62
5.4	Results . . . . .	67
5.5	Discussion . . . . .	71
5.A	The Fisher distribution . . . . .	73
<b>6</b>	<b>The 3D Moore-Rayleigh test as tool for brain morphometry</b>	<b>77</b>
6.1	Introduction . . . . .	79
6.2	Method . . . . .	81
6.3	Results . . . . .	84
6.4	Discussion . . . . .	92
<b>7</b>	<b>Quantitative morphometry on migraine mouse models</b>	<b>95</b>
7.1	Introduction . . . . .	97
7.2	Materials and methods . . . . .	98
7.3	Results . . . . .	100
7.4	Discussion . . . . .	104
<b>8</b>	<b>Summary and conclusions</b>	<b>107</b>
8.1	Summary and conclusions . . . . .	107
8.2	Future work . . . . .	109
<b>9</b>	<b>Samenvatting en aanbevelingen</b>	<b>111</b>
9.1	Samenvatting en conclusies . . . . .	111
9.2	Aanbevelingen . . . . .	113
	<b>Bibliography</b>	<b>115</b>
	<b>Publications</b>	<b>137</b>
	<b>Acknowledgements</b>	<b>139</b>
	<b>Curriculum vitae</b>	<b>141</b>
	<b>List of Figures</b>	<b>143</b>
	<b>List of Tables</b>	<b>149</b>
	<b>List of Abbreviations</b>	<b>151</b>



# CHAPTER 1

---

## Introduction

---

### 1.1 Transgenic mouse models

Mice are thankfully exploited to study biological processes that cannot be tested in a petri dish and need to be studied *in vivo* in a real organism. Mice are not only used because they are small, easy to handle, have a fast reproduction rate and are widely available, but because mice and humans share about 97.5% of their DNA [1]. The latter statement implies that many diseases in mice and humans have a similar form of progression and show similar effects. Therefore, studying biological processes in mice will give insight on human biological processes. A mouse is called ‘transgenic’ if its genetic material has been altered, for instance by the introduction of human genes into its genotype. The exploitation of transgenic mice for research of human diseases is a world-wide debate, even though parallel to the development of transgenic mice, ethical committees were set up everywhere to guarantee that all transgenic animal research is performed under strict guidelines for health and wellbeing of the mice.

#### 1.1.1 History

The first notice of mice appearing in the laboratory setting was around 1897, when the Biologist William Haacke described the effect of heritage of the coat in albino mice. Unfortunately his work is often overlooked because of his failure to supply data [2]. Therefore, the first description of genetical heritage of the color coat in mice is generally considered to be the work of the Frenchmen Lucien Cuenot, who described recessive and dominant alleles. In 1909, Clarence Cook Little developed the first inbred mouse strain to study their genotype in the hope that, one day, this

would support research on human diseases, such as cancer. Later this mouse developed into the C57BL mouse, which is now the most widely described and used genotype. The first genetically modified mice were reported in 1974 by Jaenisch and Mintz [3]. They injected mouse embryos (blastocysts) with the Simian virus 40 (SV40) DNA, a polyomavirus that has the potential to cause tumors, and showed that the mice and their offspring had inherited the SV40 DNA in their DNA. But it was not until 1982 when Brinster and Palmiter incorporated the human growth gene to a mouse model [4], that the clinical world could see the enormous advantages of exploiting transgenic mouse models for human genetic studies.

Since the early 80s, the development of standardized mouse strains (groups of mice with identical genotype) for biomedical research has expanded tremendously. Currently mouse models are available for a whole range of human diseases. Also, standardized mouse atlases were created which contain a complete set of images that have a full description of the anatomy visualized. This information is worldwide used as reference, for validation purposes and as guideline for the interpretation of one's own findings. Digital mouse atlases are currently available for the brains [5–11], the limbs [12] and even the whole body [13,14].

### 1.1.2 Generation of transgenic mice

There are several ways to modify the genotype of mice [15,16]. Here the two most common techniques are shortly explained:

#### *Microinjection*

The female reproductive cells, the oocytes are harvested mixed with sperm from the male and in a petri dish. After one spermatozoon has entered the oocyte, it takes a few hours before the pronuclei of the two cells fuse and become a so-called zygote. In that period, the linear DNA sequences of the foreign genes are typically injected by microinjection into the male pronucleus [17]. For a microinjection, special needles and cell-holders are used which are roughly 0.5 - 5  $\mu\text{m}$  in diameter (see figure 1.1 ). After the microinjection, the oocytes are placed into the uterus of a pseudopregnant female mouse. If the integration of the gene with the DNA was successful, the offspring will express the new gene.

#### *Injections of embryonic stem cells*

A blastocyst is the very first stage of the embryo, consisting of a group of cells that will later form the embryo (embryoplast) surrounded by an outer layer of cells that will become the placenta (trophoblast). Cells that are taken out from the embryoplast are called stem cells and have the capability to develop themselves into almost any type of tissue. The DNA of these cells can be modified with high precision and will, depending on the technique, result in knock-out, knock-in or conditional mice [18]. After modification of the gene, the genetically modified embryonic stem cells are placed back in the embryoplast by microinjection. Now, the cells will behave exactly as the other cells in the embryo, resulting in a chimera mouse (figure 1.1), where the phenotype expresses a mixture of the modified genotype and the normal genotype, depending on which cells were descendants of the modified stem cells and which were

not. If the embryonic stem cells have contributed to the germ cells of the chimeras, then their offspring will all express the gene. The chimera mouse is mated with a mouse with the normal genotype, a so-called wildtype, half of their offspring will be heterozygous for the modified gene and the rest are wildtype mice. The offspring of the heterozygous mice results in mice that are either wildtypes, heterozygous for the modified gene, or homozygous for the mutated gene. The mice in the latter group all express the gene and will pass it on to their offspring, allowing a reliable production of genetically modified mice.

## 1.2 Mouse brain anatomy

The mouse brain is considered a valid model for human brain diseases [19, 20], since all brain structures that occur in the human brain are also present in the mouse brain and they are similarly connected to each other, although differently organized and in different volume proportions. Especially human neurophysiology and neuropathology can be well studied in mouse models [21, 22]. Human psychiatric disorders are less commonly studied as the cerebral cortex of the mice is not as highly developed as in humans [23]. Figure 1.2 displays a comparison between the human brain and the mouse brain, where a volume rendering of the whole brain (A) shows the lack of cortical folding in mice and a slice through both brains (B) shows a few corresponding brain structures.

## 1.3 High resolution magnetic resonance imaging

Due to the increasing amount of applications for transgenic mice, small animal devices are being developed that are capable of imaging at high resolutions ( $\sim 10\text{-}50\ \mu\text{m}$ ). For magnetic resonance imaging (MRI), this resulted in scanners with a high magnetic field, from 7 T, 9.4 T up to 17 T. MRI is a highly suitably imaging modality for brain imaging, as it is not based on ionizing radiation and therefore not damaging for the subject and it can acquire the images *in vivo* as well as *ex vivo*. Furthermore, MRI allows the acquisition of functional and anatomical scans with a wide range of imaging protocols all giving different information of the brain. The most commonly used imaging protocols for mouse brain imaging are:

### *T<sub>1</sub>-weighted imaging*

This protocol with long repetition times (TR) and short echo times (TE) has low contrast between gray and white matter, and is typically used with contrast agent, for example to visualize the vessel structure.

### *T<sub>2</sub>-weighted imaging*

This protocol has short repetition times (TR) and long echo times (TE), resulting in a relatively high contrast between gray and white matter and is therefore excellent for anatomical imaging of the brain.

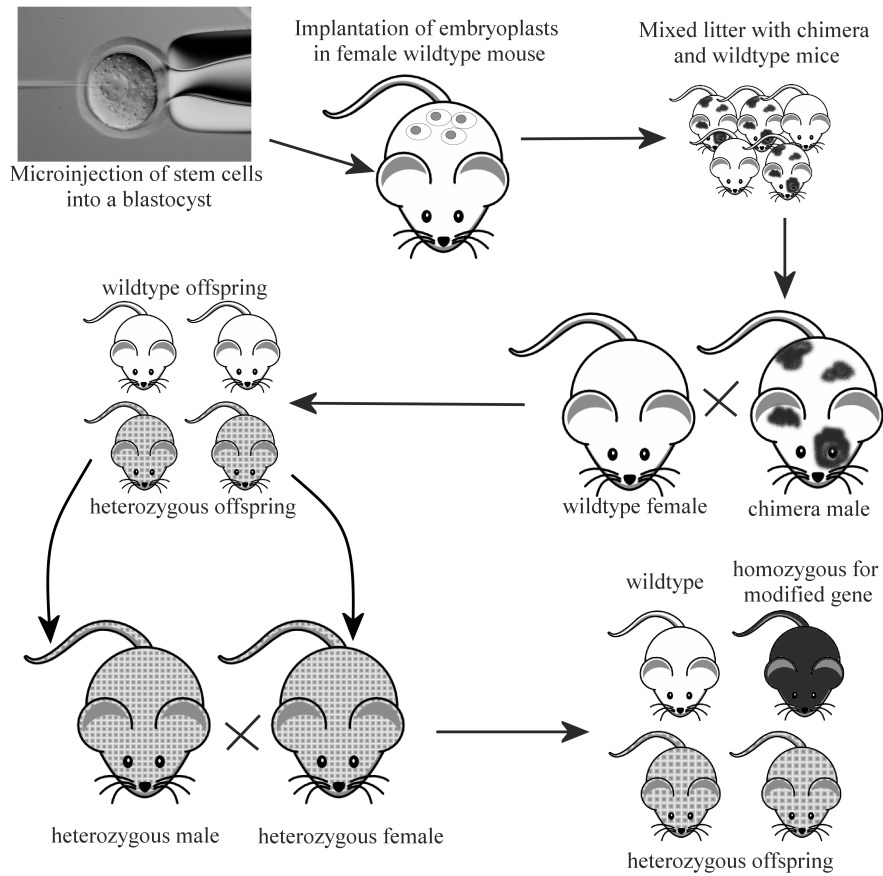


Figure 1.1: The process to generating transgenic modified mice through the implantation of embryonic stem cells, see section 1.1.2 for further details. Photography courtesy of Anne Bower and Manfred Baetscher, Transgenic Core, Oregon Health & Science University, Portland, OR. Printed with permission.

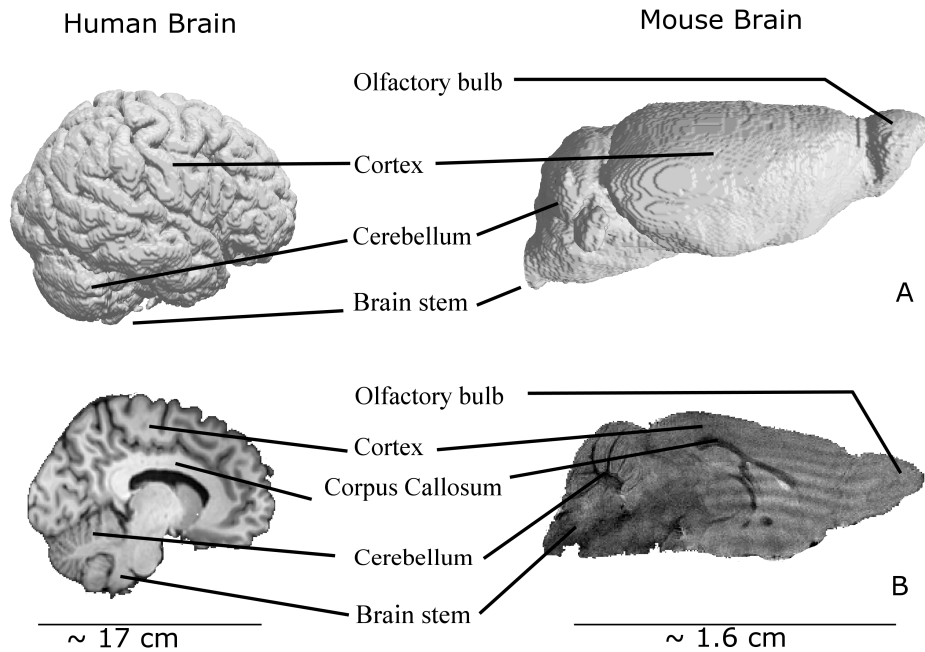


Figure 1.2: Comparison between the mouse brain and the human brain A) the outer surface and B) the internal anatomy.

#### *Blood-oxygen-level dependence (BOLD) imaging*

Protocol for BOLD imaging is sensitive for changes in the concentration of oxygenated hemoglobin, which will occur if a certain part of the brain has higher activity and requests an increased blood supply. Therefore, this technique is very suitable for functional imaging.

#### *Diffusion tensor imaging*

This protocol measures the diffusion of water in tissue. In brain tracts, the diffusion of water usually follows the direction of the tract and therefore, with diffusion tensor imaging the direction of the brain tracts can be visualized in high detail.

## 1.4 Aim of the thesis

To study the diseased brain, it is important to quantify local changes in the brain that occur as a result of the disease. The study of global or local shape variations in the brain is also defined as brain morphometry. In human brain MRI, automation of the morphometry process has already guided researchers to new insights regarding the (diseased) brain. With the development of MR systems for animal models, it is

now possible to acquire detailed anatomical *in vivo* images that offer the possibility to perform *in vivo* morphometry on the rodent brain. However, extending automated methods developed for human MRI to mouse brain MRI is not as trivial as it seems.

This thesis explores the possibilities for automated morphometry on MR images of mouse brains. The main contributions of this work are (a) to investigate the methods currently applied for quantitative morphometry in mouse brain MR images and (b) to provide analytical tools that can be used for the automated quantitative morphometry of mouse brains. Each chapter in this thesis is self-contained and therefore, some overlap between the chapters occurs.

## 1.5 Thesis outline

First, the topic of morphometry is introduced in **chapter 2** and an overview is given of the various morphometry methods and trends that are available for mouse brain MRI analysis. Furthermore, it is discussed which method is the most suitable for which situation and what the limitations and attention points are for automated morphometry. **Chapter 3** describes an application of automated morphometry on Alzheimer's Disease (AD) at an early stage in transgenic mouse brain MRI, it reviews (automated) methods in the literature. The most common transgenic mouse strains for AD are introduced, the several MR imaging parameters to detect plaques in the brain are discussed and an overview of the currently available automated methods capable of detecting AD is given.

As volumetry is the first step in quantitative morphometry, **chapter 4** presents an automated segmentation method for *in vivo* and *ex vivo* MRI, based on a hybrid method of affine registration and clustering. This clustering method is compared to manual segmentation and segmentation by nonlinear registration to evaluate its performance.

Automated morphometry is continued in the direction of deformation-based morphometry. **Chapter 5** presents the generalized Moore-Rayleigh test that tests high-dimensional vector fields for spherical symmetry and shows on simulated data how this nonparametric test can be applied to detect local brain differences between groups of transgenic mice. In **chapter 6** the Moore-Rayleigh test is further explored on experimental data of AD transgenic mice. Using synthetical and clinical data we show that the performance of the Moore-Rayleigh test outperforms the classical permutation test and significantly lowers the computational time as it is not dependent on the randomization of the data. In **chapter 7**, a clinical application of the Moore-Rayleigh test is shown in parallel with a volumetry study to phenotype a transgenic mouse model that exhibits migraine.

**Chapter 8** and **9** conclude this thesis with a summary and indications for further research in English and Dutch, respectively.

## CHAPTER 2

---

### Morphometry on rodent brains

---

A.E.H. Scheenstra  
J. Dijkstra  
L. van der Weerd

*This chapter was adapted from:*  
Volumetry and other quantitative measurements to assess the rodent brain, *In vivo NMR Imaging: Methods and Protocols*. Humana Press, USA. Ed. C. Faber and L. Schroeder. *in press*.

**Abstract:** *Morphometry is defined as studying variations in shapes and the detection of possible shape changes between groups. Evaluation of shape changes in the brain is a key step in the development of new mouse models, the monitoring of different pathologies and the measuring of environmental influences. Traditional morphometry was performed by volume measurements on manual shape delineation, the so-called volumetry. Currently, automated methods have been developed that can be roughly divided in three groups; voxel-based morphometry, deformation-based morphometry and shape-based morphometry. In this chapter we describe the different approaches for quantitative morphometry and how they can be applied to the quantitative analysis of the rodent brain.*



## 2.1 Introduction

MRI of the brain is increasingly used for standard phenotyping of transgenic mouse models, or for non-invasive monitoring of disease progression and treatment response. Quantitative analysis of the brain images is also referred to as brain morphometry, which is derived from the Greek  $\mu\omicron\phi$  (shape or form) and  $\mu\epsilon\tau\rho\omicron\nu$  (measurement) and is defined as studying variations and changes of different structures in the brain. The main research question in brain morphometry is how to determine significant differences between two groups of rodents, i.e. how to determine the brain shape differences between two groups of mice that are not the result of inter-subject variation in the brain, but caused by the differences between the two groups. For example, a diseased and a healthy group, or one group of rodents followed over time and measured at multiple time-points.

## 2.2 Volumetry

In mouse studies volumetry is traditionally the standard method to perform brain morphometry and is done by measuring the volume of the structure of interest (SOI) by delineation. Therefore, this method is often used as the gold standard in the presentation of new morphometry methods. In volumetry, a structure is delineated, either manually or automatically [24] and that segmentation is used to calculate the volume. The volume of each segmented structure is calculated by multiplying the number of voxels in the structure with the volume of a voxel. Since mice with larger brains have larger brain structures, the volumes are usually normalized to a percentage of the total brain volume before the comparison between mice can be made. Furthermore, partial volume effects will occur and therefore, volume calculation of small structures will be less accurate than that of large structures. The advantage of this method is that simple image processing methods are sufficient to perform volumetry, even though volumetry doesn't give any insight into *how* the shape changes.

## 2.3 Automated morphometry

This relatively new field of research analyzes the brain images locally to determine *where precisely* the two brain shapes differ from each other. Although the automated morphometry methods differ in the way the data is analyzed, the image processing pipeline is similar for all methods, and can be described by the following steps:

1. Describe the SOI by its features, such as the outer boundary of the structure being defined by landmarks or segmentation, intensity value, etc.
2. Extract these features for all images in the different treatment groups
3. Statistically test the features for a significant difference between the groups.

4. Present the results by means of a Statistical Parametric Map (SPM), which indicates local significance per voxel or reference point (Examples are shown in figure 6.6 and figure 7.3).
5. Possibly apply a multiple test correction (see appendix 2.A)

Based on their feature selection, the morphometry methods can be roughly divided into three groups; 1) Voxel-based morphometry (VBM) which calculates the gray and white matter density for each voxel and uses that for further analysis, 2) deformation-based morphometry (DBM) which warps all images to a standard reference and uses the resulting deformation fields for the analysis, and 3) Shape-based morphometry (SBM) which defines the shape based on the contour or landmarks. In this section the principles of each method are discussed.

### 2.3.1 Voxel-based morphometry

In VBM, the mouse brains are normalized to a reference image. A very smooth (not too accurate) non-linear registration step is applied for a better fitting of the brain structures to the template. The non-linear registration should be smooth enough to bring homologous regions as close together as possible, but not too accurate in order to avoid the homologous regions becoming identical. Afterwards the individual brains are segmented using a probabilistic method into different structures of interest based on the image intensity; thus each voxel is labeled with a posteriori probability of belonging to gray matter, white matter or cerebrospinal fluid. Statistical analysis is performed by applying a general linear model to retrieve a statistical parametrical map [25]. Incorporating a GLM has the advantage that covariates and confounders (e.g. age and total brain volume) can also be incorporated. The multiple-test correction which is applied in the software is based on the random-field theory [26]. The free SPM software package [27] is specially designed for voxel-based morphometry of human brains, and has lately been extended with a special module for rodent brains [28].

### 2.3.2 Shape-based morphometry

SBM is currently mainly applied to human brains [29–31] and is added to this chapter to complete the overview of possible methods. This method is especially useful to assess local changes within structures of interest, e.g. in the case of enlarged ventricles, to assess which parts of the ventricles are most affected [29]. To perform SBM, all structures are normalized to a standard reference image to correct for global brain size differences and brain orientations. Afterwards the SOI is segmented, either manually or automatically; only the surface of the brain structure is considered for further analysis of this segmentation. The surfaces are compared to each other using reference points and/or anatomical landmarks; these points need to be at the exact same anatomical locations.

In case of a local group difference, the spatial point clouds will differ between groups and that can be tested for significance with a statistical test, e.g. permutations

of a 3D Hotelling's  $T^2$  test and displayed in a SPM. Multiple test correction is required before any conclusion on the whole brain structure can be made.

### 2.3.3 Deformation-based morphometry

The name of DBM refers to the nonlinear (deformable) registration which is applied before the morphometry. Another term which is used within this framework is Tensor-based morphometry (TBM). The difference between TBM and DBM is found in the method of statistical analysis. To look for local differences in brain volume or shape, DBM uses the deformation vectors directly as they are obtained from the nonlinear registration of brain images, whereas TBM examines the Jacobian determinant (the spatial derivative of the deformation fields) and uses that for the statistical analysis. For both TBM and SBM, a normalization step is applied that globally registers the brains to a standard reference brain, the target image. Afterwards a nonlinear registration is applied in such way that the source image is warped exactly onto the target image. The resulting deformation field shows locally the changes that the source image had to undergo to fit the target, thereby indicating the differences between the source and target. The chosen registration method has to be as accurate as possible and preferable diffeomorphic [32–34], which means that the registration method tries to preserve the biological shape. Statistical analysis of the vector fields is performed by:

*Direct comparison (the DBM methods)*

Statistical analysis is performed on the features that are directly taken from the vectors, e.g. their magnitude [35], or their vector length and direction [36] (this thesis).

*Jacobian Calculation (The TBM methods)*

The Jacobian is calculated from the deformation vectors, which is a measure of the volume changes produced by a deformation. If the determinant of the Jacobian has a value between 0 and 1, there is possible shrinkage of the tissue, if it is larger than 1 there is an increase of tissue volume. If the determinant of the Jacobian results in a negative value then there is a biologically impossible deformation. A SPM is obtained directly by applying a statistical test or by incorporating the deformation field into general linear models that also model the global variables, such as gender and age [37–39]. Another option is to use the volumetric changes to perform volumetry measurements of a complete structure [40].

## 2.4 Method comparison

The described methods are all suitable for quantitative morphometry. All methods have been standardized by using automated normalization and segmentation according to an imaging processing pipeline and are, therefore, in principle unbiased for brain size and observer. Furthermore, all automated methods are testing each voxel separately for significant differences and are thus capable of producing an SPM. An

overview of the characteristics of the methods is given in the table 2.1.

The differences between the methods determine the choice for which quantitative morphology method is most suitable for the performed study. If a segmentation of the SOI can be easily obtained by computer algorithms or by manual delineations, one may consider SBM or volumetry. Volumetry is a good option if the research question is only to detect SOIs that are significantly different between two groups. If a local effect in the brain structure is expected, SBM may be considered as it returns locally significant differences on the surface of the structure.

VBM and DBM are both capable of analyzing the whole brain, which is very suitable for general phenotyping. The choice between VBM and DBM is more subtle and is dependent on the method available in the lab and the preference of the researcher. Both methods produce an SPM, both need smoothing to handle noise in the images, both need a perfect normalization to avoid improper conclusions and both allow the usage of general linear models for the incorporation of global parameters. However, VBM is based on a segmentation which defines composition of brain tissue in amounts of grey matter and white matter and cerebrospinal fluids, whereas DBM uses the voxel intensity range as input for the nonlinear registration. The use of deformation vectors allows DBM to perform multivariate statistics per voxel, where VBM applies univariate statistics. Univariate statistics are less realistic, as they consider only one voxel at a time without the interaction with its neighbors. However, VBM is available as a ready to use software package [27], whereas DBM is only available as free code [36].

Property	volumetry	VBM	SBM	DBM
Automated	–	✓	✓	✓
analysis per structure	✓	–	✓	–
full brain analysis	–	✓	–	✓
statistical parametrical map	–	✓	✓	✓
normalization required	–	✓	✓	✓
segmentation required	✓	✓	✓	–
multi-variate analysis	–	–	✓	✓
Ready to use software	✓	✓	–	–

Table 2.1: An overview of the characteristics between the four morphometry methods.

## 2.5 Limitations to automated morphometry

Automated image processing methods are developed to save time during the analysis, to have a more robust overall performance, i.e. reducing observer variability, and to perform more and complicated analyses: it therefore allows analyses which are impossible to perform by hand. But for automation to work properly, a fixed

(imaging) protocol is required as unexpected artifacts may influence the results of the automated method. For instance, the excision of a mouse brain in *ex vivo* brain MRI causes to large deformations in the brain that are many times larger than the expected *in vivo* shape variations between subjects [6]. Furthermore, automated methods rely on reference images, and are optimized for a certain contrast (e.g. T<sub>1</sub>-weighted or T<sub>2</sub>-weighted scans). Unexpected input, like a different orientation, or a slightly different scan protocol, or an update in the MRI scanner software may seriously hamper automated analysis. Since animal MRI is still in development, researchers and MR system developers tend to change imaging protocols continuously in order to optimize image quality. This is one of the reasons why automated analysis is not as readily and frequently used in small animal research as in clinical settings.

Interpretation of morphometry results must be performed prudently. If a significant difference has been detected, it actually implies a significant difference in intensity between the two groups. This can be due to a morphological difference between the structures of the two groups or it can be caused by one or more errors during the image processing. Since each of the pre-processing steps in the automated morphometry method such as normalization, segmentation and/or non-linear registration to a reference image may all introduce errors leading to a significant result [41, 42]. Therefore, if a significant difference is detected the raw data, the automated segmentations, and the registered data have to be cross-checked carefully to determine whether the significant effect can be explained by other causes than shape differences. A complete guideline for reporting VBM studies, which is also applicable for SBM and DBM methods, has recently been published [43].

## 2.A Multiple-test correction

In most automated quantitative morphometry methods a certain hypothesis about group difference is tested for each voxel separately resulting in a  $p$ -value for each voxel. All these tests are, unless otherwise specified, independently performed tests.

If a general conclusion on the brain is made instead of several conclusions for individual voxels, Multiple-test correction is required [44]. Multiple-testing refers to the testing of more than one hypothesis at the same time, where each test has its own error margin. Combining these independent tests without correction results in unacceptable error margins.

**Example 1.** *2 groups of brain MR images from the same population are tested for group difference. The MR images have a  $256 \times 256 \times 128$  volume with 8,388,608 voxels. If all hypotheses are tested with  $\alpha = 0.01$ , on average 83,886 incorrect rejections of the null-hypotheses might appear by chance and thus 83,886 voxels are considered incorrectly as significantly different. If we don't correct for this effect we might draw the conclusion that groups from the same population are significantly different.*

Multiple-test correction can be performed in several ways, of which the following are advised for multiple-test correction in morphometry [25, 44, 45]:

*Bonferroni correction*

This is the most stringent and most straightforward correction. Bonferroni multiple-test correction avoids false rejection of the null-hypotheses with a probability of  $\alpha$ , but thereby severely increasing the chance of a type 2 error (false negatives). To correct for multiple-tests with Bonferroni, the null hypothesis for each voxel should be rejected if  $(\alpha/n) \leq 0.01$ , where  $n$  is the number of tests (which equals the number of voxels in the MR volume that is analyzed). This test is the best method for truly independent voxels, although for brain morphometry Bonferroni correction is usually too conservative, as the voxels in the brain usually are correlated with at least neighbor voxels.

*Random field theory*

As Bonferroni correction is too conservative for locally dependent voxels, random field theory is used to determine clusters of dependent voxels so that multiple-test correction is only applied on the clusters instead of the voxels. This method requires a smooth SPM, which means that its value changes gradually without sharp transitions of probability values.

*Resampling*

The resampling method uses permutation tests to determine the corrected  $p$ -values. A permutation test iteratively randomizes the two groups and tests if the original situation is significantly different from the randomized groups [46]. In general, this method has a high accuracy higher than the random field theory, but the resampling method is computationally much more expensive than the Bonferroni correction and the random field theory correction.

*False Discovery Rate*

The false discovery rate (FDR) is defined as the ratio of expected false positives in the test [47] which can be used to threshold the SPM [48]. Since it is as straightforward as the Bonferroni correction, but less conservative, it is often applied to multiple-test correction. However, recently it has been shown that the FDR rate cannot be directly used for voxel-based morphometry studies [49, 50]

## CHAPTER 3

---

### Prospects for early detection of Alzheimer's disease from serial MR images in transgenic mouse models

---

M. Muskulus  
A.E.H. Scheenstra  
N. Braakman  
J. Dijkstra  
S. Verduyn-Lunel  
A. Alia  
H.J.M. de Groot  
J.H.C. Reiber

*This chapter was adapted from:*  
Prospects for early detection of Alzheimer's disease from serial MR images in transgenic mouse models. *Current Alzheimer research*. 2009;6(6):503-18.

**abstract:** *The existing literature on the magnetic resonance imaging of murine models of Alzheimer's disease is reviewed. Particular attention is paid to the possibilities for the early detection of the disease. To this effect, not only are relaxometric and volumetric approaches discussed, but also mathematical models for plaque distribution and aggregation. Image analysis plays a prominent role in this line of research, as stochastic image models and texture analysis have shown some success in the classification of subjects affected by Alzheimer's disease. It is concluded that relaxometric approaches seem to be a promising candidate for the task at hand, especially when combined with sophisticated image analysis, and when data from more than one time-point is available. There have been few longitudinal studies of mouse models so far, so this direction of research warrants future efforts.*



### 3.1 Introduction

Alzheimer’s disease (AD) is an age-related neurodegenerative disease characterized by structural brain changes and cognitive dysfunction. Due to the aging in western societies, AD will pose a large psychological and economical burden in the future [51]. Early detection of AD is therefore of considerable interest, since pharmacological treatment can reduce the amyloid burden and atrophy of the brain [52, 53]. The atrophy in the brain causes structural changes, which are detectable by various non-invasive imaging modalities [54, 55] and such considerations have led to the development of new imaging methodologies, for example diffusion-weighted magnetic resonance (MR) imaging [56, 57] multiphoton microscopy [58] or positron emission tomography [59].

The detection of AD by MR imaging techniques [60] is conveniently studied in standardized mouse models [61–67]. Brain mapping techniques [68] can be used to quantify changes, for example in voxel-based morphometry [27, 69], and more involved approaches estimate diffeomorphic changes in local brain structure [70] or construct local surface models [31] from volumetric measurements. Texture analysis is an interesting alternative [71] that has received little attention so far. The statistical analysis of MR images allows to discriminate between disease conditions [35, 72]. However, these analyses are often static, and do not usually incorporate knowledge about disease dynamics, molecular mechanisms [73, 74] or structural changes in time. The latter can in fact be inferred from longitudinal studies [75, 76], whereby animal models are employed favorably [77, 78].

Many extensive review papers have been written on Alzheimer’s disease in the past [65, 79–84]; it is not our intention to duplicate previous efforts. However, most reviews on AD concerned with small animal imaging focus on the development of mouse models or different scanning protocols to visualize plaques. In this paper we review the existing work on early *detection* of AD from serial MR images of transgenic mice, with special regard to the integration of dynamical information, i.e. how does (a) knowledge about AD dynamics from longitudinal studies, (b) knowledge about developmental changes in brain structure and (c) knowledge about disease processes at the molecular level help in the detection process? In particular, statistical and quantitative image analysis methods are addressed, and we subdivide them into volumetric approaches, relaxometric approaches, methods based on plaque burden evaluation, and methods based on texture analysis. Finally we give some recommendations for further research, by indicating gaps in the literature, interesting research directions and problems still to be solved.

### 3.2 Alzheimer mouse models

Several of the genes involved in the development of familial AD have been isolated in human studies. These genes have been used to develop a wide variety of transgenic mouse models, all displaying one or more of the characteristic pathological features

of the disease [79]. The most common lesions are schematically depicted in figure 3.1: Senile plaques arising from amyloid-beta ( $A\beta$ ) accumulation and inflammatory processes involving glial cells, neurofibrillary tangles (NFT) involving tau protein from the cytoskeleton of affected neurons, and vascular lesions caused by amyloid-beta deposits in cerebral arteries. The characteristics of several lines of transgenic mouse models [85–106] is given in table 3.1. A more extensive description of these lines is given in Appendix 3.A. Not all available mouse models are described, but those models which have either significantly advanced the understanding of AD pathogenesis or are otherwise in widespread use. This overview is adapted from the work of McGowan et al. [82], and expanded upon with information obtained from the Alzheimer Research Forum<sup>1</sup>.

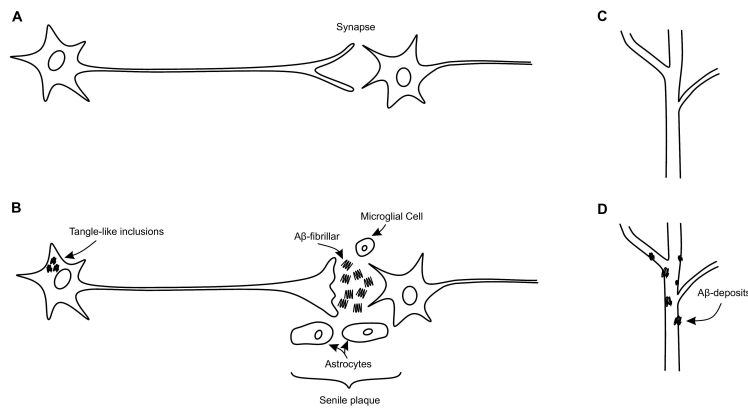


Figure 3.1: Pathological features of Alzheimer’s disease: Normal neuron and synapse (A). Affected neuron in late-stage (B). Normal cerebral artery (C). Affected cerebral vessel (D).

The work of Benveniste et al. [62] showed in 1999 that it is possible to visualize plaques in *ex vivo* samples of human brain by means of MR imaging. *In vivo* imaging of plaque deposition in human brains has so far not been successfully implemented. Visualization of plaques in mouse models at high field strengths has been successful, with first *in vivo* results reported in 2003 by Wadghiri et al. [107]. Since then, several groups have attempted to visualize plaque burden *in vivo* in different transgenic strains of mice, both with and without the aid of contrast agents. Furthermore, the development of plaques with age in individual mice has been successfully tracked using *in vivo* high resolution magnetic resonance imaging [77]. To date, the most commonly used AD models in this line of MRI research are doubly transgenic APP/PS1 strains [64–67, 78, 107–113], followed by singly transgenic APP strains [67, 77, 78, 107, 114, 115]. PS1 mouse models are occasionally used as controls,

<sup>1</sup><http://www.alzforum.org/res/com/tra/>

in addition to non-transgenic animals, as these animals have elevated  $A\beta$  levels, but no  $A\beta$  deposits.

### 3.3 Relaxometry

In addition to anatomical or pathological features, several intrinsic MR parameters can be studied to determine the effect of disease progression. In relaxometric approaches, the  $T_1$  (longitudinal, or spin-lattice) and  $T_2$  (transverse, or spin-spin) relaxation rates are commonly studied to facilitate the quantification of disease processes.  $T_1$  specifies the rate at which the net magnetization returns to its equilibrium state along the axis of the magnets bore, while  $T_2$  specifies the rate at which the net magnetization in the transverse plane returns to zero after RF excitation. Alternate relaxation parameters are  $T_2^*$  and  $T_{1\rho}$ . Unlike  $T_2$ , the parameter  $T_2^*$  is influenced by magnetic field gradient inhomogeneities and its relaxation time is shorter than the  $T_2$  relaxation time. The spin lattice relaxation time constant in the rotating frame,  $T_{1\rho}$ , determines the decay of the transverse magnetization in the presence of a "spin-lock" radiofrequency (RF) field [64].

Since both  $T_2$  and  $T_1$  relaxation times are sensitive to changes in biophysical water content it has been hypothesized that the presence of  $A\beta$  deposits in the brain has an effect on these parameters [110]. As such they might be used as independent markers for changes occurring in tissue, averaged over a region of interest (ROI). In fact, even pathological changes below the MRI resolution, i.e. at the subvoxel level, could in principle be detected, as parameter values of a single voxel are the result of an averaging process (partial volume effect). Several groups have studied the effects of the progression of AD on the transverse relaxation rate  $T_2$ ; there is a general consensus that the  $T_2$  values of affected brain tissue are lower than in controls, and decrease further as AD progresses [77, 78, 109, 110].

The analysis of relaxometric data in murine models of AD was first reported by Helpert et al. in 2004 [110]. In their work APP/PS1 mice were compared to PS1 mice and non-transgenic littermates.  $T_2$  values were found to be significantly lower in the cortex, hippocampus and corpus callosum, when comparing doubly transgenic animals to PS1 and non-transgenic mice, but  $T_1$  values did not show significant differences between the three genotypes. Falangola et al. [109] studied APP/PS1, PS1 and non-transgenic mice at two different ages. In addition to reporting a decrease in  $T_2$  in the APP/PS1 mice, compared to the others, the authors performed image registration to correctly compare specific regions of the brain between the different mice and age groups. In the study by Vanhoutte et al. [115]  $T_2^*$  values were calculated for the cortex and thalamic nuclei in APP<sub>V717I</sub> mice, which were compared to values in wild type mice.  $T_2^*$  values in the cortex were found to be the same in both groups, but decreased in the ventral thalamic nuclei of transgenics. Braakman et al. studied Tg2576 mice and non-transgenic littermates, starting at 12 months and following them until the age of 18 months [77]. The average  $T_2$  values in the cortex and hippocampus of transgenic mice were found to decrease with age. Significant

Model	Transgene (mutation)	Promoter	Phenotype					Reference
			DP	AP	NFT	ND	Cog	
PDAPP	APP <sub>V717F</sub>	PDGF	+	+	-	-	+	[91]
Tg2576	APP <sub>695(K670N,M671L)</sub>	PrP	+	+	-	-	+	[94]
APP23	APP <sub>751(K670N,M671L)</sub>	Thy1	+	+	-	+	+	[88, 89, 103]
APP <sub>717I</sub>	APP <sub>717I</sub>	Thy1	+	+	-	+	+	[97]
APP <sub>V717F</sub> × ADAM10-dn	APP <sub>V717F</sub> ADAM10-E384A-HA	Thy1	+	+	-	+	+	[100]
TgCRND8	APP <sub>695</sub> , APP <sub>V717F</sub>	PrP	+	+	-	?	+	[95]
PS1 <sub>M146V</sub> , PS1 <sub>M146L</sub>	PS1 <sub>M146V</sub> , PS1 <sub>M146L</sub>	R1 ES cells	-	-	-	-	-	[90]
PSAPP(Tg2576×PS1 <sub>M146L</sub> , PS1-A246E + APP <sub>SWE</sub> )	PS1 <sub>M146L</sub> , APP <sub>695</sub> PS1-A246E, APP <sub>695</sub>	PrP + PDGF	+	+	-	-	+	[87, 93]
APP <sub>Dutch</sub>	APP <sub>E693Q</sub>	Thy1	-	-	-	+	?	[92]
BRL-A $\beta$ 40	BRL-A $\beta$ 40	MoPrP	-	-	-	-	-	[96]
BRL-A $\beta$ 42	BRL-A $\beta$ 42	MoPrP	+	+	-	-	-	[96]
JNPL3	Taup <sub>301L</sub>	MoPrP	-	-	+	+	-	[106]
Taup <sub>301S</sub>	Taup <sub>301S</sub>	Thy1	-	-	+	+	-	[85]
TauV <sub>337M</sub>	TauV <sub>337M</sub>	PDGF	-	-	+	+	+	[104]
TauR <sub>406W</sub>	TauR <sub>406W</sub>	MoPrP	-	-	+	+	?	[105]
rTg4510	Taup <sub>301L</sub>	CAMKII	-	-	+	+	+	[101, 102]
Htau	Human PAC	Tau	-	-	-	-	-	[86]
TAPP(Tg2576×JNPL3)	APP <sub>695</sub> , Taup <sub>301L</sub>	PrP + PrP	+	+	+	?	?	[116]
3xTgAD	APP <sub>695</sub> , Taup <sub>301L</sub> , PS1 <sub>M146V</sub>	PrP + PrP	+	+	+	+	?	[98, 99]

Table 3.1: The characteristics of transgenic mouse models of AD. Transgene: PAC, P1 artificial chromosome. Promoters: PDGF, platelet-driven growth factor; PrP, prion protein; MoPrP, mouse prion protein; CAMKII, calcium/calmodulin-dependent protein kinase II. Phenotype: DP, diffuse (pre-amyloid) plaques; AP, amyloid plaques; NFT, neurofibrillary tangles; ND, neurodegeneration; Cog, cognitive impairment. For phenotype: +, positive; -, negative; ?, unknown.

decreases of  $T_2$  were not observed in controls. Borthakur et al. studied  $T_{1\rho}$  values in the cortex, hippocampus and thalamus of APP/PS1 mice and controls at ages 6, 12 and 18 months [64].  $T_{1\rho}$  values decreased in both the transgenic and nontransgenic groups as age increased, however the decrease was significantly more pronounced in the transgenic animals. El Tannir El Tayara et al. studied both  $T_1$  [117] and  $T_2$  [117,118] relaxation rates in APP/PS1 mice, with PS1 animals serving as controls. They found that  $T_2$  values in the subiculum of adult APP/PS1 mice were significantly lower than in PS1 mice and could thus serve as an early marker. Young mice (16-31 weeks) without histochemically detectable iron showed  $T_2$  changes, which may indicate that  $T_2$  variations can be induced solely by aggregated amyloid deposits. Falangola et al. studied the changes of  $T_2$  in a large group consisting of APP/PS1, APP, PS1 and non-transgenic controls [78]. This study revealed that only the APP and APP/PS1 groups show significant changes in  $T_2$  compared to non-transgenic controls. Table 3.2 presents an overview of relaxometric research in AD mouse models.

The statistical analysis of relaxometric data in its simplest form is based on summary statistics over a region of interest (ROI), which is usually much larger than the resolution achieved, encompassing a number of voxels on the order of ten or more. To compare the values of these variables between subjects and over the course of time (in one or more subjects), the images need to be registered with respect to each other. Between groups of subjects affected by AD and control subjects, there exist significant differences between relaxometric rates.  $P$ -values can be derived from the empirical standard deviation by assuming normality of the underlying population and relating this to Student’s  $t$ -distribution. Given a large enough population one can even analyze the dependence of the relaxometric data on further parameters, for example gender or behavioral data, by the more general analysis of variance (ANOVA) or general linear models. However, the assumption of normality can be problematic, especially for the limited number of mice usually included in the studies under review [119]; so one better resorts to nonparametric tests such as the Mann-Whitney U test or the computation of Spearman’s rank correlation coefficient. If three or more time points exist, linear regression is usually used, but nonparametric, nonlinear techniques can be more powerful. Permutation tests, in particular, allow the computation of exact  $p$ -values for the hypothesis that the summary statistics change in the course of time [120]. To our knowledge, the latter has not yet been applied to the analysis of relaxometric data of AD. Of course, suitable generalizations of ANOVA and linear regression also exist, in the form of generalized linear models (GLM) or mixture models [121,122].

Ultimately, i.e. for a successful clinical application, the detection of AD should be so robust, and the signal-to-noise ratio so large, that the correct choice of statistical model will be largely irrelevant. At present, however, and especially in the analysis of longitudinal studies, the choice of a correct statistical model is important to increase the sensitivity and to prevent one from drawing the wrong conclusions.

Ref.	Relax. Par.	Species	N.mice (Tg/Ntg)	N <sub>t</sub>	Stat. anal.	Results	Age	Lt.
[110]	T <sub>1</sub> , T <sub>2</sub>	APP/PS1 PS1	9+9/9	1	Student's <i>t</i> -test	T <sub>2</sub> lower in Tg mice than in Ntg. No significant changes in T <sub>1</sub> detected.	16-23m	-
[109]	T <sub>2</sub>	APP/PS1 PS1	9+9/9; 6/6	1	?	T <sub>2</sub> decreased APP/PS1 mice compared to controls	18m; 6w	-
[115]	T <sub>2</sub> *	APPV717I	4/4	1	N/A	Differences noted between Tg and Ntg	24m	-
[77]	T <sub>2</sub>	Tg2576	5/5	4	Student's <i>t</i> -test	Decrease with time detected	12-18m	+
[64]	T <sub>1,ho</sub>	APP/PS1	2/2	3	Student's <i>t</i> -test	Significant decrease if age >12 m	6, 12, 18 m	-
[117]	T <sub>1</sub> , T <sub>2</sub>	APP/PS1 PS1	10/9; 13/13	2	Pearson, Mann-Whitney, Wilcoxon tests	Negative correlation between T <sub>1</sub> and age in APP/PS1 animals. T <sub>2</sub> in the subiculum of adult APP/PS1 animals was lower than in PS1 mice	27-45w; 60-86 w	-
[118]	T <sub>2</sub>	APP/PS1 PS1	11/10	1	Mann-Whitney U test	T <sub>2</sub> is reduced in the subiculum of APP/PS1 mice; T <sub>2</sub> is an early <i>in vivo</i> marker of amyloid deposition	16-31w	-
[78]	T <sub>2</sub>	APP/PS1, APP, and PS1	64+33+ 61/48	3	Mixed model	Significant decrease in APP and APP/PS1 mice	6 w - 19m	+, -

Table 3.2: Relaxometry measurements in AD mouse models. Relax.Par, Relaxation parameters; N.mice, number of mice included in the study; Tg, transgenic; Ntg, non-transgenic; N<sub>t</sub>, Time points; Stat. anal., statistical analysis; Lt., longitudinal study; m, months; w, weeks

### 3.4 Analysis and models of plaque burden

Ever since Hardy and Higgins stated that the development of  $A\beta$  plaques is the main cause of Alzheimer's disease, leading to neurofibrillary tangles, cell loss, vascular damage, and finally resulting in dementia [123]; this theory has been discussed and supported by other findings [124–126]. As mentioned before, the development in plaque burden is still acknowledged as the primary biomarker of Alzheimer's disease. As Zhang et al. [67] showed, comparing histologically stained plaques with microimaging data (8–24 hrs acquisition time), senile plaques can in principle be reliably identified in *ex vivo*  $T_2$ -weighted MR images. However, numerous smaller plaques were not identifiable by visual inspection of the MR images. Later studies have shown that *in vivo* and *ex vivo* visualization of both individual plaques and total plaque load can be achieved by MR techniques in reasonable scan times without the aid of contrast agents [64–66, 77, 110–113, 115, 127–129]. An overview of the relevant studies of plaque burden in murine brain tissue is shown in Table 3.3.

In general, amyloid plaques are only visible on MRI scans in the later stages of the plaque development. For example, plaque sizes in 12 month old APP/PS1 mice are  $19\ \mu\text{m}$  on the average [130, 131], whereas the average voxel size in a MRI slice is around  $50\times 50\times 200\ \mu\text{m}$ , which is further discussed in [63, 111, 127]. Therefore, automated, direct detection and analysis of amyloid plaques on MRI scans is useful for analyzing the progression of amyloid deposition, but it cannot be used for early detection algorithms. Of course, indirect detection is still a possibility, since small changes in tissue formation are detectable with the MRI scanner because of the partial volume effect: insufficient image resolution leads to a mixture of the MR parameters of different tissues within a single voxel. In other words, plaques influence the recorded average relaxation rate per voxel proportionally, even in the case that the amount of amyloid deposition is smaller than the sampling volume per single voxel. However, a specific threshold in size for a plaque to be detectable at a prescribed confidence level is not known at present. The analysis of plaque burden by direct imaging could contribute to the latter by supplying the necessary data to set up a more sensitive parametric image model. To this extent, plaque burden analysis has focused on the statistical properties of senile plaques.

In principle, the locations at which plaques appear can be statistically modeled as a spatial point process [132, 133]. However, plaques are spatially extended objects that aggregate, grow and change their shape over the course of time. Stanley and co-workers therefore considered plaques as connected clusters and have found that the cluster sizes in AD human patients follow a log-normal distribution [133]. Moreover, they analyzed the spatial correlation function  $C(r)$ , i.e. the (normalized) probability of finding another plaque cluster at a distance  $r$  from a given cluster [74]. Comparison with randomized surrogate data allowed them to define a characteristic cluster size that changes from about  $14\ \mu\text{m}$  at 8 months to  $22\ \mu\text{m}$  at 12 months. Moreover, the size of individual plaques has been inferred to be roughly constant in time, with a characteristic length of  $1.3\ \mu\text{m}$ , indicating that disease progression consists mainly in accumulation and aggregation of individual plaques.

Following this analysis, Stanley et al. have built a mathematical model for the aggregation and disaggregation of senile plaques on a discrete lattice, i.e. as a random field [73]. This stochastic model also incorporates sudden plaque formation. The latter is consistent with recent evidence that plaques can form rapidly, even within 1-2 days [125]. A more detailed model, incorporating inflammatory processes as well, has been developed by Edelstein-Keshet and co-workers [134].

Shortly thereafter, a chemotactical model emphasizing the role of microglia in the aggregation of senile plaques has been investigated, that unfortunately does not capture the observed dynamics well [135]. The distribution of plaques and microglia, however, seems to be in agreement with observations [136]. For a discussion of microglia in the context of mouse models, see the reviews in [137] and [138]. Imaging of plaques has been addressed in [139], where a mathematical model for the kinetics of PET molecular imaging probes that bind to plaques is proposed.

MR images of senile plaques can be modeled by Markov random field models (or more generally, stochastic image models), where the values of each voxel are considered realizations of a probabilistic process  $X_{ij}$ , indexed by coordinates  $i$  and  $j$  in 2D. For simplicity, these processes are assumed Markovian; to be more precise, the conditional probability  $P(X_{ij}|X_{kl}, (i, j) \neq (k, l))$  is determined by the distribution of its direct neighbor voxels only:

$$P(P(X_{ij}|X_{kl}, (i, j) \neq (k, l)) := P(P(X_{ij}|X_{kl}, (i, j) \neq (k, l), |i - k| \leq 1, |j - l| \leq 1).$$

Alternatively, such a process is characterized by a Gibbs distribution, i.e. a potential energy associated with each realization (image) [140]. Medical applications of this methodology are mainly found in image segmentation up to now, e.g., of lung tissue or anatomical regions in brain images. In particular, a usable parametric random field model of plaque distributions in brain tissue is still lacking. A different approach to the analysis of plaque distributions in images is the language of fractals, where an image is considered to consist of morphological features that are self-similar, exhibiting the same structural properties at more than one scale. In [141] the authors have found that cortical blood vessel structure, evaluated with fractal-based morphological descriptors, can be correlated with AD pathology. Among other things, estimates of correlation dimension in Alzheimer patients showed smaller values than in controls.

### 3.5 Cerebral amyloid angiopathy

Alzheimer's disease is a multi-factorial disease that can be associated with cerebrovascular lesions in addition to the aforementioned plaques, the formation of NFT and brain atrophy. In fact, such lesions are often correlated with neurodegeneration. De la Torre and Mussivand suggested in 1993 that a disturbed brain microcirculation can cause Alzheimer's disease [143] and further studies confirmed that the reduced cerebral blood flow (CBF) that accompanies AD correlates well with the severity of dementia [144, 145]. A possible cause of CBF abnormalities in AD is cerebral amyloid angiopathy (CAA). This particular form of vascular pathology is caused by the deposition of  $\beta$ -amyloid protein in cerebral vessels [146–148].



Ref.	Image modality	Field Strength	Strain	N.mice (Tg/Ntg)	Age	Long.	<i>in vivo</i>
[107]	2D/3D T <sub>1</sub> SE; 2D T <sub>2</sub> SE; 2D T <sub>2</sub> * GE	7 T	APP and APP/PS1	5/5 ( <i>ex vivo</i> ) 7/7 ( <i>in vivo</i> )	15-16; 5-6m	-	+, -
[67]	T <sub>2</sub> SE	9.4 T	APP/PS1 and APP	2+1/2	15.5 m	-	-
[113]	T <sub>2</sub> FSE	7 T	APP/PS1,PS1	2+1/1	17-19 m	-	-
[110]	T <sub>2</sub> FSE	7 T	APP/PS1,PS1	9+9/9	16-23 m	-	-
[111]	T <sub>2</sub> SE, T <sub>2</sub> * GE	9.4 T	APP/PS1	?	24-26 m	-	+
[112]	T <sub>2</sub> SE	9.4 T	APP/PS1	?	3, 6, 9, 12,24 m	-	+
[115]	3D T <sub>2</sub> * GE	7 T	APP <sub>V7171</sub>	4/4	24 m	-	+
[114]	3D FSE <sup>19</sup> F, T <sub>1</sub> GE	9.4 T	Tg2576	?	16, 23 m	-	+
[108]	3D T <sub>2</sub> * GE	4.7 T	APP/PS1	?	28, 39 w	-	-
[77]	T <sub>2</sub> FSE	9.4 T	Tg2576	5/5	12-18 m	+	+
[64]	T <sub>1rho</sub> GE	4.7 T	APP/PS1	2/2, 2/2, 2/2	6, 12, 18 m	-	+
[65]	T <sub>2</sub> SE	9.4 T	APP/PS1	6	12 m	+	+
[142]	T <sub>2</sub> * GE, 3D T <sub>1</sub> GE,T <sub>2</sub> SE	4.7 T	APP/PS1, PS1	32/36 (long. 7/4)	27-103 w	+	+
[128]	2D/3D GE, 2D/3D SE, CRAZED	17.6 T	APP <sub>V7171</sub> ×ADAM10-dn	3/5	16 m	-	+, -
[129]	3D T <sub>2</sub> * GE	7 T	APP/PS1, Tg2576	20/10	6-8 m; 18-20 m	-	+

Table 3.3: MR micro-imaging of senile plaques and plaque burden in humans and mouse models. Image modality: 2D/3D, 2- or 3-dimensional; T<sub>1</sub>/T<sub>1rho</sub>/T<sub>2</sub>/T<sub>2</sub>\* applied weighting in MR imaging experiments; DW, diffusion-weighted; <sup>19</sup>F, imaging of Fluorine-19 labeled contrast agent; GE, Gradient Echo; SE, Spin Echo; FSE, Fast Spin Echo; CRAZED, COSY revamped with asymmetric z-GE detection. Long.: indicates whether the study was longitudinal

Multiphoton microscopy with a contrast agent showed that plaque development progresses seemingly linearly in Tg2576 mice [149, 150], with an average increase of 0.35% per day in vascular involvement, i.e. vessel area affected. In APP<sub>SWE</sub>/PS1 mice, CAA progresses slower with a slope of 0.17% per day [151].

Magnetic resonance angiography (MRA) can be applied to visualize vascular structures. The MRA technique differs from MRI in that the signal of stationary tissue is suppressed, and the signal from flowing blood is made visible. MRA is commonly applied to study flow artifacts or defects, to determine whether the vascular structure has been compromised. As in AD neurodegeneration is commonly correlated with CAA, MRA might provide insight into a possibly altered blood supply to specific brain regions. Only a few MRA studies in transgenic mice have so far been reported; in 2003 Beckmann et al. [152] studied 10 APP23 and 10 control animals at ages 6-7, 11 and 20 months, and observed flow voids in the majority of large brain arteries of APP mice with increasing age, including severe defects such as the absence of one of the carotid arteries. In 2004 Krucker et al. [153] used MRA to non-invasively study the arterial vascular architecture of APP23 mice. Due to the limited spatial resolution of MRA, the *in vivo* studies were complemented by analysis of the vasculature using vascular corrosion casting. Both techniques revealed age-dependent blood flow alterations and cerebrovascular abnormalities in these mice. Thal et al. [154] used MRA to show blood flow alterations in the thalamic vessels of APP23 mice. CAA-related capillary occlusion in the branches of the thalamoperforating arteries of APP23 mice corresponded to the occurrence of blood flow disturbances. Similarly, CAA-related capillary occlusion was observed in the occipital cortex of human AD subjects more frequently than in controls.

### 3.6 Volumetric methods

Brain atrophy has been pointed out as a biomarker for the development of Alzheimer's disease in human patients with Mild Cognitive Impairment (MCI) [155–159]. Most studies reported neurodegeneration in the structures of the mesial temporal lobe, such as the hippocampus, parahippocampal gyrus and amygdala, as a result of Alzheimer's disease. Brain atrophy can be quantified and followed in time by performing volumetric measurements in MRI. Voxel-based morphometry is an essential step in these types of analysis [160], and it is crucial that the necessary image registration steps are performed correctly [41]. A review paper on this topic was recently published by Ramani et al. [57]. Although there is overwhelming evidence on the utility of volumetric biomarkers from human studies, most research in the development of transgenic mouse models has focused on models which develop A $\beta$  aggregation (diffuse plaques and amyloid plaques), usually combined with an emphasis on astro- and microgliosis. Only in the last decade, small animal research has turned towards the study of neurodegeneration of specific brain regions. Recently it has been found that the progression of amyloid deposition in APP mouse strains is correlated to a decrease in neurogenesis in the hippocampal region [161].

Brain morphometry in transgenic mouse models of Alzheimer’s disease is challenging due to the small size of the structures of interest in the brain and the low contrast between these structures. Although manual segmentation is still considered as the gold standard in morphometric studies, the variability in these findings is large [162]. This problem is overcome by automated segmentation methods, which not only reduce the amount of time needed for delineation, but also improve the objectivity and repeatability of the segmentation, especially when a brain atlas is used as a template [63]. MRI enables the creation of digital atlases to describe the anatomy of mice [163, 164] by averaging normalized MRI scans of a group of animals. This has been of considerable interest in the analysis of various phenotypes [10, 162, 165] and is used in the comparative analysis of both *in vivo* [9] and *ex vivo* MR images [6, 8]. An example of such an atlas is shown in section 4 in figure 4.2.

To study brain atrophy it is a prerequisite to compare images to an atlas and several studies employ registration algorithms to automatically perform this task. Nonlinear registration is generally superior to simple affine registration, although it is much more sensitive to noise and image distortion. MR images of high quality are therefore required for nonlinear registration [91]. Another way to study neuroanatomical differences between mouse strains, is the statistical analysis of landmark points after nonlinear registration, as employed by Chen et al. [35]. Falangola et al. [109] applied nonlinear registration techniques to quantify group averages of three distinct mouse strains and proved that nonlinear registration is able to detect small differences of *in vivo* MR images. In addition to cross-sectional studies, longitudinal studies can be performed. Verma et al. created a longitudinal map of the average brain development in multiple C57BL/6J mice [166], the so-called spatio-temporal heterogeneity map of brain development and maturation. Recently, Maheswaran et al. applied nonlinear deformation analysis to both *in vivo* cross sectional as well as longitudinal studies [167].

### 3.7 Texture analysis

The texture of an image is an elusive concept that can be roughly defined as its statistical properties at different levels of scale [164]. Above we have already discussed the characterization of plaque burden in terms of *stochastic image models*, which is a particular approach. Here we discuss three more branches of texture analysis (see also [71, 119, 168, 169]). The *statistical approach* is specifically targeted at discrimination purposes. To this respect, from a given image or a ROI a number of feature descriptors are computed. The classic example is gray-level co-occurrence matrices (GLCM) in 2D [170]. Let  $(i, j)$  be a displacement vector in 2D. For each possible pair  $(r, s)$  of gray-level values, i.e. discretized relaxation rates for our purposes, its number of occurrences in the image  $X$  is counted. All such co-occurrences define a symmetric matrix:  $C(r, s | i, j) = |\{(k, l) | X_{kl} = r, X_{k \pm i, l \pm j} = s\}|$ . Two examples are shown in Fig. 3.2, where for simplicity the MR parameter has been discretized at 8 levels. For each of these matrices, distinct statistical measures can be defined. The energy of

the matrix is the sum of its squared entries  $C^2(r, s | i, j)$ , and quantifies the image inhomogeneity. The contrast is the sum of  $|r - s|^2 \cdot C(r, s | i, j)$  over all pairs  $(r, s)$ , and measures local image variations.

Other commonly used feature descriptors are entropy measures, which can also be directly estimated from images [171]. The latter has been applied to  $T_2$  images in a cuprizone mouse model, for example [172]. With respect to Alzheimer's disease, Freeborough and co-workers used a total of 17 feature descriptors, selected from an initial set of 260 descriptors, most derived from GLCMs, to classify and track the progression of the disease in  $T_1$  images of human brains [173]. Liu and colleagues used an initial set of 3456 descriptors to classify  $T_1$  images of humans [174]. The cross-validated accuracy exceeded 90 percent in both cases. Kovalev and colleagues discuss the use of discrete anisotropy measures to classify general cerebral pathologies in 3D [175], but do not perform a statistical analysis.

Another branch of texture analysis is the *signal processing approach*. In its simplest form an image is analyzed in frequency space, i.e. its discrete Fourier transform is the basis for discrimination based on the occurrence of specific frequency components or power changes. A more advanced method is the use of discrete Gabor or wavelet transforms to extract localized frequency information. An example of the latter is the classification of regions in  $T_1$ -weighted images of human knees with respect to tissue type [176].

More recently, *geometric methods* have been used to classify MR images. The main idea is to consider the image as consisting of a number of smaller texture elements whose distribution indicates changes in structural composition. This was demonstrated on human x-ray mammographic images, classifying them with respect to whether radiological findings were present or not, and this method appears suitable to the analysis of relaxometric data as well [177]. Table 3.4 summarizes the literature on the classification or detection of AD in human studies by texture analysis. Up to now this approach has not been used in mouse models, and only Freeborough et al. [173] consider a longitudinal approach (for the tracking of AD).

### 3.8 Discussion and conclusion

Summarizing the literature, we can conclude that tracking of relaxometric changes, supplemented by parametric image models and the analysis of image features, is a promising approach to the early detection of the characteristic features of Alzheimer's disease in mouse models.  $T_2$  relaxation times were uniformly found to be the best discriminator, whereas  $T_1$  could not sufficiently discriminate between mutants and their controls. Both changes in  $T_2^*$  and  $T_{1\rho}$  were found to correlate with aging as well, which warrants further research efforts.

Plaque burden analysis of *in vivo* MRI is comparable to the relaxometric approach. Figure 3.3 shows the age of mice in weeks for which features of AD are detected with *in vivo* MRI by several plaque detection methods (light grey bars) and by relaxometric methods (dark grey bars). Since the APP/PS1 mouse model features a more aggres-

Ref.	Imaging modality	N.Subjects AD/MCI/C	Time points	Methods	Accuracy
[173]	T <sub>1</sub> MRI	40/0/24 5/0/5	N.A. 2-6	17 statistical 2D feature descriptors	0.91 N.A.
[175]	T <sub>2</sub> MRI	11 14	N.A. 2	3D Texture anisotropy	N.A. N.A.
[174]	T <sub>1</sub> MRI (?)	20/20/20	N.A.	3444 + 12 statistical 2D feature descriptors	> 0.95
[141]	Histological data	?	N.A.	Morphological descriptors.	Different results in different areas of the brain

Table 3.4: Texture analysis applied to alzheimer’s disease on human data. AD: Alzheimers disease, MCI: Mild cognitive complainers, C: controls.

sive progression of pathology development, the results are grouped by APP/PS1 mice and the remaining mice (PS1, Tg2576 and further APP variants).

Direct imaging of plaque burden is very valuable in the creation and validation of mathematical models of plague aggregation. Ultimately, it is desirable to incorporate this knowledge into parametric image models, as this should allow for an increase of sensitivity in the detection of AD from relaxometric images. This is further substantiated by the success of texture analysis of MR images. Unfortunately, the few studies undertaken in this regard are phenomenological, and a truly convincing solution for the early detection of Alzheimer’s disease does not yet exist. Related to this is the important problem at which size senile plaques are detectable under a prescribed significance level.

As for the volumetric analysis of MRI data, research in human patients has shown that this approach allows the prediction of the development of AD in patients already suffering from mild cognitive impairment (which does not necessarily lead to AD). However, in small animal research this approach is still in an early phase of development. Obviously, it is difficult to detect and quantify cognitive impairment in animals (confer [178,179] though). The quantification of cerebral amyloid angiopathy by MRA also appears to be promising, especially in tracking the progression of the disease. However, this is again a mostly unexplored area. It has been demonstrated that early detection of AD is feasible by these two approaches, but the results as yet

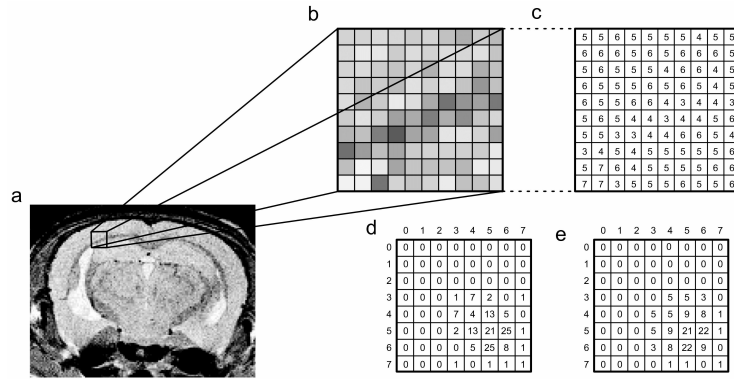


Figure 3.2: Statistical texture analysis of MR images: Original image (A). Enlarged  $10 \times 10$  subimage (B). Gray level representation with 8 levels (C). Co-occurrence matrix  $C(r, s|1, 0)$  of subimage for horizontal displacement (D). Co-occurrence matrix  $C(r, s|0, 1)$  for vertical displacement, which describes the statistical properties of the subimage with regard to local variations (E).

are not as convincing as when employing relaxometric data.

In general, MR imaging is very attractive due to its non-invasiveness and its ability to produce images of high quality, and thus very suitable to study Alzheimer's disease in transgenic mouse models. When performing *in vivo* imaging on transgenic mice environmental factors, such as the use of anesthetics, stress caused by the imaging process, or even the specific mouse strain used are all confounding factors that influence imaging results [180]. Therefore it is necessary to also study the influence of environmental factors in transgenic mouse models of AD, especially in longitudinal or cognitive studies, where the choice of a correct statistical model is important. If sufficient data is available, these effects can be modeled and estimated, for example in a GLM.

Automated analysis of MR images is a nontrivial task. The relatively large differences between scanners, the possibility of artifacts, and the large number of scanning parameters demand standardized imaging protocols and involved methods of image analysis. Automated analysis methods can overcome some of the problems associated with low spatial resolution, low signal-to-noise ratio and inter-group variability, but it seems that there is still a need for the development of new imaging protocols that are specifically targeted at the visualization of amyloid plaques and other symptoms of AD.

With regards to the literature, a striking general observation is that there are relatively few longitudinal studies, and almost no effort to utilize temporal information in the detection of AD. On one hand, it is not immediately obvious how to do this. On the other hand, the main problem in discrimination tasks is the following: there

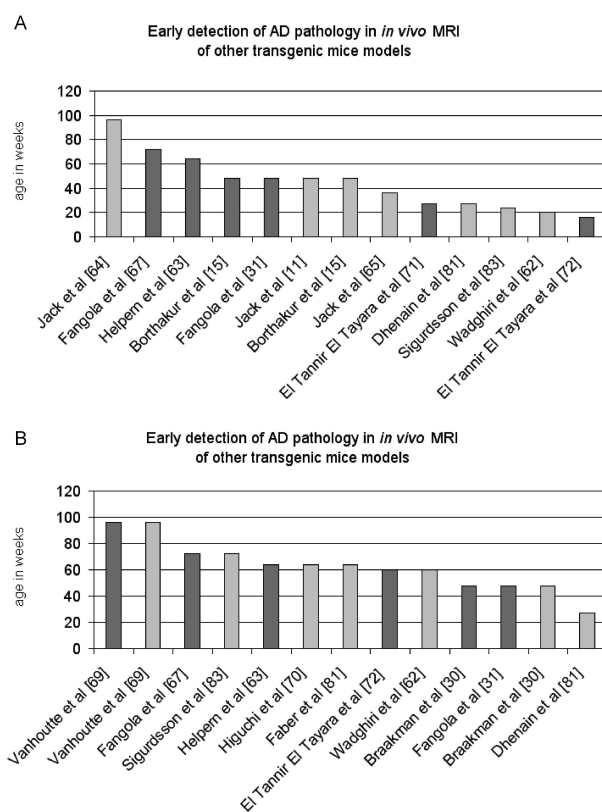


Figure 3.3: Overview of the minimum age at which Alzheimer pathology was detected in *in vivo* MRI volumes of APP/PS1 mice (A) and the remaining mouse models (B). Dark grey denotes a relaxometric method, light grey denotes that AD pathology was determined by plaque burden analysis. None of the studies used contrast agents during imaging.

do exist significant differences between groups of diseased individuals and groups of controls, as the above cited studies have shown. On the level of the individual, however, the (usually considerable) overlap between the two populations renders correct classification difficult. But by following an individual through the course of time, even if it is only a few measurements over the course of a few months, changes can be detected that are otherwise unnoticeable. We believe that this is the key to an early detection of AD, and expect that future studies will be conducted in this direction.

To conclude, small animal imaging will always be ahead of imaging in human patients, as small animals can be exposed to higher field strengths and their aging process is more rapid. This makes murine models the perfect testbed for the development of detection and screening procedures. Also, small animals provide a way to test and apply new treatment strategies and experimental medication. We should not forget, however, that the ultimate goal of our research efforts is its application to humans.

**Acknowledgements** The authors wish to thank Janneke Ravensbergen for her contributions to the review of MRA research in transgenic mouse models of CAA and AD.

### 3.A The most commonly used AD mouse models

PDAPP [91]:

The first mutant amyloid precursor protein (APP) transgenic mouse model with robust plaque pathology. These mice express a human APP cDNA with the Indiana mutation (APP<sub>V717F</sub>). Plaque pathology starts between 6-9 months in hemizygous PDAPP mice. There is synapse loss, but cell loss and NFT pathology are not observed. This model has been used widely in vaccination therapy strategies.

Tg2576 [94]:

This model expresses mutant APP<sub>SWE</sub> under control of the hamster prion promoter. Plaque pathology is observed from approximately 9 months of age onwards. These mice have cognitive deficits but show no cell loss or NFT pathology. Tg2576 is one of the most widely used transgenic models.

APP23 [88, 89, 103]:

These mice express mutant APP<sub>SWE</sub> under control of the Thy1 promoter. Prominent cerebrovascular amyloid and cerebral amyloid deposits are observed from 6 months of age onwards. Some hippocampal neuronal loss in this model is associated with amyloid plaque formation.



TgCRND8 [95]:

Mice express multiple APP mutations (Swedish plus Indiana). Cognitive deficits coincide here with rapid senile plaque development at approximately 3 months of age. The cognitive deficits can be reversed by  $A\beta$  vaccination therapy.

PS1<sub>M146V</sub> and PS1<sub>M146L</sub> [90]:

These models were the first *in vivo* demonstration that mutant presenilin 1 (PS1) selectively elevates  $A\beta$ 42 levels. No overt plaque pathology is observed though.

PSAPP (Tg2576 $\times$ PS1<sub>M146L</sub>, PS1-A246E+APP<sub>SWE</sub>) [87,93]:

A bigenic transgenic mouse model which showed that addition of the mutant PS1 transgene markedly accelerates amyloid pathology compared to singly transgenic mutant APP mice, demonstrating that the PS1-driven elevation of  $A\beta$ 42 enhances plaque pathology.

APP<sub>Dutch</sub> [92]:

Mice expressing APP with the Dutch mutation, which causes hereditary cerebral hemorrhage with amyloidosis in humans, develop severe congophilic amyloid angiopathy. The addition of a mutant PS1 transgene redistributes the amyloid pathology to the parenchyma, indicating differing roles for  $A\beta$ 40 and  $A\beta$ 42 in vascular and parenchymal amyloid pathology.

BRI-A $\beta$ 40 and BRI-A $\beta$ 42 [96]:

These mice express individual  $A\beta$  isoforms without over-expression of APP [96]. Only mice expressing  $A\beta$ 42 develop senile plaques and CAA, whereas BRI-A $\beta$ 40 mice do not develop plaques, suggesting that  $A\beta$ 42 is essential for plaque formation.

JNPL3 [106]:

These mice express 4R0N microtubule associated protein tau (MAPT) with the P301L mutation. This is the first transgenic mouse model with a marked tangle pathology and cell loss, demonstrating that tau protein alone can cause cellular damage and neuronal loss. JNPL3 mice develop motor impairments with age owing to severe pathology and motor neuron loss in the spinal cord.

Tau<sub>P301S</sub> [85]:

This line of mice expresses the shortest isoform of 4R MAPT with the P301S mutation. Homozygous mice develop severe paraparesis at 5-6 months of age with widespread neurofibrillary pathology in the brain and spinal cord, and neuronal loss in the spinal cord.

Tau<sub>V337M</sub> [104]:

Mice express low level synthesis of 4R MAPT with the V337M mutation (1/10 of endogenous mouse MAPT) driven by the promoter of platelet-derived growth factor (PDGF). The development of neurofibrillary pathology in these mice suggests the nature of tau rather than absolute intracellular tau concentrations drives pathology.

**Tau<sub>R406W</sub>** [105]:

Mice express 4R human MAPT with the R406W mutation under control of the CAMKII promoter. These mice develop MAPT inclusions in the forebrain from 18 months of age onward and have impaired associative memory.

**rTg4510** [101, 102]:

Mice have inducible MAPT using the TET-off system. Abnormal MAPT pathology occurs from one month of age on. These mice show progressive NFT pathology and severe cell loss. Cognitive deficits are evident from 2.5 months of age onwards.

**Htau** [86]:

These transgenic mice express human genomic MAPT only (mouse MAPT knocked-out). Htau mice accumulate hyperphosphorylated tau from 6 months on and develop Thio-S-positive NFT by the time they are 15 months old.

**TAPP (Tg2576×JNPL3)** [116]:

mice have increased MAPT forebrain pathology when compared to JNPL3 mice, suggesting mutant APP and/or A $\beta$  can affect downstream MAPT pathology.

**3×TgAD** [98, 99]:

This is a triple transgenic model expressing mutant APP<sub>SWE</sub> and MAPT<sub>P301L</sub> on a PS1<sub>M146V</sub> 'knock-in' background (PS1-KI). This line develops plaques from 6 months on, and MAPT pathology from the time they are 12 months old, strengthening the hypothesis that neurofibrillary pathology can be directly influenced by APP or A $\beta$ .

**APP<sub>717I</sub>** [97]:

Mice express human APP cDNA with the London mutation (APP<sub>V717I</sub>). This strain displays decreased exploration, increased neophobia and increased male aggressiveness. Pathological features include amyloid plaques and cerebrovascular angiopathy with an onset around 10-12 months, and cholinergic fiber distortion.

**APP<sub>V717I</sub>×ADAM10-dn** [100]:

Double transgenic mice expressing both APP<sub>V717I</sub> and a proteinase of the ADAM (a disintegrin and metalloproteinase) family. Expression of ADAM10-dn leads to an enhancement of the number and size of amyloid plaques in the brains of these double-transgenic mice. However, compared to APP<sub>V717I</sub> mice, they exhibit improved performance in the Morris water maze test.

## CHAPTER 4

---

### Automated segmentation of mouse brains

---

A.E.H. Scheenstra  
R.C.G. van de Ven  
L. van der Weerd  
A.M.J.M. van den Maagdenberg  
J. Dijkstra  
J.H.C. Reiber

*This chapter was adapted from:*  
automated segmentation of *in vivo* and *ex vivo* mouse brain magnetic resonance images. *Molecular Imaging* 2009; **8**(1):35-44.

**abstract:** *Segmentation of MRI data is required for many applications, such as the comparison of different structures or time-points, and for annotation purposes. Currently, the gold standard for automated image segmentation is nonlinear atlas-based segmentation. However, these methods are either not sufficient or highly time consuming for mouse brains. This is due to the low signal-to-noise ratio and low contrast between structures compared to other applications. We present a novel generic approach to reduce processing time for segmentation of various structures of mouse brains, in vivo as well as ex vivo. The segmentation consists of a rough affine registration to a template followed by a clustering approach to refine the rough segmentation near the edges. Compared to manual segmentations, the presented segmentation method has an average kappa index of 0.7 for 7 out of 12 structures in in vivo MRI and 11 out of 12 structures in ex vivo MRI. Furthermore, we found that these results were equal to the performance of a nonlinear segmentation method, but had the advantage of being 8 times faster. The presented automatic segmentation method is quick, intuitive and can be used for annotation and volume quantification of structures.*

## 4.1 Introduction

The versatility of MRI techniques makes animal MRI suitable for the identification of new disease biomarkers and evaluation of novel diagnostic or therapeutic agents, similar to clinical MRI [61, 181]. Studying neurological disorders in mouse models often requires segmentation to perform either phenotyping or morphometry. Although sophisticated automated methods assist in the analysis of the full mouse brain [28, 35], segmentation or delineation of the structures of interest (SOI) is necessary to evaluate *which* structure is involved and exactly *how* that structure changes.

Manual segmentations are, although tedious, generally considered as the golden standard for brain annotations. Automation of the segmentation process has some advantages above manual delineations, such as repeatability and standardization. Since animal MR scanners are still in the developmental phase, automated segmentation in mouse MR images is still very challenging, in contrast to the automated segmentation of human brain MR images [182]. Most algorithms developed for the human brain segmentation are not directly applicable to mouse brain images, as Tohka et al. [183] recently presented. These segmentation problems in mouse brain MRI are mostly due to artifacts caused by the MRI scanner, deformations caused by the excision of the brain in the case of *ex vivo* imaging and, most importantly, less contrast between brain structures and a lower signal-to noise ratio compared to human MRI.

Segmentation of mouse brain MRI for experimental studies is generally performed by nonlinear registration of an annotated atlas to a subject, for which the segmentation is manually refined afterwards [6, 8, 9, 38, 165]. In these studies, no segmentation performance is reported. A completely automated segmentation method based on nonlinear registration was presented by Rohling et al., which reached a segmentation accuracy of 90% overlap with manual contours in bee brain MRI [184]. The method consists of the nonlinear registration to several atlases which are combined by an Expectation-Maximization classification method. The success rate of this method is very dependent on the amount of atlases available. Another promising fully automated segmentation method is based on probabilistic intensity information or intensity patterns of various *ex vivo* imaging protocols, which were known beforehand [162, 185, 186]. With this approach the automated segmentation had on average 90% overlap with manually drawn contours. Its advantage is that no computational expensive registration methods are required, although the usage of various imaging protocols might be time consuming as well.

In this paper, a new, fast, and automatic segmentation method is presented that produces segmented images of *in vivo* and *ex vivo* mouse brains based on a single atlas, imaged by a single imaging protocol. We first applied a fast affine atlas registration to a template to obtain a rough initial segmentation that was then refined by a clustering algorithm. As already stated by Tohka et al. [183] regular clustering algorithms, such as fuzzy k-Means Clustering [187] and Markov random field models [188] fail to segment the volume properly, mainly because there is lack of contrast between the structures. Therefore, a more specialized clustering algorithm is required, which we present in this study. The presented clustering algorithm combines intensity values,

the class labels of the neighboring voxels, and edge information. This information is given to the clustering algorithm by means of a template.

The algorithm is tested on *in vivo* and *ex vivo* mouse brain MRI volumes. For both volumes different structures are segmented based on the visibility and contrast of the structures in the volumes. Furthermore, the performance of the algorithm is validated by comparison to manually drawn expert contours. The *in vivo* MRI segmentations are also compared to automated segmentations obtained by a nonlinear segmentation method. For this approach, the MRI volume is nonlinearly registered to the atlas by the Demons registration method [189].

## 4.2 Materials and methods

### 4.2.1 Experimental setup

C57Bl/6J mice ( $n = 5$ ) were first imaged by MR *in vivo* on a Bruker 9.4 Tesla scanner using a  $T_2$ -weighted multi-slice spin echo sequence with TR/TE=6000/35 ms (4 averages). The *in vivo* volume had a matrix size of  $256 \times 256$ , with 40 slices, resulting in a resolution of  $97.6 \times 97.6 \times 200 \mu\text{m}$  per voxel. The total scan time was 102 minutes. Afterwards, mice were sacrificed and the brain their brains were the skull and perfusion-fixed with 4% phosphate-buffered paraformaldehyde (PFA). Prior to *ex vivo* MR imaging, brains were incubated for 8 hours in 4% PFA containing 12.5 mM gadolinium-tetraazacyclododecanetetraacetic acid (Gd-DOTA , Dotarem, Guerbet, Roissy, France). *Ex vivo* imaging was performed using a  $T_1$ -weighted 3D-gradient echo protocol, with TR/TE=17/7.6 ms and flip angle 25 degrees. The total scan time was 10 hours. The *ex vivo* volume had a matrix size of  $256 \times 256 \times 256$  and an isotropic resolution of  $78.1 \mu\text{m}$  per voxel. Figure 4.1 displays the pipeline of the presented segmentation algorithm with its two main steps: the registration to an atlas and the clustering. Also, the required input for the algorithm is displayed. In the following, the various brain structures of interest are denoted as classes. The automated segmentation for each image took on a single Pentium-IV 3.4 GHz processor approximately 30 minutes for the *ex vivo* volume and 15 minutes for the *in vivo* volume. The difference in calculation time between *ex vivo* and *in vivo* volumes is due to the differences in number of voxels.

### 4.2.2 Template creation

In atlas-based segmentation, or model-guided segmentation, a new MRI volume can be segmented if it is registered to an atlas. The atlas contains all prior information on the average volume and spatial organization, which is useful to avoid biologically impossible solutions. The best representative atlas for image segmentation and normalization is an unbiased atlas, which means an atlas that is constructed by averaging scans of multiple subjects in an independent coordinate system and is not dependent on inter-subject changes [5]. If insufficient subjects are available for the creation of an unbiased atlas, an approximation can be made as presented by Thompson and

Toga [163], who mapped an unknown brain to a database of normal brains to acquire an accurate segmentation.

In this study, a limited number of subjects were available which excluded the possibility to create an average unbiased atlas. So, we had to work with a template; a single segmented volume that was selected from the *in vivo* and *ex vivo* images from the dataset. Although an unbiased atlas is desirable, a template would suffice for this purpose. The clustering algorithm is applied after the affine registration, adjusting the initial segmentation until a perfect individual segmentation is reached. Due to the differences in intensity values, contrast and noise between *in vivo* and *ex vivo* MR images, we used both *in vivo* and *ex vivo* images of the template. Furthermore, the number of manually segmented structures was also dependent on the visibility of those structures. For the *ex vivo* atlas, 15 structures were segmented: The cortex, midbrain-hindbrain, cerebellum, olfactory areas, hippocampal formation, caudoputamen, thalamus, corpus callosum, hypothalamus, fornix system, corticospinal tract, substantia nigra, ventricles, anterior commissure - olfactory limb, and the anterior commissure - temporal limb. For the *in vivo* atlas 12 structures were segmented: The cortex, midbrain-hindbrain, cerebellum, olfactory areas, hippocampal formation, caudoputamen, corpus callosum, fornix system, substantia nigra, ventricles, anterior commissure - olfactory limb, and the anterior commissure - temporal limb. A coronal, sagittal and transversal view of the atlas is given in figure 4.2.

### 4.2.3 Registration

Each volume in the dataset is affine registered to a manual segmented template which provides an initial segmentation. Since the registration is an intermediate step in the segmentation algorithm, a fast and rough registration of the template to the new volume is sufficient. For this purpose, we used a registration algorithm composed of an affine transform with mutual information as image metric that was optimized by a regular step gradient optimizer [190, 191] as implemented in National Library of Medicine Insight Segmentation and Registration Toolkit (itk) [192]. When the registration has finished, the segmentation of the atlas is affine transformed and mapped on the new MRI volume as an initial segmentation. In addition to acquiring an initial segmentation, the atlas was also used to derive prior information on the intensity distribution for each class as input for the clustering algorithm. Furthermore, the initial segmentation is used to remove the skull and surrounding tissue of the *in vivo* MRI volume, so the clustering algorithm will not be distracted by those.

### 4.2.4 Edge-based clustering

After the atlas-based registration is completed, the clustering algorithm is applied. This clustering is necessary, since the affine registration results in an initial segmentation that only accounts for global differences between the new volume and the atlas image. The clustering corrects the segmentation for local changes caused by inter-subject variation and deformations in the *ex vivo* mouse brain caused by the physical

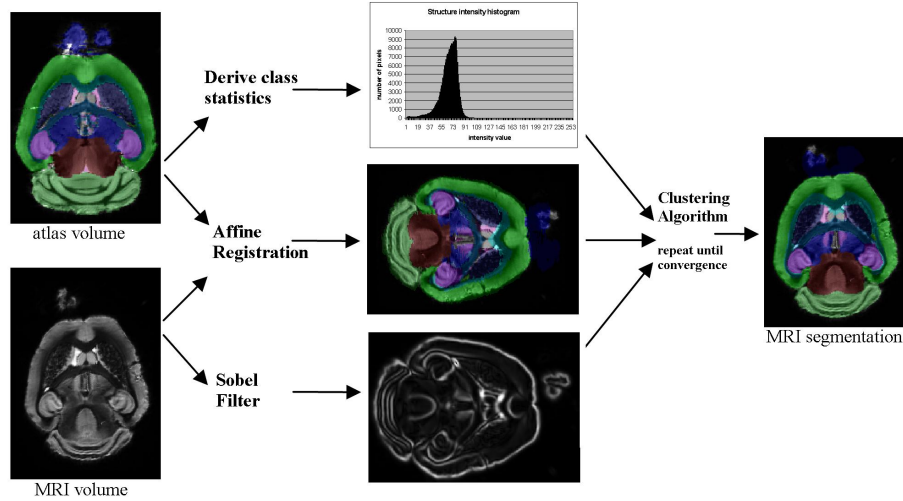


Figure 4.1: The segmentation pipeline for the mouse brain segmentation algorithm. The Atlas volume is first registered to the new MRI volume, resulting in an initial segmentation. Furthermore, the intensity distributions per class are derived from the atlas and the edge information from the new MRI volume is extracted. The clustering algorithm is performed in the second step for a final segmentation; therefore it uses the class statistics, initial segmentation and edge information as input.

excision and preparation of the brain.

It is assumed that in an MR image, each voxel  $x$  has a probability  $p$  to be a part of a certain brain structure (class  $c$ ). The presented clustering algorithm uses information on the intensity distribution and information retrieved from the  $N$  neighbors of  $x$  to evaluate for each class  $c$  and assigns  $x$  to the class with the highest probability. This process is iterated until it converges to a stable solution. The convergence level is defined as the minimum percentage of voxels changing label in a single iteration. This predefined percentage of voxels has to be set by the user. The algorithm usually finishes between the 5 and 10 iterations, dependent on the convergence threshold set by the user and the quality of the initial segmentation. The latter is provided by the first step of the algorithm, the affine registration, as described in the section above.

The presented algorithm needs three inputs, as can be seen in figure 4.1; (a) an initial segmentation, as given by the global atlas-based registration; (b) the intensity distributions per class as derived from the atlas, and (c) the edge information of the new MRI volume. As can be seen in formula 4.1 the clustering is separated in two main components; the first one,  $P_{intensity}$ , is based on the intensity distribution of the various classes and second part ( $P_{neighborhood}(c|x, n \in N)$ ) is based on information



retrieved from the neighborhood of voxel  $x$ . The knowledge on the intensity distributions is derived from the atlas. The initial segmentation and the edge information are used to calculate the neighborhood influence.

$$p(c|x, n \in N) = (1 - \alpha)P_{intensity}(c|x) + \alpha P_{neighborhood}(c|x, n \in N) \quad (4.1)$$

with  $0 \leq \alpha \leq 1$ . The weight  $\alpha$  is used to tune the algorithm for the various contrast-to-noise ratios and signal-to-noise ratios, depending on the imaging protocols of the MRI scanner. If the image volume has very high contrast, the emphasis may lie on the probability from the intensity, so  $\alpha$  should be smaller than 0.5. If the data is very noisy, the probability calculated from the intensity is less reliable. In this case,  $\alpha$  should be put higher than 0.5 since the edges can still be found correctly by using an anisotropic smoothing filter. The first part of the clustering algorithm, the  $P_{intensity}(c|x)$ , is used to incorporate the intensity distribution of the various classes. It measures the relative distance of each voxel  $x$  of the complete volume  $X$  to the class mean intensity of each class  $c$  ( $\bar{x}_c$ ), where the shortest distance has the highest probability of assigning the  $x$  to  $c$ :

$$P_{intensity}(c|x) = 1 - \frac{(x - \bar{x}_c)}{\sum_{(x \in \bar{x}_c)^2} (x - \bar{x}_c)} \quad (4.2)$$

The second part of the probability function, the  $P_{neighborhood}(c|x, n \in N)$ , models the dependency on the neighboring voxels. The influence of the neighbors is weighted by the edge information obtained from the original image, since the initial segmentation is usually erroneous near the edges, especially when the segmentation is found by global atlas registration. Therefore, the neighbors that are located inside a structure have more influence than the neighbors located on or close to a (strong) edge. The edges are found by a standard Sobel filter  $S(x)$  and, afterwards, the intensities of the image are scaled to range from 0 to 1. For this algorithm, we use a second order neighborhood, which means all voxels located next to  $x$  in a horizontal, vertical, and diagonal direction are included in  $N$ , thus resulting in a neighborhood of  $N = 3^3 = 27$  voxels, including  $x$  itself. The  $n_c$  in formula 4.3 symbolizes that for each class  $c$ , neighbors can only contribute if they are also classified to  $c$ .

$$P_{intensity}(c|x, n \in N) = \sum_{n_c \in N} \frac{1 - S(n_c)}{N} \quad (4.3)$$

#### 4.2.5 Validation

For validation purposes, all brains were manually segmented in concordance of two experts, who used the LONI mouse brain atlas [10] as guidance. This is a standardized mouse brain atlas from the Laboratory of Neuro Imaging at the University of California. The structures which were selected for manual and automated segmentation, were selected by the experts based on their visibility in the MR images. The

results of the automated segmentation method were validated by comparing them to the manual segmentations of experts by means of the kappa index  $\kappa$ , as given in eq. (4.4). The kappa index is a measure that represents a ratio of the amount of overlap to the total number of voxels of an automatically segmented brain structure  $V_a$  and a manually segmented brain structure  $V_m$ .

$$\kappa = \frac{2(V_a \cap V_m)}{V_a + V_m} \quad (4.4)$$

This measure is robust to changes in volume size and therefore very suitable to compare the automated and manual segmentation. A  $\kappa$  of 1.0 indicates total overlap of two volumes, where a  $\kappa$  of 0.0 shows no overlap at all. In an inter-observer study [162] it was shown that an automated segmentation method was performed equally as well as human observers if the kappa indices between 0.7 and 1.0 can be achieved. An overall validation score of the algorithm is obtained by averaging the kappa indices per structure for all volumes in the dataset.

## 4.3 Results

### 4.3.1 Automated segmentation results

For the segmentation of the *ex vivo* images, the algorithm used on average 6 iterations to reach the threshold when less than 0.5% of the voxels changed label, while for the *in vivo* segmentation only 3 iterations were needed for convergence to a 5% threshold. The different settings for the convergence threshold is a consequence of the different voxel sizes of the mouse brain in the *ex vivo* and *in vivo* images; which is respectively 127,655 voxels and 833,800 voxels. The automated segmentation for each image took on a single Pentium-IV 3.4 GHz processor approximately 20 minutes for the *ex vivo* volume and 15 minutes for the *in vivo* volume. The difference in calculation time is due to the differences in number of voxels.

The average kappa indices are calculated for all automatically segmented structures and displayed in Table 4.1. Also given are the volumes of the segmented structures in voxels. As stated in the previous section, an automatically segmented structure with  $\kappa$  larger than 0.7 can be assumed to be segmented with reasonable accuracy. If we consider the *in vivo* automated segmentation, the algorithm reached a satisfying result for 7 of the 12 structures with an average  $\kappa$  of 0.7. In the case of the *ex vivo* segmentations, 12 out of 15 structures are correctly segmented with an average  $\kappa$  of 0.85. These results imply that the automated segmentation method for *ex vivo* MR images is comparable to manual segmentations, as Ali et al. has shown with an intra-observer study which reached an average  $\kappa$  of 0.85 [162]. The automated segmentation method from Sharief et al. [186] outperforms the presented method with an overall kappa index of 0.95. However, their method is only applicable to *ex vivo* MRI, since they use various imaging protocols whereas the presented method is also applicable on *in vivo* MRI.

structure name	<i>in vivo</i> segmentation			<i>ex vivo</i> segmentation		
	Volume (mm <sup>3</sup> )	N. voxels	$\kappa$	Volume (mm <sup>3</sup> )	N. voxels	$\kappa$
cortex	157.07	82443	0.884±0.005	124.11	260280	0.884 ± 0.001
midbrain-hindbrain	94.68	49696	0.918 ± 0.09	76.29	159995	0.924 ± 0.027
cerebellum	51.92	27250	0.896 ± 0.005	50.93	106801	0.904 ± 0.014
olfactory areas	24.95	13098	0.822 ± 0.010	22.57	47341	0.739 ± 0.230
hippocampal formation	21.67	11377	0.826 ± 0.014	22.06	46268	0.916 ± 0.019
caudoptamen	20.15	10577	0.813 ± 0.020	20.2	42353	0.906 ± 0.021
thalamus	—	—	—	26.35	55260	0.888 ± 0.078
corpus callosum	13.12	6887	0.578 ± 0.028	18.72	39260	0.808 ± 0.006
hypothalamus	—	—	—	9.96	20884	0.86 ± 0.072
fornix system	4.85	2548	0.53 ± 0.038	6.14	12886	0.721 ± 0.038
corticospinal tract	—	—	—	4.38	9183	0.709 ± 0.033
substantia nigra	1.93	1012	0.687 ± 0.021	1.26	2634	0.729 ± 0.181
ventricles	8.43	4427	0.778 ± 0.039	2.66	5579	0.508 ± 0.051
anterior commissure olfactory limb	0.52	275	0.425 ± 0.216	1.45	3047	0.557 ± 0.039
anterior commissure temporal limb	0.36	190	0.322 ± 0.110	0.22	461	0.401 ± 0.062

Table 4.1: The sizes in mm<sup>3</sup> and in voxels for the several brain structures. Furthermore, for the separate structures are the *in vivo* and *ex vivo* segmentation results given in average kappa indices and standard deviation. The missing values of the *in vivo* segmentation column represent structures for which no proper expert segmentation could be obtained and thus were excluded from the atlas and automated segmentation.

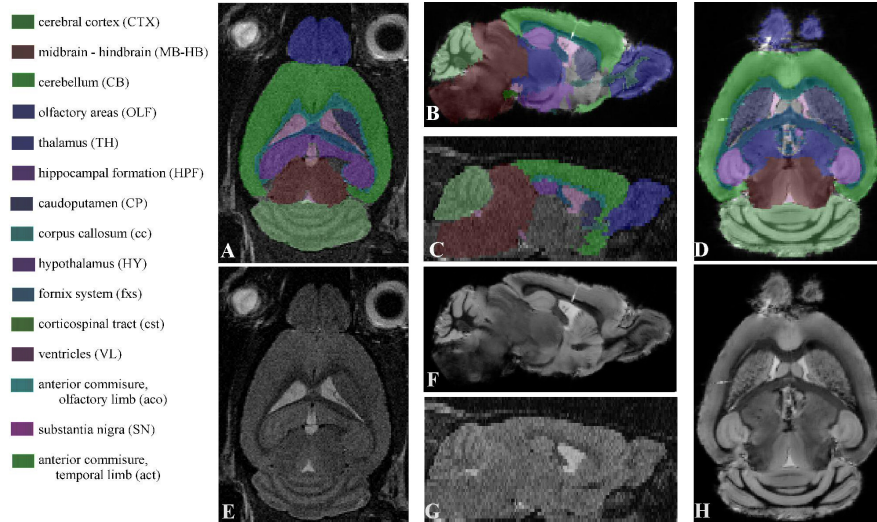


Figure 4.2: The MRI atlas used for *in vivo*(E,G) and *ex vivo*(F,H) MR images with their manual segmentations (A,B,C, and D) and corresponding names and abbreviations. For a better understanding, the abbreviations for the brain structures are in upper case where the abbreviations for brain tracts are in lower case.

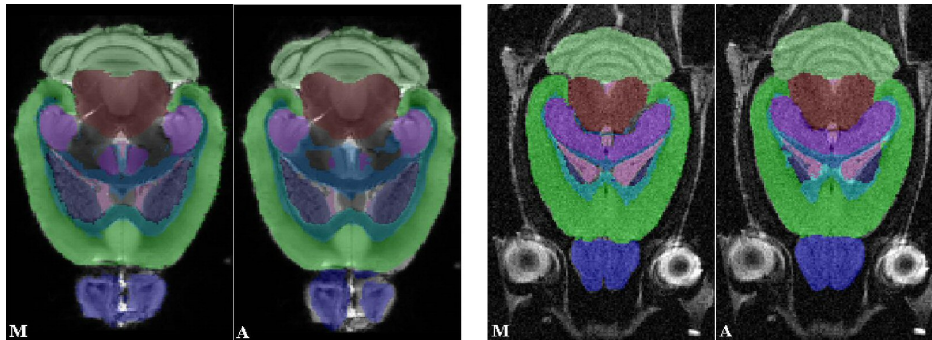


Figure 4.3: A visual comparison between the manual (M) and automated (A) segmentation for *in vivo* MRI (left) and *ex vivo* MRI volumes (right). The colour-coding of the various classes correspond to the legend as given in figure 4.2.

If the number of voxels are compared to the final  $\kappa$ , it can be seen that the performance of the algorithm decreases with the size of the structure to be segmented. This is especially true for the brain tracts included in the segmentation algorithm: the corpus callosum, corticospinal tract, and the anterior commissures. To illustrate the differences in segmentation for the *in vivo* and *ex vivo* images, an automatically and manually segmented slice are displayed in figure 4.3. Most structures have an overall better segmentation result in the *ex vivo* images, due to a better contrast-to-noise ratio and a higher resolution. The effect of these parameters is clearly visible if one considers e.g. the corpus callosum. However, some brain structures suffer from major deformations caused by the excision of the brain, impairing an accurate automatic segmentation of *ex vivo*. Examples are the olfactory areas, which are easily damaged during excision, or the ventricles which often collapse post mortem and therefore have a smaller volume leading to worse segmentation results for the *ex vivo* images. In figure 4.3, one can clearly see differences in proportion for the ventricles in *in vivo* and *ex vivo* images.

In figure 4.4 we presented the kappa indices of the *ex vivo* segmentations after the first step of the algorithm (the affine atlas-based registration) and its second step (the clustering algorithm). This figure shows that only for the three brain tracts, the fornix system and anterior commissures, a decrease in  $\kappa$  is obtained after the clustering is performed. For the anterior commissures, this decrease is also found significant. The  $\kappa$  increases for all other structures and although this increase seems unimportant and small in the figure, we found a significant increase in  $\kappa$  for all structures except the cortex, midbrain-hindbrain, and caudate putamen. For these three structures the initial segmentation is already quite accurate, leading to minor adjustments by the clustering algorithm. These minor corrections may not be a significant improvement, but are still important since these small changes are actually corrections for the inter-subject variations. The structures, for which a significant improvement of segmentation was found, are the structures which have a less accurate initial segmentation compared to the cortex, midbrain-hindbrain and caudate putamen and therefore need more correction by the clustering algorithm.

### 4.3.2 Nonlinear atlas-based segmentation

To compare the performance of the proposed algorithm to atlas-based nonlinear segmentation, nonlinear registration of the annotated template to the subjects was performed by means of the symmetric Demons algorithm as implemented in itk [192]. This nonlinear registration method is based on a thermodynamic concept of diffusion [189].

The same procedure was followed for the presented segmentation method; the same *in vivo* image was used as atlas, as shown in figure 4.2, while the segmentation algorithm was evaluated on the other *in vivo* images. Before applying the Demons algorithm, the *in vivo* images were affine registered by the same algorithm as used for the atlas-based registration of the newly presented method. The average kappa indices are retrieved by comparing the results from the automated segmentation to

the manual segmentation of the *in vivo* mouse brain volumes. The results of the nonlinear registration to the atlas can be found in figure 4.5. As can be seen in the figure, the results of the algorithm are comparable with the results of the demons algorithm, where the clustering method reached convergence in 10 minutes and the Demons algorithm reached convergence in 2 hours on the same computer.

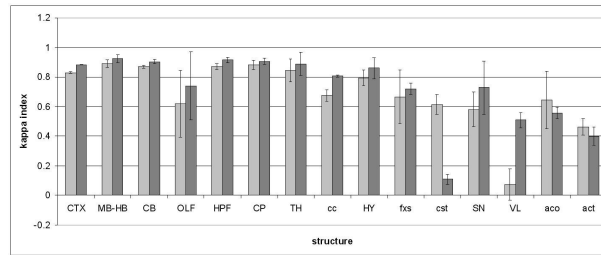


Figure 4.4: The increase in  $\kappa$  between the two steps of the presented algorithm for the *ex vivo* segmentation results. The light grey bar denotes the average  $\kappa$  after the affine atlas-based registration step, whereas the dark grey bar displays the average  $\kappa$  after the clustering step. Furthermore, the standard deviations are given for each bar to indicate the robustness of the algorithm. The abbreviations of the various structures are explained in figure 4.2

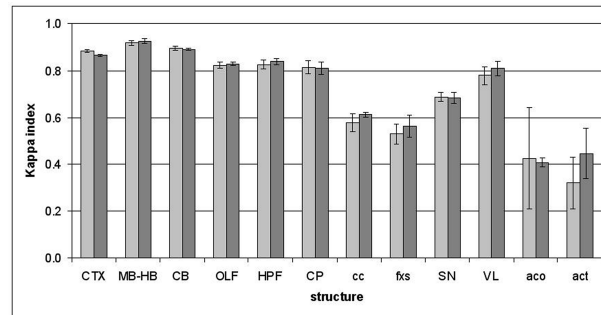


Figure 4.5: The average kappa indices of the Demons algorithm (light grey bars) compared to the average  $\kappa$  of the presented method (dark grey bars) for the *in vivo* mouse brain segmentation. Furthermore, the standard deviations are given for each bar to indicate the robustness of the algorithm. The abbreviations of the various structures are explained in figure 4.2

## 4.4 Discussion

As described in the results, the algorithm was found to segment the larger brain structures, e.g. cortex or cerebellum correctly. For these structures, the algorithm performs equally to the kappa indices from literature [162]. The results are also comparable for one of the most challenging structures in the *ex vivo* brain, the corpus callosum. This structure is challenging to segment because it is assumed to be large with an average of 37,990 voxels (18.11 mm<sup>3</sup>), but its flat and thin shape resembles more of a small structure. The algorithm has a high segmentation performance on structures which are affected by noise or deformations and have low contrast. This performance can be obtained by the usage of neighborhood voxels in combination with the edge information of the unknown brain image. A drawback of this method is, however, that the algorithm has less accurate segmentation on structures having a width of the image resolution. This can be seen in figure 4.4, where the three brain tracts, the fornix system and anterior commissures show a decrease in kappa index after the clustering is performed. This is due to (a) the usage of a single image as atlas for the segmentation. Small inter-subject variations result in a misregistration for the smaller structures in the brain, in such a way that there is no overlap and thus no seed point for the algorithm to segment this structure, (b) the partial volume effects that reduce the contrast between the neighboring structures. The voxels on the boundary of a structure have an intensity that is similar to its neighboring structure and have a higher chance at incorrect classification.

By comparing the *in vivo* to *ex vivo* segmentations, the algorithm returns a better segmentation for the *ex vivo* images, except for the olfactory areas and ventricles. The superior segmentation of the *ex vivo* segmentation can be explained by the lower resolution of the *in vivo* volume, resulting in more structures for which their widths are about the resolution of the image. As mentioned above, the algorithm encounters more difficulties when segmenting structures with a limited number of voxels. The superior segmentation of the olfactory areas and ventricles in the *in vivo* volumes can be explained by the local deformations occurring during the extraction of the brain from the skull. The olfactory areas are very loosely attached to the brain and in most cases were damaged or completely removed in the procedure, while the ventricles collapse if the brain is fixed. The latter is shown by table 4.1, where the volumes in mm<sup>3</sup> are given for all structures. Although the segmentation algorithm can compensate for these changes to some level, these deformations still cause some errors. The boundaries of some brain structures, e.g. the transition of the midbrain-hindbrain to the thalamus, are difficult to determine due to little contrast differences in the image between the structures. For these structures, the manual segmentation is also subjective and differs for each mouse brain. In this study, two experts were used to validate the manual segmentation and obtain a more objective segmentation. However, more information on the user variability is needed for these structures before some conclusions can be drawn on the quality of the automated segmentation of these structures.

The automated segmentation of the structures with poor edge information is very

dependent on the atlas, since the clustering is guided by the prior information given by the manual segmentation of the atlas. For these structures, the main differences between the automated segmentation and the manual segmentation are the differences in the transformed expert segmentation of the atlas mapped on the image on one hand and the manual segmentation of the new image as validation on the other hand. This raises the need for a publically available average atlas that also includes inter-variability and intra-variability information for these structures, as already developed for human brains [193]. We found that the presented segmentation method has similar performance as the nonlinear segmentation method. If compared to the expert segmentations, the performance of the presented method is more consistent. Manual segmentation is still considered the most reliable method although the intra-observer variation is on average higher than in automated algorithms. This is due to different interpretations of the various structures, as well as tiredness and weariness of the observer. Therefore, an automated segmentation algorithm not only reduces the amount of time needed to segment, but also improves the objectivity of the segmentation. Especially, when there is good contrast between structures, the automated segmentation algorithm will return a good and objective segmentation which is also repeatable.

The algorithm has one limitation caused by the extraction of prior knowledge on intensities and edges from the atlas. If the imaging protocols differ, incorrect intensity distributions per brain structure are derived from the atlas and do not represent the intensity distribution per structure. Since these distributions are used to guide the clustering, the clustering will result in an incorrect segmentation. So, it is required that the atlas is either acquired with the same imaging protocol as the image dataset, or has to be preprocessed by some intensity transform to map the intensities on the protocol of the new image dataset. In practice, the last method is the most likely choice, although errors made in the intensity mapping will induce errors in the segmentation of the volumes. If an atlas - or example segmentation - can be obtained, it is more likely that a better segmentation result is reached. Future work will also include a study on the segmentation of other types of MRI. We will investigate the performance of this segmentation method for other images, since no specific brain tissue information is used and consequently all the posterior information for the clustering is derived directly from the atlas. In summary, the presented method is a quick and promising segmentation method for mouse brain images, especially when major deformations of the tissue are absent. The smaller, local deformations in the brain tissue are corrected by the adapted clustering algorithm as a complement of the linear registration. This collaboration of both segmentation algorithms result in a quick and accurate segmentation method for *in vivo* and *ex vivo* mouse brain MRI, despite its low signal-to-noise ratio and artifacts. Finally, since no prior information has been used in this segmentation algorithm, this algorithm is highly generic and can be applied on various images without any difficulties.



## 4.5 Conclusion

The main objective of this study was to find a new, fast, and fully automatic segmentation method that produces segmented images of *in vivo* and *ex vivo* mouse brains based on a single atlas, imaged by a single imaging protocol. The presented method consists of an affine atlas-based registration combined with an edge refining clustering algorithm, where the clustering is supplemented by edge information and statistical information derived from the anatomical atlas. It is shown that the addition of the clustering algorithm improves the segmentation and is able to compensate for some nonlinear deformations in the *ex vivo* mouse brain. Where fully automated and highly accurate segmentation methods for *in vivo* and *ex vivo* mouse brains are extremely time consuming, e.g. by nonlinear registration, the presented method is quick and yet accurate enough for the segmentation of the principle structures needed for the registration.



## CHAPTER 5

---

### The generalized Moore-Rayleigh test

---

M. Muskulus  
A.E.H. Scheenstra  
R. Nabuurs  
J. Dijkstra  
L. van der Weerd  
S. Verduyn Lunel

*This chapter was adapted from:  
The generalized Moore-Rayleigh test. Submitted to computational statistics and data  
analysis*

**abstract:** *The Rayleigh test is a popular one-sample test of randomness for directional data on the unit circle. Based on the Rayleigh test, Moore developed a nonparametric test for two-dimensional vector data that takes vector lengths into account as well, which is generalized to arbitrary dimensions. In the important case of three-dimensional vector data the asymptotic distribution can be given in closed form as a finite combinatorial sum. This reduces the computational effort considerably. In particular, when analyzing deformation fields arising in nonlinear brain registration, the generalized Moore-Rayleigh test offers an efficient alternative to conventional permutation testing for the initial screening of voxels. Simulation results for a few multivariate distributions are given and the test is applied to brain scans of hydrocephalic transgenic mice. Compared with the permutation version of Hotelling's  $T^2$  test its increased power allows for improved localization of brain regions with significant deformations.*

## 5.1 Introduction

Consider the following illustrating example. In the deformation-based morphometry, individual brain volumes are mapped to a reference brain image by a nonlinear transformation to assess inter- or intra-variability of the brain structures [6]. The nonlinear image registration results in a three-dimensional vector field of displacement vectors. The significance of local deformations between groups of subjects, usually a treatment and a control group, can be tested by either considering the Jacobian of the deformation field, or testing the displacement vectors directly [194]. In the latter case, if one assumes that variations between subjects are given by a Gaussian random field, Hotelling’s  $T^2$  statistic can be used to test for significant differences between groups [195]. Its value is the squared sample Mahalanobis distance, estimated from the pooled covariance matrix, and the test assumes normality of the population of deformation vectors and equal covariances for the two groups. If these assumptions are not met, the  $T^2$  test is known to fail gracefully, i.e. it will still be approximately conservative and the loss in power for the alternative will not be too dramatic for moderate violations of the assumptions. However, it is preferable to analyze deformation fields nonparametrically. Permutation tests, with their minimal assumptions, are the usual method of choice for this two-sample problem [35, 196]. However, they also rely on a test statistic that is evaluated for each labelling (“permutation”), and the null hypothesis is that this statistic is distributed symmetrically around zero. The usual choice for the statistic is again Hotelling’s  $T^2$ , so permutation tests are not nonparametric, but rather result in *adjusted* significance probabilities [197].

For example, as shown in [198], the adjusted one-dimensional version of the  $T^2$  test, i.e. the permutation version of the classic  $t$ -test, is the uniformly most powerful test for the Gaussian alternatives with fixed variance, but fails to be uniformly most powerful against other alternatives. A more serious practical problem is that, even for small sample sizes, the number of permutations to consider for an exact test is prohibitively large. Especially so, if the number of voxels, i.e. the number of tests, is on the order of hundreds of thousands, as is common in neuroimaging applications. Therefore, in current analyses one often limits the data to only 10,000 or less random relabelings per voxel, at the expense of increasing the simulation error. Moreover, correcting for multiple comparisons imposes severe lower bounds on the numbers of relabelings needed per voxel for testing at realistic significance levels, i.e. on the sample size and simulation time. Particularly for small sample sizes that occur in prospective studies in mice, permutation tests cannot resolve low enough significance probabilities to allow for strong control of the family-wise error. Even the modern, liberal approach of limiting the False Discovery Rate [47, 50] does often not lead to useful results in these datasets [199].

In this paper we describe a new nonparametric statistical test that allows to efficiently perform a large number of such tests on vector data. The two-sample version of the test is not provably conservative, but its advantage is that it can be used for the initial screening of voxels. It is sensitive enough to work even under the conservative Bonferroni correction. Voxels where the null hypothesis is rejected can then be ana-

lyzed further by this test under the permutation distribution of the data; alternatively a different test statistic can be employed. This problem of testing one or more groups of vectors for distributional differences does not only arise in neuroimaging, but also in a number of other disciplines and diverse contexts, e.g. in geostatistics, human movement sciences, astronomy and biology. In the two-dimensional case, a natural nonparametric test for such problems has been given by [200], which we describe next. After generalizing this test to arbitrary dimensions, in Section 5.2.2 we focus on the three-dimensional case, being the most important one for applications.

## 5.2 The Moore-Rayleigh test

Let  $X = (X_1, \dots, X_N)$  be a finite sample of real  $k$ -vector-valued random variables  $X_i = (X_{i,1}, \dots, X_{i,k})$ . If we assume that the  $X_i$  are independently drawn from a common absolutely continuous probability distribution with density  $f : \mathbb{R}^k \rightarrow [0, \infty)$ , then the null hypothesis is:

$H_0$  : The probability density  $f$  is *spherically symmetric*.

Consequently, this implies that the density  $f$  is spherically decomposable. It factors into the product of a radial density  $p_r : [0, \infty) \rightarrow [0, \infty)$  and the uniform distribution on each hypersphere  $rS^{k-1} = \{x \in \mathbb{R}^k \mid \|x\| = r\}$ , such that  $f(x) = p_r(\|x\|)/\text{vol}(\|x\|S^{k-1})$ . We can then write  $X_i = R_i U_i$ , where  $R_i \sim p_r$  and  $U_i$  is distributed uniformly on the  $k$ -dimensional unit sphere  $S^{k-1}$ . The latter distribution can be realized as the projection of a  $k$ -dimensional diagonal Gaussian distribution with equal variance in each dimension. The sum  $\sum_{i=1}^N X_i$ , where the  $X_i$  are independently distributed according to a common, spherically symmetric distribution, is easy to interpret. It corresponds to a Rayleigh random flight [201] with  $N$  steps, whose lengths are distributed according to  $p_r$ . Scaling the vector-valued random variables  $X$  by the ranks of their lengths, the distribution of the resultant vector

$$S_N = \sum_{i=1}^N \frac{iX_{(i)}}{\|X_{(i)}\|}, \quad (5.1)$$

where  $X_{(i)}$  denotes the  $i$ -th largest vector in the sample (with ties being arbitrarily resolved), is independent of  $p_r$ ; consequently, a test based on  $S_N$  is nonparametric. The test statistic of interest here is the asymptotically scaled length of the resultant,

$$R_N^* = \frac{\|S_N\|}{N^{3/2}}. \quad (5.2)$$

A large value of  $R_N^*$  for a given sample  $X$  from an unknown distribution (not necessarily absolutely continuous) indicates a deviation from spherical symmetry. This test was introduced by [200], who treated the two-dimensional case numerically, and has been used in neuroscience [202–204], human movement science [205] and avian

biology [206–209].

In contrast to the Rayleigh test of uniformity [210], where the  $X_i$  are constrained to lie on (alternatively, are projected onto) the unit sphere, in the Moore-Rayleigh test also the vector length influences the test statistic. This follows the observation of [211], that differences in scale between two distributions will be mostly evident in their (radial) tails, i.e. when moving away from the mean. The interpretation of  $R_N^*$  is not so easy as in the Rayleigh test, however, where the test statistic is a measure of *spherical variance*. Consider the projections

$$S_{N,j} = \sum_{i=1}^N \frac{iX_{(i),j}}{\|X_{(i)}\|}, \quad (j = 1, \dots, k). \quad (5.3)$$

A direct calculation shows that under the null hypothesis the variance of  $X_{(i),j}/\|X_{(i)}\|$  is  $1/k$ , and that

$$\sigma^2 = \text{var}(S_{N,j}) = N(N+1)(2N+1)/(6k). \quad (5.4)$$

As  $E(S_{N,j})^3 = 0$  and  $\sigma^2 < \infty$ , the Lyapunov version of the Central Limit Theorem implies that the random variables  $S_{N,j}$  approach Gaussian  $\mathcal{N}(0, \sigma^2)$  distributions for large sample sizes  $N$ . Although the random variables  $\|S_{N,j}\|$  are obviously not independent, by the same argument as in [212] the corresponding distribution of  $\|S_N\|^2/\sigma^2$  asymptotically approaches a  $\chi_k^2$  distribution. Let  $\alpha_N = N^{3/2}$ . The exact null distribution of  $R_N = \alpha_N R_N^*$  in  $k$  dimensions,  $k \geq 2$ , is given by

$$\text{pr}(R_N \leq \alpha_N r; k) = r \left[ \Gamma\left(\frac{k}{2}\right) \right]^{N-1} \int_0^\infty \left(\frac{rt}{2}\right)^{\frac{k-2}{2}} J_{\frac{k}{2}}(rt) \prod_{n=1}^N \frac{J_{\frac{k-2}{2}}(nt)}{(nt/2)^{\frac{k-2}{2}}} dt, \quad (5.5)$$

where  $J_l$  denotes the Bessel function of order  $l$ ; see [213].

### 5.2.1 The one-dimensional case

In one dimension, the Moore-Rayleigh statistic for the null hypothesis corresponds to a symmetric random walk with linearly growing steps,

$$S_N = \sum_{i=1}^N \gamma_i i, \quad \text{where } \gamma_i = \pm 1 \text{ with equal probability.} \quad (5.6)$$

**Proposition 1.** *The probability mass function  $\text{pr}(S_N = r) \stackrel{\text{def}}{=} p(r, N)/2^N$  is given by the recurrence*

$$p(r, N) = p(r-n, N-1) + p(r+n, N-1) \quad (5.7)$$

with initial condition  $p(0, 0) = 1$  and  $p(r, 0) = 0$  for  $r \neq 0$ .

Rewriting Eq. 5.6 as

$$\sum_{\{\gamma_i=+1\}} i = \frac{1}{2} \left( S_N + \frac{1}{2} N(N+1) \right), \quad (5.8)$$

where the sum runs over all step sizes  $i \in \{1, \dots, N\}$  that have positive sign  $\gamma_i$ , shows that the numbers  $p(r, N)$  have a well-known combinatorial interpretation.

**Proposition 2.** *The numbers  $p(r, N)$  count the number of partitions of  $\frac{1}{2}(r + \frac{1}{2}N(N+1))$  with distinct parts less or equal to  $N$ .*

As before, denote the length of the resultant by  $R_N = \|S_N\|$ . Its probability function  $\text{pr}(R_N = r)$  is given by

$$\text{pr}(R_N = r) = \begin{cases} p(r, N)/2^{N-1} & \text{if } r > 0, \\ p(0, N)/2^N & \text{if } r = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.9)$$

In the sequel, we also need the random signs defined by

$$\epsilon_N = \prod_{i=1}^N \gamma_i, \quad (5.10)$$

conditional on the resultant  $S_N$ : Let  $\epsilon_{r,N}$  denote the average sign of the partitions of  $\frac{1}{2}(r + \frac{1}{2}N(N+1))$  with distinct terms less or equal to  $N$ , i.e.

$$\epsilon_{r,N} \stackrel{\text{def}}{=} E(\epsilon_N \mid S_N = r). \quad (5.11)$$

Anticipating the two-sample Moore-Rayleigh test discussed in Section 5.3, we note the following:

**Remark 1** (Relation to the Wilcoxon signed-rank test). *In the Wilcoxon signed-rank test for two paired samples  $X$  and  $Y$  of equal size  $|X| = |Y| = N$ , the null hypothesis is that the paired differences  $Z_i = Y_i - X_i$  are distributed (independently and identically) symmetrically around zero [214]. The test statistic is the sum  $W_+ = \sum_{i=1}^N iI(Z_i > 0)$ , where  $I(\cdot)$  is an indicator function. Under the null hypothesis we have that  $\text{pr}(Z_i > 0) = \text{pr}(Z_i < 0) = \frac{1}{2}$ . Assuming that  $\text{pr}(X_i = Y_i) = 0$ , which is fulfilled with probability 1 for continuous distributions, we can then identify  $I(Z_i > 0) - I(Z_i < 0)$  with a random sign  $\gamma_i$ , such that*

$$\begin{aligned} \sum_{i=1}^N \gamma_i i &= \sum_{i=1}^N iI(Z_i > 0) - \sum_{i=1}^N (1 - I(Z_i > 0))i \\ &= 2W_+ - \frac{1}{2}N(N+1). \end{aligned}$$



Therefore, testing for symmetry of the  $Z_i$  under the one-dimensional Moore-Rayleigh test is equivalent to the signed-rank Wilcoxon two-sample test of  $X$  and  $Y$ , with

$$\text{pr}(W_+ = r) = \text{pr}(S_N = 2r - \frac{1}{2}N(N+1), N).$$

This approach easily generalizes to more than one dimension.

**Remark 2** (Testing for radial dependence). Assume the density  $f$  decomposes spherically, such that  $X_i = R_i U_i$ , with  $R_i \sim p_r$  and  $U_i \sim u$ , where  $p_r(r) = \text{pr}(|X_i| = r)$  and  $u(x) = \text{pr}(X_i/|X_i| = x)$ . In one dimension,  $u$  can only attain the values  $\{-1, +1\}$  and  $u(\mp 1) = \text{pr}(X_i \leq 0)$ . If the mean of  $f$  is zero, i.e.  $E(X_i) = 0$ , then  $\text{pr}(X_i > 0) = \text{pr}(X_i < 0) = 1/2$ , and this implies that  $f$  is (spherically) symmetric. The Moore-Rayleigh test, under the assumption that  $X_i = R_i U_i$ , therefore tests the null hypothesis that  $E(X_i) = 0$ . On the other hand, assume that  $E(X_i) \neq 0$ . If the Moore-Rayleigh test finds a significant departure from uniformity, then this leads to the rejection of the hypothesis that the density  $f$  decomposes in such way, i.e. to accept the alternative that the common distribution of the random variables  $X_i$  is conditional on the length  $|X_i|$ . In practice, centering  $X = (X_1, \dots, X_N)$  by the sample mean, the Moore-Rayleigh test could be used to detect such radial dependence. However, its power would be quite limited and it seems likely that directly testing for differences in the two tails  $\{X_i > x\}$  and  $\{X_i < -x\}$  will be more powerful.

### 5.2.2 The three-dimensional case

Taking derivatives, the distribution function of  $R_N = \alpha_N R_N^*$ , given in Eq. (5.5), reduces to the density

$$\text{pr}(R_N = r) = \frac{2r}{\pi} \int_0^\infty t \frac{\sin rt / \alpha_N}{r} \prod_{n=1}^N \frac{\sin nt}{nt} dt \quad (5.12)$$

in the three-dimensional case ( $k = 3$ ). This formula can alternatively be derived by using characteristic functions (see Eq. 16 in [201]). The oscillating integral in Eq. (5.12) can be evaluated by numerical quadrature, but it is difficult to calculate its tail accurately. Another approach to evaluate this integral is based on a finite series representation, following an idea originally due to G. Pólya. Let  $N_{\max} = N(N+1)/2$ . If we expand  $\sin(nt) = (e^{nt} - e^{-nt})/2i$  and integrate the oscillating integral in Eq. (5.12) by parts  $N-2$  times as in [215], a simple but tedious calculation (which we omit) results in the following representation:

**Theorem 1.** The probability density of  $R_N^*$  under the null hypothesis can be evaluated as

$$\text{pr}(R_N^* = r) = \frac{2rN^3}{N!(N-2)!} \sum_{\substack{k \in \mathbb{N}: \\ \alpha_N r < k \leq N_{\max}}} \epsilon_{k,N} (\alpha_N r - k)^{N-2}, \quad (5.13)$$

where  $\epsilon_{k,N}$  is given by Eq. 5.11.

This is a generalization of Treolar's representation for the random flight with equal step sizes [216]. We see that, interestingly, the density of the three-dimensional case can be expressed in terms of statistical properties of the one-dimensional case. Integrating Eq. 5.13 term-by-term from  $r$  to infinity, we have the following corollary.

**Corollary 1.** *The cumulative distribution function of  $R_N^*$  under the null hypothesis can be evaluated as*

$$\text{pr}(R_N^* \leq r) = 1 - \frac{2}{N!N!} \sum_{\substack{k \in \mathbb{N}: \\ \alpha_N r < k \leq N_{\max}}} \epsilon_{k,N} (\alpha_N r - k)^{N-1} (\alpha_N r (1 - N) - k). \quad (5.14)$$

In particular,  $\text{pr}(R_N^* > (N + 1)/(2\sqrt{N})) = 0$ .

Note that because of the representation (5.14) for smaller  $r$  successively more and more terms enter the sum in the calculation of  $\text{pr}(R_N^* > r)$ . The numerical accuracy is therefore higher for larger  $r$ , i.e. in the tail of the distribution. The representations (5.13) and (5.14) therefore allow the efficient computation of exact significance probabilities for the test statistic  $R_N^*$  for small to moderately large sample sizes  $N$  (e.g., for  $N \lesssim 60$  under double precision IEEE 754 arithmetic). This restriction on the sample size is only due to numerical accuracy; for larger  $N$  approximations of the Gamma function can be used. In figure 5.1 the distribution of  $R_N^*$ , for some values of  $N$ , is plotted and compared with the asymptotic  $\chi_3^2$  distribution. In Table 5.1 the values of the quantile function are listed. These values have been calculated by numerically inverting Eq. (5.14) with a bisection method and are conservatively rounded.

**Remark 3** (What is tested by the Moore-Rayleigh test?). *As in Remark 2, assume that  $X_i = R_i U_i$ , with  $R_i \sim p_r$  and  $U_i \sim u$ , where  $p_r(r) = \text{pr}(|X_i| = r)$  and  $u(x) = \text{pr}(X_i/|X_i| = x)$  are arbitrary. If  $E(X_i) = 0$ , this implies  $E(U_i) = 0$ , and suggests that  $\sum_i U_i \approx 0$  for a sample. More precisely, an upper bound for the variance of the test statistic  $R_N^*$  is realized by the one-dimensional Moore-Rayleigh null hypothesis, whose distribution is similar to the null hypothesis of the three-dimensional case (confer figure 5.6). Therefore, as in the one-dimensional case, the Moore-Rayleigh test under the assumption of radial decomposability tests mostly for differences in location. Note that symmetry of the  $U_i$ , i.e.  $\text{pr}(U_i = u) = \text{pr}(U_i = -u)$ , implies that  $E[\sum_i U_i] = 0$ . Thus, under the assumption of decomposability, testing for spherical symmetry and testing for symmetry are approximately equivalent, i.e. the Moore-Rayleigh test will not be sensitive to deviations from spherical uniformity if the underlying distribution is merely symmetric or mean-centered. This is actually an advantage when the Moore-Rayleigh test is considered as a two-sample test (see below).*

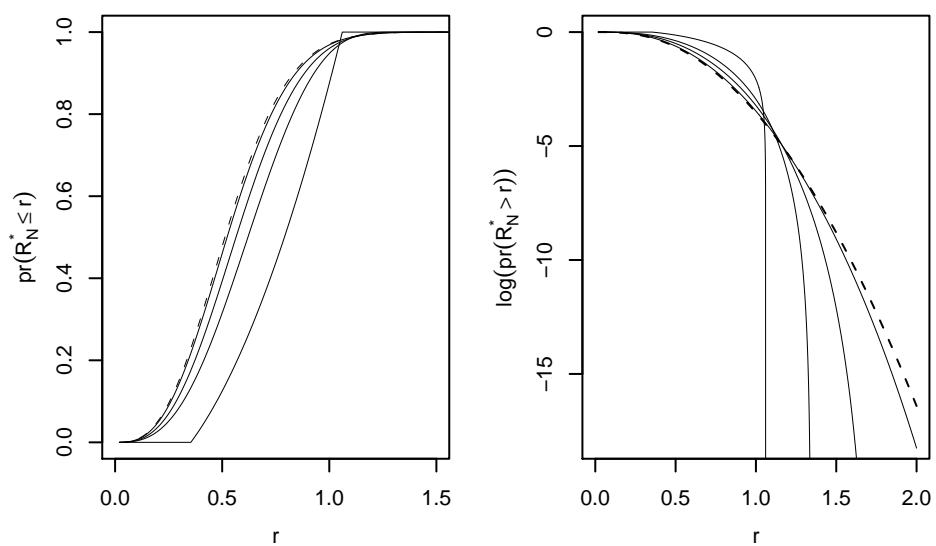


Figure 5.1: Left: Distribution function for  $N = 2, 5, 10$  and  $50$  (solid, from right to left), compared to the asymptotic  $\chi_3^2$  distribution (dashed). Right: Logarithm of significance probabilities.

### 5.2.3 Power estimates

To evaluate the performance of the three-dimensional Moore-Rayleigh test (MR3), power functions for a number of distributions were obtained by Monte-Carlo simulation. These show the fraction of rejections of the null hypothesis for a specific distribution, significance level  $\alpha$ , and sample size  $N$ . The left panel of figure 5.2 shows the power function for a family of diagonal Gaussian distributions with unit variances, shifted away from zero (along the  $z$ -axis) a constant distance  $\mu \geq 0$ . Each point power estimate was obtained by 1,000 realizations of the distributions and represents the fraction of significance probabilities (“ $p$ -values”) less than the nominal significance level  $\alpha$ . The test was performed on  $N = 10$  randomly drawn samples, and is compared to Hotelling’s (non-randomized)  $T^2$  one-sample test of location [217], as implemented in the R package ICSNP<sup>1</sup>, and to the spherical uniformity permutation test of [218], under  $10^4$  resamplings. The test statistic of the latter is an U-estimator of the difference between two probability distributions of vectors, calculated by a Gaussian kernel with a bandwidth parameter. The choice of the proper bandwidth is the subject of ongoing research; we show results for the two bandwidths  $b_1 = 0.25$  and  $b_2 = 2.5$ , and denote the corresponding tests by “Diks1” and “Diks2”, respectively. In comparison with the  $T^2$  test, MR3 shows larger power, an effect that is more pronounced for lower significance levels. It is thus a more sensitive measure of changes in location. Note that this does not contradict the well-known optimality of Hotelling’s  $T^2$  test for the family of multivariate Gaussian distributions, since in the calculation of  $T^2$  the covariance matrix needs to be estimated from the data. In the special case of equal covariances considered here, the Moore-Rayleigh test can therefore exhibit larger power. Also note that the test of Diks & Tong can be more powerful than the MR3 test, but as its results depend strongly on the bandwidth parameter, it is difficult to apply it routinely.

In figure 5.3, power functions are shown for a family of diagonal Gaussian distributions where the standard deviation of one axis was varied from  $\sigma = 0.1$  to  $\sigma = 5.0$  in steps of 0.1, the other standard deviations were kept at unity. As expected from Remark 3, the MR3 test performs poorly for this specific violation of spherical symmetry. The remaining symmetry in the distribution means that although sample points are now increasingly less concentrated on one axis, on average their contributions to the resultant length still mostly cancel each other. Analogously, the  $T^2$  test has only nominal power for the anisotropic multivariate Gaussian, being a test of location only. Note that MR3 shows slightly more power than the nominal significance levels  $\alpha$  for  $\sigma \neq 1$ , as do the Diks1 and Diks2 tests.

To assess the effect of asymmetry of the sample distribution, we employ the Fisher distribution, also known as the Fisher–Bingham three-parameter distribution. This is the  $k = 3$  case of the  $k$ -dimensional von–Mises Fisher distributions commonly used in directional statistics [210]. Details of its computation are given in the Appendix. We denote the Fisher distribution with concentration parameter  $\lambda$  (and with the

<sup>1</sup>K. Nordhausen, S. Sirkia, H. Oja, and D.E. Tyler. ICNSP: Tools for Multivariate Nonparametrics, R-package version 1.0-2 (2007)

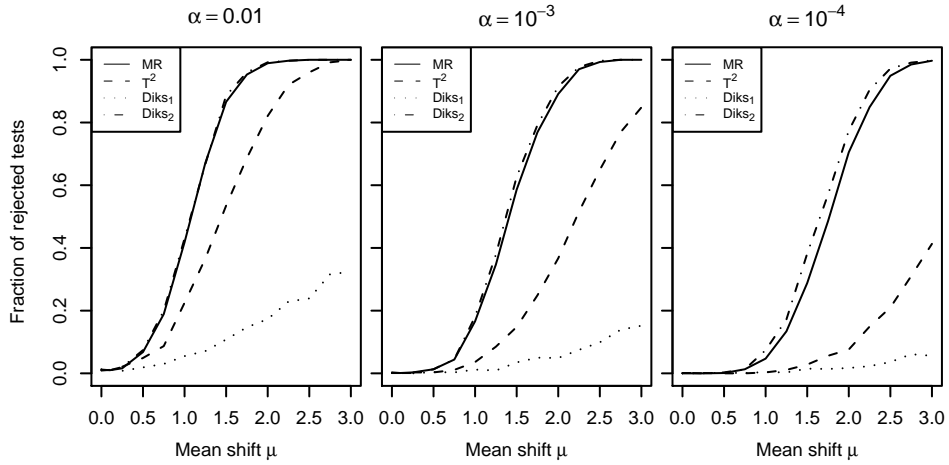


Figure 5.2: Estimated power functions for the family of Gaussian distributions with covariance matrix the identity and mean shifted a distance  $\mu$  away from zero. Sample size  $N = 10$ .

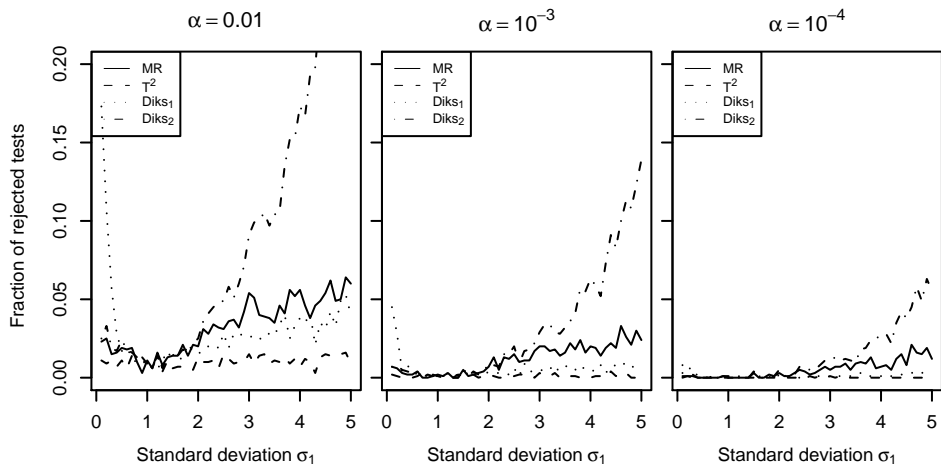


Figure 5.3: Estimated power functions for the family of Gaussian distributions, varying the standard deviation  $\sigma$  of a single axis. Sample size  $N = 10$ . Note the small range of the power.

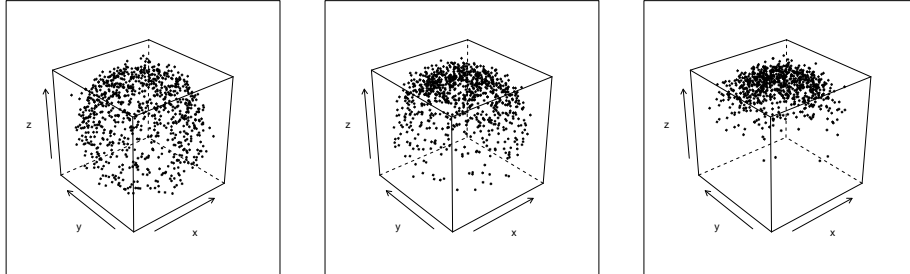


Figure 5.4: Scattered Fisher distribution, visualized by 1,000 randomly drawn points in the unit-cube. Left: Concentration parameter  $\lambda = 1$ . Middle: Concentration parameter  $\lambda = 2.5$ . Right: Concentration parameter  $\lambda = 5$ .

choice  $\xi = e_3$ ) by  $F3_\lambda$ . To avoid degeneracies due to its singular character, the  $F3_\lambda$  distribution is multiplied by  $1 - Z$ , where  $Z \sim \mathcal{N}(0, 0.1)$ . Figure 5.4 shows three examples of  $N = 1,000$  random variates obtained from these “scattered” Fisher distributions for distinct values of the concentration parameter  $\lambda$ , with increasingly larger deviation from the uniform distribution. The power of MR3 for the family of scattered Fisher distributions, varying the concentration parameter, is comparable to the power of the other tests (not shown). Let us now consider a mixture, where the samples are chosen either (i) from the uniform distribution on the unit sphere, or (ii) from the scattered Fisher distribution  $2F3_5$ . The probability  $0 \leq p \leq 1$  for each sample vector to be chosen from the second distribution is the parameter of this family of mixture distributions, with larger  $p$  indicating stronger deviations from uniformity for the larger vectors. Figure 5.5 depicts the estimated power for this family under variation of the mixture probability  $p$ . Compared to the  $T^2$  test, the MR3 test is seen to be more sensitive to these specific departures from uniformity. It should be noted that reversing the situation, e.g., by considering  $F3_5/2$  instead of  $2F3_5$ , such that the smaller vectors exhibit deviations from uniformity, the power of MR3 becomes less than that of the  $T^2$  test (not shown).

### 5.3 The two-sample test

The most interesting application of the Moore-Rayleigh test is the two-sample problem. There, we are given two vector-valued random variables

$$X = (X_1, \dots, X_N) \quad \text{and} \quad Y = (Y_1, \dots, Y_N), \quad (5.15)$$

and we assume that they are identically and independently distributed with densities  $f$  and  $g$ , respectively. The differences  $Y_j - X_i$  are then distributed according to the

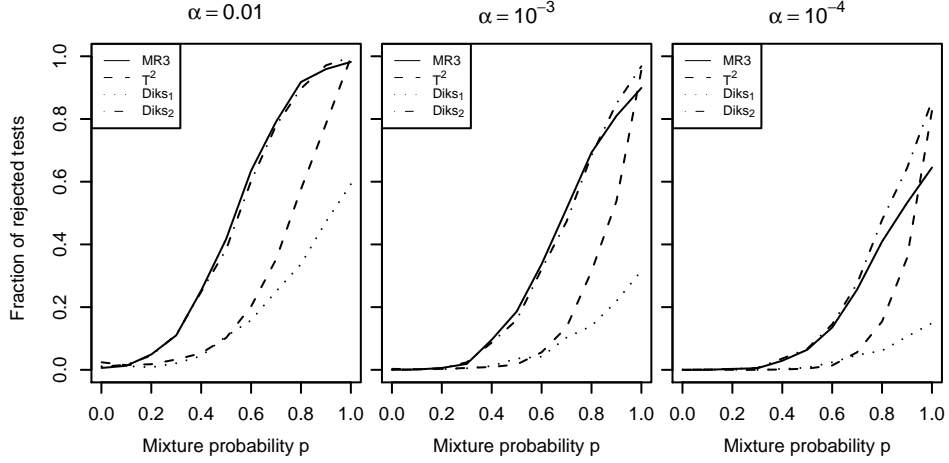


Figure 5.5: Estimated power functions for the mixture of the scattered Fisher distribution  $2F3_5$  with the uniform distribution on the sphere  $S^2$ , varying the mixture probability  $p$  that a sample vector arises from the first distribution. Sample size  $N = 10$ .

convolution  $g * (-f)$ , whose density is

$$\text{pr}(Y - X = x) = \int \text{pr}(Y = u) \text{pr}(X = u + x) d^k u. \quad (5.16)$$

Under the null hypothesis that the  $X_i$  and  $Y_j$  come from a common probability density  $f$ , this reduces to the *symmetrization* of  $f$ , with density

$$\text{pr}(Y - X = x) = \int \text{pr}(X = u) \text{pr}(X = u + x) d^k u. \quad (5.17)$$

If the probability density  $f$  is *spherically symmetric* around its mean  $\mu$ , i.e. uniform on each hypersphere  $\{x \mid \|x - \mu\| = r\}$ , then Eq. (5.14) gives the significance probability of a deviation from the null hypothesis. In particular, this applies when  $f$  is assumed to arise from a multivariate normal distribution, justifying the use of the Moore-Rayleigh statistic in many practical situations.

### 5.3.1 Testing for symmetry

In general, however, the distribution of  $h = f * (-f)$  is merely *symmetric*, i.e.  $h(x) = h(-x)$  for all  $x \in \mathbb{R}^k$ . This follows from

$$\int \text{pr}(X = u) \text{pr}(X = u + x) \text{d}^k u = \int \text{pr}(X = u) \text{pr}(X = u - x) \text{d}^k u. \quad (5.18)$$

The following demonstrates the difference.

**Example 2.** Consider the symmetric singular distribution  $B_x \stackrel{\text{def}}{=} \frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x}$ , where  $\delta_x$  is the Dirac measure concentrated at the point  $x \in \mathbb{R}^k$ . The distribution  $B_x$  leads to an embedding of the one-dimensional Moore-Rayleigh null distribution in three-dimensional space. Its realizations take values  $x$  and  $-x$  with equal probability, and it is not spherically symmetric. As it is,  $B_x$  is neither absolutely continuous, nor can it arise as the symmetrization of a distribution. Nevertheless, it is a model for a distribution that can arise in practice: First, the delta distributions can be approximated, e.g., by a series of Gaussian distributions with decreasing variance. Secondly, consider the singular distribution  $B^x$  that is concentrated on a line  $\{\lambda x \mid \lambda \in \mathbb{R}\} \subseteq \mathbb{R}^k$  through the origin. Applying the Moore-Rayleigh test to  $B^x$  is equivalent to calculating the test statistic from  $B_1$ , since  $B^x$  is invariant under symmetrization and is projected, before ranking, to the sphere  $S^0 = \{-1, +1\}$ .

The distribution  $B_1$  is a representative of the class of “fastest growing” random flights in three dimensions, since any other distribution of increments has less or equal probability to reach the highest values of the test statistic. On the other hand, the uniform distribution on the sphere, which represents the null hypothesis of the Moore-Rayleigh test statistic  $R_N^*$ , will attain lower values of  $R_N^*$  with higher probability, as the uniform random walk can do “orthogonal” steps that increase the distance from the origin faster than in  $B_1$  (on the average). To be specific, if the finite sample  $X$  is distributed according to  $B_1$ , the  $n$ -th step of the scaled random walk either increases or decreases the distance from the origin by  $n$  (when crossing the origin, there is an obvious correction to this). However, if the  $n$ -th step were taken in a direction that is orthogonal to the resultant obtained so far, the distance will increase from  $R$  to  $\sqrt{R^2 + n^2} \approx R + n/(2R)$ , with probability 1 (conditional on the orthogonality). Figure 5.6 compares significance probabilities for  $B_1$  with those of the uniform random flight that represents the null hypothesis of the Moore-Rayleigh test, for  $N = 10$  sample points. There exists a value of the test statistic where the two curves cross (at about  $p = 0.20$ ), and after which the distribution function (significance probability) of the one-dimensional random walk  $B_1$  lies below (above) the one for the uniform random flight. The two-sample Moore-Rayleigh test, interpreted as a goodness-of-fit test, is therefore liberal, although it escaped Moore from his attention. This has also proven for the Wilcoxon signed-rank test, the equivalent of the one-dimensional Moore-Rayleigh test [219, 220]. These findings have casted doubt on the applicability of the test in this setting.



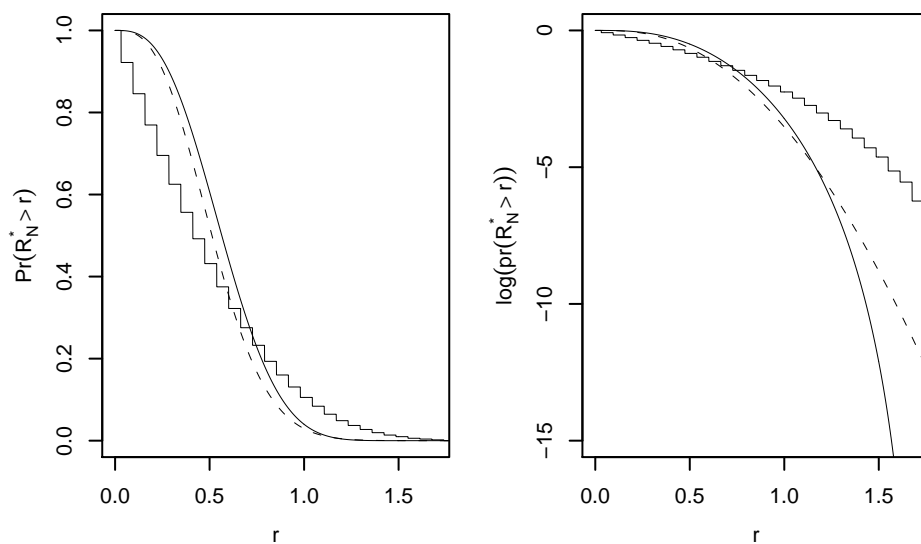


Figure 5.6: Comparison of significance probabilities for resultant lengths of spherically symmetric (smooth curve) and one-dimensional symmetric random walk (piecewise-linear curve) in three dimensions for  $N = 10$  steps. Dotted curve shows the asymptotic case.

The optimal upper bound for a conservative significance probability would be

$$G_N^*(r) = \sup_{\Psi_N} \text{pr}(|S_N| \geq r), \quad (5.19)$$

where the supremum is taken over the set  $\Psi_N$  of all possible symmetric probability distributions for  $N$  increments. More precisely, these increments are not independent but arise from a mixture of independent distributions by the order distribution (due to the ranking of vector lengths) of their radial projections. Even if one restricts this to the class where only independent, not necessarily identical symmetric probability distributions for each step are considered, this is a difficult problem. First steps in this direction have been made by [221], where the three-dimensional problem is reduced to a similar problem in one dimension by the familiar tangent-normal decomposition of the sphere. Apart from that, there has not been much progress in determining the envelope in Eq. 5.19. Even in the one-dimensional case it is not clear what the “fastest” random flight with linearly bounded increments is. If a liberal test is admissible for the specific problem at hand, e.g., in exploratory data analysis, MR3 offers an efficient two-sample test. Moreover, the Remarks and figure 5.3 suggest that the significance probabilities are only liberal for relatively large values of the test statistic. Studies with synthetic data seem to confirm that the MR3 test fails gracefully, if at all, for distributions expected in biomedical imaging practice [36]. Since the assumed null hypothesis is stronger than mere symmetry, MR3 can also be used for negative testing, i.e., if the null hypothesis of the uniform random flight cannot be rejected for a sample of difference vectors, then the modified null hypothesis that  $g*(-f)$  is symmetric, not necessarily spherically symmetric, cannot be rejected. For the null hypothesis of mere symmetry, there does not exist an accessible sufficient statistic and existing tests are either only asymptotically nonparametric or require further randomization of the underlying distribution [218, 222–226], so the MR3 test offers a simpler and much more efficient alternative, albeit with the disadvantage that it is potentially liberal.

### 5.3.2 Further issues

A different issue with the Moore-Rayleigh test arises in the (usual) case of unpaired samples. The two sample test we have presented up to now assumes paired vectors, and this approach reduces the symmetry group of the null hypothesis from the group of permutations to the much smaller group of reflection symmetry of the given pairs. The main reason here is simplicity in applications and reproducibility of the test statistic. If there is no natural pairing, it seems advisable to randomly pair samples, as e.g. [200] advocates. However, a drawback is that the test statistic then becomes a random variable, and replications of the test will result in distinct significance probabilities. This is undesirable, for example, in a clinical context. Bootstrapping the test, i.e. considering the mean of the test statistic  $R_N^*$  obtained during a large enough number of resamples from the empirical distributions, is a natural way to obtain more or less replicable significance probabilities, but on the expense of computational time. It

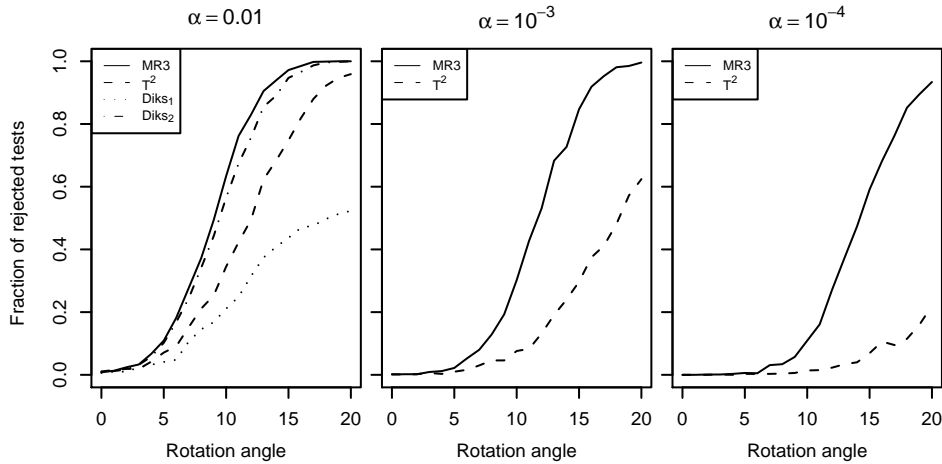


Figure 5.7: Estimated power for translated (10 standard deviations), then rotated diagonal Gaussians with unit variances as a function of relative rotation angle. Sample size  $N = 10$ .

is also not precisely known at present what the convergence properties of such an estimator are. A different approach would be to pair samples based on a measure of optimality. This seems natural enough, but has the undesirable feature that the test might become biased, e.g., too sensitive in case the sample points are matched by the method of least-squares or the Wasserstein distance. Therefore, as a practical solution in a context where reproducibility is desired, we propose to pair samples based on their ranks, such that  $X_{(i)}$  is matched with  $Y_{(i)}$ ,  $i = 1, 2, \dots, N$  (with ties resolved arbitrarily). Under the null hypothesis, the decomposability of the common distribution of  $X$  and  $Y$  guarantees the asymptotic unbiasedness of this approach, although for finite samples a slight bias is expected.

## 5.4 Results

### 5.4.1 Simulation results

In this section we show the results of a number of numerical simulations for the two-sample problem and compare them with Hotelling's  $T^2$  test and the Diks1 and Diks2 tests. Throughout, we use random matching of samples and  $N = 10$ . Figure 5.7 shows the results for two standard Gaussian distributions that were first translated in the same direction by ten standard deviations, and then one of them was rotated against the other (with the origin as the center of rotation), for 1,000 realizations. The Moore-Rayleigh test performs well: Its power for the trivial rotation is nominal,

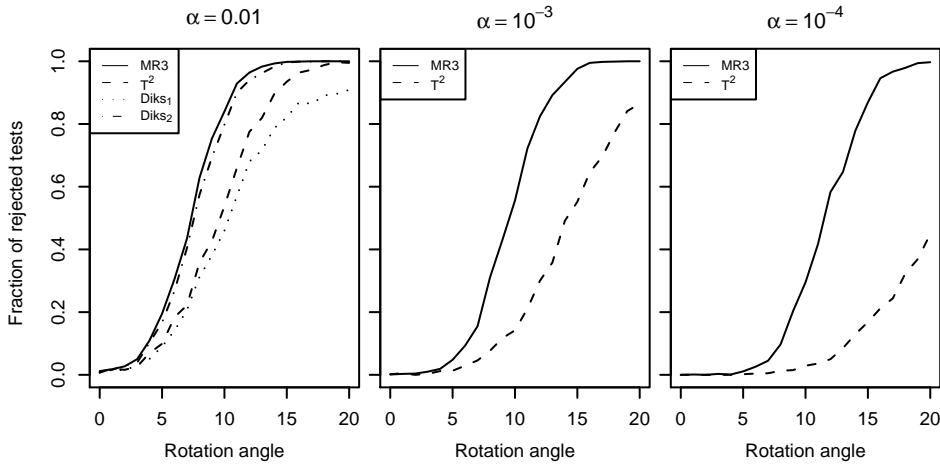


Figure 5.8: Estimated power functions for translated (10 standard deviations), then rotated diagonal Gaussians with unit variance against the scattered Fisher distribution (scaled by a unit Gaussian) as a function of relative rotation angle. Sample size  $N = 10$ . Note the small power at angle zero.

and for larger rotation angles higher than the power of the  $T^2$  test. Similar results are obtained when rotating Fisher distributions (not shown). Note that the Diks1/Diks2 tests are performed on the group of symmetric sign changes (of order  $2^{10}$ ), in contrast to the previous section where the full symmetry group of all rotations (of infinite order) was used, and do not resolve significance probabilities smaller than  $1/1,000$ , i.e. their power is zero for the lower significance levels, and therefore not indicated. Figure 5.8 compares the Gaussian distribution with the distribution  $R \cdot F_{3\lambda}$ ,  $R \sim \mathcal{N}(0, 1)$ , when both distributions are first translated and then rotated against each other, with similar results. Finally, figure 5.9 shows adjusted  $p$ -values, for  $10^4$  permutations and 100 realizations each. The Moore-Rayleigh test again shows slightly better power than the  $T^2$  test. More importantly, there is not much difference with the unadjusted power functions. These results are based on 100 realizations only, to speed up the considerable amount of computations, which accounts for the visible fluctuations.

### 5.4.2 Synthetic data

As remarked in the introduction, an important field of application of the Moore-Rayleigh test is the morphometric analysis of magnetic resonance (MR) images. Therefore, a synthetic data was generated simulating two groups of images with inter- and intra-variations: The Moore-Rayleigh test was validated on a synthetic  $50 \times 50 \times 80$  three-dimensional image domain. Five spherical deformations were added

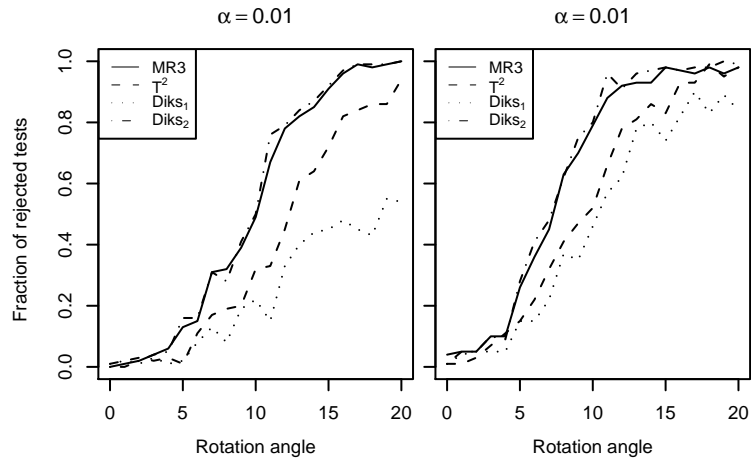


Figure 5.9: Adjusted estimated power functions. Left: translated (10 standard deviations), then rotated diagonal Gaussians with unit variances as a function of relative rotation angle. Right: translated (10 standard deviations) then rotated diagonal Gaussians with unit variance against the scattered Fisher distribution (scaled by a unit Gaussian) as a function of relative rotation angle. Sample size  $N = 10$ . Results based on 100 realizations of  $10^4$  permutations each.

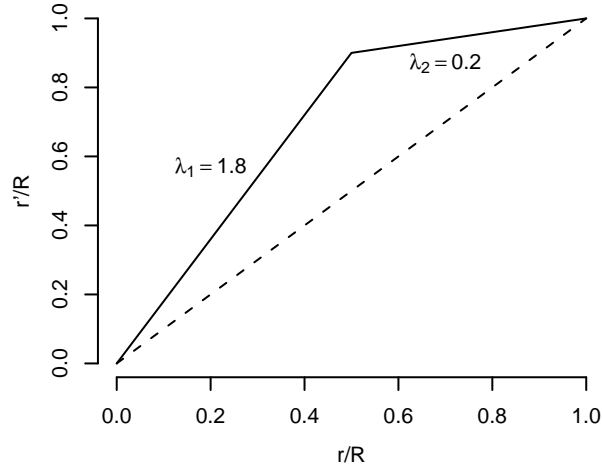


Figure 5.10: Generation of spherical deformations. The plot shows the behaviour of the deformation field in a one-dimensional projection along a radial direction. The volume at normalized distance  $r/R$  from the centerpoint of the sphere (radius  $R$ ) is mapped radially to  $r'/R$ . For  $r < R/2$  the length of the deformation vectors expands linearly,  $r' = \lambda_1 r$ , attaining its maximum at half radius  $r = R/2$ . For  $r > R/2$  the length shrinks linearly by  $\lambda_2 = 2 - \lambda_1$ , ensuring continuity at the boundary. The stippled line shows the case of no deformation ( $\lambda_1 = 1$ ).

in two distinct regions, introducing characteristic localized changes. The volume in each sphere was mapped radially, linearly expanding by a factor  $\lambda_1 = 1.8$  from the centerpoint to half radius distance, and then contracting linearly by  $\lambda_2 = 2 - \lambda_1$ , resulting in a one-to-one transformation of each spherical volume. Although the transformations were not smooth, interpolation at subpixel level guaranteed that they were indeed local diffeomorphisms. Figure 5.10 shows the transformation along a radial direction. A Gaussian noise process (zero mean,  $SD = 1.0$ ) was added to the deformed image, for a total of 15 distinct realizations, thereby simulating natural variation in brain structure.

Figure 5.11(A) shows the average lengths of deformation vectors in this image set. In the upper part (Region 1) one sphere of radius 9 voxels ( $S_2$ ) and two spheres of radius 6 voxels ( $S_1$  and  $S_3$ ) were created at successive distances of 12.5 voxels between their center points, creating a more complex deformation due to partial overlap in the superposition of deformation fields. Two spheres in the lower part (Region A) with radii 6 voxels ( $S_4$ , left) and 9 voxels ( $S_5$ , right) were created at a distance of 25 voxels. A second group of 15 images was created, with a reduced radius of 6 voxels for the spheres  $S_2$  and  $S_5$ . Figure 5.11(B) depicts the absolute differences in deformation vector lengths between the average deformation fields of both groups in the central

slice. For the evaluation of the statistical tests, ground truth, i.e. voxels for which the null hypothesis of no group difference should be rejected, was taken to be the total volume of the two spheres  $S_2$  and  $S_5$ . This approximation allowed the estimation of *precision* and *recall* from the numbers of true positives (TP), false positives (FP, type I error) and false negatives (FN, type II error), where

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

The results are shown in figure 5.12 and table 5.2 for different significance levels  $\alpha$ . The rightmost level  $\alpha = 2.5 \cdot 10^{-7}$  corresponds to 0.05 under Bonferroni correction with 200,000 voxels. The performance of all four tests is comparable, with the Moore-Rayleigh test exhibiting better recall and precision rates than the other tests. Note that the results of the permutation version of Hotellings  $T^2$  test are limited by the number of relabelling ( $N = 10,000$ ), such that Bonferroni correction for multiple comparisons did not result in any significant voxels.

## 5.5 Discussion

It is possible to test spherical symmetry in three dimensions with high numerical accuracy by using the combinatorial sum representation given in Eq. (5.14). In combination with Kahan summation [227], this representation makes it feasible to routinely calculate  $p$ -values for finite sample sizes that allow to assess statistical significance. Even for hundreds of thousands of multiple comparisons with a Bonferroni correction, as is common practice in neuroscientific imaging applications, the proposed approach is effective. Permutation methods, although theoretically preferred, are difficult to use in this setting due to practical limitations. The standard approaches to cope with these limitations, based on either saddle-point approximations to permutation tests [228] or on permutation tests for linear test statistics, where the conditional characteristic function can be rewritten as a convergent approximating series [229], are not directly applicable because these statistics usually do not arise in these practical problems or are too involved in the multivariate case. An alternative might be the use of optimal (Bayesian) stopping rules in the resampling process [230, 231]. However, small sample sizes can still seriously restrict the possible range of the significance probabilities. In the special case of the two-sample problem, the distribution of the null hypothesis is conditional on the unknown distribution of the data, and the generalized Moore-Rayleigh test is only approximately valid, a feature that all other (non-randomized) tests of symmetry exhibit. In Section 5.4.2 we evaluated the properties of this generalized Moore-Rayleigh test empirically with simulated imaging data of known ground-truth and by comparison with other nonparametric tests; for further comparisons see [36]. Even though the test is theoretically liberal, it seems to work well when applied for deformation-based morphometry, as it is not overtly sensitive to the difference between symmetry and spherical symmetry. An exact test is furthermore available by the permutation variant of the Moore-Rayleigh test, with

N	Probability			-Log(Probability)						
	0.100	0.010	0.001	4	5	6	9	12	15	18
2	1.013	1.056	1.061							
3	0.973	1.100	1.138	1.150	1.153	1.155				
4	0.948	1.116	1.189	1.222	1.237	1.244	1.250			
5	0.930	1.124	1.221	1.275	1.304	1.321	1.338	1.341	1.342	
6	0.916	1.129	1.245	1.314	1.357	1.384	1.418	1.427	1.429	
7	0.905	1.132	1.262	1.344	1.398	1.435	1.488	1.505	1.510	1.511
8	0.897	1.133	1.275	1.368	1.432	1.477	1.549	1.576	1.586	1.588
9	0.890	1.134	1.284	1.387	1.460	1.513	1.603	1.640	1.656	1.659
10	0.885	1.134	1.292	1.402	1.483	1.543	1.649	1.698	1.720	1.726
12	0.877	1.133	1.303	1.426	1.519	1.590	1.727	1.797	1.834	1.844
14	0.871	1.133	1.310	1.443	1.545	1.626	1.788	1.879	1.931	1.946
16	0.866	1.132	1.316	1.455	1.565	1.654	1.838	1.947	2.013	2.033
18	0.863	1.132	1.320	1.464	1.580	1.675	1.878	2.003	2.083	2.108
20	0.860	1.131	1.323	1.472	1.593	1.693	1.911	2.051	2.144	2.174
30	0.851	1.129	1.331	1.493	1.629	1.746	2.016	2.209	2.350	2.399
40	0.847	1.128	1.335	1.503	1.647	1.771	2.071	2.294	2.467	2.529
50	0.844	1.127	1.337	1.509	1.657	1.787	2.103	2.347	2.540	2.612
60	0.843	1.126	1.338	1.513	1.664	1.797	2.125	2.382	2.590	2.668
$\infty$	0.834	1.123	1.345	1.532	1.697	1.846	2.233	2.559	2.847	

Table 5.1: Critical values of Moore-Rayleigh statistic in 3D for various sample sizes N.

Test	$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.005$		$\alpha = 0.001$	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
MR3	0.07	0.81	0.21	0.63	0.59	0.39	1	0.04
HT2	0.07	0.77	0.21	0.54	0.56	0.28	1	0.01
permuted HT2	0.07	0.77	0.21	0.54	0.57	0.28	0	0
Diks1	0.03	0.38	0.1	0.23	0.35	0.11	0.69	0.04
Diks2	0.07	0.80	0.22	0.59	0.56	0.31	0.77	0.08

Table 5.2: Precision and recall for the synthetic dataset.



slightly improved power when compared with conventional permutation testing. This can be used in a second stage after initial screening with the fast, unadjusted Moore-Rayleigh test. Although such screening could also be realized by the  $T^2$  test, the MR3 test seems better suited to this problem due to its enhanced power, which allows for strong control of the family-wise error. In contrast, the  $T^2$  test does often not allow the localization of individual voxels, as demonstrated in the example on deformation morphometry in brain scans. It should be noted that we have only considered the conservative Bonferroni correction here, for simplicity, but it is expected that the MR3 test remains a more sensitive instrument also under modern step-down multiple comparison procedures (as described in, e.g., [196]). An implementation of the Moore-Rayleigh test as a package for the statistical computing environment R<sup>2</sup> or implemented in C++ is available on request.

## 5.A The Fisher distribution

The Fisher distribution is a singular distribution on the hypersphere  $S^{k-1}$  whose density  $f(x), x \in \mathbb{R}^k$ , is proportional to  $e^{\lambda \xi^t x}$ , where  $\xi^t$  denotes the transpose of  $\xi$ . The mean direction  $\xi$  is constrained to be a unit vector, and  $\lambda \geq 0$  is a concentration parameter. Without restricting generality, we let  $\xi = e_k$  be the unit vector in the  $k$ -th dimension, so  $f \sim e^{\lambda x_k}$  only depends on the last coordinate, and we are left with a one-parameter family of distributions. Following [232] and [233], a random variate distributed according to the von-Mises Fisher distribution is obtained by generating a random variate  $W$  for the last coordinate, by the density proportional to

$$e^{\lambda w} (1 - w^2)^{(k-3)/2}, \quad w \in (-1, 1), \quad k \geq 2,$$

and a  $k - 1$ -dimensional variate  $V$  uniformly distributed on the hypersphere  $S^{k-2}$ . The vector

$$X = (\sqrt{1 - W^2} \cdot V^t, W) \in \mathbb{R}^k$$

then has the desired density. In  $k = 3$  dimensions the former can be achieved by integrating the distribution function of  $W$  directly. Choosing a uniform variate  $U$  on the interval  $[-1, 1]$ , a random variate  $W$  is clearly given by

$$W = \frac{1}{\lambda} \log(2U \sinh \lambda + e^{-\lambda}).$$

---

<sup>2</sup>The code can be downloaded from <http://folk.ntnu.no/muskulus/>

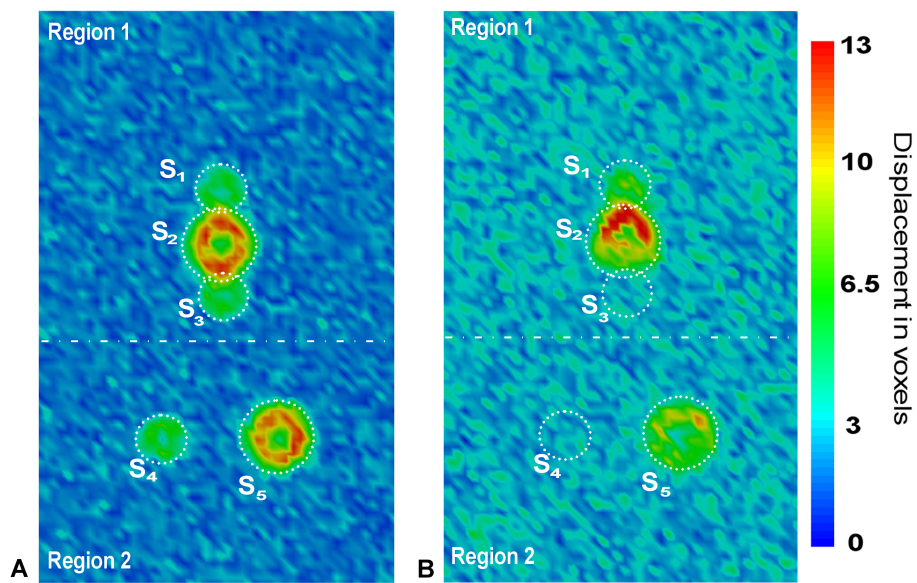


Figure 5.11: Validation with a synthetic dataset. A slice from the deformation field of  $50 \times 50 \times 80$  voxels (isotropic spacing of 1.00 mm), showing the five spherical deformations that were added to it (see text for details). The color indicates the length of the deformation vectors per voxel. A: Mean deformation field for the first group. B: Difference between deformation fields for the two groups (smaller deformations in spheres  $S_2$  and  $S_5$  in the second group).

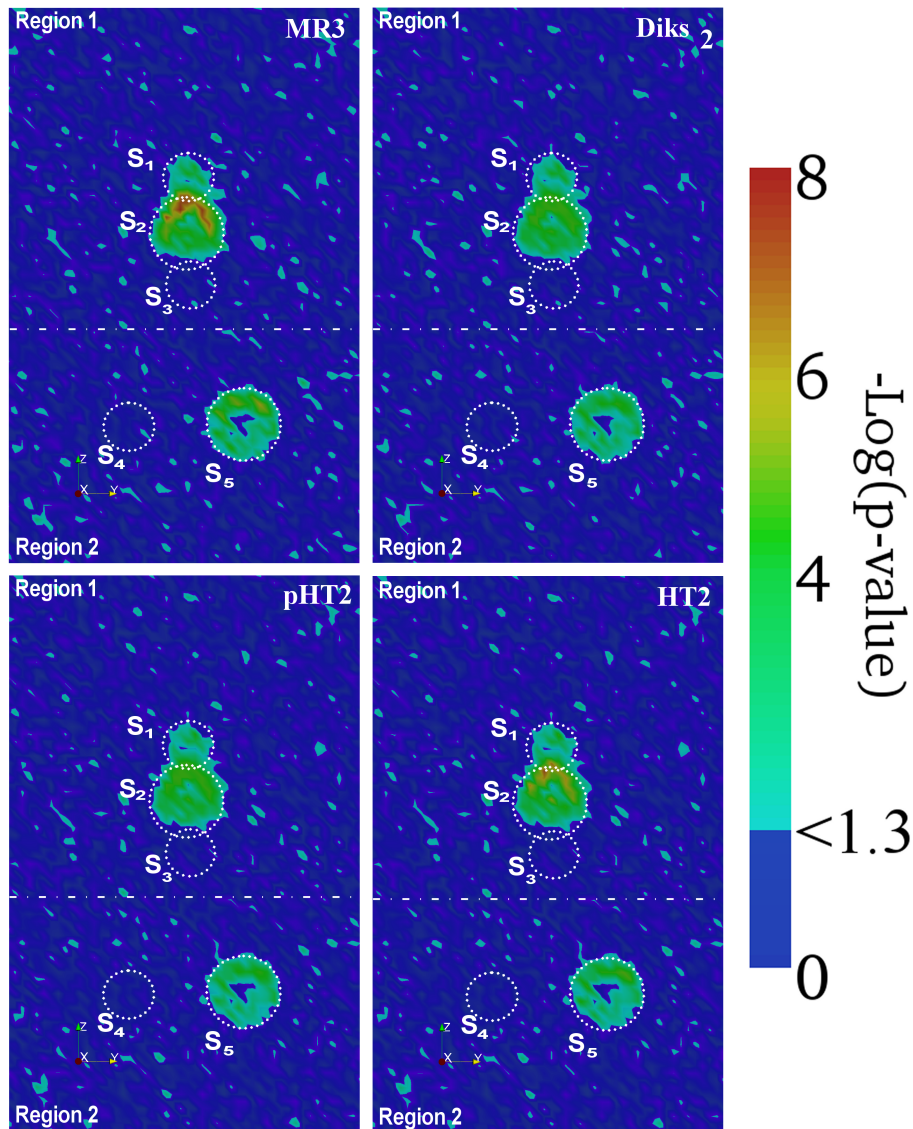


Figure 5.12: Validation with a synthetic dataset. Negative logarithms of significance probabilities for two-sample Moore-Rayleigh test (MR3), the Diks2 test (see text for details), the permutations of the Hotellings  $T^2$  test (pHT2), and the Hotellings  $T^2$  test (HT2).



## CHAPTER 6

---

### The 3D Moore-Rayleigh test: A novel nonparametric test for rapid deformation-based morphometry

---

A.E.H. Scheenstra  
M. Muskulus  
L. Ferrarini  
N. Beckmann  
A.M.J.M. van den Maagdenberg  
S. Verduyn Lunel  
L. van der Weerd  
J. Dijkstra  
J.H.C.Reiber

*This chapter was adapted from:*  
The 3D Moore-Rayleigh test: A novel nonparametric test for rapid Deformation-based Morphometry. *Submitted to NeuroImage*

**abstract:** *Non-rigid registration of MR images to a common reference image results in deformation fields, from which anatomical differences can be statistically assessed, within and between studies. Without further assumptions on the underlying distributions, nonparametric tests are needed and usually the analysis of deformation fields is performed by permutation tests. However, permutation tests are computationally expensive and have limitations by sample size and number of iterations. In this paper, we consider a single nonparametric test as an alternative for permutation tests; the 3D Moore-Rayleigh test. As its distribution function is available in closed form, permutation testing can be avoided. Furthermore, the test incorporates both the directions and magnitude of the deformation vectors to attain a high power. Using synthetic and clinical data we show that the performance of the Moore-Rayleigh test outperforms the classical permutation test and significantly lowers the computational time as it is not dependent on the randomization of the data.*

## 6.1 Introduction

Quantifying group differences between cases and controls gives insight in the development and progression of diseases. In brain research, Magnetic Resonance Imaging (MRI) is considered a useful tool for studying various aspects of diseases. Among others, brain morphometry has been exploited to indicate differences within and between groups of patients and controls [234]. Brain morphometry can be divided in three different groups: volume-based morphometry (VBM), shape-based morphometry (SBM), and deformation-based morphometry (DBM) [235]. SBM quantifies differences by shape parameterizations of the segmented brain structures [236–239], which is generally restricted to a few structures. VBM, also referred to as voxel-based morphometry [27, 240], is capable of analyzing the whole brain by quantifying the ratio of grey-matter, white-matter, and cerebrospinal fluid in each voxel.

Although both methods work very well for human subjects, in mice the segmentation of brain structures into gray and white matter is still challenging [28]. DBM offers an elegant solution to avoid segmentation by exploiting nonlinear registration to a common average and analyzing the resulting deformation fields [241], where the Jacobian matrix of a deformation vector encodes for the volume change of a voxel and is therefore a measure for shrinking or expansion of brain structures. DBM has been exploited to determine intra-group variation at a single timepoint [6, 242] or in longitudinal studies [243]. Inter-group variation can be determined by, e.g. the application of Hotelling’s  $T^2$  test (HT2) on the Jacobian [194] or directly on the deformation field vectors [195]. Studholme et al. presented a multi-variate linear model to detect the relationship between environmental variables (such as diagnosis, age and risk factors) and local brain shape changes [244]. Inter-group variation is assessed by the analysis of covariance (ANCOVA) of the multi-variate linear model in alcohol abuse patients [245] or in transgenic mice models [38]. Multi-variate parametric tests make assumptions of the underlying distributions, such as normality or equal variances. When these assumptions are not met, parametric tests might become unreliable.

Nonparametric test, with their minimal assumptions, provide an alternative for quantifying group differences in DBM. Permutation tests, for instance, provide an exact test for the quantification of group differences, with the only assumption that the groups are exchangeable under the null hypothesis. The performance of the Permutation tests depends on the selection of the test statistic, which has to be chosen so it best represents and separates the data: For instance, in the analysis of 3D deformation fields, the HT2 has been applied for the quantification of the local group differences [39, 194] or the partial least squares method on the Jacobian [246]. A disadvantage of the permutation tests is that their minimal significance probability ( $p$ -value) is bound by the number of iterations  $i$  and the sample size of groups ( $n$  and  $m$ ): Randomly relabeling the two groups results in maximal  $K = \frac{(n+m)!}{2(n!m!)}$  different combinations. If  $K \leq i$ , the original setting occurs by chance around  $\frac{i}{K} + 1$  times, resulting in a minimal  $p$ -value of  $p \gtrsim \frac{i+K}{K \cdot (i+1)}$ . This formula also shows the dependency of the minimal  $p$ -value on the number of iterations: In the case that  $n \rightarrow \infty$  and

$m \rightarrow \infty$ , the  $p$ -value is still limited by  $\frac{1}{i+1}$ .

**Example 3.** Consider 4 samples from a standard normal distribution in 2D:

$$X = \{(1.7, -0.2), (0.7, -1.2), (1.2, 1.9), (1.1, 3.0)\}$$

$$Y = \{(3.7, -0.2), (2.7, -1.2), (3.2, 1.9), (2.1, 3.0)\}$$

Here,  $Y$  is equal to  $X$  with a shift in the  $x$ -coordinate of 2 units. Testing for a group difference would be found significant by the HT2 with a  $p$ -value of 0.0068. If significance would be tested with permutations of the HT2, this differences would be found not significant with a  $p$ -value of  $\sim 0.0287$ , which can be determined by the following calculation: With the sample sizes of  $n = m = 4$ , the maximal amount of combinations that can be made is  $K = \frac{(8)!}{2(4!4!)} = 35$ . Therefore, if permutations with  $i = 10,000$  relabelings are applied, the original setting, as given above, appears by chance approximately  $\frac{i}{K} + 1 = \frac{10,000}{35} + 1 = 287$  times. Thus,  $p \approx \frac{i+k}{k \cdot (i+1)} 10,035 / (35 \cdot 10,001) = 0.0287$ .

The example shows the nature and disadvantage of permutation tests. For an accurate calculation of the probability to reject the null hypothesis a sufficiently large sample size is needed, but when the sample size increases, the number of permutations will become the limiting factor and has to be increased too. When the sample size has grown sufficiently large ( $(n = m) \geq 100$ ), the permutation test can be approximated by the standard  $t$ -test [198]. As sample sizes of 100 subjects are rare in an experimental setting, the highly computationally expensive permutation tests need to be performed. To lower the computational load of permutation testing, optimal stopping rules have been developed, although not (yet) put to practice [230, 247].

Despite the computational load, permutation testing on mean-based statistics, like Hotelling's  $T^2$  test, are popular in DBM. Firstly, because it still gives sufficient power to perform DBM, secondly, because the test fails gracefully in the case of a violation against the assumptions of the test statistic, and thirdly, by lack of another method of similar or improved power. Permutation tests with nonparametric test statistics, such as the Brunner-Munzel statistic [248], have already been considered for the quantitative groupwise comparison in functional Magnetic Resonance Images. Despite the nonparametric nature of the Brunner-Munzel statistic, Rorden et al. found that it could only outperform permutations with a  $t$ -test in a few cases as the  $t$ -test is relatively robust for violations in the assumptions.

In previous work [249], we generalized the two-dimensional Moore-Rayleigh test [200] to arbitrary dimensions. Furthermore, we have given the three-dimensional Moore-Rayleigh test (MR3) in closed form and we discussed the implications of the test in the two-sample problem. On synthetic data we have shown that the MR3 can be used for deformation-based morphometry and is robust for noise and applied registration algorithm [36]. In summary, it's advantages are threefold: (a) randomization of the statistic is not necessarily since the Moore-Rayleigh test is fully nonparametric; (b) the vector length and direction are directly incorporated in the test. (c) the resulting  $p$ -values of the Moore-Rayleigh test might be very low, which allows correction



for multiple tests, including the highly conservative Bonferroni correction.

In this work, we demonstrate how the MR3 can be exploited for regional quantitative morphometry of mouse brain MRI and how it compares to permutations of the Hotelling’s  $T^2$  test (pHT2). First, we shortly explain the principle of the MR3 and how it can be exploited for DBM. Afterwards we illustrate the behavior of the MR3 and the pHT2 by simulating situations that might occur in a deformation field and compare the power of both tests. Afterwards we apply both tests on real data, the transgenic APP23 mouse and its wild types. At the end of the paper, the Moore-Rayleigh test and its implications for deformation based morphometry are discussed.

## 6.2 Method

### 6.2.1 Formation of deformation fields

Consider two sets of 3D MR images taken from different populations of equal size. The first step in the analysis is to affinely register the images to an average  $A$ . This normalization step brings all images to the same coordinate system and removes all non-specific anatomical differences, like global orientation and scaling. From now on, we consider only the normalized sets of images  $I = (i_1, \dots, i_n)$  and  $J = (j_1, \dots, j_m)$ . A non-rigid registration defines the relation between the average and an image  $i$ , which is found by the minimum of a similarity measure  $\rho$ :

$$\mathcal{T}_i = \min \rho(i, A) \tag{6.1}$$

Assuming that there is no misregistration,  $\mathcal{T}_i$  indicates the local anatomical differences between  $i$  and  $A$ , which are coded by vectors in  $\mathbb{R}^3$ . Non-rigidly registering image sets  $I$  and  $J$  to the  $A$  results in two sets of deformation fields  $\mathcal{T}_I = (\mathcal{T}_{i_1}, \dots, \mathcal{T}_{i_n})$  and  $\mathcal{T}_J = (\mathcal{T}_{j_1}, \dots, \mathcal{T}_{j_m})$  for each homologous point in the average.

**Assumption 1.** *If all images in  $I$  belong to the same population as the subjects used to generate  $A$  (or  $A$  is the result of averaging the subjects from  $I$ ),  $\mathcal{T}_I$  represents the intra-group variation and noise. This results in vectors that are randomly spread around each point in  $A$ .*

**Assumption 2.** *If all images in  $J$  do **not** belong to the same class of the subjects used to build average  $A$ ,  $\mathcal{T}_J$  represents besides the intra-group variation and noise, also the inter-group variation between  $J$  and  $A$ . In case of a groupwise shape difference at a certain point  $x$ , the deformation vectors are likely to be not randomly spread around  $x$ , but indicate the direction of the shape difference.*

### 6.2.2 3D Moore-Rayleigh test

The 3D Moore-Rayleigh test (MR3) is a nonparametric test for spherical symmetry<sup>1</sup> of a group of three-dimensional real-valued vectors [36, 249]. Assumption 1 and 2 justify the MR3 on the separate voxels of the deformation fields:

To calculate the MR3 statistic for a single group of vectors, the first step is to scale the  $N$  vectors by the rank of their lengths, where  $X_{(1)}$  is the smallest vector of length 1 and  $X_{(N)}$  is the largest vector of length  $N$ . Under the null hypothesis, this ensures the nonparametric properties of the test. The summation of the ranked vectors is considered a Rayleigh random flight [201] with increasing steps:

$$R_N = \sum_{n=1}^N \frac{nX_{(n)}}{\|X_{(n)}\|} \quad (6.2)$$

The length of the resultant,  $\|R_N\|$  has a range of  $[0, \sum_{n=1}^N n]$ , as  $X_1, \dots, X_N$  are scaled by their ranks. Under the null hypothesis,  $\|R_N\|$  will be small and large values indicate the random flight is biased, i.e. most vectors point in the same direction. This test statistic is subject to the Moore-Rayleigh test, as described in more detail in section 5.2.2.

To test if a group of images  $I$  is significantly different from  $A$ , each separate voxel is tested by the Moore-Rayleigh test, as given in equation 6.2. For each voxel the following null hypothesis ( $H_0$ ) can be formed.

*$H_0$  : the probability density distribution of the deformation vectors is spherically symmetric, i.e. the elements of the vectors are drawn from a multivariate normal distribution. .*

The  $p$ -value of the Moore-Rayleigh test indicates the likeliness that, under the null hypothesis, the  $\|R_N\|$  for a certain voxel is returned and thus according to assumption 1 the likeliness that at this location no shape difference between  $I$  and  $A$  is found.

### 6.2.3 3D Moore-Rayleigh test for groupwise comparison

To test two  $I$  and  $J$  for group differences, their voxelwise relationship to  $A$  can be exploited:

**Proposition 3.** *If  $\mathcal{T}_I$  and  $\mathcal{T}_J$  are drawn from a spherically symmetric distribution, then the differences between  $I$  and  $J$  are also spherically symmetric distributed.*

This proposition implies that in case  $J$  is from the same population as  $I$  and  $A$ , the difference vectors between  $I$  and  $J$  must also be randomly distributed, as illustrated in figure 6.1(a). In case  $J$  is from a *different* population with a local shape difference in voxel  $x$ , the difference vectors will not randomly be spread around the origin, as shown in figure 6.1(b). It has to be noted that in theory these difference

<sup>1</sup>Vectors in a vector field are spherical symmetrical, if their magnitude and orientation (inward or outward) is uniformly distributed and only depending on the distance to the origin.

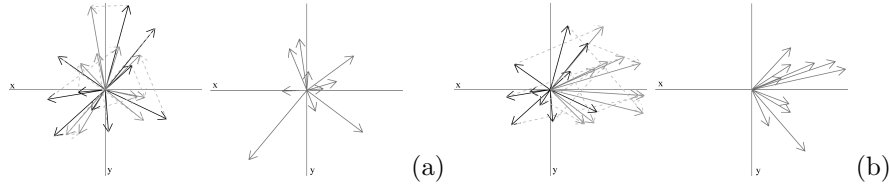


Figure 6.1: The deformation fields of a control group (black) and the test group (dark grey) for one voxel ( $n = m = 10$ ), where the difference vectors, see text for details, are given in dotted arrows. The right frame shows the difference vectors translated to the origin to show the spread of the differences. This situation is given for two differences in means:  $\Delta\mu = 0$  (a) and  $\Delta\mu = 2\sigma$  (b).

vectors are merely symmetric instead of spherical symmetric and therefore, the Moore-Rayleigh test might be potentially liberal [249], although this could not be detected by simulations on synthetic data [36].

The application of the MR3 on the vector fields is only valid if the right average has been created, as described in remark 4. If this requirement is not met, systematic differences between the control group and the average might be picked up and might be wrongly interpreted as differences between the control and the test group. An average can be created by simultaneously and iteratively registering a population of images to generate a true and new average [5], or by using the leave-one-out method, which ensures unbiased warping of controls to an average. The images of the mutants are afterwards matched to the whole average of controls.

**Remark 4.** *The average that is used to test for significant group differences between two populations **must** be created from subjects belonging to the same population as the control group.*

The vectors in the deformations fields are voxelwise tested by the Moore-Rayleigh test on the following null hypothesis:

$H_0$  : *There is no group difference between  $I$  and  $J$ , therefore the probability density distribution of the differences of the deformation vectors is spherically symmetric.*

The differences between  $I$  and  $J$  can be calculated in multiple ways [249]. For an asymptotically unbiased approach, the vectors can be matched by bootstrapping: For the same voxel, the vectors are randomly matched resulting in a value for  $\|R_N\|$ . This process is repeated  $i$  times, resulting in  $i$  values for  $R_N$ . The MR3 is calculated for the median of the resulting  $\|R_N\|$ s. Although this method is unbiased, it converts the test statistic  $\|R_N\|$  to a random variable and the Moore-Rayleigh test will not be repeatable. Therefore we propose to match the deformation vectors in each voxel according to their rank, where the largest vector in  $I$  ( $I_{(N)}$ ) is matched to the largest vector in  $J$  ( $J_{(N)}$ ), as shown in equation 6.3. Under the null hypothesis of spherical symmetry, this approach is unbiased for large sample sizes. However, in finite data set

still a small selection bias might occur in the pairing of the vectors for some voxels.

$$R_N = \sum_{n=1}^N \frac{n(I_{(n)} - J_{(n)})}{\|(I_{(n)} - J_{(n)})\|} \quad (6.3)$$

If two groups have unequal sample sizes it is advised to match on the highest rank is, as the largest deformation vectors usually code for group difference. A disadvantage of this approach is that the test might have a higher sensitivity for outliers especially in *ex vivo* imaging [6]. Therefore, in data with a high noise level, one can also decide to match on the one but highest rank in the case of unequal sample sizes. In this paper, however, we address only the case of equal sample sizes  $N = \min(n, m)$ .

### 6.2.4 Violations against the assumptions

Nonparametric methods make only minimal assumptions on the underlying data, such as continuity or symmetry (exchangeability). To justify the application of the Moore-Rayleigh test on clinical data, one main assumption is made on the data: It is assumed that if the average is representative for the the control group, the deformation vectors of nonlinear registration of the controls to the average leads to spherically symmetric distributed vectors.

In mouse studies, the genotypes of the subjects can be controlled and are, therefore, very similar. Yet, due to dehydration of the mice prior or during scanning or other influences, might cause this assumptions to be violated [180]. Violations against the assumptions of the MR3 can be easily found with the application of the one sample MR3 on the deformation vectors of the control group. If for a voxel a  $p$ -value is found that is smaller than a predefined critical value  $\alpha$ , the null hypothesis of spherical symmetry is rejected. For these voxels the assumptions are violated and the the MR3 cannot be applied directly as it might be too liberal. These particular voxels can be handled according to the following procedures:

#### *Exclusion*

By excluding the voxels that violate against the null hypothesis from the analysis, it is certain that no incorrect conclusions are drawn.

#### *Randomization*

The framework of permutation testing provides an exact test without prior assumptions on the data besides exchangeability. Therefore, the  $\|R_N\|$  can be used as test statistic in permutation tests, thus allowing to incorporate the MR3 without further assumptions on the data.

## 6.3 Results

In this section we evaluate the power of the Moore-Rayleigh test with the vectors paired by their rank (MR3) and the permutation test using the Hotelling's  $T^2$  test as

test statistic with 10,000 relabelings (pHT2) for comparison purposes [39] on simulation data in section 6.3.1 and on experimental data in section 6.3.2.

### 6.3.1 simulation data

The type 2 error is also called the power of a statistical test, indicated by  $\beta$  and is defined as the fraction of rejections of the null hypothesis at a certain significance level  $\alpha$ . The power of the proposed tests is evaluated by monte carlo simulations on simple simulations of practical situations, with  $\alpha$  fixed at 0.01.

#### Discriminating power

Assuming that the intra-variation of the control group and the mutants are spherically symmetric, 2 gaussian points clouds in 3D were simulated defined by a mean ( $\mu$ ) and standard deviation ( $\sigma$ ). Each point  $x$  in a cloud represent a deformation vector from the origin to  $x$ . A group difference is simulated by a difference in the means of the two point clouds, i.e.  $\Delta\mu > 0$ . The discriminating power of MR3 and the pHT2 was calculated for various differences in mean ( $0 \leq \Delta\mu \leq 10\sigma$ ), while the standard deviation was kept constant ( $\sigma=1.0$ ) for both groups. For each value of  $\Delta\mu$ ,  $N$  points were randomly drawn from the two distributions, which were afterwards tested for group difference by the three tests, where  $N = \{5, 7, 10, 20\}$ . This was repeated a 1,000 times. The power for  $\Delta\mu$  is estimated by the ratio of the number of rejections of the null hypothesis (group difference detection) to the number of iterations.

#### Violations to the assumptions of spherical symmetry

To show the robustness of the MR3 to violations to the null hypothesis, two gaussian disks were simulated with  $\sigma_x = \frac{1}{2}\sigma_y = \frac{1}{2}\sigma_z$  with sample sizes  $N = \{5, 7, 10, 20\}$ . Two power tests were performed: The first one tests the performance when the main direction of variance is in the direction of the group difference, therefore  $\Delta\mu$  was varied from 0 (completely overlap) to  $10\sigma_x$  (completely separated) in the  $x$  direction. The other test was analyzing the performance when the main direction of variance was *perpendicular* to the direction of the group difference, therefore  $\Delta\mu$  was varied from 0 to  $5\sigma_y$  (completely separated) in the  $y$ -direction.

#### Results

In figure 6.2 the power functions for the MR3 and pHT2 are shown for sample sizes of  $N = \{5, 7, 10, 20\}$ . In case of no violations against the null hypothesis, the MR3 has a better performance than the pHT2 when the sample sizes are small. This effect fades out if the samples per group are larger than 10. In case the assumptions of spherical symmetry are violated, shown in figure 6.3, both the MR2 as the pHT2 seem to fail gracefully, i.e. the tests will become more conservative. Interestingly, in case the violations against the assumptions are in *perpendicular* direction of the group differences (figure 6.3(B, D, F, and H) ), the MR3 becomes more conservative

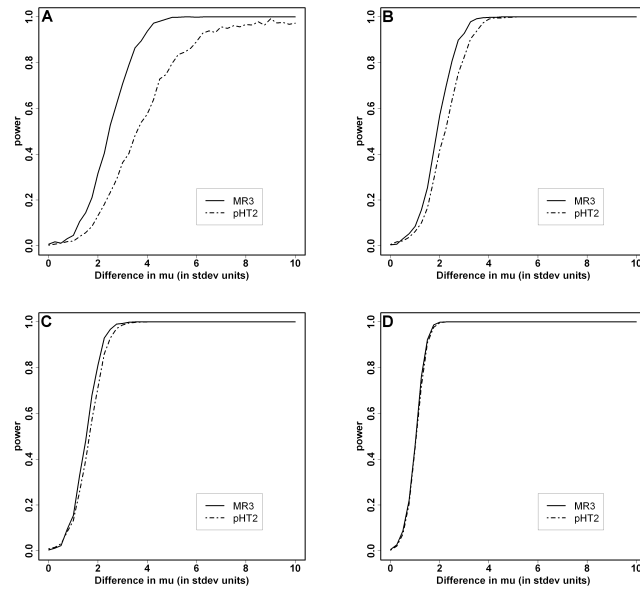


Figure 6.2: The estimated power for the discrimination of two gaussian spheres with  $\sigma = 1$  and relative displacement of  $\Delta\mu$  with  $N = 5$  (A),  $N = 7$  (B),  $N = 7$  (C) and  $N = 20$  (D).

than the pHT2. In the case that the violations against the assumptions are *in the same* direction of the group differences, the MR3 is outperforming the pHT2. In the case of extreme small sample sizes (shown in figure 6.2 (A) and figure 6.3 (A and B)) the pHT2 have limited power .

### 6.3.2 Experimental application

#### Image formation

In this study, we compared the APP23 mouse brain with control mice. The APP23 is a transgenic mouse model for Alzheimer’s Disease, which expresses the deposit of amyloid plaques in the brain and its vascular structure, resulting in brain atrophy in the cortex, hippocampal region and cerebellum [89,103]. Atrophy in these regions causes enlargement of the ventricles from 6 months onwards, for more details see section 3.2. Due to this well known pathology, this dataset is suitable for validation purposes of the MR3 and to compare its performance to the performance of the pHT2.

For the experiments, 6 APP23 mice and 6 controls from the same or parallel litters were considered. Mice were hemizygous for the transgene of interest. Animals were housed under standard conditions and fed a standard diet and water supply *ad libitum*. Animal handling, care and experimental use have been performed in line with the Swiss Federal Law for animal protection (animal licenses BS No. 1094 and 1283). The mice were imaged at the age of  $19.7 \pm 0.45$  months, their age-matched non-transgenic littermates served as controls. For the MRI investigations, the animals were anesthetized with 1.3% isoflurane (Abbott, Cham, Switzerland) in a mixture of oxygen/N<sup>2</sup>O (1:2) administered via a face mask. The body temperature of the mice was kept at 37 degrees Celcius. No stereotactic holding was used. Measurements were performed with a Biospec 47/40 spectrometer (Bruker, Medical Systems, Ettlingen, Germany) with field strength of 4.7 T, equipped with an actively shielded gradient system. A (3D) gradient-echo sequence was used with the following imaging parameters: TE=8ms, TR=40 ms, 2 averages, matrix=256×192×48, image resolution=2.8×7.5×30mm.

As objective we set to detect brain shape differences between the APP23 mice and the controls. So, we adopted as null hypothesis:

$H_0$  : *There is no difference between the brains in APP23 mice and their controls*

First, all images were normalized by an affine transform to correct for all non-significant anatomical differences, as global orientation, global shape and brain. The MR images were very noisy with low contrast in the brain, where only the ventricles were clearly visible as shown in figure 6.4 (A and B).

To reduce the amount of misregistration that negatively influences the results of the automated morphometry [41,42] a mask was created to extract the ventricles, since the main objective of this study was to measure shape differences of the ventricle area. Doing so, the registration algorithm was optimized for ventricle normalization,

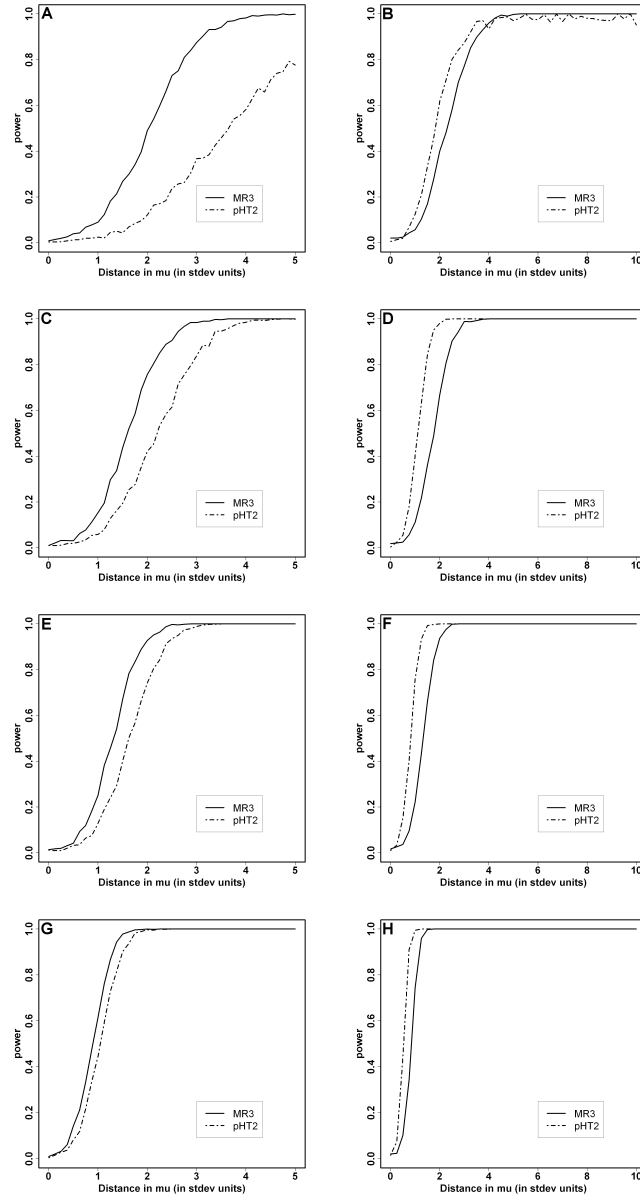


Figure 6.3: The estimated power for the influence of violations against the assumption of spherical symmetry for two gaussian spheres with  $\sigma_x = \frac{1}{2}\sigma_y = \frac{1}{2}\sigma_z$  with sample sizes  $N = 5$  (A and B),  $N = 7$  (C and D),  $N = 7$  (E and F) and  $N = 20$  (G and H), where the left column shows the displacement of one cloud in the  $x$ -direction and the right column shows the displacement of a point cloud in the  $y$ -direction.



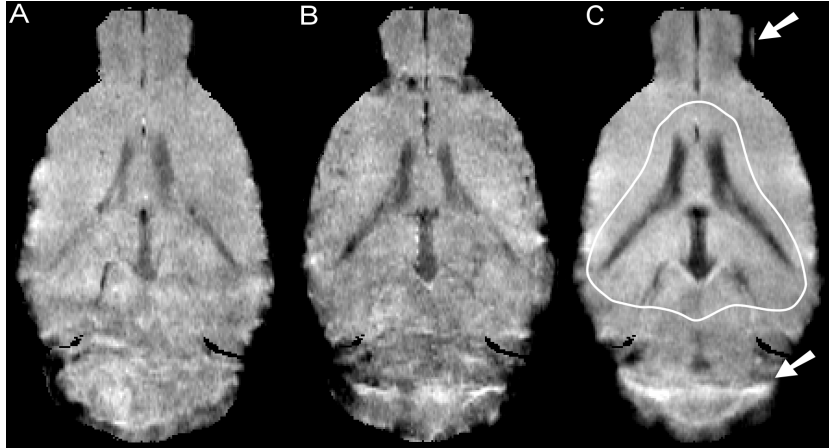


Figure 6.4: A slice from the 3D volume of a single control subject (A), the APP23 mouse (B), and the control average (C). Arrows indicate regions which show inhomogeneities, misregistration or incorrect masking. Furthermore, the region around the ventricles is delineated as region of interest.

discarding all other brain areas. All mice were affine registered with mutual information to an in-house developed mouse average. The normalization process resamples all images to the same dimensions ( $X=160$ ,  $Y=132$ ,  $Z=255$ ) and an isotropic voxel size of 0.06mm, allowing voxel wise comparison between the different scans. Afterwards, the whole brain is extracted from the image by masking the background, skull and redundant brain tissue. The resulting average of the normalized wild types including the mask for the region of interest (ROI) is shown in figure 6.4 (C). The white arrows indicate regions outside the ROI that are misaligned in order to ensure the best fit for the areas inside ROI. For this reason, all areas outside the ROI are therefore excluded from further analysis. As a result, the null hypothesis changes to:

$H_0$  : *There is no difference between the ventricles areas in APP23 mice and their controls*

As described in section 6.2.3, the formation of the average is a critical step in the analysis. If the average is not from the same population as the control group, the assumptions of spherical symmetry (thus intra-group variation and noise only) might be violated. Therefore we generate the average from the control group itself. As the sample sizes are small, a leave-1-out strategy was chosen for the warping of the controls: Each control mouse was nonlinearly registered to the average of the residuals. The APP23 mice were afterwards warped to the average of all controls. The nonlinear registration was performed using the symmetric Demons algorithm [189], as implemented in ITK [192]. The resulting vector fields of the Demons algorithm were

used to test for group differences with the pHT2 and the MR3. In addition to that, a one-sample Moore-Rayleigh test was applied on the two separate groups to screen the deformation fields for violations against the assumptions.

## Results

Figure 6.5 shows the result for the one sample case of Moore-Rayleigh test for the controls and the APP23 mice overlaid on the average. Only voxels that are found significant ( $\alpha < 0.05$ ) are overlaid on the average, for these voxels it is likely that the assumptions of the MR3 are violated, i.e. are not spherically symmetric. The image shows that the assumptions for spherical symmetry (assumption 1) holds for most regions in the control brain, which justifies the application of the Moore-Rayleigh test. The large amount of voxels for which the null hypothesis is rejected in the APP23 mice inside the ROI suggests already a group difference, which can be quantified by the two-sample Moore-Rayleigh test.

The significant voxels that are located outside the ROI are possibly due to intensity inhomogeneities or misregistration and are already excluded from the analysis. However, the significant voxels that are located inside the ROI will not be tested with the MR3 as indicated in section 6.2.3, but will be tested by permutation test with the MR3 as test statistic as described in section 6.2.4.

Both the pHT2 and the MR3 were used to test the APP23 mice and their controls for group differences. The results are shown in figure 6.6. Both test find similar significant differences in the ventricle region. By comparing the  $p$ -values of the MR3 and the pHT2 in the ventricle region, we find significantly lower  $p$ -values for the MR3, which are overall a factor 10 smaller compared to the pHT2. Calculation time on the same computer for the MR3 and pHT2 was 10 minutes and 5 hours, respectively.

Figure 6.6 also demonstrates the effect of misregistration and image artifacts on the results of morphometry. In the regions outside the ROI, especially in the cerebellum (lower part of the image), large significant region is detected, that likely due to misregistration and other image artifacts.

## Generalization issues

To generalize the results from voxel level to brain level, multi-test correction has to be applied. Bonferroni correction has the strongest control of the Type 1 error, leading to the least false positives. In this study the ventricles and surrounding tissue were segmented from the volume (209,232 voxels), therefore Bonferroni correction results in a rejection of the  $H_0$  of equal ventricles if the  $p$ -value is smaller than  $0.05/209,232 = 2 \cdot 10^{-7}$  for one or more voxels inside the ROI. In the case of Bonferroni correction, the MR3 has 34 significant voxels left, allowing the acceptance of the alternative hypothesis that there is a difference in the ventricles of the APP23 mice and their controls. To achieve the same result with the permutation test, over 5,000,000 iterations are needed. Therefore other multiple-test correction method are proposed, such using permutations [250] or the random field theory [45].

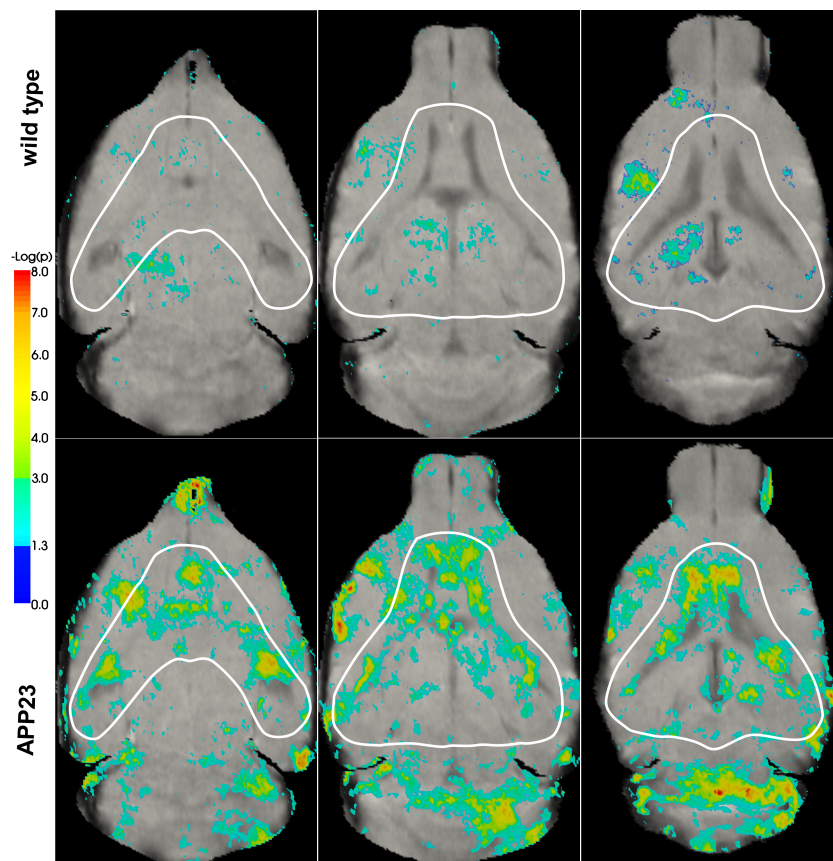


Figure 6.5: The probability that a voxel is spherically symmetric for the wild types (top row) and the APP23 mice (bottom row) for three different slices through the brain. The areas which were not found spherically symmetric at a significance level of  $\alpha < 0.05$  were color coded by their  $p$ -value.

In this study, generalization of the null hypothesis to the *whole* brain is not possible, as the registration is not optimal in regions other than the ventricles. In theory however, if Bonferroni correction was applied for the whole brain (1.9 million voxels), the  $p$ -values of some voxels in the ROI would still found significant after Bonferroni correction of  $2 \cdot 10^{-9} = 0.05/1.9 \cdot 10^6$ , i.e. 5 voxels for MR3 and therefore the null hypothesis for shape difference in the whole brain between the APP23 mice and their controls would be rejected.

## 6.4 Discussion

This work presents the nonparametric three-dimensional Moore-Rayleigh test (MR3) as alternative test for permutation testing in deformation-based morphometry. The results on simulated data show that the MR3 outperforms permutation tests on Hotelling's  $T^2$  test (pHT2) in the detection of difference of mean in small sample sizes, since the pHT2 has to estimate the covariance matrix from the data. The results on real data show that the MR3 and the pHT2 classify similar regions in the brain as locally significant different, where MR3 has more power in detecting these regions.

The  $p$ -value of permutation tests is limited for small sample sizes and by the number of iterations, which explains the lower power for the pHT2 compared to the MR3. But another important limitation of the pHT2 is that Hotelling's  $T^2$  test needs to estimate the covariance matrix for the two groups of vectors, which might be inaccurate for small sample sizes. As already shown in section 6.3.1, the power of the MR3 is higher than the permutation test for small sample sizes as the MR3 has no limit on its  $p$ -value *and* uses the whole vector information in contrast to permutation testing and no covariance has to be estimated from the data by the Moore-Rayleigh test. Permutations of the Hotelling's  $T^2$  test will only outperform the MR3 when the violations against the assumptions are in the direction of the group differences. If the tests are compared by power and speed, the MR test is outperforming the pHT2.

If the assumptions of the MR3 are violated, our results show that the Moore-Rayleigh test fails gracefully and becomes more conservative. Interestingly, the opposite has been proven for the Wilcoxon signed-rank test, the equivalent of the one-dimensional Moore-Rayleigh test. Violations against the assumptions of the Wilcoxon signed-rank test will make the test excessively liberal [219, 220]. Since it is theoretically expected that the MR3 might be liberal [249], an alternative test has been proposed: If the assumptions of the MR3 are violated for a certain voxel, that voxel is tested for significance differences by permutation testing with the MR3 as test statistic.

In experimental data, the Moore-Rayleigh test has proven to be a powerful tool to quantify group differences and the results are comparable to the permutation tests on Hotelling's  $T^2$  statistics as currently used in human brain analysis [39]. Furthermore, the regions detected by both tests are located on the areas of the ventricles that are connected to either the hippocampus, the thalamus, or the caudoputamen, which

correspond to the pathology of the APP23 mice [89].

The applicability of voxel wise discrimination methods is a topic for discussion over the years. There are two main problems at the interpretation of voxel wise morphometry:

First, registration errors influence the result and lead to incorrect rejections of the the null hypothesis [41]. Therefore, it is recommended to validate statistical test results on the data before reporting clinical implications [251]. Our findings on clinical data confirms these statements, where the cerebellum region was not perfectly aligned for some cases. This initial misregistration resulted in many false positives, found by both the MR test and the pHT2. Already in previous work we found no significant difference of significance detection by changing deformation algorithm [36], although nonlinear registration algorithms differ in performance [252].

Secondly, generalizations of the conclusions at voxel level to structure level or even brain level should be done with great caution [42]. The most important critic is that voxels are assumed to be independent, while in fact they are correlated to each other. From a statistical point of view, the generalization issues of voxel wise morphometry can be interpreted as a multiple-hypotheses testing problem with a strong control of the familywise error. The Bonferroni test has the strongest control of the type 1 error and is therefore the most reliable test for avoiding false positives. However since most tests are dependent on relabelings of the permutations, other methods can be applied [44]. Since the Moore-Rayleigh test is not bounded by relabelings it has no lower bound on its  $p$ -values, which allows the application of Bonferroni correction. Therefore, the conclusions of the MR test can be generalized in the safest way possible to structure level and even to brain level.

In conclusion, we presented the 3D Moore-Rayleigh test as a single nonparametric statistical test for deformation-based morphometry that offers an elegant alternative for the computational expensive permutation testing. The test is not only faster to compute, but also more suitable for small sample sizes and multiple-test correction can be applied without problems.

### 6.4.1 Implementation

The code for the various MR tests in R and C++ have been made publicly available at the website <http://folk.ntnu.no/muskulus/>.

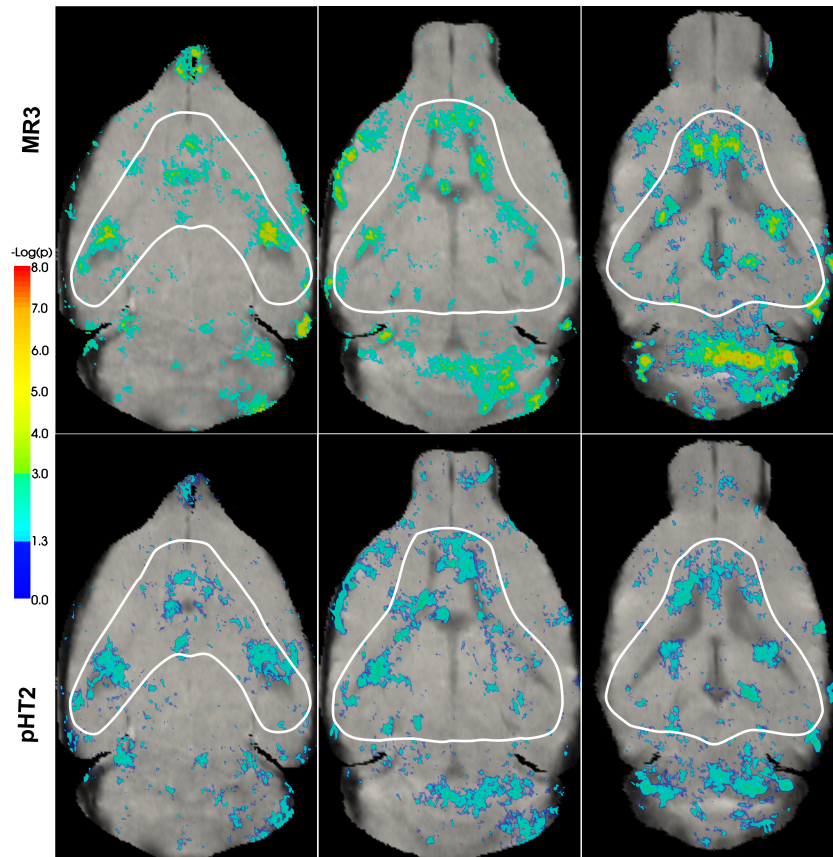


Figure 6.6: The probability per voxel that there is locally a group difference between wild type mice and APP23 mice for three slices through the brain. The top row shows the results from the permutations of the Hotelling's  $T^2$  test (pHT2) and the bottom row shows the results from the single Moore-Rayleigh test (MR3). The areas which were not found spherically symmetric at a significance level of  $\alpha = 0.05$  were not printed.

## CHAPTER 7

---

### Morphological assessment and validation of the transgenic Cacna1a knockin migraine mouse model by *in vivo* MRI

---

A.E.H. Scheenstra  
R.C.G. van de Ven  
R.R. Frants  
J.H.C.Reiber  
M.A. van Buchem  
F. Admiraal-Behloul  
M.D. Ferrari  
J. Dijkstra  
A.M.J.M. van den Maagdenberg  
L. van der Weerd

*This chapter was adapted from:*  
Morphological assessment in a transgenic Cacna1a knockin migraine mouse model by  
*in vivo* MRI. *In preparation*

**abstract:** *Non-rigid registration of MR images to a common reference image results in deformation fields, from which anatomical differences can be statistically assessed, within and between studies. Without further assumptions on the underlying distributions, nonparametric tests are needed and usually the analysis of deformation fields is performed by permutation tests. However, permutation tests are computationally expensive and have limitations by sample size and number of iterations. In this paper, we consider a single nonparametric test as an alternative for permutation tests; the 3D Moore-Rayleigh test. As its distribution function is available in closed form, permutation testing can be avoided. Furthermore, the test incorporates both the directions and magnitude of the deformation vectors to attain a high power. Using synthetic and clinical data we show the performance of the Moore-Rayleigh test outperforms the classical permutation test and significantly lowers the computational time as it is not dependent on the randomization of the data.*



## 7.1 Introduction

Migraine is a neurological, paroxysmal disorder affecting up to 16% in the general population. Patients suffer from throbbing, often unilateral headaches lasting 4 to 72 hours that are accompanied by nausea, vomiting and/or photo- and phonophobia [253]. Of all migraine patients about one-third suffers from migraine with aura (MA), characterized by an aura that consists of visual disturbances, but sensory-, motor- or speech-related phenomena can occur as well [254]. Familial hemiplegic migraine type-1 (FHM1) is an autosomal dominant subtype of MA, caused by mutations in the CACNA1A gene. This gene encodes the pore forming  $\alpha_1$ -subunit of  $\text{Ca}_v2.1$  (P/Q-type) calcium channels.  $\text{Ca}_v2.1$  channels open upon membrane depolarization, allowing  $\text{Ca}^{2+}$  influx and subsequent neurotransmitter release.

Thus far, 3 FHM genes have been identified [255]. The FHM1 CACNA1A gene encodes the pore-forming  $\alpha_1$ -subunit of neuronal, voltage-gated  $\text{Ca}_v2.1$  (previously known as P/Q-type) calcium channels [256,257]. Two knock-in FHM1 mouse models, one carrying the human pathogenic R192Q missense mutation [258,259], the other the S218L CACNA1A mutation [260–262], were generated. In patients, the R192Q mutation causes pure FHM without other associated neurological features [256], whereas the S218L mutation causes a severe migraine phenotype with excessive and often fatal cerebral edema [263]. When expressed in transfected cultured neurons, both mutations shift channel opening toward more negative membrane potentials and delay channel inactivation; the S218L mutation causes more pronounced single-channel gain of function than R192Q [258,264]. As a result, channels open with smaller depolarization and stay open longer, allowing more  $\text{Ca}^{2+}$  to enter presynaptic terminals. FHM1 mouse models exhibit a reduced threshold for electrically evoked cortical spreading disease (CSD) and increased SD velocity [259,265]. The S218L FHM1 mutation, but not the R192Q FHM1 mutation, increased the probability of multiple CSD events in response to only a single threshold stimulus. Whereas WT and R192Q mice generally experienced a single CSD event on a single stimulus, the S218L mice showed a gene dosage-dependent increased probability of having successive CSD events [262].

The neuropathology of MA patients shows some MRI abnormalities, like cerebellar atrophy, thickening of the somatosensory cortex and sub-clinical white matter abnormalities in the cerebellum and brainstem correlating with attack frequency [266,267]. However, the nature of these MRI abnormalities is not known, and their relationship to this disorder is still not well understood. In light of these findings in patients, the present study investigates the morphological phenotypes of FHM1 mouse models, using *in vivo* MRI in conjunction with semi-automated volumetry and deformation-based morphometry [36] to quantify group differences. Thus, the goal of this investigation is to validate the FHM1 mouse model morphologically as a valid model for future migraine studies.

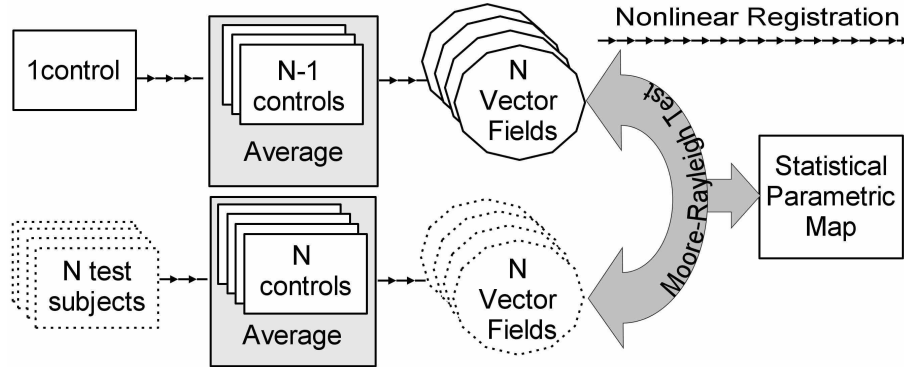


Figure 7.1: The schematic protocol for post processing of the normalized MR images for a cross-sectional study, so the two groups (control group and test group) can be tested for local significant differences with the Moore-Rayleigh test.

## 7.2 Materials and methods

### 7.2.1 Specimen preparation

Female FHM1 mutant mice, homozygous for R192Q [259] or S218L [260] mutation in the mouse *Cacna1a* gene (encoding the  $\alpha 1A$  pore-forming subunit of  $Ca_v2.1$  channels), were compared with wild type (WT) littermates. The dataset consisted of 7 WT mice, 7 R192Q mice and 5 S218L mice at an age of  $15.5 \pm 2.5$  months,  $17.8 \pm 2.2$  months and  $12.6 \pm 1.1$  months, respectively. For *in vivo* MR imaging, mice were initially anaesthetized with 4% isoflurane in  $O_2$ /air (50/50%) and maintained on 1.5% isoflurane during the procedure. The respiratory rate was monitored via an air-pressure cushion and Biotrig software (Bruker, Rheinstetten, Germany). After the MRI acquisitions, the brains were dehydrated and embedded into paraffin. Hematoxylin-Eosin and Klüver-Barrera staining were performed on  $5 \mu\text{m}$ -thick sections, using standard protocols. All animal handling and experiments were performed in accordance with the guidelines of the universities and national legislation.

### 7.2.2 Magnetic resonance imaging

*In vivo* imaging of the mice in interictal state was performed on a Bruker 9.4 T vertical 89-mm-bore magnet (Bruker BioSpin, Rheinstetten, Germany) with a Bruker Micro2.5 gradient system and a transmit/receive birdcage radio frequency coil with an inner diameter of 30 mm. Bruker ParaVision 3.0 software was used for image acquisition. *In vivo* anatomical images were acquired using a  $T_2$ -weighted multi-slice spin echo sequence. Imaging parameters were: TE = 35 ms, TR = 6 s, FOV = 25.6 mm, matrix =  $256 \times 256$ , 40 slices of 0.2 mm thickness, with 4 averages. Total scan

time was 102 minutes.

### 7.2.3 Histological evaluation

After MR imaging, the animals were sacrificed by intracardial perfusion with 4% PFA and processed for standard histochemical staining with hematoxylin and eosin and Klüver-Barrera. All sections were cut in coronal direction with a thickness of 5  $\mu\text{m}$ .

### 7.2.4 Inspection correlated to human studies

First all *in vivo* MRI of the mice were visually examined for morphological abnormalities by a mouse brain expert and a radiologist, both blinded for genotype. Special attention was given to hyper- and hypointense lesions, as in human studies sub-clinical white matter abnormalities in the cerebellum and brainstem correlate with attack frequency. The comparison of histological brain tissue of R192Q mice S812L mice with the wild types did not demonstrate any obvious structural abnormalities (not shown).

### 7.2.5 Image normalization

All *in vivo* MR images were normalized by affine registration (12 parameters) to an in-house developed reference brain<sup>1</sup> to correct for all non-significant anatomical differences, such as global orientation, global shape and brain. The normalization process resamples all images to the same dimensions ( $160 \times 132 \times 255$ ) with an isotropic voxel size of 56  $\mu\text{m}$ , allowing voxel wise comparison between the different scans. Afterwards, the whole brain is extracted from the image by masking the background, skull and redundant head tissue. Unless otherwise specified, the normalized and masked *in vivo* MR scans were used for all further analysis.

### 7.2.6 Automated morphometry

The three groups of mice were compared to each other (WT versus R192Q, WT versus S218L and R192Q versus S218L) by deformation-based morphometry: For each cross-sectional study, one group of mice was selected as control group defining the baseline brain shape. The other group was considered as test group, which was tested for equal brain size and shape relative to the control group. The process of the preparation of the data is illustrated in figure 7.1. Average images were created from the normalized MR images from control group for each cross-sectional study. As the sample sizes were small, a leave-1-out strategy was chosen for the warping of the controls to avoid correlations: Each control mouse was nonlinearly registered to the average of the residuals. The normalized MR images of the test group were afterwards warped to the average that was constructed from all control mice. The nonlinear registration was performed using the symmetric Demons algorithm [189],

---

<sup>1</sup>An average created from *in vivo* MR images of 7 normalized C57Bl/6J mice.

as implemented in itk [192]. The resulting vector fields indicate at each voxel the spatial relationship to the corresponding voxel in the average. Statistical analysis was applied to quantify group differences in the spatial displacement to the average at voxel level by the application of the 3D Moore-Rayleigh test [36,249]. As a significant group difference in spatial displacement implies a significant brain shape difference, we defined hypotheses for the cross sectional studies: There is no brain shape difference between:

1. the wild type R192Q mice (control group) and the R192Q mutants (test group)
2. the R192Q mutants (control group) and the S182L mutants (test group)
3. the wild type R192Q mice (control group) and the S182L mutants (test group)

As the significance detection was performed for each voxel independently, corrections for multiple comparisons were required before conclusions can be generalized that there are differences between the total brains of the groups. In this study, Bonferroni multiple-test correct is applied, since Bonferroni correction has the strongest control of the family-wise error rate, i.e. the probability that after correction for multiple-tests still somewhere in the brain a false positive (an incorrect significant difference) is detected. With this correction, the resulting  $p$ -values from the Moore-Rayleigh test are adjusted by multiplication with the number of tests performed per cross-sectional study (which equals the number of voxels in the normalized brain volume), while the significance level  $\alpha$  remains at 0.05.

### 7.2.7 Volumetry

As the nonlinear registrations were performed for each individual, volumetric measurements were assessed by the segmentation of the three averages from the control groups, as shown in figure 7.2. This segmentation is afterwards warped on top of the individual subjects, after which the individual segmentations were manually corrected. Volume measurements of the total ventricular space, the cortex, the hippocampus and the cerebellum were performed by multiplying voxel-count by voxel size, which were calculated as percentage of the total brain volume. Significant volume changes of the brain structures were calculated by the nested Kruskal-Wallis test, which is the nonparametric alternative for the MANOVA test.

## 7.3 Results

The first step was the visual inspection of the MR volumes and histological sections. No morphologic abnormalities and clearly visible hyper- and hypointense lesions were found (not shown). To detect any other significant differences between the brains of the different groups, the volumes were analysed on a voxel-by-voxel basis using deformation-based morphometry. With this test, some significant differences between the brains were detected.

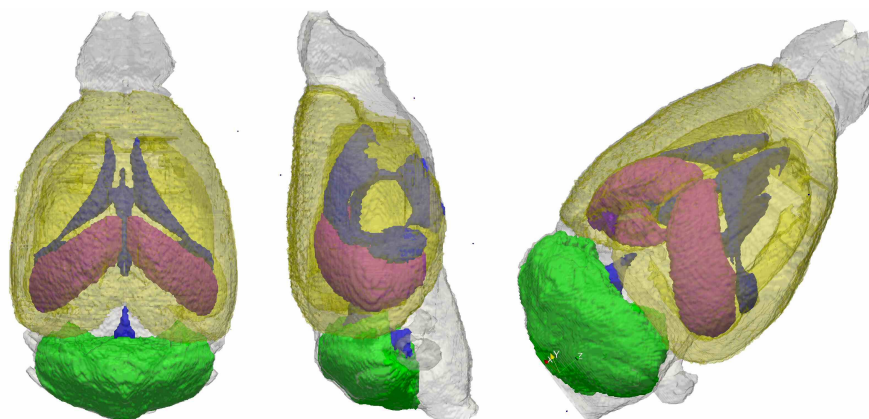


Figure 7.2: Volume rendering of the manual segmented cerebellum (green), the hippocampal regions (purple), the cortex (yellow) and the ventricles (blue)

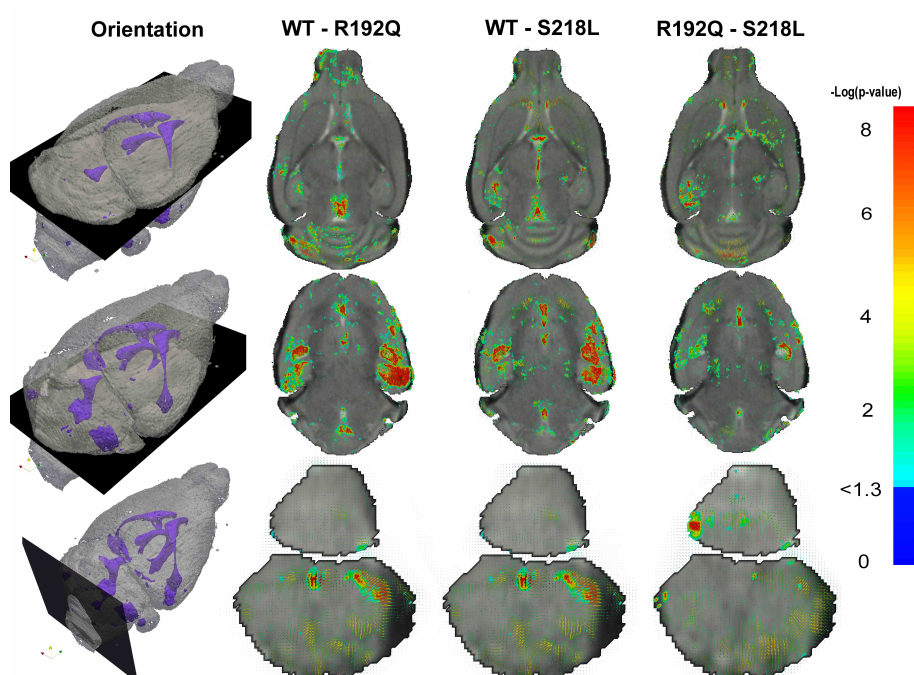


Figure 7.3: The deformation-based morphometry results for the three cross-sectional studies. The statistical parametric map indicating the significance level per voxel for three slices in the brain. The MR image is the average of each control group.

Figure 7.3 shows the areas that are found significantly different between the three groups. In the figure, three slices through the brain are overlapped on top of the average of the control group. The significance level is indicated with a logarithmic scale shown at the right hand side of the image, and coded in such a way that all red voxels are found significant after Bonferroni multiple-test correction. Main significantly different areas are found around the cerebellum, the hippocampal area, the ventricles and some small areas in the brainstem. Figure 7.4 shows a volume rendering of the brain with the areas indicated that are still found significant after Bonferroni correction. In this figure, the ventricles are shown for a better visual interpretation of the volume.

Based on these results, the MR images were again visually inspected to exclude any type of MR artifacts or misregistration to avoid incorrect conclusions [41, 42]. Some small discrepancies were detected in the averages of the MR images, as displayed in Figure 7.5. It seems that the averaging of the subjects, introduced subtle image intensity enhancing in the lower hippocampal regions and ventricular areas. Since this image enhancement is not present in the single MR images, this might have caused false positives in the hippocampal regions and the ventricles. The significant areas in the cerebellum could not be explained by any visible discrepancies in the MR images. Therefore, the averages of the different groups were placed side to side to visually check for structural differences that could explain the significant areas. Visual comparison of the averages shows a cerebellar shift in the caudal direction, as shown in Figure 7.6.

This led us to a final analysis on the mouse brain, performed by traditional volumetry on the structures that were found significant by DBM, i.e. cortex, cerebellum, hippocampal area and the total ventricular space were segmented. The brainstem was excluded for volumetry as it is located at the boundary of the image and can therefore not be equally segmented for all images, leading to unreliable volumetry results. The structures used for volumetry are shown in figure 7.2. The volumetric measurements (in % of total brain volume) are listed in Table 7.1.

No significant differences were found between the groups for all structures with the Kruskal-Wallis test ( $p$ -value = 0.16), therefore it was not statistically correct to continue with testing the separate brain structures for group differences, thus no significant structures can be reported based on volumetry.

Brain structure	Wild type	R192Q	S218L
Cortex	30.37 ± 0.32	29.55 ± 0.44	31.23 ± 0.71
Cerebellum	10.89 ± 0.37	11.14 ± 0.26	10.84 ± 0.62
Hippocampus	5.10 ± 0.28	5.24 ± 0.17	5.26 ± 0.20
Ventricles	1.51 ± 0.31	1.54 ± 0.21	1.27 ± 0.14

Table 7.1: The normalized brain structure volumes expressed in % of the total brain volume ± the standard deviations.

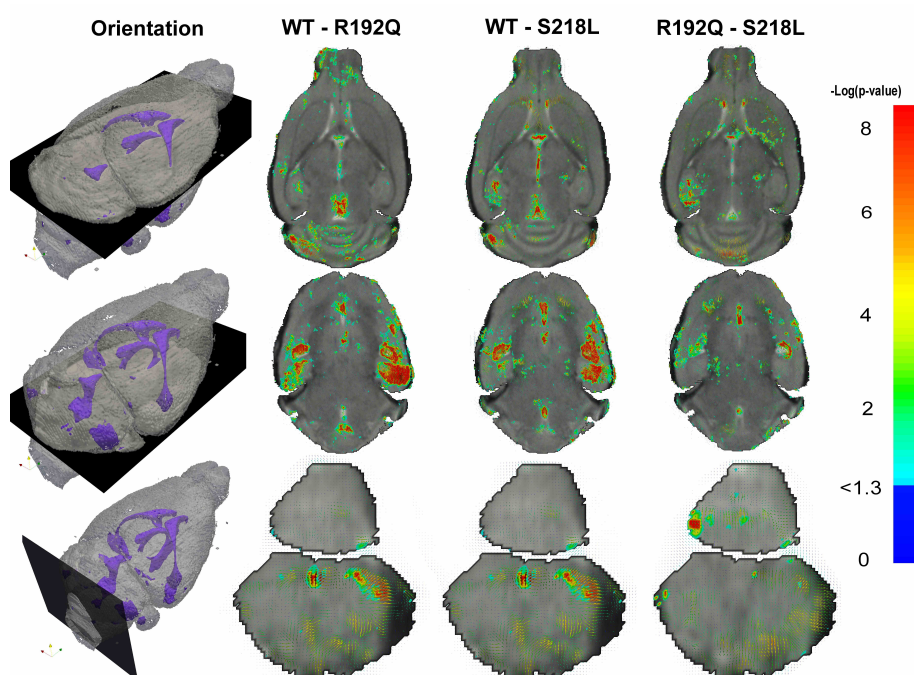


Figure 7.4: A volume visualization of the regions that were found significant under Bonferroni correction (red). For a better interpretation of the spatial orientation in the brain, the ventricles are also shown (blue).

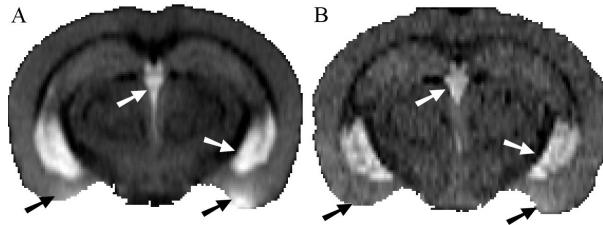


Figure 7.5: A comparison of an MR slice of the control average and a regular subject. The arrows indicate areas where some intensity enhancement due to the averaging are visible.

## 7.4 Discussion

In this work, we morphologically validated the transgenic *Cacna1a* knockin strains as migraine mouse model. Unbiased whole-brain analysis of *in vivo* MRI of FHM1 mice using deformation-based morphometry combined with visual inspection of the volumes showed a slight increase in cerebral volume and a caudal displacement. In addition, with deformation field analysis, changes in the cerebellum, cortex and brainstem were found. While at the same time, no large volumetric differences were detected in the anatomy of the three groups.

So far, morphometry studies in human migraine patients show conflicting results: One case-matched study reports no morphological changes between patients suffering from migraine with aura (11 patients) and without aura (17 patients) [268], while another case-matched study on 27 patients with migraine finds some minor significant gray matter changes [269]. From both studies it can be concluded that the groups are too small and heterogeneous to make any strong conclusion on significant structural brain changes [270]. By the exploitation of transgenic mouse models, it is possible to generate homogeneous groups that are highly suited for voxel-by-voxel morphometry measurements. Since migraine is a functional disease where the number of anatomical changes is correlated to the number of attacks, it is not surprising to find no significant volumetric differences between the brain structures of the groups of mice, since the sample sizes of the wild types (7 mice), R192Q mice (7 mice) and S218L mice (5 mice) were too small to detect any subtle volumetric differences between the brain structures.

Using deformation-based morphometry, we found some significant areas in the cortex, brainstem and cerebellum. Combined with visual inspection, a dorsal shift in the cerebellum was found. Our results on volumetry show that displacement of the cerebellum is probably not due to enlargement of the ventricles. In addition, volumetric measurements on the cerebellum suggest an increased volume of the cerebellum in R192Q mice, although the differences were not significant. Interestingly, subclinical cerebellar impairment has been reported in migraine patients [271]. A later study on



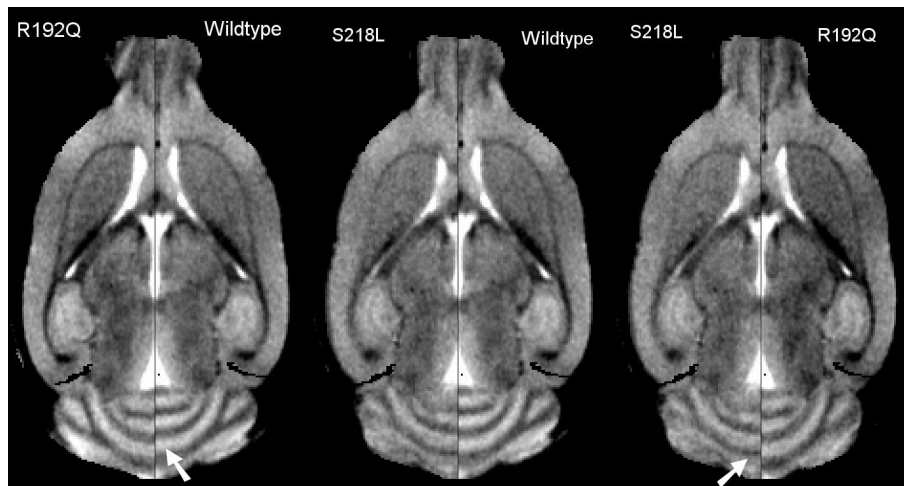


Figure 7.6: A visual comparison of the *in vivo* MR averages for each cross-sectional study, showing the caudal shift of the cerebellum indicated with a white arrow.

cerebral volume in relation to migraine does not find a significant relation between cerebral volume and patients with migraine with aura [272]. However, it is generally accepted that the migraine aura is due to CSD, a wave of transient intense spike activity that progresses slowly along the cortex and is followed by a long-lasting neuronal suppression [273–275]. In rodents, spreading depression can occur in the cortex, but also in brainstem [276] and cerebellum [277]. Earlier results showed that both the R192Q and S218L FHM1 mice exhibit an increased susceptibility to CSD [259]. CSD induces blood-brain barrier disruption and can lead to edema in rats [278]. Therefore, a possible explanation for this small cerebellar volume increase and caudal displacement is edema, which is not visible on histological sections due to tissue fixation and dehydration that precedes paraffin embedding.

Neuroimaging studies indicate that also the brainstem plays a role in migraine [279–282]. An involvement of the brainstem is further supported by the fact that: 1) lesions in the brainstem can cause migraine [283–285]; 2) electrical stimulation of the brainstem can cause headache [286, 287]; and 3) migraineurs have an increased iron deposition in the brainstem Periaqueductal gray (PAG), possibly due to a high metabolic activity in migraine [288]. In this study, deformation field analysis in R192Q mice indicates morphological changes in the brainstem, specifically in the region of the trigeminal nucleus caudalis (TNC), area postrema and nucleus tractus solitarius (NTS). Moreover, trigeminovascular activation during a migraine attack possibly triggers nausea and vomiting via activation of the nucleus tractus solitarius and area postrema. Our results support evidence for a role of the brainstem in migraine.

In MA patients with high attack frequency, sub-clinical white matter abnormalities were found particularly in regions of the cerebellum and brainstem [266, 267]. This indicates that the migraine aura may be involved in the pathogenesis of these lesions. We hypothesized that if lesions occur in FHM1 mice, lesion load would increase with age. Therefore, we analyzed aged FHM1 mice. In this study, white matter MRI abnormalities, similar to those found in migraine patients [267], were not found with *in vivo* MRI investigations. However, the frequency of spontaneous attacks in these mice is low (as in FHM1 patients), and most likely the low number of spontaneous attacks in these mice during their relatively short lifespan is insufficient to cause significant white matter damage. Also, the amount of white matter in mice is much smaller than in humans, which may render the mouse model less vulnerable to white matter damage.

MRI of migraine patients with and without aura also showed that migraineurs on average have a thicker somatosensory cortex interictally<sup>2</sup> than non-migraineurs [289]. The most significant thickness changes were noticed in the caudal somatosensory cortex, where the trigeminal area, including head and face, is somatotopically represented<sup>3</sup>. To the best of our knowledge, no cortical thickness in the somatosensory system has been measured in the FHM1 mice. In this study, some significant changes in the caudal somatosensory cortex were reported by the deformation analysis although it was also noted that some artifacts were present in the MR images. Future somatosensory thickness measurements are required to confirm this observation of somatosensory cortex changes in FHM1 mice.

In conclusion, our results show that the FHM1 mouse models are valuable models to study migraine pathophysiology. *In vivo* MRI in combination with automated deformation analysis has the potential to facilitate longitudinal analysis of neuropathology *in vivo* because live mice can be scanned on different occasions, allowing a natural history of neuropathology rather than a single, terminal, time point.

---

<sup>2</sup>Interictal: between migraine attacks

<sup>3</sup>Somatotopical arrangement: The correspondence between the position of a receptor in any part of the body and the corresponding area of the cerebral cortex that is activated by it. The size of the area in the cortex is directly correlated to the number of receptors in that body part

#### 8.1 Summary and conclusions

About 95% of the human DNA equals the DNA in mice and therefore, mice offer a perfect environment to test new treatments, develop new screening procedures or to apply experimental medications. To assist murine research, small-animal scanners such as high-resolution CT or MR scanners are of considerable importance. Especially MR scanners are considered invaluable when studying the various aspects of neuroanatomy, neuropathology and neurophysiology.

To assess group differences between mutants and their controls, brain morphometry can be exploited. As the morphological differences between the brains are subtle and large groups are required for the analysis, automation of this analysis is inevitable. Automated methods decrease the inter- and intra-observer variation, so that smaller groups of mice are required to reach similar results as manual analysis. Unfortunately, the development of automated image processing methods for high-resolutions images lags behind the development of similar techniques in human research. This problem mainly occurs because small-animal scanners were derived from human scanners and, as a result, became available later in time; furthermore, high resolution images are associated with higher noise levels.

The main objective of this thesis is to review existing automated methods for quantitative morphometry of the murine brain and to develop new, repeatable and fast automated methods for the detection of group differences using MR images.

In **chapter 1** a general introduction on transgenic mice is given with a brief history and a description of the ways transgenic mice can be generated. Furthermore the main differences and similarities between the human brain and the murine

brain are described. **Chapter 2** discusses which quantitative morphometry methods can be exploited for mouse brain MRI and classifies them in volumetry, voxel-based morphometry (VBM), shape-based morphometry (SBM) and deformation-based morphometry (DBM). In connection with chapter 2, **Chapter 3** indicates how automated morphometry in mouse studies can assist in research to human diseases, in this case Alzheimer's Disease (AD). The characteristics of the most common lines of transgenic mouse models and the various imaging techniques for AD in mice are discussed. It indicates that the bulk of research on AD has been focussed on detecting and imaging AD at earlier stages using MRI. Although an increasing number of publications show the power of combining these imaging techniques with statistical analysis, the application of automated morphometry methods in AD research in mice is still very preliminary. The latter is remarkable since in human studies automated morphometry is considered an established method.

**Chapter 4** presents a fully automated segmentation method for several brain structures in *in vivo* and *ex vivo* MRI to lower the burden of manual segmentation. The method obtains an initial segmentation of the brain structures by applying affine registration (12 parameters) to a manually delineated reference image. This initial segmentation was further refined by a clustering algorithm, which used the structure intensity and the presence of an image gradient in the neighborhood of the voxels. The volumes of the automated segmentations were compared to the manual delineation and the results of segmentation by nonlinear registration. The results show that the proposed method had an equally good performance as manual segmentation and nonlinear registration.

**Chapter 5** presents a generalization to arbitrary dimensions of the nonparametric Moore-Rayleigh test for randomness of vector data, as earlier presented by Moore for two-dimensional vector data. For the three-dimensional Moore-Rayleigh test (MR3), the asymptotic distribution in closed form and the two-sample MR3 were given to allow the application of the MR3 for DBM. It was theoretically shown that the two-sample MR3 might be potential liberal, although experiments on simulation data and on synthetic data showed that the MR3 had equal performance as permutations of the Hotelling's  $T^2$  test, a default method in human morphometry, although the MR3 is not dependent on the computationally expensive permutations. **Chapter 6** explores empirically the capabilities of the MR3 for the application of DBM and how its performance compares to the performance of permutations on the Hotelling's  $T^2$  test (pHT2). Since the MR3 can be potential liberal and alternative approach for the MR3 was presented. Experiments on simulated data and on mouse brain MRI showed that the MR3 is suitable for DBM. Furthermore, because the  $p$ -value of the MR3 is not limited by the number of relabelings, it allows multiple-test correction. **Chapter 7** shows an application of the MR3, applied on two mouse models of migraine, the R192Q knock-in mice and the S218L knock-in mouse. The MR3 was exploited to screen the brain MR images for regional differences between the groups. The detected locations were afterwards validated by volumetry on the brain structures and visual inspection of histological sections.

The regional differences between the migraine mouse models and their wild types,

as detected by the MR3, are in accordance with morphological findings in human migraine studies, although no significant group differences between the brains were detected by volumetry and no defects were found on histology.

In conclusion, quantitative and local morphometry of mouse brain MRI is a relatively new field of research, where automated methods can be exploited to rapidly provide accurate and repeatable results. In this thesis we reviewed several existing methods and applications of quantitative morphometry to brain MR images and proposed two new methods to quantitatively analyze mouse brain MRI. Both methods have been validated and proved to be reliable, accurate and repeatable, without the loss of computational time. Having said so, we have realized our goals stated in section 1.4.

## 8.2 Future work

This thesis presents automated methods for quantitative morphometry in the mouse brain by exploiting deformation fields. In this thesis it is assumed that the deformation fields code for the anatomical differences between two images. In practise, however, all nonlinear registration methods are validated on synthetically degenerated images or by eyeball validation, as other methods are lacking. On that ground, we can only conclude that these methods are capable of deforming an image so that it is similar to another one, but it still remains unclear if brain tissue truly deforms as modeled by nonlinear registration. Therefore further research on the validation of non-linear registration is necessary.

With the introduction of automated methods, the processing time of image analysis is decreased, and therefore other bottlenecks in mice research were addressed. For example, coils for multiple-mouse imaging have been made available [56] and mice centers with a capacity of 70.000 mice are being built<sup>1</sup>. However, this data-explosion might have a negative effect on the results of automated morphometry. Morphometry results are obtained by means of an image processing pipeline existing of a normalization step, a feature extraction step and possibly other image processing steps, all introducing their own sources of error. These errors might be visible in the deformation field as false positives. Therefore, it is still necessary to manually check all results on possible errors introduced during the pipeline. Earlier detection and identification of these errors during image processing will likely result in higher specificity. A possible solution for this problem is to develop computer aided methods in combination with efficient visualization methods that point the user to the images or image areas that might need manual correction. Good examples of these methods can be found in human computer aided diagnostic (CAD) approaches [290,291], e.g. the screening for lung tumors in X-ray images.

Finally, it would be recommended that for the development of future automated morphometry methods multiple MR sequences should be used possibly in combina-

---

<sup>1</sup>Personal Communication of M. Sonka during the pre-meeting of the IEEE International Symposium on Biomedical Imaging (ISBI) 2010, Rotterdam (13 April 2010).

tion with statistical analysis, instead of adding extra imaging modalities. Although currently more imaging techniques are being made available that all add extra information to the analysis, the question remains if this extra information would lead to more accurate morphometry results. Imaging of mice on several scanners introduces extra intra-subject variations, either because the mice are transported between scanners or, in case the mice remain anaesthetized and fixated during transport, because extra stress and dehydration [180]. MRI is particularly suited for anatomical imaging of soft tissue, does not use ionizing radiation and it is still maturing: Stronger magnets, better coils and other hardware techniques are being made available and will likely have a much larger impact on the quality and accuracy of morphometry, as for instance already shown in brain segmentation [162] or in early Alzheimer's detection [77]. By further exploiting the wide range of possibilities that MR imaging, offers automated quantitative morphometry will mature to a reliable and repeatable method that will be used in the daily practice.

---

## Samenvatting en aanbevelingen

---

### 9.1 Samenvatting en conclusies

Ongeveer 97.5% van het menselijk DNA is gelijk aan dat van de muizen. Daarom bieden muizen een perfecte testomgeving om nieuwe behandelingen te testen, screeningsprocedures te ontwikkelen, of om experimentele medicatie toe te passen. Om die reden zijn muizenscanners, zoals hoge resolutie CT of MR scanners van groot belang. Vooral de MR scanner wordt als zeer nuttig beschouwd als het aankomt op het bestuderen van de verschillende aspecten van de hersenanatomie, hersenpathologie en hersenfysiologie.

Om de verschillen tussen groepen van transgene muizen en de niet-transgene, controle muizen te bepalen wordt gebruik gemaakt van hersenmorfometrie. Omdat de morfologische verschillen tussen hersenen subtiel zijn en daardoor grote groepen muizen nodig zijn om de significante verschillen aan te tonen, is het onvermijdelijk dit proces te automatiseren. Dit heeft verder tot positief gevolg dat de inter- en intra-observer variatie omlaag zal gaan, waardoor de kwantitatieve analyse nauwkeuriger wordt en er minder muizen nodig zullen zijn om dezelfde resultaten te behalen als met handmatige kwantitatieve analyse. Helaas loopt de ontwikkeling van automatische beeldverwerkingsmethoden voor hoge resolutie beelden achter op de ontwikkeling van dezelfde beeldverwerkingstechnieken voor humaan onderzoek. Dit is voornamelijk doordat de scanners voor kleine dieren zijn ontwikkeld naar voorbeeld van humane scanners; hierdoor kwamen deze scanners pas veel later beschikbaar. Daar komt bij dat in een hogere beeldresolutie ook meer ruis aanwezig is.

Het hoofddoel van dit proefschrift is het in kaart brengen van bestaande automatische kwantitatieve beeldverwerkingsmethoden voor muizenhersenen en om nieuwe,

herhaalbare en snelle kwantitatieve morfometrie methoden voor de automatische detectie van groepsverschillen van muizenhersenen opgenomen met de MR scanner te ontwikkelen.

In **hoofdstuk 1** wordt een algemene introductie over transgene muizen gegeven. Er wordt kort ingegaan op de geschiedenis en de meest bekende methodes om transgene muizen te ontwikkelen en de verschillen tussen mensen en muizenhersenen worden kort geïllustreerd. **Hoofdstuk 2** behandelt de verschillende kwantitatieve morfometrie methoden voor MR beelden van muizenhersenen, welke worden opgesplitst in volumetrie, voxel-gebaseerde morfometrie (VBM), vorm-gebaseerde morfometrie (SBM) en deformatie-gebaseerde morfometrie (DBM). Aansluitend op de algemene introductie op morfometrie, geeft **hoofdstuk 3** een voorbeeld hoe automatische morfometrie bij muizenonderzoek kan helpen bij onderzoek naar humane ziektes, in dit geval de ziekte van Alzheimer (AD). De karakteristieken van de meest gebruikelijke muizenmodellen voor AD worden beschreven samen met de verschillende beeldvormingstechnieken om AD in de muizenhersenen aan te tonen. Het laat zien dat het grootste gedeelte van onderzoek is gedaan naar visualisatie van de AD in een vroeg stadium met MR scanners. Hoewel een groeiend aantal publicaties laat zien hoe krachtig de combinatie van beeldvorming en statistische analyse kan zijn, staat de ontwikkeling van automatische kwantitatieve morfometrie methoden echter nog in de kinderschoenen. Het laatste is opmerkelijk omdat automatische morfometrie voor humaan AD onderzoek al als een standaard methode wordt beschouwd.

In **Hoofdstuk 4** wordt een volledig automatische segmentatie methode gepresenteerd van verschillende hersenstructuren in *in vivo* en *ex vivo* MR beelden om de werkbelasting van de handmatige intekeningen te verlagen. Er wordt voorgesteld om een initiële segmentatie van de structuren in de hersenen te verkrijgen door een affine registratie (12 parameters) met een handmatig gesegmenteerde referentie beeld. Daarna kan deze initiële segmentatie worden verfijnd door een clustering algoritme, waarbij gebruik wordt gemaakt van voorkennis over de beeldintensiteit van de structuur en de locale beeldgradiënt. De volumes van de automatisch gesegmenteerde structuren worden vergeleken met handmatige segmentaties en met segmentaties door niet-lineaire registraties met een reeds ingetekend volume. De resultaten laten zien dat de voorgestelde methode net zo goed presteert als handmatige segmentaties en als niet-lineaire registratie.

In **Hoofdstuk 5** wordt de Moore-Rayleigh test, een niet-parametrische test voor de willekeurige verdeling van vector data, gegeneraliseerd van tweedimensionale vector data naar vector data in elke mogelijke dimensie. Verder wordt de driedimensionale Moore-Rayleigh test (MR3) in closed-form gegeven, samen met de MR3 voor twee onafhankelijke groepen, waardoor de MR3 gebruikt kan worden voor DBM. Er wordt theoretisch bewezen dat de MR3 voor twee groepen potentieel liberaal kan zijn, dat betekent dat de test een grotere kans heeft om onterecht de nulhypothese te verwerpen. Echter, op gesimuleerde data wordt aangetoond dat de test in de praktijk net zo goed presteert als een DBM methode gebaseerd op permutaties van de Hotelling's  $T^2$  test (pHT2), een standaard methode gebruikt bij beelden van het menselijk brein. In tegenstelling tot the pHT2 is de MR3 niet afhankelijk is van de permutaties, die



een lange berekeningstijd nodig hebben. In **Hoofdstuk 6** worden de mogelijkheden van de MR3 voor de toepassing voor DBM empirisch onderzocht en vergeleken met de permutaties van de Hotellings  $T^2$  test. Experimenten op gesimuleerde data en MR beelden van muizenhersen tonen aan dat de MR3 geschikt is voor DBM. De ondergrens van de  $p$ -waarde van de MR3 is, in tegenstelling tot permutatie testen, niet gelimiteerd door het aantal permutaties dat wordt uitgevoerd, waardoor deze test ook gebruikt kan worden als de  $p$ -waarden gecorrigeerd moeten worden voor het gebruik van meerdere statistische testen. **Hoofdstuk 7** laat een toepassing van de MR3 op twee muismodellen voor migraine, de R192Q en de S218L knock-in muizen, zien. De MR3 wordt gebruikt om de MR beelden van de hersenen te screenen op afwijkingen die vervolgens door volumetrie en histologie verder worden onderzocht. Hoewel de volumetrie metingen geen verschillen tussen de hersenstructuren laten zien en op de histologie ook geen afwijkingen worden gevonden, zijn de lokale significante gebieden in de hersenstructuren die door de MR3 worden gevonden wel te relateren aan de morfometrische bevindingen van studies naar migraine bij mensen.

In conclusie, kwantitatieve en regionale morfometrie in MR beelden van muizenhersen is een relatief nieuw onderzoeksgebied, waarbij automatische methodes gebruikt worden om in weinig tijd objectieve, nauwkeurige en herhaalbare resultaten te behalen. In dit proefschrift zijn verschillende bestaande morfometrie methodes vergeleken en zijn twee nieuwe methodes voor morfometrie in MR beelden van muizenhersen ontwikkeld. Beide methoden zijn gevalideerd en hebben bewezen betrouwbaar, nauwkeurig en herhaalbaar te zijn zonder dat er tijd om het algoritme uit te voeren lager is dan al bestaande algoritmen. Hiermee zijn de doelstellingen gehaald zoals die in hoofdstuk 1.4 gedefinieerd zijn.

## 9.2 Aanbevelingen

Dit proefschrift presenteert automatische methoden voor kwantitatieve morfometrie in muizenhersen die gebruik maken van deformatie velden. Hierbij wordt aangenomen dat deze deformatie velden de anatomische verschillen tussen twee beelden correct modelleren. Echter in de praktijk zijn alle bestaande niet-lineaire registratie methoden, bij gebrek aan betere methoden, slechts gevalideerd op synthetisch verkregen beelden of door middel van visuele validatie. Op basis van deze validatie kan geconcludeerd worden dat deze methoden in staat zijn om een beeld zodanig te vervormen dat het precies op het andere past, maar of deze deformaties werkelijk hetzelfde zijn als de deformaties zoals die in het hersenweefsel voorkomen blijft onduidelijk. Om deze reden is verder onderzoek naar de validatie van niet-lineaire registratie noodzakelijk.

Nu de verwerkingstijd van de beeldanalyse omlaag is gegaan met de komst van automatische beeldverwerkingsmethoden, worden andere knelpunten aangepakt. Zo worden er bijvoorbeeld MR spoelen ontwikkeld waarbij meerdere muizen tegelijkertijd kunnen worden onderzocht [56] en worden er muizencentra met een capaciteit tot

70.000 muizen opgezet<sup>1</sup>. Maar deze data-explosie kan een nadelig effect hebben op de resultaten van de automatische morfometrie. De resultaten worden namelijk berekend via een keten van beeldverwerkings methoden, namelijk normalisatie, eventueel registratie, extractie van de eigenschappen gebruikt voor classificatie, en andere beeldverwerkingstechnieken. Deze technieken introduceren allemaal hun eigen fouten. Deze fouten zijn soms niet, maar vaak ook wel zichtbaar als een fout positieve uitslag van de test. Daarom is het nog steeds noodzakelijk om handmatig alle resultaten te controleren op mogelijke fouten die geïntroduceerd zijn door de verschillende beeldverwerkings methoden. Bij een te grote dataset is er een kans dat dit achterwege wordt gelaten of onnauwkeurig wordt gedaan. Het in kaart brengen van de invloed van deze fouten en hun tijdige detectie zou de nauwkeurigheid van veel methodes ten goede komen. Een mogelijke oplossing is om computer gestuurde methodes in combinatie met efficiënte visualisatie methoden te ontwikkelen om de gebruiker op de beelden, of gebieden in de beelden, te wijzen die mogelijk een fout kunnen bevatten. In een dergelijke setting hoeft alleen nog maar een subset van beelden handmatig gecontroleerd te worden. Een goed voorbeeld van de werking van zulke methodes is de toepassing van computer gestuurde diagnostische (CAD) systemen [290, 291], bijvoorbeeld voor het screenen op long tumoren in röntgen beelden.

Een laatste aanbeveling gaat in de richting van het verder benutten van de MR sequenties en de eventuele statistische analyse daarvan, in plaats van het toevoegen van extra modaliteiten voor beeldanalyse. Hoewel met de huidige technieken steeds meer scanners beschikbaar komen die allemaal extra informatie toevoegen aan de analyse, is de vraag of dit de kwantitatieve morfometrie zou helpen. Het scannen van muizen op verschillende toestellen introduceert ofwel extra intra-subject variatie bij het verplaatsen ofwel extra stress en uitdroging voor muizen die worden verdoofd en gefixeerd om de positie te behouden [180]. MRI is zeer geschikt voor het visualiseren van zachte weefsels, het is niet schadelijk voor de gezondheid en het is nog steeds in ontwikkeling. Sterkere magneten, betere spoelen en verbeterde hardware technieken hebben waarschijnlijk een veel grotere impact op de kwaliteit en nauwkeurigheid van morfometrie, zoals aangetoond in de segmentatie van de hersenen [162] of in de vroege detectie van AD [77]. Het verder benutten van het brede toepassingsgebied van de MR scanner zou de automatische kwantitatieve morfometrie kunnen helpen zich verder te ontwikkelen tot een betrouwbare methode die in de dagelijkse routine gebruikt kan worden.

---

<sup>1</sup>Persoonlijke communicatie van M. Sonka tijdens de officiële bijeenkomst ter introductie op de IEEE International Symposium on Biomedical Imaging (ISBI) 2010, Rotterdam (13 April 2010).

---

## Bibliography

---

- [1] J.G. Fox, S.W. Barthold, M.T. Davisson, C.E. Newcomer, F.W. Quimby, and A.L. Smith. *The Mouse in Biomedical Research: Diseases*. Academic Press, 2nd edition, 2007.
- [2] M.A. Klinger. The life history of the mouse in genetics. *Nature*, 420(December):510–11, 2002.
- [3] R. Jaenisch and B. Mintz. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. In *Proceedings of the national academy of sciences*, pages 1250–54, 1974.
- [4] R.L. Brinster and R.D. Palmiter. Transgenic mice containing growth hormone fusion genes. *Philosophical transactions of the Royal Society of London. series B, biological sciences*, 307(1132):309–12, December 1984.
- [5] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.
- [6] N. Kovacevic, J.T. Henderson, E. Chan, N. Lifshitz, J. Bishop, A.C. Evans, R.M. Henkelman, and X.J. Chen. A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability. *Cerebral cortex*, 15(5):639–45, 2004.
- [7] E.S. Lein, M.J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A.F. Boe, M.S. Boguski, and Et Al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–76, 2007.
- [8] Y. Ma, P.R. Hof, S.C. Grant, S.J. Blackband, R. Bennett, L. Slatest, M.D. McGuigan, and H. Benveniste. A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Neuroscience*, 135(4):1203–15, 2005.
- [9] Y. Ma, D. Smith, P.R. Hof, B. Foerster, S. Hamilton, S.J. Blackband, M. Yu, and H. Benveniste. In vivo 3D digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Frontiers: in neuroanatomy*, 2(1):1–10, 2008.
- [10] A. MacKenzie-Graham, E. Lee, I.D. Dinov, M. Bota, D.W. Shattuck, S. Ruffins, H. Yuan, F. Konstantinidis, A. Pitiot, Y. Ding, G. Hu, R.E. Jacobs, and A.W. Toga. A

## Bibliography

---

- multimodal, multidimensional atlas of the C57BL/6J mouse brain. *Journal of anatomy*, 204(2):93–102, February 2004.
- [11] A.W. Toga and J. Mazziotta. *Brain Mapping: The Methods*. Academic Press, 2002.
- [12] A. Delaurier, N. Burton, M. Bennett, R. Baldock, D. Davidson, T.J. Mohun, and M.P. Logan. The Mouse Limb Anatomy Atlas: an interactive 3D tool for studying embryonic limb patterning. *BMC developmental biology*, 8(83):1–7, January 2008.
- [13] M. Baiker, J. Milles, J. Dijkstra, T.D. Henning, A.W. Weber, I. Que, E.L. Kaijzel, C.W.G.M. Löwik, J.H.C. Reiber, and B.P.F. Lelieveldt. Atlas-based whole-body segmentation of mice from low-contrast Micro-CT data. *Medical image analysis*, 14(6):723–37, 2010.
- [14] B. Dogdas, D. Stout, A.F. Chatziioannou, and R.M. Leahy. Digimouse: a 3D whole body mouse atlas from CT and cryosection data. *Physics in medicine and biology*, 52(3):577–87, February 2007.
- [15] M.H. Hofker and J. van Deursen. *Transgenic mouse methods and protocols*. Humana Press, 2003.
- [16] J.P. Sundberg and T. Ichiki. *Genetically engineered mice handbook*. CRC Press, 2006.
- [17] A. Nagy, M. Gertsenstein, K. Vintersten, and R. Behringer. *Production of Transgenic Mice*, chapter 7. Cold Spring Harbor Laboratory Press, 3rd edition, 2003.
- [18] M.J. Tymms and I. Kola. *Gene knockout protocols*. Humana Press, 2001.
- [19] G.S. Fisch and J. Flint. *Transgenic and knockout models of neuropsychiatric disorders*. Humana Press, 2006.
- [20] A. Lajtha, G.B. Baker, S. Dunn, and A. Holt. *Handbook of neurochemistry and molecular neurobiology: Practical neurochemistry methods*. Springer, 3rd edition, 2007.
- [21] M. Hafezparast, A. Ahmad-Annur, N.W. Wood, S.J. Tabrizi, and E.M.C. Fisher. Mouse models for neurological disease. *Lancet neurology*, 1(4):215–24, August 2002.
- [22] B. Popko. *Mouse Models in the Study of Genetic Neurological Disorders (Advances in Neurochemistry)*. Springer, 1999.
- [23] J.F. Cryan and A. Holmes. The ascent of mouse: advances in modelling human depression and anxiety. *Nature reviews. Drug discovery*, 4(9):775–90, September 2005.
- [24] I.N. Bankman. *Segmentation*, pages 71–258. USA: Academic Press, 2nd edition, 2009.
- [25] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, and R. Frackowia. Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, 2:189–210, 1995.
- [26] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73, January 1996.
- [27] J. Ashburner and K.J. Friston. Voxel-based morphometry - the methods. *Neuroimage*, 11(6 Pt 1):805–21, 2000.
- [28] S.J. Sawiak, N.I. Wood, G.B. Williams, A.J. Morton, and T.A. Carpenter. Voxel-based morphometry in the R6/2 transgenic mouse reveals differences between genotypes not seen with manual 2D morphometry. *Neurobiology of Disease*, 33(1):20–27, 2009.

- 
- [29] L. Ferrarini, W.M. Palm, H. Olofsen, R. van Der Landen, M.A. van Buchem, J.H.C. Reiber, and F. Admiraal-Behloul. Ventricular shape biomarkers for Alzheimer's disease in clinical MR images. *Magnetic resonance in medicine*, 59(2):260–7, February 2008.
- [30] A. Kelemen, G. Székely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE transactions on medical imaging*, 18(10):828–39, October 1999.
- [31] P.M. Thompson, K.M. Hayashi, G.I. De Zubicaray, A.L. Janke, S.E. Rose, J. Semple, M.S. Hong, D.H. Herman, D. Gravano, D.M. Doddrell, and A.W. Toga. Mapping cortical changes in Alzheimer's disease, brain development, and schizophrenia. *NeuroImage*, 23:S2–S18, August 2004.
- [32] J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, October 2007.
- [33] D. Rueckert, P. Aljabar, R.A. Heckemann, J.V. Hajnal, and A. Hammers. Diffeomorphic registration using B-splines. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 702–9, January 2006.
- [34] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*, 45(1 Suppl):S61–72, March 2009.
- [35] X.J. Chen, N. Kovacevic, N.J. Lobaugh, J.G. Sled, R.M. Henkelman, and J.T. Henderson. Neuroanatomical differences between mouse strains as shown by high-resolution 3D MRI. *NeuroImage*, 29:99–105, 2005.
- [36] A.E.H. Scheenstra, M. Muskulus, M. Staring, A.M.J.M. van Den Maagdenberg, S. Verduyn-Lunel, J.H.C. Reiber, L. van Der Weerd, and J. Dijkstra. The 3D Moore-Rayleigh test for the quantitative groupwise comparison of MR brain images. In *Proceedings of information processing in medical imaging*, pages 564–75, January 2009.
- [37] A. Badaea, G.A. Johnson, and J.L. Jankowsky. Remote sites of structural atrophy predict later amyloid formation in a mouse model of Alzheimer's disease. *NeuroImage*, 50(2):416–27, April 2010.
- [38] J.C. Lau, J.P. Lerch, J.G. Sled, R.M. Henkelman, A.C. Evans, and B.J. Bedell. Longitudinal neuroanatomical changes determined by deformation-based morphometry in a mouse model of Alzheimer's disease. *NeuroImage*, 42(1):19–27, August 2008.
- [39] P.M. Thompson and A.W. Toga. Computing and Visualization in Science: A framework for computational anatomy. *Computing and visualization in science*, 34:13–34, 2002.
- [40] S. Maheswaran, H. Barjat, S.T. Bate, P. Aljabar, D.L.G. Hill, L. Tilling, N. Upton, M.F. James, J.V. Hajnal, and D. Rueckert. Analysis of serial magnetic resonance images of mouse brains using image registration. *NeuroImage*, 44(3):692–700, February 2009.
- [41] F.L. Bookstein. Voxel-Based Morphometry should Not Be Used with Imperfectly Registered Images. *NeuroImage*, 14:1454–62, 2001.
- [42] C. Davatzikos. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, 23(1):17–20, 2004.

## Bibliography

---

- [43] G.R. Ridgway, S.M.D. Henley, J.D. Rohrer, R.I. Scahill, J.D. Warren, and N.C. Fox. Ten simple rules for reporting voxel-based morphometry studies. *NeuroImage*, 40(4):1429–35, May 2008.
- [44] J.P. Shaffer. Multiple Hypothesis Testing. *Annual review of psychology*, 46(1):561–84, January 1995.
- [45] D. Pantazis, T.E. Nichols, S. Baillet, and R.M. Leahy. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *NeuroImage*, 25(2):383–94, 2005.
- [46] A.P. Holmes, R.C. Blair, J.D. Watson, and I. Ford. Nonparametric analysis of statistic images from functional mapping experiments. *Journal of cerebral blood flow and metabolism*, 16(1):7–22, January 1996.
- [47] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the royal statistical society. series B (methodological)*, 57(1):289–300, 1995.
- [48] C.R. Genovese, N.A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15(4):870–8, April 2002.
- [49] J.R. Chumbley and K.J. Friston. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage*, 44(1):62–70, January 2009.
- [50] A. Schwartzman, R.F. Dougherty, J. Lee, D. Ghahremani, and Jonathan E Taylor. Empirical null and false discovery rate analysis in neuroimaging. *NeuroImage*, 44(1):71–82, January 2009.
- [51] L. Jönsson, M.E. Jönhagen, L. Kilander, H. Soininen, M. Hallikainen, G. Waldemar, H. Nygaard, N. Andreasen, B. Winblad, and A. Wimo. Determinants of costs of care for patients with Alzheimer’s disease. *International journal of geriatric psychiatry*, 21(5):449–59, 2006.
- [52] D.K. Lahiri, M.R. Farlow, N.H. Greig, and K. Sambamurti. Current drug targets for Alzheimer’s disease treatment. *Drug development research*, 56(3):267–81, July 2002.
- [53] B.P. Leifer. Early diagnosis of Alzheimer’s disease: clinical and economic benefits. *Journal of the american geriatrics society*, 51(5 Suppl Dementia):S281–8, May 2003.
- [54] R.L. Buckner, A.Z. Snyder, B.J. Shannon, G. LaRossa, R. Sachs, A.F. Fotenos, Y.I. Sheline, W.E. Klunk, C.A. Mathis, J.C. Morris, and M.A. Mintun. Molecular, structural, and functional characterization of Alzheimer’s disease: evidence for a relationship between default activity, amyloid, and memory. *The journal of Neuroscience*, 25(34):7709–17, August 2005.
- [55] S. Lehericy, M. Marjanska, L. Mesrob, M. Sarazin, and S. Kinkingnehun. Magnetic resonance imaging of Alzheimer’s disease. *European radiology*, 17(2):347–62, February 2007.
- [56] B.J. Nieman, J. Bishop, J. Dazai, N.A. Bock, J.P. Lerch, A. Feintuch, X.J. Chen, J.G. Sled, and R.M. Henkelman. MR technology for biological studies in mice. *NMR in biomedicine*, 20(3):291–303, May 2007.
- [57] A. Ramani, J.H. Jensen, and J.A. Helpert. Quantitative MR imaging in Alzheimer disease. *Radiology*, 241(1):26–44, 2006.

- 
- [58] J. Skoch, B.T. Hyman, and B.J. Bacskai. Preclinical characterization of amyloid imaging probes with multiphoton microscopy. *Journal of Alzheimer's disease*, 9(3 Suppl):401–7, January 2006.
- [59] A. Nordberg. PET imaging of amyloid in Alzheimer's disease. *The lancet neurology*, 3(9):519–27, September 2004.
- [60] L. Glodzik-Sobanska, H. Rusinek, L. Mosconi, Y. Li, J. Zhan, S. de Santi, A. Convit, K. Rich, M. Brys, and M.J. de Leon. The role of quantitative structural imaging in the early diagnosis of Alzheimer's disease. *Neuroimaging clinics of north america*, 15(4):803–26, November 2005.
- [61] S.A. Anderson and J.A. Frank. MRI of mouse models of neurological disorders. *NMR in biomedicine*, 20(3):200–15, May 2007.
- [62] H. Benveniste, G. Einstein, K.R. Kim, C. Hulette, and G.A. Johnson. Detection of neuritic plaques in Alzheimer's disease by magnetic resonance microscopy. *Proceedings of the national academy of sciences of the United States of America*, 96(24):14079–84, November 1999.
- [63] H. Benveniste, Y.U. Ma, J. Dhawan, A. Gifford, S.D. Smith, I. Feinstein, C. Du, S.C. Grant, and P.R. Hof. Anatomical and Functional Phenotyping of Mice Models of Alzheimer's Disease by MR Microscopy. *Annals of the new york academy of sciences*, 1097:12–29, 2007.
- [64] A. Borthakur, T. Gur, A.J. Wheaton, M. Corbo, J.Q. Trojanowski, V.M. Lee, and R. Reddy. In vivo measurement of plaque burden in a mouse model of Alzheimer's disease. *Journal of magnetic resonance imaging*, 24(5):1011–7, November 2006.
- [65] C.R. Jack, M. Marjanska, T.M. Wengenack, D.A. Reyes, G.L. Curran, J. Lin, G.M. Preboske, J.F. Poduslo, and M. Garwood. Magnetic resonance imaging of Alzheimer's pathology in the brains of living transgenic mice: a new tool in Alzheimer's disease research. *The neuroscientist*, 13(1):38–48, February 2007.
- [66] T.M. Wengenack, C.R. Jack, M. Garwood, and J.F. Poduslo. MR microimaging of amyloid plaques in Alzheimer's disease transgenic mice. *European journal of nuclear medicine and molecular imaging*, 35 Suppl 1:S82–8, March 2008.
- [67] J. Zhang, P. Yarowsky, M.N. Gordon, G. Di Carlo, S. Munireddy, P.C.M. van Zijl, and S. Mori. Detection of amyloid plaques in mouse models of Alzheimer's disease by magnetic resonance imaging. *Magnetic resonance in medicine*, 51(3):452–7, March 2004.
- [68] L.G. Apostolova and P.M. Thompson. Brain mapping as a tool to study neurodegeneration. *Neurotherapeutics*, 4(3):387–400, July 2007.
- [69] Y. Hirata, H. Matsuda, K. Nemoto, T. Ohnishi, K. Hirao, F. Yamashita, T. Asada, S. Iwabuchi, and H. Samejima. Voxel-based morphometry to discriminate early Alzheimer's disease from controls. *Neuroscience letters*, 382(3):269–74, July 2005.
- [70] S.J. Teipel, C. Born, M. Ewers, A.L.W. Bokde, M.F. Reiser, H. Möller, and H. Hampel. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *NeuroImage*, 38(1):13–24, October 2007.
- [71] G. Castellano, L. Bonilha, L.M. Li, and F. Cendes. Texture analysis of medical images. *Clinical radiology*, 59(12):1061–9, December 2004.

## Bibliography

---

- [72] C.E. Thomaz, F.L.S. Duran, G.F. Busatto, D.F. Gillies, and D. Rueckert. Multivariate Statistical Differences of MRI Samples of the Human Brain. *Journal of mathematical imaging and vision*, 29(2-3):95–106, December 2007.
- [73] L. Cruz, B. Urbanc, S.V. Buldyrev, R. Christie, T. Gómez-Isla, S. Havlin, M. McNamara, H.E. Stanley, and B.T. Hyman. Aggregation and disaggregation of senile plaques in Alzheimer disease. *Proceedings of the national academy of sciences of the United States of America*, 94(14):7612–6, July 1997.
- [74] B. Urbanc, L. Cruz, S.V. Buldyrev, S. Havlin, M.C. Irizarry, H.E. Stanley, and B.T. Hyman. Dynamics of Plaque Formation in Alzheimer’s Disease. *Biophysical journal*, 76(3):1330–4, March 1999.
- [75] R.I. Scahill, J.M. Schott, J.M. Stevens, M.N. Rossor, and N.C. Fox. Mapping the evolution of regional atrophy in Alzheimer’s disease: unbiased analysis of fluid-registered serial MRI. *Proceedings of the national academy of sciences of the United States of America*, 99(7):4703–7, April 2002.
- [76] P.M. Thompson, K.M. Hayashi, G. de Zubicaray, A.L. Janke, S.E. Rose, J. Semple, D. Herman, M.S. Hong, S.S. Dittmer, D.M. Doddrell, and A.W. Toga. Dynamics of Gray Matter Loss in Alzheimer’s Disease. *The journal of neuroscience*, 23(3):994–1005, 2003.
- [77] N. Braakman, J. Matysik, S.G. van Duinen, F. Verbeek, R. Schliebs, H.J.M. de Groot, and A. Alia. Longitudinal assessment of Alzheimer’s beta-amyloid plaque development in transgenic mice monitored by in vivo magnetic resonance microimaging. *Journal of magnetic resonance imaging*, 24(3):530–6, September 2006.
- [78] M.F. Falangola, V.V. Dyakin, S.P. Lee, A. Bogart, J.S. Babb, K. Duff, R. Nixon, and J.A. Helpner. Quantitative MRI reveals aging-associated T2 changes in mouse models of Alzheimer’s disease. *NMR in biomedicine*, 20(3):343–51, May 2007.
- [79] C. Duyckaerts, M. Potier, and B. Delatour. Alzheimer disease models and human neuropathology: similarities and differences. *Acta neuropathologica*, 115:5–38, 2008.
- [80] D.C. German and A.J. Eisch. Mouse models of Alzheimer’s disease: insight into treatment. *Reviews in the neurosciences*, 15(5):353–69, January 2004.
- [81] G.A. Higgins and H. Jacobsen. Transgenic mouse models of Alzheimer’s disease: phenotype and application. *Behavioural pharmacology*, 14(5-6):419–38, 2003.
- [82] E. McGowan, J. Eriksen, and M. Hutton. A decade of modeling Alzheimer’s disease in transgenic mice. *Trends in genetics*, 22(5):281–9, May 2006.
- [83] T.L. Spires and B.T. Hyman. Transgenic models of Alzheimer’s disease: learning from animals. *the journal of the American society for experimental neurotherapeutics*, 2(3):423–37, July 2005.
- [84] T. van Dooren, I. Dewachter, P. Borghgraef, and F. van Leuven. Transgenic mouse models for APP processing and Alzheimer’s disease: early and late defects. *Sub-cellular biochemistry*, 38:45–63, January 2005.
- [85] B. Allen, E. Ingram, M. Takao, M.J. Smith, R. Jakes, K. Virdee, H. Yoshida, M. Holzer, M. Craxton, P.C. Emson, C. Atzori, A. Migheli, R.A. Crowther, B. Ghetti, M.G. Spillantini, and M. Goedert. Abundant tau filaments and nonapoptotic neurodegeneration in transgenic mice expressing human P301S tau protein. *The journal of neuroscience*, 22(21):9340–51, November 2002.



- 
- [86] C. Andorfer, Y. Kress, M. Espinoza, R. de Silva, K.L. Tucker, Y. Barde, K. Duff, and P. Davies. Hyperphosphorylation and aggregation of tau in mice expressing normal human tau isoforms. *Journal of neurochemistry*, 86(3):582–90, August 2003.
- [87] D.R. Borchelt, T. Ratovitski, J. van Lare, M.K. Lee, V. Gonzales, N.A. Jenkins, N.G. Copeland, D.L. Price, and S.S. Sisodia. Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. *Neuron*, 19(4):939–45, October 1997.
- [88] M.E. Calhoun, K.H. Wiederhold, D. Abramowski, A.L. Phinney, A. Probst, C. Sturchler-Pierrat, M. Staufenbiel, B. Sommer, and M. Jucker. Neuron loss in APP transgenic mice. *Nature*, 395(6704):755–6, October 1998.
- [89] M.E. Calhoun, P. Burgermeister, A.L. Phinney, M. Stalder, M. Tolnay, K.H. Wiederhold, D. Abramowski, C. Sturchler-Pierrat, B. Sommer, M. Staufenbiel, and M. Jucker. Neuronal overexpression of mutant amyloid precursor protein results in prominent deposition of cerebrovascular amyloid. *Proceedings of the national academy of sciences of the United States of America*, 96(24):14088–93, November 1999.
- [90] K. Duff, C. Eckman, C. Zehr, X. Yu, C.M. Prada, J. Perez-tur, M. Hutton, L. Buee, Y. Harigaya, D. Yager, D. Morgan, M.N. Gordon, L. Holcomb, L. Refolo, B. Zenk, J. Hardy, and S. Younkin. Increased amyloid-beta42(43) in brains of mice expressing mutant presenilin 1. *Nature*, 383(6602):710–3, October 1996.
- [91] D. Games, D. Adams, R. Alessandrini, R. Barbour, P. Berthelette, C. Blackwell, T. Carr, J. Clemens, T. Donaldson, and F. Gillespie. Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. *Nature*, 373(6514):523–7, February 1995.
- [92] M.C. Herzog, D.T. Winkler, P. Burgermeister, M. Pfeifer, E. Kohler, S.D. Schmidt, S. Danner, D. Abramowski, C. Stürchler-Pierrat, K. Bürki, S.G. van Duinen, M.L.C. Maat-Schieman, M. Staufenbiel, P.M. Mathews, and M. Jucker. Abeta is targeted to the vasculature in a mouse model of hereditary cerebral hemorrhage with amyloidosis. *Nature neuroscience*, 7(9):954–60, September 2004.
- [93] L. Holcomb, M.N. Gordon, E. McGowan, X. Yu, S. Benkovic, P. Jantzen, K. Wright, I. Saad, R. Mueller, D. Morgan, S. Sanders, C. Zehr, K. O’Campo, J. Hardy, C.M. Prada, C. Eckman, S. Younkin, K. Hsiao, and K. Duff. Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. *Nature medicine*, 4(1):97–100, January 1998.
- [94] K. Hsiao, P. Chapman, S. Nilsen, C. Eckman, Y. Harigaya, S. Younkin, F. Yang, and G. Cole. Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. *Science*, 274(5284):99–102, October 1996.
- [95] C. Janus, J. Pearson, J. McLaurin, P.M. Mathews, Y. Jiang, S.D. Schmidt, M.A. Chishti, P. Horne, D. Heslin, J. French, H.T. Mount, R.A. Nixon, M. Mercken, C. Bergeron, P.E. Fraser, P. St George-Hyslop, and D. Westaway. A beta peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer’s disease. *Nature*, 408(6815):979–82, 2000.
- [96] E. McGowan, F. Pickford, J. Kim, L. Onstead, J. Eriksen, C. Yu, L. Skipper, M.P. Murphy, J. Beard, P. Das, K. Jansen, M. Delucia, W. Lin, G. Dolios, R. Wang, C.B. Eckman, D.W. Dickson, M. Hutton, J. Hardy, and T. Golde. Abeta42 is essential for parenchymal and vascular amyloid deposition in mice. *Neuron*, 47(2):191–9, July 2005.

## Bibliography

---

- [97] D. Moechars, I. Dewachter, K. Lorent, D. Reversé, V. Baekelandt, A. Naidu, I. Tesseur, K. Spittaels, C.V. Haute, F. Checler, E. Godaux, B. Cordell, and F. Van Leuven. Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. *The journal of biological chemistry*, 274(10):6483–92, March 1999.
- [98] S. Oddo, A. Caccamo, J.D. Shepherd, M.P. Murphy, T.E. Golde, R. Kaye, R. Metherate, M.P. Mattson, Y. Akbari, and F.M. LaFerla. Triple-transgenic model of Alzheimer’s disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. *Neuron*, 39(3):409–21, July 2003.
- [99] S. Oddo, A. Caccamo, M. Kitazawa, B.P. Tseng, and F.M. LaFerla. Amyloid deposition precedes tangle formation in a triple transgenic model of Alzheimer’s disease. *Neurobiology of aging*, 24(8):1063–70, December 2003.
- [100] R. Postina, A. Schroeder, I. Dewachter, J. Bohl, U. Schmitt, E. Kojro, C. Prinzen, K. Endres, C. Hiemke, M. Blessing, P. Flamez, A. Dequenue, E. Godaux, F. van Leuven, and F. Fahrenholz. A disintegrin-metalloproteinase prevents amyloid plaque formation and hippocampal defects in an Alzheimer disease mouse model. *The journal of clinical investigation*, 113(10):1456–64, May 2004.
- [101] M. Ramsden, L. Kotilinek, C. Forster, J. Paulson, E. McGowan, K. SantaCruz, A. Guimaraes, M. Yue, J. Lewis, G. Carlson, M. Hutton, and K.H. Ashe. Age-dependent neurofibrillary tangle formation, neuron loss, and memory impairment in a mouse model of human tauopathy (P301L). *The journal of neuroscience*, 25(46):10637–47, November 2005.
- [102] K. Santacruz, J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingelsson, A. Guimaraes, M. DeTure, M. Ramsden, E. McGowan, C. Forster, M. Yue, J. Orne, C. Janus, A. Mariash, M. Kuskowski, B. Hyman, M. Hutton, and K.H. Ashe. Tau suppression in a neurodegenerative mouse model improves memory function. *Science*, 309(5733):476–81, July 2005.
- [103] C. Sturchler-Pierrat, D. Abramowski, M. Duke, K.H. Wiederhold, C. Mistl, S. Rothacher, B. Ledermann, K. Bürki, P. Frey, P.A. Paganetti, C. Waridel, M.E. Calhoun, M. Jucker, A. Probst, M. Staufenbiel, and B. Sommer. Two amyloid precursor protein transgenic mouse models with Alzheimer’s disease-like pathology. *Proceedings of the national academy of sciences of the United States of America*, 94(24):13287–92, November 1997.
- [104] K. Tanemura, T. Akagi, M. Murayama, N. Kikuchi, O. Murayama, T. Hashikawa, Y. Yoshiike, J.M. Park, K. Matsuda, S. Nakao, X. Sun, S. Sato, H. Yamaguchi, and A. Takashima. Formation of filamentous tau aggregations in transgenic mice expressing V337M human tau. *Neurobiology of disease*, 8(6):1036–45, December 2001.
- [105] Y. Tatebayashi, T. Miyasaka, D. Chui, T. Akagi, K. Mishima, K. Iwasaki, M. Fujiwara, K. Tanemura, M. Murayama, K. Ishiguro, E. Planel, S. Sato, T. Hashikawa, and A. Takashima. Tau filament formation and associative memory deficit in aged mice expressing mutant (R406W) human tau. *Proceedings of the national academy of sciences of the United States of America*, 99(21):13896–901, October 2002.
- [106] B. Zhang, M. Higuchi, Y. Yoshiyama, T. Ishihara, M.S. Forman, D. Martinez, S. Joyce, J.Q. Trojanowski, and V.M. Lee. Retarded axonal transport of R406W mutant tau in transgenic mice with a neurodegenerative tauopathy. *The journal of Neuroscience*, 24(19):4657–67, May 2004.

- [107] Y.Z. Wadghiri, E.M. Sigurdsson, M. Sadowski, J.I. Elliott, Y. Li, H. Scholtzova, C.Y. Tang, G. Aguinaldo, M. Pappolla, K. Duff, T. Wisniewski, and D.H. Turnbull. Detection of Alzheimer's amyloid in transgenic mice using magnetic resonance microimaging. *Magnetic resonance in medicine*, 50(2):293–302, August 2003.
- [108] M. Dhenain, B. Delatour, C. Walczak, and A. Volk. Passive staining: a novel ex vivo MRI protocol to detect amyloid deposits in mouse models of Alzheimer's disease. *Magnetic resonance in medicine*, 55(3):687–93, March 2006.
- [109] M.F. Falangola, B.A. Ardekani, S. Lee, J.S. Babb, A. Bogart, V.V. Dyakin, R. Nixon, K. Duff, and J.A. Helpert. Application of a non-linear image registration algorithm to quantitative analysis of T2 relaxation time in transgenic mouse models of AD pathology. *Journal of neuroscience methods*, 144:91–97, 2005.
- [110] J.A. Helpert, S. Lee, M.F. Falangola, V.V. Dyakin, A. Bogart, B. Ardekani, K. Duff, C. Branch, T. Wisniewski, M.J. de Leon, O. Wolf, J. O'Shea, and R.A. Nixon. MRI assessment of neuropathology in a transgenic mouse model of Alzheimer's disease. *Magnetic resonance in medicine*, 51(4):794–8, April 2004.
- [111] C.R. Jack, M. Garwood, T.M. Wengenack, B. Borowski, G.L. Curran, J. Lin, G. Adriany, O.H.J. Gröhn, R. Grimm, and J.F. Poduslo. In vivo visualization of Alzheimer's amyloid plaques by magnetic resonance imaging in transgenic mice without a contrast agent. *Magnetic resonance in medicine*, 52(6):1263–71, December 2004.
- [112] C.R. Jack, T.M. Wengenack, D.A. Reyes, M. Garwood, G.L. Curran, B.J. Borowski, J. Lin, G.M. Preboske, S.S. Holasek, G. Adriany, and J.F. Poduslo. In vivo magnetic resonance microimaging of individual amyloid plaques in Alzheimer's transgenic mice. *The journal of Neuroscience*, 25(43):10041–8, October 2005.
- [113] S. Lee, M.F. Falangola, R.A. Nixon, K. Duff, and J.A. Helpert. Visualization of beta-amyloid plaques in a transgenic mouse model of Alzheimer's disease using MR microscopy without contrast reagents. *Magnetic resonance in medicine*, 52(3):538–44, September 2004.
- [114] M. Higuchi, N. Iwata, Y. Matsuba, K. Sato, K. Sasamoto, and T.C. Saido. 19F and 1H MRI detection of amyloid beta plaques in vivo. *Nature neuroscience*, 8(4):527–33, April 2005.
- [115] G. Vanhoutte, I. Dewachter, P. Borghgraef, F. Van Leuven, and A. Van Der Linden. Noninvasive in vivo MRI detection of neuritic plaques associated with iron in APP[V717I] transgenic mice, a model for Alzheimer's disease. *Magnetic resonance in medicine*, 53(3):607–13, March 2005.
- [116] J. Lewis, D.W. Dickson, W.L. Lin, L. Chisholm, A. Corral, G. Jones, S.H. Yen, N. Sahara, L. Skipper, D. Yager, C. Eckman, J. Hardy, M. Hutton, and E. McGowan. Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP. *Science*, 293(5534):1487–91, August 2001.
- [117] N. El Tannir El Tayara, B. Delatour, C. Le Cudennec, M. Guégan, A. Volk, and M. Dhenain. Age-related evolution of amyloid burden, iron load, and MR relaxation times in a transgenic mouse model of Alzheimer's disease. *Neurobiology of disease*, 22(1):199–208, April 2006.
- [118] N. El Tannir El Tayara, A. Volk, M. Dhenain, and B. Delatour. Transverse relaxation time reflects brain amyloidosis in young APP/PS1 transgenic mice. *Magnetic resonance in medicine*, 58(1):179–84, July 2007.

## Bibliography

---

- [119] T. Lumley, P. Diehr, S. Emerson, and L. Chen. The importance of the normality assumption in large public health data sets. *Annual review of public health*, 23:151–69, January 2002.
- [120] P.W. Mielke Jr. and K.J. Berry. *Permutation Methods: A Distance Function Approach (Springer Series in Statistics)*. Springer-Verlag, 2007.
- [121] P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman and Hall/CRC, 2 edition, 1989.
- [122] C.E. McCulloch, S.R. Searle, and J.M. Neuhaus. *Generalized, Linear, and Mixed Models (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2000.
- [123] J.A. Hardy and G.A. Higgins. Alzheimer’s disease: the amyloid cascade hypothesis. *Science*, 256(5054):184–5, April 1992.
- [124] J. Hardy. Has the amyloid cascade hypothesis for Alzheimer’s disease been proved? *Current Alzheimer research*, 3(1):71–3, February 2006.
- [125] M. Meyer-Luehmann, T.L. Spires-Jones, C. Prada, M. Garcia-Alloza, A. de Calignon, A. Rozkalne, J. Koenigsknecht-Talboo, D.M. Holtzman, B.J. Bacskai, and B.T. Hyman. Rapid appearance and local toxicity of amyloid-beta plaques in a mouse model of Alzheimer’s disease. *Nature*, 451(7179):720–4, March 2008.
- [126] B. Sommer. Alzheimer’s disease and the amyloid cascade hypothesis: ten years on. *Current opinion in pharmacology*, 2(1):87–92, February 2002.
- [127] M. Dhenain, N. Privat, C. Duyckaerts, and R.E. Jacobs. Senile plaques do not induce susceptibility effects in T2\*-weighted MR microscopic images. *NMR in biomedicine*, 15(3):197–203, May 2002.
- [128] C. Faber, B. Zahneisen, F. Tippmann, and A. Schroeder. Gradient-Echo and CRAZED Imaging for Minute Detection of Alzheimer Plaques in an APPV717I x ADAM10-dn Mouse Model. *Magnetic resonance in medicine*, 703:696–703, 2007.
- [129] E.M. Sigurdsson, Y.Z. Wadghiri, L. Mosconi, J.A. Blind, E. Knudsen, A. Asuni, H. Scholtzova, W.H. Tsui, Y. Li, M. Sadowski, D.H. Turnbull, M.J. De Leon, and T. Wisniewski. A non-toxic ligand for voxel-based MRI analysis of plaques in AD transgenic mice. *Neurobiology of aging*, 29:836–47, 2008.
- [130] T.M. Wengenack, G.L. Curran, and J.F. Poduslo. Targeting alzheimer amyloid plaques in vivo. *Nature biotechnology*, 18(8):868–72, August 2000.
- [131] T.M. Wengenack, S. Whelan, G.L. Curran, K.E. Duff, and J.F. Poduslo. Quantitative histological analysis of amyloid deposition in Alzheimer’s double transgenic mouse brain. *Neuroscience*, 101(4):939–44, January 2000.
- [132] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer, New York, 2nd edition, 2008.
- [133] B.T. Hyman, H.L. West, G.W. Rebeck, S.V. Buldyrev, R.N. Mantegna, M. Ukleja, S. Havlin, and H.E. Stanley. Quantitative analysis of senile plaques in Alzheimer disease: observation of log-normal size distribution and molecular epidemiology of differences associated with apolipoprotein E genotype and trisomy 21 (Down syndrome). *Proceedings of the national academy of sciences of the United States of America*, 92(8):3586–90, April 1995.

- 
- [134] L. Edelstein-Keshet and A. Spiros. Exploring the formation of Alzheimer's disease senile plaques in silico. *Journal of theoretical biology*, 216(3):301–26, June 2002.
- [135] M. Luca, A. Chavez-Ross, L. Edelstein-Keshet, and A. Mogilner. Chemotactic Signaling, Microglia, and Alzheimer's Disease Senile Plaques: Is There a Connection? *Bulletin of mathematical biology*, 65(4):693–730, July 2003.
- [136] S. Itagaki, P.L. McGeer, H. Akiyama, S. Zhu, and D. Selkoe. Relationship of microglia and astrocytes to amyloid deposits of Alzheimer's disease. *Journal of neuroimmunology*, 24(3):173–82, October 1989.
- [137] D.W. Dickson. Microglia in Alzheimer's disease and transgenic models. How close the fit? *The American journal of pathology*, 154(6):1627–31, June 1999.
- [138] S.E. Hickman, E.K. Allison, and J. El Khoury. Microglial Dysfunction and Defective Beta-Amyloid Clearance Pathways in Aging Alzheimers Disease Mice. *Journal of neuroscience*, 28(33):8354–60, 2008.
- [139] K. Shoghi-Jadid, J.R. Barrio, V. Kepe, and S. Huang. Exploring a mathematical model for the kinetics of beta-amyloid molecular imaging probes through a critical analysis of plaque pathology. *Molecular imaging and biology*, 8(3):151–62, 2006.
- [140] U. Grenander and M.I. Miller. *Pattern theory: from representation to inference*. Oxford University Press, New York, 2007.
- [141] D. Cornforth and H. Jelinek. Automated classification reveals morphological factors associated with dementia. *Applied soft computing*, 8(1):182–90, 2008.
- [142] M. Dhenain, N. El Tannir El Tayara, T. Wu, M. Guegan, A. Volk, C. Quintana, and B. Delatour. Characterization of in vivo MRI detectable thalamic amyloid plaques from APP / PS1 mice. *Neurobiology of aging*, 30(1):41–53, 2009.
- [143] J.C. de la Torre and T. Mussivand. Can disturbed brain microcirculation cause Alzheimer's disease? *Neurological research*, 15(3):146–53, June 1993.
- [144] A. Komatani, K. Yamaguchi, Y. Sugai, T. Takanashi, M. Kera, M. Shinohara, and S. Kawakatsu. Assessment of demented patients by dynamic SPECT of inhaled xenon-133. *Journal of nuclear medicine*, 29(10):1621–6, October 1988.
- [145] T. Ohnishi, H. Hoshi, S. Nagamachi, S. Jinnouchi, L.G. Flores, S. Futami, and K. Watanabe. High-resolution SPECT to assess hippocampal perfusion in neuropsychiatric diseases. *Journal of nuclear medicine*, 36(7):1163–9, July 1995.
- [146] C.P. Chao, A.L. Kotsenas, and D.F. Broderick. Cerebral amyloid angiopathy: CT and MR imaging findings. *Radiographics*, 26(5):1517–31, 2006.
- [147] E. Kouznetsova, M. Klingner, D. Sorger, O. Sabri, U. Grossmann, J. Steinbach, M. Scheunemann, and R. Schliebs. Developmental and amyloid plaque-related changes in cerebral cortical capillaries in transgenic Tg2576 Alzheimer mice. *International journal of developmental neuroscience*, 24(2-3):187–93, 2006.
- [148] M. Marjanska, G.L. Curran, T.M. Wengenack, P. Henry, R.L. Bliss, J.F. Poduslo, C.R. Jack, K. Ugurbil, and M. Garwood. Monitoring disease progression in transgenic mouse models of Alzheimer's disease with proton magnetic resonance spectroscopy. *Proceedings of the national academy of sciences of the United States of America*, 102(33):11906–10, August 2005.

## Bibliography

---

- [149] S.B. Domnitz, E.M. Robbins, A.W. Hoang, M. Garcia-Alloza, B.T. Hyman, G.W. Rebeck, S.M. Greenberg, B.J. Bacskai, and M.P. Frosch. Progression of cerebral amyloid angiopathy in transgenic mouse models of Alzheimer disease. *Journal of neuropathology and experimental neurology*, 64(7):588–94, July 2005.
- [150] E.M. Robbins, R.A. Betensky, S.B. Domnitz, S.M. Purcell, M. Garcia-Alloza, C. Greenberg, G.W. Rebeck, B.T. Hyman, S.M. Greenberg, M.P. Frosch, and B.J. Bacskai. Kinetics of cerebral amyloid angiopathy progression in a transgenic mouse model of Alzheimer disease. *The journal of Neuroscience*, 26(2):365–71, January 2006.
- [151] M. Garcia-Alloza, E.M. Robbins, S.X. Zhang-Nunes, S.M. Purcell, R.A. Betensky, S. Raju, C. Prada, S.M. Greenberg, B.J. Bacskai, and M.P. Frosch. Characterization of amyloid deposition in the APP<sup>swe</sup>/PS1<sup>dE9</sup> mouse model of Alzheimer disease. *Neurobiology of disease*, 24(3):516–24, December 2006.
- [152] N. Beckmann, A. Schuler, T. Mueggler, E.P. Meyer, K. Wiederhold, M. Staufenbiel, and T. Krucker. Age-dependent cerebrovascular abnormalities and blood flow disturbances in APP23 mice modeling Alzheimer’s disease. *The journal of neuroscience*, 23(24):8453–9, September 2003.
- [153] T. Krucker, A. Schuler, E.P. Meyer, M. Staufenbiel, and N. Beckmann. Magnetic resonance angiography and vascular corrosion casting as tools in biomedical research: application to transgenic mice modeling Alzheimer’s disease. *Neurological research*, 26(5):507–16, July 2004.
- [154] D.R. Thal, W.S.T. Griffin, R.A.I. de Vos, and E. Ghebremedhin. Cerebral amyloid angiopathy and its relationship to Alzheimer’s disease. *Acta neuropathologica*, 115(6):599–609, 2008.
- [155] L.G. Apostolova and P.M. Thompson. Mapping progressive brain structural changes in early Alzheimer’s disease and mild cognitive impairment. *Neuropsychologia*, 46(6):1597–612, January 2008.
- [156] C.R. Jack, M.A. Bernstein, N.C. Fox, P. Thompson, G.E. Alexander, D. Harvey, B. Borowski, P.J. Britson, J. L Whitwell, C. Ward, A.M. Dale, J.P. Felmlee, J.L. Gunter, D.L.G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C.S. DeCarli, G. Krueger, H.A. Ward, G.J. Metzger, K.T. Scott, R. Mallozzi, D. Blezek, J. Levy, J.P. Debbins, A.S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M.W. Weiner. The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of magnetic resonance imaging*, 27(4):685–91, April 2008.
- [157] W.J. Jagust, L. Zheng, D.J. Harvey, W.J. Mack, H.V. Vinters, M.W. Weiner, W.G. Ellis, C. Zarow, D. Mungas, B.R. Reed, J.H. Kramer, N. Schuff, C. DeCarli, and H.C. Chui. Neuropathological basis of magnetic resonance images in aging and dementia. *Annals of neurology*, 63(1):72–80, January 2008.
- [158] R.J. Killiany, T. Gomez-Isla, M. Moss, R. Kikinis, T. Sandor, F. Jolesz, R. Tanzi, K. Jones, B.T. Hyman, and M.S. Albert. Use of structural magnetic resonance imaging to predict who will get Alzheimer’s disease. *Annals of neurology*, 47(4):430–9, April 2000.
- [159] M.L. Ries, C.M. Carlsson, H.A. Rowley, M.A. Sager, C.E. Gleason, S. Asthana, and S.C. Johnson. Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. *Journal of the american geriatrics society*, 56(5):920–34, May 2008.

- 
- [160] J. Ashburner and K.J. Friston. Why Voxel-Based Morphometry Should Be Used. *NeuroImage*, 14:1238–43, 2001.
- [161] C. Zhang, E. McNeil, L. Dressler, and R. Siman. Long-lasting impairment in hippocampal neurogenesis associated with amyloid deposition in a knock-in mouse model of familial Alzheimer’s disease. *Experimental neurology*, 204(1):77–87, March 2007.
- [162] A.A. Ali, A.M. Dale, A. Badea, and G.A. Johnson. Automated segmentation of neuroanatomical structures in multispectral MR microscopy of the mouse brain. *NeuroImage*, 27:425–35, 2005.
- [163] P.M. Thompson and A.W. Toga. Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. *Medical image analysis*, 1(4):271–94, September 1997.
- [164] R.P. Woods, S.T. Grafton, J.D. Watson, N.L. Sicotte, and J.C. Mazziotta. Automated image registration: II. Intersubject validation of linear and nonlinear models. *Journal of computer assisted tomography*, 22(1):153–65, 1998.
- [165] N.A. Bock, N. Kovacevic, T.V. Lipina, J.C. Roder, S.L. Ackerman, and R.M. Henkelman. In Vivo Magnetic Resonance Imaging and Semiautomated Image Analysis Extend the Brain Phenotype for cdf / cdf Mice. *The journal of Neuroscience*, 26(17):4455–59, 2006.
- [166] R. Verma, S. Mori, D. Shen, P. Yarowsky, J. Zhang, and C. Davatzikos. Spatiotemporal maturation patterns of murine brain quantified by diffusion tensor MRI and deformation-based morphometry. *Proceedings of the national academy of sciences of the United States of America*, 102(19):6978–83, 2005.
- [167] S. Maheswaran, D.L.G. Hill, M.F. James, L. Tilling, N. Upton, J. Hajnal, and D. Rueckert. Deformation based morphometry analysis of serial magnetic resonance images of mouse brains. In *Biomedical image registration*, pages 58–65. Springer-Verlag Berlin Heidelberg, 2006.
- [168] I.N. Bankman, T. Spisz, and S. Pavlopoulos. *Two-dimensional shape and texture analysis*, page 984. Academic Press, 2nd edition, 2009.
- [169] M. Hajek, M. Dezortova, A. Materka, and R. Lerski. *Texture analysis for magnetic resonance imaging*. Med4 publishing, 2006.
- [170] C.M. Haralick. Statistical and structural approach to textures. In *Proceedings of the IEEE*, pages 786–804, 1979.
- [171] T. Tasdizen, S.P. Awate, R.T. Whitaker, and N.L. Foster. MRI tissue classification with neighborhood statistics: a nonparametric, entropy-minimizing approach. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 517–25, January 2005.
- [172] O. Yu, J. Steibel, Y. Mauss, B. Guignard, B. Eclancher, J. Chambron, and D. Grucker. Remyelination assessment by MRI texture analysis in a cuprizone mouse model. *Magnetic resonance imaging*, 22(8):1139–44, October 2004.
- [173] P.A. Freeborough and N.C. Fox. MR image texture analysis applied to the diagnosis and tracking of Alzheimer’s disease. *IEEE transactions on medical imaging*, 17(3):475–9, June 1998.

- [174] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C.S. Carter, V.A. Stenger, S. Davis, H. Aizenstein, J.T. Becker, O.L. Lopez, and C.C. Meltzer. Discriminative MR Image Feature Analysis for Automatic Schizophrenia and Alzheimer's Disease Classification. In *International conference on medical image computing and computer-assisted intervention*, pages 393–401. Springer Berlin / Heidelberg, 2004.
- [175] V.A. Kovalev, M. Petrou, and Y.S. Bondar. Texture anisotropy in 3-D images. *IEEE transactions on image processing*, 8(3):346–60, 1999.
- [176] C.C. Reyes-Aldasoro and A. Bhalerao. Volumetric texture description and discriminant feature selection for MRI. In *Proceedings of IEEE information processing in medical imaging*, volume 18, pages 282–93, July 2003.
- [177] V. Megalooikonomou, J. Zhang, D. Kontos, and P.R. Bakic. Analysis of texture patterns in medical images with an application to breast imaging. In M.L. Giger and N.L. Karssemeijer, editors, *Proceedings of SPIE medical imaging*, volume 6514, page 651421, 2007.
- [178] G.W. Arendash and D.L. King. Intra- and intertask relationships in a behavioral test battery given to Tg2576 transgenic mice and controls. *Physiology & behavior*, 75(5):643–52, 2002.
- [179] E. Dere, J.P. Huston, and M.A. De Souza-Silva. Episodic-like memory in mice: simultaneous assessment of object, place and temporal order memory. *Brain research protocols*, 16(1-3):10–9, December 2005.
- [180] I.J. Hildebrandt, H. Su, and W.A. Weber. Anesthesia and other considerations for in vivo imaging of small animals. *ILAR journal*, 49(1):17–26, January 2008.
- [181] D.H. Turnbull and S. Mori. MRI in mouse developmental biology. *NMR in biomedicine*, 20(3):265–74, May 2007.
- [182] A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath. Brain functional localization: A survey of image registration techniques. *IEEE transactions on medical imaging*, 26(4):427–51, 2007.
- [183] J. Tohka, E. Krestyannikov, I.D. Dinov, A.M. Graham, D.W. Shattuck, U. Ruotsalainen, and A.W. Toga. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE transactions on medical imaging*, 26(5):696–711, 2007.
- [184] T. Rohlfing, D.B. Russakoff, and C.R. Maurer. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE transactions on medical imaging*, 23(8):983–94, August 2004.
- [185] G.A. Johnson, A.A. Ali-Sharief, A. Badea, J. Brandenburg, G. Cofer, B. Fubara, S. Gewalt, L.W. Hedlund, and L. Upchurch. High-throughput morphologic phenotyping of the mouse brain with magnetic resonance histology. *NeuroImage*, 37(1):82–89, 2007.
- [186] A.A. Sharief, A. Badea, A.M. Dale, and G.A. Johnson. Automated segmentation of the actively stained mouse brain using multi-spectral MR microscopy. *NeuroImage*, 39(1):136–45, 2008.
- [187] R.O. Duda, E.H. Peter, and D.G. Stork. *Pattern Classification*. John Wiley and Sons, INC, 2nd edition, 2000.



- 
- [188] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Journal of applied statistics*, 20(5):25–62, 1993.
- [189] J.P. Thirion. Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical image analysis*, 2(3):243–60, September 1998.
- [190] D.L.G. Hill, P.G. Batchelor, M. Holden, and D.J. Hawkes. Medical image registration. *Physics in medicine and biology*, 46:R1–R45, 2001.
- [191] J.B.A. Maintz and M.A. Viergever. A Survey of Medical Image Registration. *Medical image analysis*, 2(1):1–37, 1998.
- [192] T.S. Yoo. *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*. AK Peters Ltd, 2004, 2004.
- [193] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi. Multi-modal image set registration and atlas formation. *Medical image analysis*, 10(3):440–51, June 2006.
- [194] M.K. Chung, K.J. Worsley, T. Paus, C. Cherif, and D.L. Collins. A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14(3):595–606, 2001.
- [195] J. Cao and K.J. Worsley. The detection of local shape changes via the geometry of Hotelling’s T2 fields. *Annals of statistics*, 27:925–42, 1999.
- [196] T.E. Nichols and A.P. Holmes. *Non-parametric procedures*, chapter 21, pages 253–72. Academic Press, 2006.
- [197] A.C. Davison and D.V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.
- [198] E.L. Lehman. *Testing Statistical Hypothesis*. John Wiley & Sons, 2nd edition, 1986.
- [199] J.R. Chumbley and K.J. Friston. False discovery rate revisited : FDR and topological inference using Gaussian random fields. *NeuroImage*, 44(1):62–70, 2009.
- [200] B.R. Moore. A Modification of the Rayleigh Test for Vector Data. *Biometrika*, 67(1):175–80, 1980.
- [201] J. Dutka. On the problem of random flights. *Archive for history of exact sciences*, 32(3):351–375, 1985.
- [202] Y. Kajikawa and T.A. Hackett. Entropy analysis of neuronal spike train synchrony. *Journal of neuroscience methods*, 149:90–93, 2005.
- [203] J.J. Tukker, P. Fuentealba, K. Hartwich, P. Somogyi, and T. Klausberger. Cell type-specific tuning of hippocampal interneuron firing during gamma oscillations in vivo. *The journal of neuroscience*, 27(31):8184–9, August 2007.
- [204] A.G. Richardson, G. Lassi-Tucci, C. Padoa-Schioppa, and E. Bizzi. Neuronal activity in the cingulate motor areas during adaptation to a new dynamic environment. *Journal of neurophysiology*, 99(3):1253, 2008.
- [205] R.J. van Beers, P. Haggard, and D.M. Wolpert. The role of execution noise in movement variability. *Journal of neurophysiology*, 91(2):1050–63, February 2004.
- [206] K.P. Able and M.A. Able. Development of sunset orientation in a migratory bird: no calibration by the magnetic field. *Animal behaviour*, 53(2):363–368, February 1997.
- [207] M.K. McNaught and I.P.F. Owens. Interspecific variation in plumage colour among birds: species recognition or light environment? *Journal of evolutionary biology*, 15(4):505–14, June 2002.

## Bibliography

---

- [208] N.H.K. Burton. Nest orientation and hatching success in the tree pipit *Anthus trivialis*. *Journal of avian biology*, 37(4):312–17, July 2006.
- [209] N. Chernetsov, W. Chromik, P.T. Dolata, P. Profus, and P. Tryjanowsky. Sex-related natal dispersal of White Storks (*Ciconia ciconia*) in Poland: How far and where to? *The auk*, 123(4):1103–09, 2006.
- [210] K.V. Mardia and P.E. Jupp. *Directional statistics*. John Wiley & Sons, 2000.
- [211] J.L. Gastwirth. Percentile Modifications of Two Sample Rank Tests. *Journal of the American statistical association*, 60(312):1127–41, 1965.
- [212] M.A. Stephens. Exact and approximate tests for directions. I. *Biometrika*, 49(3-4):463–77, 1962.
- [213] R.D. Lord. The Use of the Hankel Transform in Statistics I. General Theory and Examples. *Biometrika*, 41(1):44–55, 1954.
- [214] F. Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics bulletin*, 1(6):80–3, 1945.
- [215] D. Borwein and J.M. Borwein. Some Remarkable Properties of Sinc and Related Integrals. *The ramanujan journal*, 5(1):73–89, March 2001.
- [216] S. Dvorak. Treloar’s distribution and its numerical implementation. *Journal of physics A: general physics*, 5(1):78, 1972.
- [217] H. Hotelling. The Generalization of Student’s Ratio. *The annals of mathematical statistics*, 2(3):360–78, 1931.
- [218] C. Diks. A test for symmetries of multivariate probability distributions. *Biometrika*, 86(3):605–14, September 1999.
- [219] J.W. Pratt. Robustness of Some Procedures for the Two-Sample Location Problem. *Journal of the American statistical association*, 59(307):665–80, 1964.
- [220] M.W. Fagerland and L. Sandvik. The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in medicine*, 28(10):1487–97, May 2009.
- [221] J.F.C. Kingman. Random walks with spherical symmetry. *Acta mathematica*, 109(1):11–53, July 1963.
- [222] S. Aki. On nonparametric tests for symmetry in  $R^m$ . *Annals of the institute of statistical mathematics*, 45(4):787–800, 1993.
- [223] P.E. Jupp. A nonparametric correlation coefficient and a two-sample test for random vectors or directions. *Biometrika*, 74(4):887–90, December 1987.
- [224] N. Henze, B. Klar, and S.G. Meintanis. Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *Journal of multivariate analysis*, 87(2):275–97, 2003.
- [225] V.A. Fernández, M.D.J. Gamero, and J.M. García. A test for the two-sample problem based on empirical characteristic functions. *Computational statistics & data analysis*, 52(7):3730–48, 2008.
- [226] J. Ngatchou-Wandji. Testing for symmetry in multivariate distributions. *Statistical methodology*, 6(3):230–50, May 2009.
- [227] D. Goldberg. What every computer scientist should know about floating-point arithmetic. *ACM computing surveys*, 23(1):5–48, 1991.

- 
- [228] J. Robinson. Saddlepoint Approximations for Permutation Tests and Confidence Intervals. *Journal of the Royal statistical society. Series B (Methodological)*, 44(1):91–101, 1982.
- [229] P. Gill. Efficient calculation of p-values in linear-statistic permutation significance tests. *Journal of statistical computation and simulation*, 77(1):55–61, January 2007.
- [230] J. Besag and P. Clifford. Sequential Monte Carlo p-values. *Biometrika*, 78(2):301–304, 1991.
- [231] M.P. Fay, H. Kim, and M. Hachey. On Using Truncated Sequential Probability Ratio Test Boundaries for Monte Carlo Implementation of Hypothesis Tests. *Journal of computational and graphical statistics*, 16(4):946–67, 2007.
- [232] G. Ulrich. Computer Generation of Distributions on the m-Sphere. *Journal of the Royal statistical society. Series C (Applied Statistics)*, 33(2):158–63, 1984.
- [233] A.T.A Wood. Simulation of the von mises fisher distribution. *Communications in statistics - simulation and computation*, 23(1):157–64, 1994.
- [234] P.A. Bandettini. What's New in Neuroimaging Methods? *Annals of the new york academy of sciences*, 1156:260–293, 2009.
- [235] D. Wang, L. Shi, W.C.W. Chu, T. Paus, J.C.Y. Cheng, and P. Ann. A comparison of morphometric techniques for studying the shape of the corpus callosum in adolescent idiopathic scoliosis. *NeuroImage*, 45(3):738–748, 2009.
- [236] T.F. Cootes, C.J. Twining, K.O. Babalola, and C.J. Taylor. Diffeomorphic statistical shape models. *Image and vision computing*, 26(3):326–32, 2008.
- [237] Walter M Ferrarini, L.Palm, H. Olofsen, M.A. van Buchem, J.H.C. Reiber, and F. Admiraal-Behloul. Shape differences of the brain ventricles in Alzheimer's disease. *NeuroImage*, 32(3):1060–9, September 2006.
- [238] M. Styner, G. Gerig, J. Lieberman, D. Jones, and D. Weinberger. Statistical shape analysis of neuroanatomical structures based on medial models. *Medical image analysis*, 7(3):207–220, 2003.
- [239] T.B. Terriberry, S.C. Joshi, and G. Gerig. Hypothesis testing with nonlinear shape models. In *Proceedings of information processing in medical imaging*, pages 15–26, 2005.
- [240] A. Mechelli, C.J. Price, K.J. Friston, and J. Ashburner. Voxel-based morphometry of the human brain: methods and applications. *Current medical imaging reviews*, 1(2):105–13, 2005.
- [241] C. Davatzikos and N. Bryan. Using a deformable surface model to obtain a shape representation of the cortex. *IEEE transactions on medical imaging*, 15(6):785–95, January 1996.
- [242] S. Durrleman, X. Pennec, A. Trouvé, P. Thompson, and N. Ayache. Inferring brain variability from diffeomorphic deformations of currents : An integrative approach. *Medical image analysis*, 12:626–637, 2008.
- [243] P. Aljabar, K.K. Bhatia, M. Murgasova, J.V. Hajnal, J.P. Boardman, L. Srinivasan, M.A. Rutherford, L.E. Dyet, A.D. Edwards, and D. Rueckert. Assessment of brain growth in early childhood using deformation-based morphometry. *NeuroImage*, 39(1):348–58, January 2008.

## Bibliography

---

- [244] C. Studholme, V. Cardenas, R. Blumenfeld, N. Schuff, H.J. Rosen, B. Miller, and M. Weiner. Deformation tensor morphometry of semantic dementia with quantitative validation. *NeuroImage*, 21(4):1387–98, April 2004.
- [245] V.A. Cardenas, C. Studholme, S. Gazdzinski, T.C. Durazzo, and D.J. Meyerhoff. Deformation-based morphometry of brain changes in alcohol dependence and abstinence. *NeuroImage*, 34(3):879–87, February 2007.
- [246] T. Chen and D. Metaxas. A hybrid framework for 3D medical image segmentation. *Medical image analysis*, 9:547–565, 2005.
- [247] M.P. Fay and D.A. Follmann. Designing Monte Carlo Implementations of Permutation or Bootstrap Hypothesis Tests. *The american statistician*, 56(1):63–70, February 2002.
- [248] C. Rorden, L. Bonilha, and T.E. Nichols. Rank-order versus mean based statistics for neuroimaging. *Neuroimage*, 35(4):1531–1537, 2007.
- [249] M. Muskulus, A.E.H. Scheenstra, and S. Verduyn-Lunel. A generalization of the Moore-Rayleigh test for testing symmetry of vector data and two-sample problems, 2009.
- [250] K.J. Friston, A.P. Holmes, C.J. Price, C. Bu, and K.J. Worsley. Multisubject fMRI Studies and Conjunction Analyses. *NeuroImage*, 396:385–396, 1999.
- [251] J.S. Allen, J. Bruss, S. Mehta, T. Grabowski, C. Kice, and H. Damasio. Effects of spatial transformation on Regional Brain Volume Estimates. *NeuroImage*, 2008.
- [252] A. Klein, J. Andersson, B.A. Ardekani, J. Ashburner, B. Avants, M. Chiang, G.E. Christensen, D.L. Collins, J. Gee, P. Hellier, J. Hyun, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R.P. Woods, J.J. Mann, and R.V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, 2009.
- [253] J. Olesen. The International Classification of Headache Disorders. *Headache*, 48(5):691–3, May 2008.
- [254] M.D. Ferrari. The economic burden of migraine to society. *Pharmacoeconomics*, 13(6):667–76, June 1998.
- [255] A.M.J.M. van den Maagdenberg, J. Haan, G.M. Terwindt, and M.D. Ferrari. Migraine: gene mutations and functional consequences. *Current opinion in neurology*, 20(3):299–305, June 2007.
- [256] R.A. Ophoff, G.M. Terwindt, M.N. Vergouwe, R. van Eijk, P.J. Oefner, S.M. Hoffman, J.E. Lamerdin, H.W. Mohrenweiser, D.E. Bulman, M. Ferrari, J. Haan, D. Lindhout, G.J. van Ommen, M.H. Hofker, M.D. Ferrari, and R.R. Frants. Familial Hemiplegic Migraine and Episodic Ataxia Type-2 Are Caused by Mutations in the Ca<sup>2+</sup> Channel Gene CACNL1A4. *Cell*, 87(3):543–52, November 1996.
- [257] D. Pietrobon. Migraine: new molecular mechanisms. *The neuroscientist*, 11(4):373–86, August 2005.
- [258] A. Tottene, F. Pivotto, T. Fellin, T. Cesetti, A.M.J.M. van Den Maagdenberg, and D. Pietrobon. Specific kinetic alterations of human CaV2.1 calcium channels produced by mutation S218L causing familial hemiplegic migraine and delayed cerebral edema and coma after minor head trauma. *The journal of biological chemistry*, 280(18):17678–86, May 2005.

- [259] A.M.J.M. van Den Maagdenberg, D. Pietrobon, T. Pizzorusso, S. Kaja, L.A.M. Broos, T. Cesetti, R.C.G. van De Ven, A. Tottene, J. van Der Kaa, J.J. Plomp, R.R. Frants, and M.D. Ferrari. A Cacna1a knockin migraine mouse model with increased susceptibility to cortical spreading depression. *Neuron*, 41(5):701–10, March 2004.
- [260] K. Eikermann-Haerter, E. Dileköz, C. Kudo, S.I. Savitz, C. Waeber, M.J. Baum, M.D. Ferrari, A.M.J.M. van Den Maagdenberg, M.A. Moskowitz, and C. Ayata. Genetic and hormonal factors modulate spreading depression and transient hemiparesis in mouse models of familial hemiplegic migraine type 1. *The journal of clinical investigation*, 119(1):99–109, January 2009.
- [261] B. Todorov, R.C.G. van de Ven, S. Kaja, L.A.M. Broos, S.J. Verbeek, J.J. Plomp, M.D. Ferrari, R.R. Frants, and A.M.J.M. van den Maagdenberg. Conditional inactivation of the Cacna1a gene in transgenic mice. *Genesis*, 44(12):589–94, December 2006.
- [262] A.M.J.M. van den Maagdenberg, T. Pizzorusso, S. Kaja, N. Terpolilli, M. Shapovalova, F.E. Hoebeek, C.F. Barrett, L. Gherardini, R.C.G. van de Ven, B. Todorov, L.A.M. Broos, A. Tottene, Z. Gao, M. Fodor, C.I. De Zeeuw, R.R. Frants, N. Plesnila, J.J. Plomp, D. Pietrobon, and M.D. Ferrari. High cortical spreading depression susceptibility and migraine-associated symptoms in Ca(v)2.1 S218L mice. *Annals of neurology*, 67(1):85–98, January 2010.
- [263] E.E. Kors, G.M. Terwindt, F.L. Vermeulen, R.B. Fitzsimons, P.E. Jardine, P. Heywood, S. Love, A.M.J.M. van den Maagdenberg, J. Haan, R.R. Frants, and M.D. Ferrari. Delayed cerebral edema and fatal coma after minor head trauma: role of the CACNA1A calcium channel subunit gene and relationship with familial hemiplegic migraine. *Annals of neurology*, 49(6):753–60, June 2001.
- [264] A. Tottene, T. Fellin, S. Pagnutti, S. Luvisetto, J. Striessnig, C. Fletcher, and D. Pietrobon. Familial hemiplegic migraine mutations increase Ca(2+) influx through single human CaV2.1 channels and decrease maximal CaV2.1 current density in neurons. *Proceedings of the national academy of sciences of the United States of America*, 99(20):13284–9, October 2002.
- [265] D. Pietrobon. Familial hemiplegic migraine. *Neurotherapeutics*, 4(2):274–84, April 2007.
- [266] M.C. Kruit, L.J. Launer, M.D. Ferrari, and M.A. van Buchem. Brain stem and cerebellar hyperintense lesions in migraine. *Stroke*, 37(4):1109–12, April 2006.
- [267] M.C. Kruit, M.A. van Buchem, P.A.M. Hofman, J.T.N. Bakkers, G.M. Terwindt, M.D. Ferrari, and L.J. Launer. Migraine as a risk factor for subclinical brain lesions. *the journal of the American medical association*, 291(4):427–34, January 2004.
- [268] M.S. Matharu, C.D. Good, A. May, A. Bahra, and P.J. Goadsby. No change in the structure of the brain in migraine: a voxel-based morphometric study. *European journal of neurology*, 10(1):53–7, January 2003.
- [269] W. Valfrè, I. Rainero, M. Bergui, and L. Pinessi. Voxel-based morphometry reveals gray matter abnormalities in migraine. *Headache*, 48(1):109–17, January 2008.
- [270] A. May and M. Matharu. New insights into migraine: application of functional and structural imaging. *Current opinion in neurology*, 20(3):306–9, June 2007.
- [271] P.S. Sándor, A. Mascia, L. Seidel, V. de Pasqua, and J. Schoenen. Subclinical cerebellar impairment in the common types of migraine: a three-dimensional analysis of reaching movements. *Annals of neurology*, 49(5):668–72, May 2001.

## Bibliography

---

- [272] O. Yilmaz-Kusbeci, N. Gocmen-Mas, A. Yucel, H.S. Karabekir, T. Ertekin, and A.C. Yazici. Evaluation of Cerebellar and Cerebral Volume in Migraine with Aura: A Stereological Study (ahead of print). *Cerebellum*, March 2010.
- [273] S.M. Bowyer, K.S. Aurora, J.E. Moran, N. Tepley, and K.M. Welch. Magnetoencephalographic fields from patients with spontaneous and induced migraine aura. *Annals of neurology*, 50(5):582–7, November 2001.
- [274] F.M. Cutrer, A. O'Donnell, and M. Sanchez del Rio. Functional neuroimaging: enhanced understanding of migraine pathophysiology. *Neurology*, 55(9 Suppl 2):S36–45, January 2000.
- [275] N. Hadjikhani, M. Sanchez Del Rio, O. Wu, D. Schwartz, D. Bakker, B. Fischl, K.K. Kwong, F.M. Cutrer, B.R. Rosen, R.B. Tootell, A.G. Sorensen, and M.A. Moskowitz. Mechanisms of migraine aura revealed by functional MRI in human visual cortex. *Proceedings of the national academy of sciences of the United States of America*, 98(8):4687–92, April 2001.
- [276] F. Richter, S. Rupprecht, A. Lehmenkühler, and H. Schaible. Spreading depression can be elicited in brain stem of immature but not adult rats. *Journal of neurophysiology*, 90(4):2163–70, October 2003.
- [277] T.J. Ebner and G. Chen. Spreading acidification and depression in the cerebellar cortex. *The neuroscientist*, 9(1):37–45, February 2003.
- [278] Y. Gurses-Ozdemir, J. Qiu, N. Matsuoka, H. Bolay, D. Bermpohl, H. Jin, X. Wang, G.A. Rosenberg, E.H. Lo, and M.A. Moskowitz. Cortical spreading depression activates and upregulates MMP-9. *The journal of clinical investigation*, 113(10):1447–55, May 2004.
- [279] S.K. Afridi, M.S. Matharu, L. Lee, H. Kaube, K.J. Friston, R.S.J. Frackowiak, and P.J. Goadsby. A PET study exploring the laterality of brainstem activation in migraine using glyceryl trinitrate. *Brain*, 128(Pt 4):932–9, April 2005.
- [280] A. Bahra, M.S. Matharu, C. Buchel, R.S. Frackowiak, and P.J. Goadsby. Brainstem activation specific to migraine headache. *Lancet*, 357(9261):1016–7, March 2001.
- [281] M.A. Rocca, A. Ceccarelli, A. Falini, B. Colombo, P. Tortorella, L. Bernasconi, G. Comi, G. Scotti, and M. Filippi. Brain gray matter changes in migraine patients with T2-visible lesions: a 3-T MRI study. *Stroke*, 37(7):1765–70, July 2006.
- [282] C. Weiller, A. May, V. Limmroth, M. Jüptner, H. Kaube, R.V. Schayck, H.H. Coenen, and H.C. Diener. Brain stem activation in spontaneous human migraine attacks. *Nature medicine*, 1(7):658–60, July 1995.
- [283] J.R. Gee, J. Chang, A.B. Dublin, and N. Vijayan. The association of brainstem lesions with migraine-like headache: an imaging study of multiple sclerosis. *Headache*, 45(6):670–7, June 2005.
- [284] P.J. Goadsby. New directions in migraine research. *Journal of clinical neuroscience*, 9(4):368–73, July 2002.
- [285] Z. Katsarava, T. Egelhof, H. Kaube, H. Diener, and V. Limmroth. Symptomatic migraine and sensitization of trigeminal nociception associated with contralateral pontine cavernoma. *Pain*, 105(1-2):381–4, September 2003.
- [286] N.H. Raskin, Y. Hosobuchi, and S. Lamb. Headache may arise from perturbation of brain. *Headache*, 27(8):416–20, September 1987.

- [287] F. Veloso, K. Kumar, and C. Toth. Headache secondary to deep brain implantation. *Headache*, 38(7):507–15, 1998.
- [288] K.M. Welch, V. Nagesh, S.K. Aurora, and N. Gelman. Periaqueductal gray matter dysfunction in migraine: cause or the burden of illness? *Headache*, 41(7):629–37, 2001.
- [289] A.F.M. DaSilva, C. Granziera, J. Snyder, and N. Hadjikhani. Thickening in the somatosensory cortex of patients with migraine. *Neurology*, 69(21):1990–5, 2007.
- [290] K. Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5):198–211, 2007.
- [291] B.J. Erickson and B. Bartholmai. Computer-aided detection and diagnosis at the start of the third millennium. *Journal of digital imaging*, 15(2):59–68, 2002.





## International journal papers

**A.E.H. Scheenstra**, R.C.G. van de Ven, L. van der Weerd, A.M.J.M. van Den Maagdenberg, J. Dijkstra, and J.H.C. Reiber. Automated segmentation of *in vivo* and *ex vivo* mouse brain magnetic resonance images. *Molecular imaging*. **8**(1) pages 35-44 (2009)

M. Muskulus, **A.E.H. Scheenstra**, N. Braakman, J. Dijkstra, S. Verduyn-Lunel, A. Alia, H.J.M. de Groot, and J.H.C. Reiber. Prospects for early detection of Alzheimer's disease from serial MR images in transgenic mice. *Current alzheimer research*. **6**(6) pages 503-18 (2009)

## Peer-Reviewed conference papers

**A.E.H. Scheenstra**, A.C.C. Ruifrok, and R.C. Veltkamp. A Survey of 3D Face Recognition Methods. In *Audio- and video-based biometric person authentication (AVBPA)*. pages 891-9 (2005)

**A.E.H. Scheenstra**, J. Dijkstra, R.C.G. van de Ven, L. van der Weerd, and J.H.C. Reiber. Automated segmentation of the *ex vivo* mouse brain. In: *Proceedings of SPIE medical imaging*. page 651106 (2007)

**A.E.H. Scheenstra**, J. Dijkstra, R.C.G. van de Ven, L. van der Weerd, and J.H.C. Reiber. Automated Edge-Driven Markov Random Field Segmentation of *ex vivo* Mouse Brain MRM Images. In: *International symposium of biomedical imaging (ISBI)*. pages 1292-5 (2007)

**A.E.H. Scheenstra**, M. Muskulus, M. Staring, A.M.J.M. van den Maagdenberg, S. Verduyn-Lunel, J.H.C. Reiber, L. van Der Weerd, and J. Dijkstra. The 3D Moore-Rayleigh test for the quantitative groupwise comparison of MR brain images. In: *Proceedings of information processing in medical imaging (IPMI)*. pages 564-75 (2009)

---

## Book chapters

A.C.C. Ruifrok, **A.E.H. Scheenstra**, J. Bijhold, and R.C. Veltkamp. Facial image comparison using 3D techniques. In: *Facial Reconstruction. BKA Research Series nr. 35*. Luchterhand Publishers, Ed. T. Buzug, K.M. Sigl, J. Bongartz, K. Pruffer. pages 192-8 (2007)

**A.E.H. Scheenstra**, J. Dijkstra, and L. van der Weerd. Volumetry and other quantitative measurements to assess the rodent brain. In: *In vivo NMR Imaging: Methods and Protocols*. Humana Press, USA. Ed. C. Faber and L. Schroeder. *in press*.

## Abstracts

A.C.C. Ruifrok, **A.E.H. Scheenstra**, J. Bijhold, and R.C. Veltkamp. Facial Image Comparison Using 3D Techniques. In *Proceedings 2nd International Conference on Reconstruction of Soft Facial Parts*. (2005)

**A.E.H. Scheenstra**, J. Dijkstra, R.C.G. van de Ven, L. van der Weerd, and J.H.C. Reiber. Automated registration of histology sections with *ex vivo* MRM volumes. In *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*. page 2012 (2006)

**A.E.H. Scheenstra**, J. Dijkstra, R.C.G. van de Ven, L. van der Weerd, and J.H.C. Reiber. Automated Edge-Driven Markov Random Field Segmentation of *ex vivo* Mouse Brain MRM Images. In *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*. page 624 (2007)

L. Ferrarini, **A.E.H. Scheenstra**, G.B. Frisoni, M. Muskulus, M. Pievani, R. Ganzola, J.H.C. Reiber, J. Dijkstra and J. Milles, Morphological changes in the hippocampus predict MCI conversion to AD: An MR-based comparison between Moore-Rayleigh and permutation tests. In *Proceedings International Conference on Alzheimer's Disease*. (2009)

D. Suciu, **A.E.H. Scheenstra**, J. Dijkstra, M.S. Oitzl, and L. van der Weerd. Effects of Continuously High Levels of Corticosteroids on Mouse Hippocampus a Longitudinal *in vivo* MRI Study. In *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*. page 2378 (2010)

**A.E.H. Scheenstra**, D. Suciu, M. Muskulus, J.H.C. Reiber, M.S. Oitzl, L. van der Weerd, and J. Dijkstra. Quantitative and Local Mouse Brain Morphometry in Longitudinal MRI Studies. In *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)* page 3137 (2010)

---

## Acknowledgements

---

This thesis describes the work which was performed between 2005 and 2010 under the supervision of Prof. dr. ir. J.H.C. Reiber and dr. ir. J. Dijkstra, at the *Laboratorium voor Klinische en Experimentele Beeldverwerking* (LKEB), a division of the Radiology department at the Leiden University Medical Center (LUMC) in The Netherlands. During the course of my PhD study, many have contributed to this work and I would like to express my gratitude to them:

Louise van der Weerd, it has been a pleasure working with you and learning from you. Michael Muskulus, Niels Braakman, Arn van den Maagdenberg, Rob van der Ven, Rob Nabuurs, and Dana Suciu, thank you all for the pleasant collaborations, for giving me insights on statistics, Alzheimer's Disease, migraine, mice, for sharing MR images, and for giving feedback on my papers. Boudewijn Lelieveldt, Luca Ferrarini, Julien Milles and Michèle Huijberts, sometimes our discussions were work related, sometimes they weren't. However, they were always vivid and funny. I enjoyed all of them. Marijn van Stralen, we send so many e-mails with the subject 'MISP' that they were marked as spam. Thank you for the co-foundation of the Medical Imaging Symposium for PhD-students. At the same time, I would like to thank Prof. Reiber and Prof. Niessen, for believing in this crazy idea and giving us time and finances to set up this symposium.

I would like to thank the Terminaalzaal and all members of the notorious cocktail party, for the not-so-work-related discussions and for accepting me for who I am, Alize. Emmanuelle, Maribel and Noortje, thank you for adding the girly touch to the LKEB. Marielle, Marcel, Maddalena, Mary, Josephine, Lillian, and Chris, thank you for volunteering to check my spelling and grammar.

Joyce, Maaike, Jantien, Sabine, Inge, Lieselot, Astrid, Sylvia, Margriet, and Corine, I owe you all my thanks for feeding me tea and chocolate, cooking me dinners, for listening to all my stories, and for allowing me to disappear from time to time, so I could finish my thesis. Furthermore, my gratitude goes to the Amsterdam Lions and the Dutch National Lacrosse Team, among others Marielle, Lillian, Mary, Josephine, Astrid and Rosa, for sharing my passion for lacrosse and for the hours of fun on and off the field. And a special thanks to Lillian for an amazing time while organizing

---

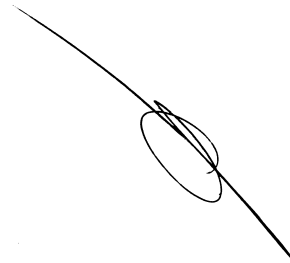
many things for the NLB board. Also, I would like to thank the NKI-AvL RT research group for the nice atmosphere during during my wannabe-postdoc period, with a special thanks to Alessia, Yenny, Maddalena, Angelo, and Edoardo for sharing my addiction for caffeine, sweets, and caffè corretto.

Sabine and Marcel, thanks for being my *paranimfen*. I have no doubt that you can laugh away all my stress, my agitation, and my tendency towards control freakness just before the defense.

Last, but not least, I would like to thank my family for being there for me. Papa and Mama, nobody could have supported me more than you have done. Renske, congratulations by winning our "who will finish the PhD first" competition, I loved our battle and hope we keep having this sisterly competitions. Maaïke, I know you will beat the two of us by having the *best* PhD-thesis. Gurbe, you will always be my most favorite brother. Opa, Grootvader and Omie thank you for your unconditional love (*Opa, Grootvader and Omie, bedankt voor al jullie onvoorwaardelijke liefde*). The family Pistidda, thank you for welcoming me with open arms into your family (*I Pistidda, grazie di tutto cuore di avermi accolto a braccia aperte*)

Alessio, thank you for teaching me how to (occasionally) relax and slow down. Impatient as ever, I cannot wait to start the rest of our lives together.

Alize Scheenstra

A handwritten signature in black ink, consisting of a long diagonal stroke that curves downwards and ends in a loop.

---

## Curriculum vitae

---

Alize Elske Hiltje Scheenstra was born on 4 November 1981 in Gouda. In 1999, she received her secondary school degree from the Coornhert Gymnasium in Gouda. That same year she started her study in Computer Science at the University of Utrecht, where she received the M.Sc degree in Biomedical Image Sciences in 2005. Her graduation project was performed under the supervision of Prof. dr. R.C. Veltkamp and carried out at the Dutch Forensic Institute on landmark-based 3D face recognition.

From March 2005 until September 2009 she was employed at the *Laboratorium voor Klinische en Experimentele Beeldverwerking* (LKEB) in the Leiden University Medical Center (LUMC). She performed research to the quantitative morphometry on mouse brain MRI under the supervision of Prof. dr. ir. J.H.C. Reiber and dr. ir. J. Dijkstra, of which the results are presented in this thesis. For her work on the 3D Moore-Rayleigh test for quantitative local morphometry she received the IPMI best poster award. During that period she co-founded with Marijn van Straalen the Medical Imaging Symposium for PhD-students, which was the precursor of the Nederlands Forum for Biomedical Imaging (NFBI; <http://www.nfbi.nl>).

Currently she is employed as postdoc at the radiotherapy department of the Dutch cancer institute (NKI) under the supervision of Prof. dr. M. van Herk and dr. ir. J.-J. Sonke, where she is working on the regional quantification of radiation induced lung damage for stereotactic body radiation therapy.



---

## List of Figures

---

1.1	The process to generating transgenic modified mice through the implantation of embryonic stem cells, see section 1.1.2 for further details. Photography courtesy of Anne Bower and Manfred Baetscher, Transgenic Core, Oregon Health & Science University, Portland, OR. Printed with permission. . . . .	4
1.2	Comparison between the mouse brain and the human brain A) the outer surface and B) the internal anatomy. . . . .	5
3.1	Pathological features of Alzheimer’s disease: Normal neuron and synapse (A). Affected neuron in late-stage (B). Normal cerebral artery (C). Affected cerebral vessel (D). . . . .	18
3.2	Statistical texture analysis of MR images: Original image (A). Enlarged 10×10 subimage (B). Gray level representation with 8 levels (C). Co-occurrence matrix $C(r, s 1, 0)$ of subimage for horizontal displacement (D). Co-occurrence matrix $C(r, s 0, 1)$ for vertical displacement, which describes the statistical properties of the subimage with regard to local variations (E). . . . .	30
3.3	Overview of the minimum age at which Alzheimer pathology was detected in <i>in vivo</i> MRI volumes of APP/PS1 mice (A) and the remaining mouse models (B). Dark grey denotes a relaxometric method, light grey denotes that AD pathology was determined by plaque burden analysis. None of the studies used contrast agents during imaging. . . . .	31

---

4.1	The segmentation pipeline for the mouse brain segmentation algorithm. The Atlas volume is first registered to the new MRI volume, resulting in an initial segmentation. Furthermore, the intensity distributions per class are derived from the atlas and the edge information from the new MRI volume is extracted. The clustering algorithm is performed in the second step for a final segmentation; therefore it uses the class statistics, initial segmentation and edge information as input. . . . .	40
4.2	The MRI atlas used for <i>in vivo</i> (E,G) and <i>ex vivo</i> (F,H) MR images with their manual segmentations (A,B,C, and D) and corresponding names and abbreviations. For a better understanding, the abbreviations for the brain structures are in upper case where the abbreviations for brain tracts are in lower case. . . . .	44
4.3	A visual comparison between the manual (M) and automated (A) segmentation for <i>in vivo</i> MRI (left) and <i>ex vivo</i> MRI volumes (right). The colour-coding of the various classes correspond to the legend as given in figure 4.2. . . . .	44
4.4	The increase in $\kappa$ between the two steps of the presented algorithm for the <i>ex vivo</i> segmentation results. The light grey bar denotes the average $\kappa$ after the affine atlas-based registration step, whereas the dark grey bar displays the average $\kappa$ after the clustering step. Furthermore, the standard deviations are given for each bar to indicate the robustness of the algorithm. The abbreviations of the various structures are explained in figure 4.2 . . . . .	46
4.5	The average kappa indices of the Demons algorithm (light grey bars) compared to the average $\kappa$ of the presented method (dark grey bars) for the <i>in vivo</i> mouse brain segmentation. Furthermore, the standard deviations are given for each bar to indicate the robustness of the algorithm. The abbreviations of the various structures are explained in figure 4.2 . . . . .	46
5.1	Left: Distribution function for $N = 2, 5, 10$ and $50$ (solid, from right to left), compared to the asymptotic $\chi^2_3$ distribution (dashed). Right: Logarithm of significance probabilities. . . . .	59
5.2	Estimated power functions for the family of Gaussian distributions with covariance matrix the identity and mean shifted a distance $\mu$ away from zero. Sample size $N = 10$ . . . . .	61
5.3	Estimated power functions for the family of Gaussian distributions, varying the standard deviation $\sigma$ of a single axis. Sample size $N = 10$ . Note the small range of the power. . . . .	61
5.4	Scattered Fisher distribution, visualized by 1,000 randomly drawn points in the unit-cube. Left: Concentration parameter $\lambda = 1$ . Middle: Concentration parameter $\lambda = 2.5$ . Right: Concentration parameter $\lambda = 5$ . . . . .	62



---

5.5	Estimated power functions for the mixture of the scattered Fisher distribution $2F3_5$ with the uniform distribution on the sphere $S^2$ , varying the mixture probability $p$ that a sample vector arises from the first distribution. Sample size $N = 10$ . . . . .	63
5.6	Comparison of significance probabilities for resultant lengths of spherically symmetric (smooth curve) and one-dimensional symmetric random walk (piecewise-linear curve) in three dimensions for $N = 10$ steps. Dotted curve shows the asymptotic case. . . . .	65
5.7	Estimated power for translated (10 standard deviations), then rotated diagonal Gaussians with unit variances as a function of relative rotation angle. Sample size $N = 10$ . . . . .	67
5.8	Estimated power functions for translated (10 standard deviations), then rotated diagonal Gaussians with unit variance against the scattered Fisher distribution (scaled by a unit Gaussian) as a function of relative rotation angle. Sample size $N = 10$ . Note the small power at angle zero. . . . .	68
5.9	Adjusted estimated power functions. Left: translated (10 standard deviations), then rotated diagonal Gaussians with unit variances as a function of relative rotation angle. Right: translated (10 standard deviations) then rotated diagonal Gaussians with unit variance against the scattered Fisher distribution (scaled by a unit Gaussian) as a function of relative rotation angle. Sample size $N = 10$ . Results based on 100 realizations of $10^4$ permutations each. . . . .	69
5.10	Generation of spherical deformations. The plot shows the behaviour of the deformation field in a one-dimensional projection along a radial direction. The volume at normalized distance $r/R$ from the centerpoint of the sphere (radius $R$ ) is mapped radially to $r'/R$ . For $r < R/2$ the length of the deformation vectors expands linearly, $r' = \lambda_1 r$ , attaining its maximum at half radius $r = R/2$ . For $r > R/2$ the length shrinks linearly by $\lambda_2 = 2 - \lambda_1$ , ensuring continuity at the boundary. The stippled line shows the case of no deformation ( $\lambda_1 = 1$ ). . . . .	70
5.11	Validation with a synthetic dataset. A slice from the deformation field of $50 \times 50 \times 80$ voxels (isotropic spacing of 1.00 mm), showing the five spherical deformations that were added to it (see text for details). The color indicates the length of the deformation vectors per voxel. A: Mean deformation field for the first group. B: Difference between deformation fields for the two groups (smaller deformations in spheres $S_2$ and $S_5$ in the second group). . . . .	74
5.12	Validation with a synthetic dataset. Negative logarithms of significance probabilities for two-sample Moore-Rayleigh test (MR3), the Diks2 test (see text for details), the permutations of the Hotellings $T^2$ test (pHT2), and the Hotellings $T^2$ test (HT2). . . . .	75

---

6.1	The deformation fields of a control group (black) and the test group (dark grey) for one voxel ( $n = m = 10$ ), where the difference vectors, see text for details, are given in dotted arrows. The right frame shows the difference vectors translated to the origin to show the spread of the differences. This situation is given for two differences in means: $\Delta\mu = 0$ (a) and $\Delta\mu = 2\sigma$ (b). . . . .	83
6.2	The estimated power for the discrimination of two gaussian spheres with $\sigma = 1$ and relative displacement of $\Delta\mu$ with $N = 5$ (A), $N = 7$ (B), $N = 7$ (C) and $N = 20$ (D). . . . .	86
6.3	The estimated power for the influence of violations against the assumption of spherical symmetry for two gaussian spheres with $\sigma_x = \frac{1}{2}\sigma_y = \frac{1}{2}\sigma_z$ with sample sizes $N = 5$ (A and B), $N = 7$ (C and D), $N = 7$ (E and F) and $N = 20$ (G and H), where the left column shows the displacement of one cloud in the $x$ -direction and the right column shows the displacement of a point cloud in the $y$ -direction. . . . .	88
6.4	A slice from the 3D volume of a single control subject (A), the APP23 mouse (B), and the control average (C). Arrows indicate regions which show inhomogeneities, misregistration or incorrect masking. Furthermore, the region around the ventricles is delineated as region of interest. . . . .	89
6.5	The probability that a voxel is spherically symmetric for the wild types (top row) and the APP23 mice (bottom row) for three different slices through the brain. The area's which were not found spherically symmetric at a significance level of $\alpha < 0.05$ were color coded by their $p$ -value. . . . .	91
6.6	The probability per voxel that there is locally a group difference between wild type mice and APP23 mice for three slices through the brain. The top row shows the results from the permutations of the Hotelling's $T^2$ test (pHT2) and the bottom row shows the results from the single Moore-Rayleigh test (MR3). The area's which were not found spherically symmetric at a significance level of $\alpha = 0.05$ were not printed. . . . .	94
7.1	The schematic protocol for post processing of the normalized MR images for a cross-sectional study, so the two groups (control group and test group) can be tested for local significant differences with the Moore-Rayleigh test. . . . .	98
7.2	Volume rendering of the manual segmented cerebellum (green), the hippocampal regions (purple), the cortex (yellow) and the ventricles (blue) . . . . .	101
7.3	The deformation-based morphometry results for the three cross-sectional studies. The statistical parametric map indicating the significance level per voxel for three slices in the brain. The MR image is the average of each control group. . . . .	101

---

7.4	A volume visualization of the regions that were found significant under Bonferroni correction (red). For a better interpretation of the spatial orientation in the brain, the ventricles are also shown (blue). . . . .	103
7.5	A comparison of an MR slice of the control average and a regular subject. The arrows indicate areas where some intensity enhancement due to the averaging are visible. . . . .	104
7.6	A visual comparison of the <i>in vivo</i> MR averages for each cross-sectional study, showing the caudal shift of the cerebellum indicated with a white arrow. . . . .	105



---

## List of Tables

---

2.1	An overview of the characteristics between the four morphometry methods. . . . .	12
3.1	The characteristics of transgenic mouse models of AD. Transgene: PAC, P1 artificial chromosome. Promoters: PDGF, platelet-driven growth factor; PrP, prion protein; MoPrP, mouse prion protein; CAMKII, calcium/calmodulin-dependent protein kinase II. Phenotype; DP, diffuse (pre-amyloid) plaques; AP, amyloid plaques; NFT, neurofibrillary tangles; ND, neurodegeneration; Cog, cognitive impairment. For phenotype: +, positive; -, negative; ?, unknown. . . . .	20
3.2	Relaxometry measurements in AD mouse models. Relax.Par, Relaxation parameters; N.mice, number of mice included in the study; Tg, transgenic; Ntg, non-transgenic; N <sub>t</sub> , Time points; Stat. anal., statistical analysis; Lt., longitudinal study; m, months; w, weeks . . . . .	22
3.3	MR micro-imaging of senile plaques and plaque burden in humans and mouse models. Image modality: 2D/3D, 2- or 3-dimensional; T <sub>1</sub> /T <sub>1rho</sub> /T <sub>2</sub> /T <sub>2</sub> <sup>*</sup> , applied weighting in MR imaging experiments; DW, diffusion-weighted; <sup>19</sup> F, imaging of Fluorine-19 labeled contrast agent; GE, Gradient Echo; SE, Spin Echo; FSE, Fast Spin Echo; CRAZED, COSY revamped with asymmetric z-GE detection. Long.: indicates whether the study was longitudinal . . . . .	25
3.4	Texture analysis applied to alzheimer's disease on human data. AD: Alzheimers disease, MCI: Mild cognitive complainers, C: controls. . . .	29

---

4.1	The sizes in mm <sup>3</sup> and in voxels for the several brain structures. Furthermore, for the separate structures are the <i>in vivo</i> and <i>ex vivo</i> segmentation results given in average kappa indices and standard deviation. The missing values of the <i>in vivo</i> segmentation column represent structures for which no proper expert segmentation could be obtain and thus were excluded from the atlas and automated segmentation. . . . .	43
5.1	Critical values of Moore-Rayleigh statistic in 3D for various sample sizes N. . . . .	72
5.2	Precision and recall for the synthetic dataset. . . . .	72
7.1	The normalized brain structure volumes expressed in % of the total brain volume $\pm$ the standard deviations. . . . .	102

---

## List of Abbreviations

---

AD	Alzheimer's disease
DBM	Deformation-Based Morphometry
HT2	Hotelling's $T^2$ test
MRI	Magnetic Resonance Imaging
MR3	three-dimensional Moore-Rayleigh test
pHT2	Permutation test with the HT2 as test statistic
ROI	Region of Interest
SBM	Shape-Based Morphometry
SPM	Statistical Parametric Map
SOI	Structure of Interest
SPM	Statistical Parametric Map
TBM	Tensor-Based Morphometry
TE	Echo Time
TR	Repetition Time
WT	Wild type