

MRI Brain Perfusion Data: Comparing PCA and ICA in Parkinson's Disease

Van der Veer Institute for Parkinson's and Brain Research,
Christchurch, New Zealand

Zurich University of Applied Sciences, Winterthur, Switzerland

Sina Rüeger¹
sina.rueeger@gmail.com

June 25, 2010

¹Master of Science in Engineering MSE, Projectthesis II

This report compares two statistical methods applied in MRI perfusion data of Parkinson's disease people and healthy controls. Parkinson's disease (PD) is a neurodegenerative disorder of unknown cause and of unknown pathogenetic mechanism. Symptoms are motoric and/or cognitive impairment. Unfortunately the pathogenesis is not enough explored, also the diagnosis is difficult and not reliable. One is therefore interested in finding a reliable biomarker and learn more about affected organs and the process of the disease. Previous studies have shown that the brain and its cerebral blood perfusion is of particular interest.

We analyzed the cerebral blood perfusion of 26 healthy participants and 60 Parkinson's disease people. The cerebral blood flow data was collected with ASL MRI technique. MRI images contain (depending on the resolution) more than 1000 values, for each voxel one. These voxel are treated in a data set as variables such as age or gender. Thus each person has more than 1000 variables and the difficulty of high dimensionality of a data set appears. To determine which voxel discriminates Parkinson's disease and healthy persons, a statistical test can be used or, to perform this test for all voxels at once, a projection method can be applied. We applied two different projection methods - principal component analysis and independent component analysis - and compared them. A projection method returns linear combinations. Each subjects data can be expressed by a linear combination of *weights* and *components*. The weights per subject are then used as new variables to distinguish Parkinson's disease and healthy participants in a logistic regression with variable selection. The resulting patterns are then combined to define a typical Parkinson's disease brain-perfusion pattern. Bootstrap is performed to calculate the variance for the Parkinson's disease pattern, which is used for testing purposes. Cross-validation gives the performance of the biomarker.

Goal of this report was to perform the analysis with *principal component analysis* and *independent component analysis* and recommend one of these methods for this type of analysis.

Results showed that *independent component analysis* is outperforming *principal component analysis* regarding the numbers of misclassification in cross-validation.

This report is a result of my own work. If I write "we" in this report, then I mean me and the people from the Van der Veer Institute which supervised my work and gave support.

I would like to thank Tracy Melzer for his teamwork, assistance and teaching. Dr. Michael MacAskill and Prof. Tim Anderson offered the opportunity to write this thesis at the Van der Veer Institute. I would like to thank them for the confidence they had and for the support they provided. And last but not least I would like to thank Prof. Andreas Ruckstuhl who was very helpful with equations [3.9](#) to [3.16](#).

The report is written

- for persons at the Van der Veer Institute to understand and reproduce my work, (therefore the report is rather comprehensive)
- and for Prof. Marianne Müller from the University of Applied Sciences who will evaluate my work.

Contents

1. Introduction	6
2. Data	9
2.1. Variables	9
2.2. Missing Data	14
3. Methods	15
3.1. Data Preparation	20
3.2. Projection Methods	21
3.3. Classification	34
3.4. Building Pattern	40
3.5. Assessment of Model	41
3.6. Comparing PCA and ICA	43
4. Results	44
4.1. Data Preparation	44
4.2. Projection Methods	44
4.3. Classification	47
4.4. Building Pattern	66
4.5. Assessment of Model	71
4.6. Comparing PCA and ICA	82
4.7. Subset: Exclusion of Controls	87
5. Discussion	110
5.1. Outlook	111
A. Appendix A: Misc	113
A.1. Details to Cognitionlevels	113
A.2. Details to Mosaicplot	113
B. Appendix B: Pseudo Code	117
B.1. Data preparation	117
B.2. Projection Methods: ICA	119
B.3. Classification	119
B.4. Assessment of Model: Cross-validation	120
B.5. Assessment of Model: Bootstrap	122

C. Appendix C: Results	124
C.1. Data preparation	124
C.2. Projection Methods: PCA	131
C.3. Projection Methods: ICA	138
C.4. Linear Regression of Components and Co-Information	145

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder of unknown cause [2] and also of unknown pathogenetic mechanism. After Alzheimer's disease, Parkinson's disease is the second most common neurodegenerative disorder [1] and it is expected that the prevalence will increase with the increasing populations age. Symptoms of Parkinson's disease are the classic triad: tremor, rigidity and akinesia (motor impairment) as well as dementia (cognitive impairment). Mental impairment occur in particular in elderly people.

A reliable and easily applicable diagnostic test or marker for Parkinson's disease is not yet available [1]. The diagnosis is made based on clinical criteria: A response of the symptoms to the drug levodopa as well as asymmetric symptoms onset. Diagnosis might be difficult and not reliable due to different reasons. The clinical criteria can only indicate a *probable* Parkinson's disease. A definite diagnosis requires post-mortem confirmation. Misdiagnosis is common because the symptoms of parkinsonism can have a number of different causes (drugs or other diseases). Also the diagnosis of Parkinson's disease is difficult in early stages of the disease.

About 0-3% of the entire population in industrialized countries are suffering from Parkinson's disease [1]. Parkinson's diseased people under 50 years are rare, also some studies assume a higher risk to suffer Parkinson's disease for men than women.

Recent evidence indicates that people with Parkinson's disease have abnormal cerebral blood flow¹ in certain brain regions compared to healthy people. Starting point for this work was the article of Eidelberg (2000) [16] where principal component analysis and medical imaging (PET) were used to distinguish healthy controls and Parkinson's disease participants and to generate an image which emphasizes the brain regions related to motor and cognitive symptoms. This approach leads (a) to a biomarker for PD and (b) identifies parts of the brain which can be seen as related to PD (also called *pattern recognition*).

In this work we go a step further by using a different medical imaging technique and by applying a different statistical method. The focus of this work is to recommend one statistical method.

MRI We are using perfusion-weighted MRI images instead of PET to extract cerebral blood flow with medical imaging. Apart from the fact, that MRI is noninvasive

¹Technically cerebral blood *perfusion* was measured: Perfusion = $\frac{\text{Cerebral blood flow}}{\text{Tissue mass}}$.

whereas PET needs injections, there are also other advantages for patients/participants and the research team: it produces reproducible images, structural images are possible and the scanning duration is fast.

PCA & ICA We are performing principal component analysis *and* independent component analysis on the collected cerebral blood flow data. Although independent component analysis can be seen as a generalized form of a principal component analysis it has a different generative model behind and requires more complex computation - complex in terms of computation time and of selectable parameters.

Comparing We recommend one of the applied statistical methods - principal or independent component analysis - for this type of data and application.

The main idea of pattern recognition in Parkinson's disease can be explained with an example of face recognition.

Imagine photos are taken of people. Every person has a nose, mouth, ears, eyes, hair or no hair, different shaped face, and men might have a beard or a mustache. We want to develop an algorithm which recognizes either faces again or is able to recognize certain features (e.g. big nose or long hair or the gender). This report employed the latter option in the identification of different groups. For example if people with beards are the group of interest, a projection algorithm is applied to facilitate classification. There are several projection algorithms; principal component analysis and independent component analysis are often applied in face recognition. Both methods try to find components so that each person's face can be constructed with a mixture of different components. One component has the same spatial dimension than as the original photo, but captures just one feature of a face. The goal of principal component analysis is to find components that capture the maximal amount of variance (means to minimize the mean square error), whereas the goal of independent component analysis is to find components that are as independent from each other as possible. Both methods extract weights (or mixture) for each component and each person ("how much" from each component). Coming back to our beard example - if there is any component where the weights for each person are different for people with beards than without, then the beard-component has been found. People in the future who exhibit high weights for the beard-component will be classified as people with beards.

The article *Eigenfaces for Recognition* [19] by Turk & Pentland (1991) explains principal component analysis in face recognition very well.

Projection methods have a generative model behind. *Generative* means that we are assuming that the data can be expressed with a linear combination. With projection methods we can extract the so far unknown coefficients in the linear combination and the components by giving the data as input. Which linear combination is extracted depends on the type of the projection method. Principal component analysis produces others results as independent component analysis due to a different optimization criteria. The vectors in the linear combination can also be seen as *latent variables*. Latent variables are variables which can not be directly observed or measured.

Let us translate the face recognition example to brain imaging data. As well as people's faces, we can also look at different brain characteristics - in our case cerebral blood flow data. Components are extracted by applying a projection method. Each person will have a certain linear combination of the components. In this work we aimed to characterize the mixture of components in Parkinson's disease and control participants. We also compared early stage PD with more severe PD in terms of the expression of identified components, while identifying typical PD perfusion patterns. Basically we are assuming that PD's and controls have different scalars in the linear combination.

The goal of this work were as follows:

1. Identify perfusion patterns which optimally distinguish Parkinson's disease participants from healthy controls.
2. Combine the patterns to define a typical Parkinson's disease brain-perfusion-pattern.
3. Describe, employ and compare the two possible methods for this type of pattern recognition analysis: principal component analysis (PCA) and independent component analysis (ICA).
4. Decide whether PCA or ICA is recommended for this type of analysis.

29 healthy controls and 60 subjects with Parkinson's disease (hereinafter named as *controls* and *PD's*) participated in the study. Cerebral blood flow was measured with ASL fMRI technique. Collected were also several measures about the cognitive, motoric and disease status of the participants as well as standard measures such as gender and age. Details to data can be find in chapter 2 (page 9).

The model to describe the cerebral blood flow measures is explained in chapter 3 (page 15). It needs several methods to built the model and to validate it. These methods are described in detail in this chapter as well.

The results are then listed in chapter 4 (page 44) and are discussed in chapter 5 (page 110).

2. Data

65 subjects with varying levels of cognitive and motor impairment Parkinson's disease and 30 healthy controls participated in the study. Two types of data were collected: whole brain cerebral blood flow measures for each person and additional information included *Age*, *Sex*, *Montreal Cognition Assessment (MoCA)*, *Hoehn & Yahr Score (H&Y)*, *Unified Parkinson's Disease Rating Scale (UPDRS)*, *Duration* and *Level of Cognition*.

2.1. Variables

Figure 2.2 shows a graph for variables *Age*, *Sex* and *MoCA*. Figure 2.3 shows the relations only for the variables collected for Parkinson's disease people and MoCA. Six of all 95 subjects were excluded from the analysis due to large values in cerebral blood flow data or massive spiral artifacts, see details in section 2.2. The following summaries of the data do not contain these six subjects.

Group Categorical variable with levels *PD* (Parkinson's disease people) (60) and *Control* (29).

Age Measured in years, range from 45 to 84.

Sex Categorical variable with levels *male* (61) and *female* (28).

MoCA Montreal Cognition Assessment measures the global cognitive status [6], [7]. High values imply unimpaired, low values impaired cognition, range from 0 to 30.

Hoehn & Yahr Hoehn & Yahr measures the disease severity in PD [4]. Categorical variable with levels from 0 (unimpaired) to 5 (impaired) (levels: 1, 1.5, 2, 2.5, 3, 4, 5)

Duration Duration of PD in years (only PD).

UPDRS Unified Parkinson's Disease Rating Scale [3]. Mainly for measuring the motor symptoms in PD. High values imply more severe motor impairment.

Cognition Only Parkinson's disease people. Categorical variable to classify the cognitive impairment with levels *PDU* (cognitively unimpaired, 33), *MCI* (mild cognitive impairment, 16), *PDD* (demented, 11). Details to the classification of cognition level are in appendix A.1.

Subject number Subjects were named by the scanning number (4 or 5 digit).

Cerebral blood flow data CBF This resulted in a $(79 \times 95 \times 69)$ matrix of CBF values, where one point in the 3D space is called a voxel, stored as a NIfTI image [8]. Cerebral blood flow in the brain was measured with *ASL perfusion-weighted MRI*. Pooley (2005) [9] gives a review about fundamental physics of magnetic resonance imaging.

The raw cerebral blood flow data was collected with a 3T GE HDx scanner. All images were preprocessed with algorithms of SPM5 [10] and custom MATLAB scripts. The preprocessing steps (to get from raw to usable cerebral blood flow images) were as follows: *reorientation* (centering the image), *co-registering* (aligning the CBF image to the structural image), *normalizing* (bringing the images in a normalized space), *segmentation* (producing probabilities for white matter, gray matter and cerebral spinal fluid) and *smoothing*. Due to spiral artifacts the voxels in *z* direction could not be used from 1 to 14 and from 66 to 69. The image file was stored in NIfTI format [8].

The brain can be divided in gray matter, white matter and cerebral spinal fluid. Gray matter contains the neuronal cell bodies, white matter the myelinated axons. Gray matter has more blood perfusion and is therefore of greater importance for this analysis. Therefore a gray matter mask was used to exclude white matter and cerebral spinal fluid from the analysis. A gray matter mask is the average gray matter image across all subjects¹ with values ranging from 0 to 1 (probability that voxel is gray matter). White matter or non-brain tissue is denoted by a voxel value of 0 and 1 is gray matter. Voxels under the threshold of 0.1 were excluded from the analysis. Figure 2.1b shows a cerebral blood flow image of a representative healthy control where bright areas indicate high perfusion and surrounding dark regions represent non-brain tissue. Figure 2.1a shows the gray matter mask. Bright regions indicate gray matter; black/dark regions represent white matter or areas where no data was collected (non-brain tissue). We applied the gray matter mask to exclude CBF contributions from white matter and noise from non-brain regions.

¹Average for gray matter mask was taken from 30 controls, 10 PDU, 10 MCI and 10 PDD.

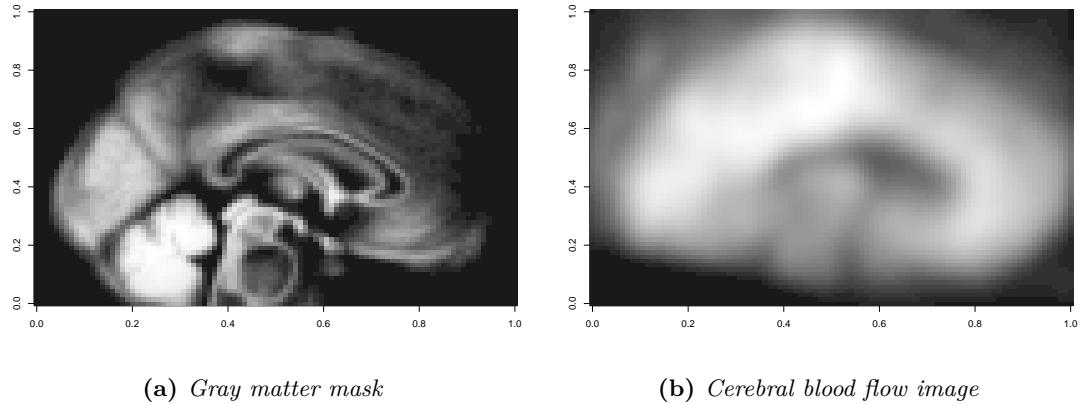
(a) *Gray matter mask*(b) *Cerebral blood flow image*

Figure 2.1.: (a) *Gray matter image*: Black parts of the images are white matter or non-brain tissue, which are excluded from the analysis. (b) *Cerebral blood flow image*: Black pixels are non-scanned areas or contain non-brain tissue (e.g. the ventricles in the middle of the image).

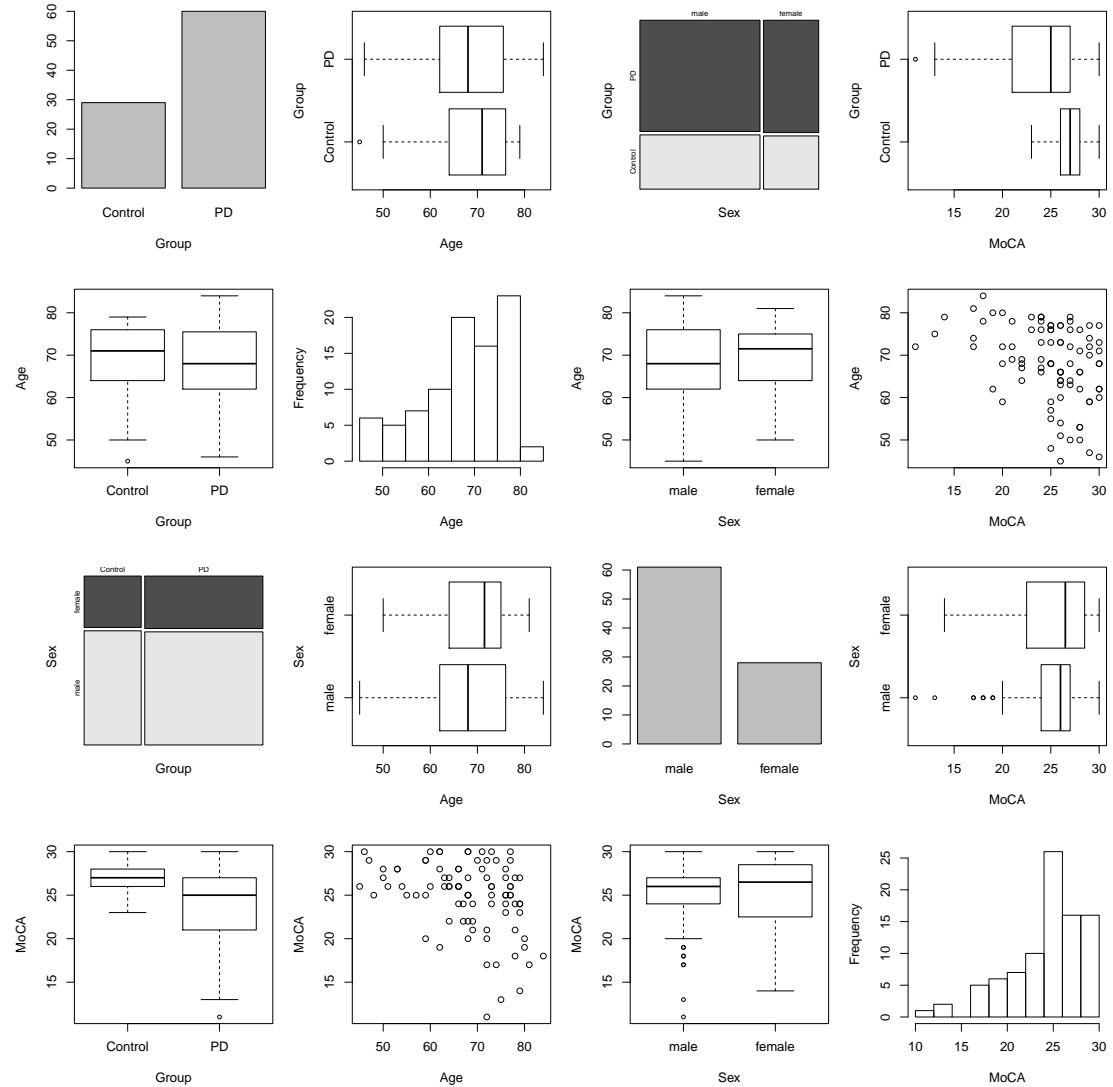


Figure 2.2.: Pairs plot of the variables Group, Age, Sex and MoCA. The proportion of male and female is about the same in the control and Parkinson's disease group. No MoCA values under 23 were obtained for control subjects, but Parkinson's disease subjects had values from 11 to 30. The age is comparable in the Parkinson's disease and control group as well as in the male and female group. Low MoCA values (<18, impaired) were only obtained for subjects with 70 years or elder.

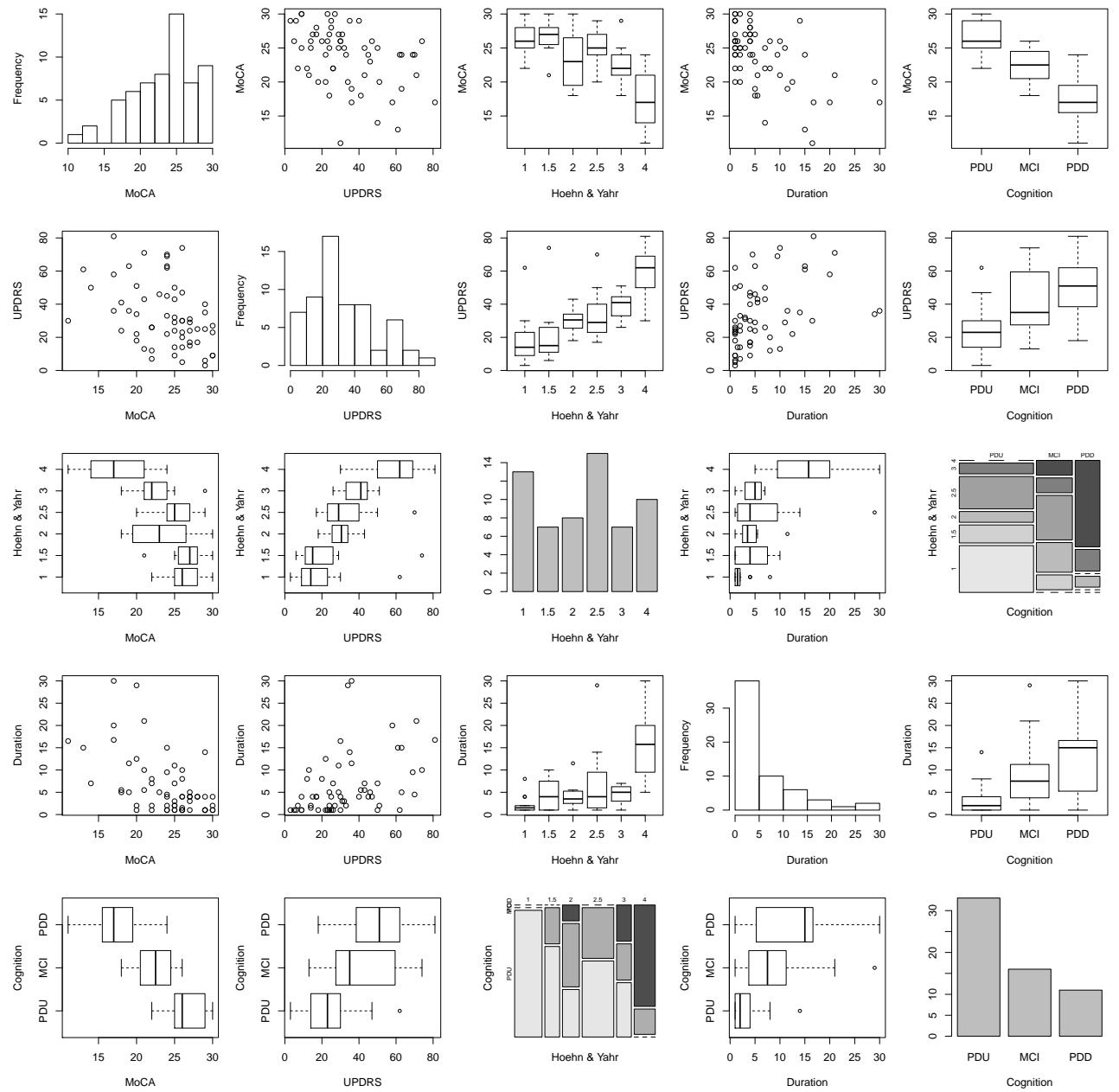


Figure 2.3.: Pairs plot of the variables MoCA, UPDRS, Hoehn & Yahr, Duration and Level of Cognition. High MoCA (cognitively unimpaired) scores imply low UPDRS score, high Hoehn & Yahr level and short duration. This means, if the cognition is impaired, then the disease severity has progressed and motor symptoms exist, whereas if the cognition is unimpaired, the disease is not severe yet and only slight motor symptoms exist. The higher UPDRS (motor impairment) the higher the Hoehn & Yahr level. Subjects with higher duration tend to have a higher UPDRS value and a higher Hoehn & Yahr level. Most PDU subjects have level 1 in Hoehn & Yahr, most PDD subjects have level 4.

2.2. Missing Data

Six subjects (subject number 5584, 6576, 8012, 9553, 11718 and 13032) were excluded from the analysis due to abnormal head positioning (signal loss, causing missing value) or large spiral artifacts (arising from the image reconstruction), see table 2.1 for an overview.

Subject	Group	Cognition	Age	Sex	MoCA	UPDRS	Hoehn & Yahr	Duration
8012	PD	MCI	78	male	19	35	4	5.00
11718	PD	MCI	73	male	24	57	4	10.00
5584	PD	PDU	63	male	21	36	3	3.00
9553	PD	PDU	62	male	26	36	2.5	2.00
13032	PD	PDU	59	male	26	23	2	2.00
6576	Control	Control	63	male	28	NA	NA	NA

Table 2.1.: Excluded subjects 8012, 11718, 9553, 13032, 5584 and 6576.

3. Methods

There are a few statistical methods to identify and compare brain-derived values in healthy and diseased states (here Parkinson's disease). The 1st method is the oldest one: choosing *regions of interest (ROI)* and applying basic statistical analysis, such as comparing the cerebral blood flow (CBF) values of two groups with a t-test. The 2nd approach is called *voxelbased* or *voxelwise* analysis [12], [13]. It is a logistic regression for each voxel where the target variable is the group label per subject (PD or control) and the explanatory variables are the CBF value of the voxel, gender and age. This logistic regression is then repeated for every voxel. The logistic regressions of voxels with statistically significant CBF values are then rated as PD-related regions. The fact that typically more than 100,000 voxel could be involved makes this approach difficult - computationally (running more than 100,000 regressions) and statistically, because making statements can be demanding (multiple comparisons problem). The 3rd method is called *scaled subprofile model* (SSM) and is basically an analysis of covariance with a modified principal component analysis (PCA) followed by a classification of subject group using the result of the PCA. This method was employed by Alexander & Moeller (1994) [11] and recently reviewed by Eidelberg (2009) [16]. A 4th approach is generalizing the 3rd approach and apply one of several projection methods instead of PCA. In this report, we applied both independent component analysis (ICA) and principal component analysis (PCA).

These four approaches are computationally intensive because the brain has a few thousand voxels (depending on resolution). Therefore, development of these techniques has advanced with computational improvement.

This chapter describes two approaches to analyzing CBF data and how they can be compared: **principal component analysis** (PCA) and **independent component analysis** (ICA). Both approaches are combined with classification of PD's and controls, and are applied with the subjects as variables.

Figure 3.1 and the following enumeration gives an overview of the model and methods used, with a detailed discussion later.

I. Data preparation Logarithmic transformation and demeaning of the data by subject and voxel.

II. Projection Methods: identify pattern ICA and PCA are applied to extract the mixing values and the components. The matrix containing the mixing values has as many rows as components and as many columns as subjects.

III. Classification: identify Parkinson's disease related components A logistic regression is applied to distinguish PD and controls by the mixing values. Model selection is done with stepwise Akaike's information criterion. The performance of the logistic regression was assessed with a leave-one-out cross-validation.

IV. Build Parkinson's disease related pattern The logistic regression model determined with model selection tells us which variables are necessary to optimally distinguish PD from controls. We can use the linear combination of the components and associated coefficients to build a PD-related pattern.

Some notation: each step from I. to IV. is a *method*, but all steps from I. to IV. together are called *model*.

V. Assessment of model The entire model was assessed with a leave-one-out cross-validation. The variance for each voxel-estimation was determined with bootstrap. Knowing the estimation and the variance for each voxel makes it possible to test the voxel whether it is different from 0.

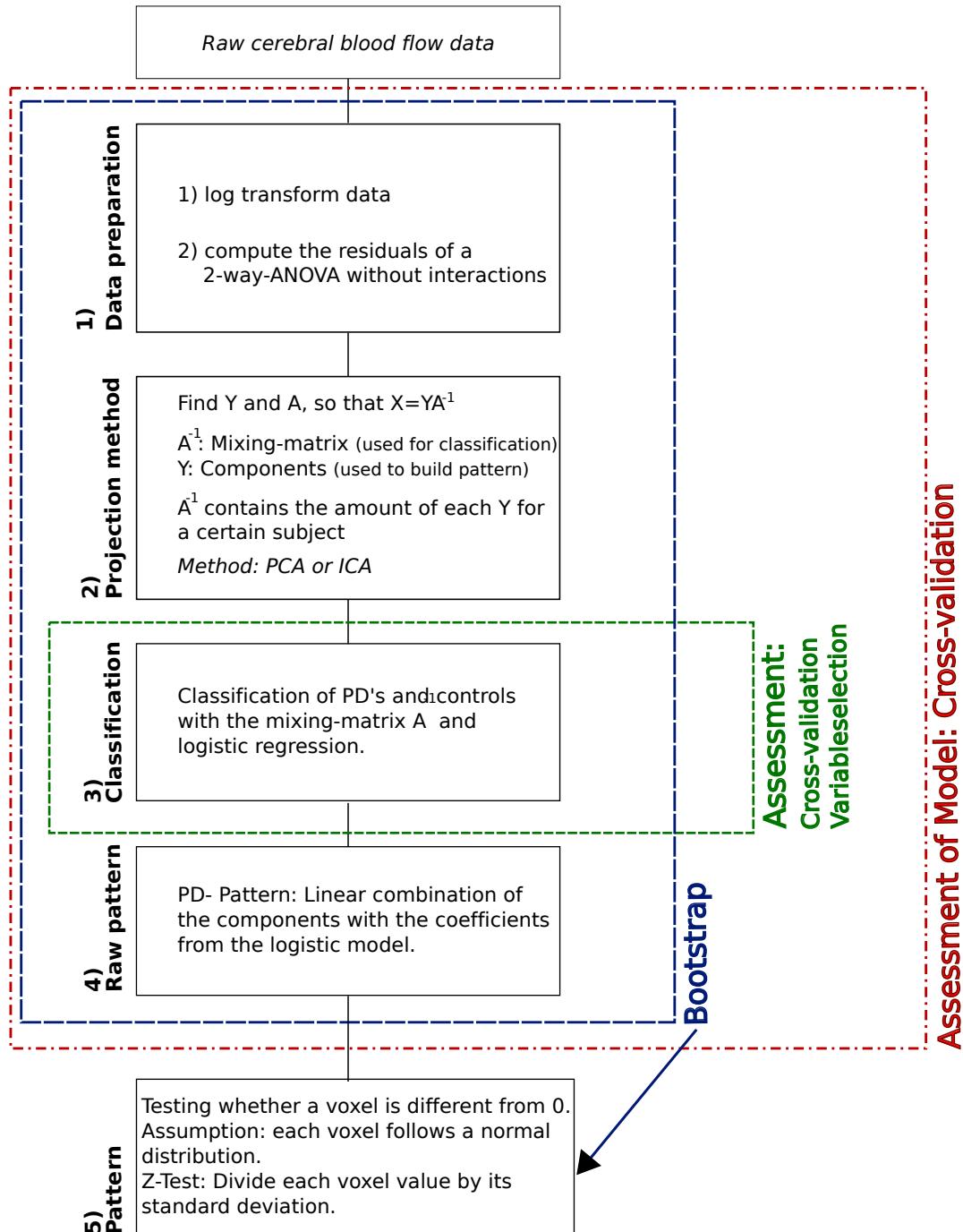


Figure 3.1.: Overview of model.

Literature review

Understandable and clearly written papers about the methodology or application of PCA/ICA are:

- Turk & Pentland [19]: "must-read article" to familiarize oneself with the application of projection methods (in particular PCA) in face recognition or medical imaging.
- Bell & Sejnowski (1995) [21] and Oja & Hyvärinen (2000) [22] provide accessible descriptions of ICA. Bell & Sejnowski (1995) wrote about mutual information and it is needed in source separation and blind deconvolution. Oja & Hyvärinen (2000) give a general overview of ICA with details to the fast fixed-point algorithm, fastICA. Hyvärinen (1999) [23] presents a more general, understandable introduction to ICA, PCA, factor analysis and projection pursuit for the reader who is unfamiliar with projection methods.
- Draper et al. (2003) [15] give an understandable overview about the methodology of PCA and ICA and their comparison.

Projection methods were first applied to imaging data in face recognition (Turk & Pentland (1991) [19]). Medical and face images are similar and eventually Alexander & Moeller (1994) [11] applied PCA in medical imaging data. See Eidelberg (2009) [16] for a recent review of PCA in medical imaging.

An early ICA article by Bell & Sejnowski (1995) [21] proposed the idea of *separating unknown mixtures* and *information maximization*. Later Hyvärinen & Oja (1997) [24] published the fast fixed-point ICA algorithm which is now widely used. Similar ICA approaches are frequently employed in functional MRI analysis (McKeown et al. (2003) [18]).

PCA and ICA have not been compared with ASL MRI data or in PD with any other imaging modality, but McKeown, Sejnowski et al. (1998) [17] compared ICA and PCA in functional MRI. Baek & Draper et al. (2002) [14] compared PCA and ICA on a known and available face image data set (FERET data set). Similar article from Draper & Baek et al. (2003) [15] dealt with different applications of PCA and ICA in face recognition.

Abbreviations

ASL Arterial spin labeling

CBF Cerebral blood flow

ICA Independent component analysis

MRI Magnetic resonance imaging

PET Positron emission tomography

PCA Principal component analysis

PD Parkinson's disease

R Name of the statistic software

SRP Subject residual profile

SSM Scaled subprofile model

Symbols	Dimension	
I	$i = 1 \dots I$	Number of Subjects
J	$j = 1 \dots J$	Number of Voxels
K	$k = 1 \dots K$	Number of Components, $K \leq I$
X	$(J \times I)$	SRP of log(CBF) data
Y	$(J \times K)$	Components
E	$(I \times K)$	Eigenvectors, orthogonal
v	$(1 \times K)$	Eigenvalues
V	$(K \times K)$	Diagonal matrix with eigenvalues
A^{-1}	$(K \times I)$	Mixing Matrix
A	$(I \times K)$	Unmixing Matrix

3.1. Data Preparation

Cerebral blood flow images (CBF) were acquired from each subject in a single 8 minutes scan. The projection method was not implemented directly on raw CBF itself, but with the subject-residual-profile (SRP) of the logarithmic CBF, which are the residuals of the 2-way-ANOVA without interactions between subjects and voxels.

Two possibilities to calculate the SRP matrix:

Possibility 1 With the estimated effects $\hat{\mu}$, \hat{S}_i and \hat{V}_j of an ANOVA model as in equation 3.1. $\hat{\mu}$ is the global mean, \hat{S}_i are the coefficients for each subject i and \hat{V}_j are the coefficients for each voxel j .

$$\log\langle CBF_{ji} \rangle = \mu + S_i + V_j + \varepsilon_{ji} \quad (3.1)$$

The subject-residual-profile is then

$$SRP_{ji} = \widehat{\varepsilon_{ji}} = \log\langle CBF_{ji} \rangle - \hat{\mu} - \hat{S}_i - \hat{V}_j \quad (3.2)$$

Possibility 2 First calculate the averages of the logarithm CBF matrix for each subject i and subtract them from the logarithm CBF matrix:

$$SRP_{ji}^* = \log\langle CBF_{ji} \rangle - \overline{\log\langle CBF_i \rangle} \quad (3.3)$$

Then calculate from the new matrix the averages for each voxel j and subtract them from the new matrix:

$$SRP_{ji} = SRP_{ji}^* - \overline{SRP_j^*} \quad (3.4)$$

If X is mentioned later then SRP_{ji} is used.

3.2. Projection Methods

Projection methods try to transform given variables in such a manner, that the transformation and the transformed variables are of a particular interest. We will have a look at two projection methods: independent component analysis (ICA) and principal component analysis (PCA).

Applying a projection method can have different purposes. *Visualizing high dimensional* data is one of them. It can be difficult to visualize more than two dimensions in one graph, hence reducing multi-dimension data is often required. If the new representation of the data set is appropriate, one might discover patterns or clusters. Second purpose is *feature extraction* or *pattern recognition*, which assume that the observations are a mixture of certain features. These features can be detected with a projection method. Another application for projection methods - which we won't discuss in this report - is *blind source separation*, an important field in time series. For *dimension reduction*, it is important that the least amount of information is lost. "Information" here means the variability within the data set. The goal is therefore to find a corresponding (faithful) representation of the data in terms of minimizing the mean square error. Whereas in *feature extraction* it is more important to find a useful (meaningful) representation [23] of the data. For dimension reduction the standard method is principal component analysis. For feature extraction independent component analysis is recommended, primary because the goal of ICA is to identify "clusters" of related data.

"Projection" methods take their name from "transformation": transforming or projecting original data in a new space. The axes in the new space are referred to as principal or independent components. The data is assumed to be a linear mixture of uncorrelated (PCA) or independent (ICA) components. The mixture amount of each component expressed by each subject is sometimes called the loading or score. If the Parkinson's disease subjects express a different amount of a certain component than the control subjects, this component can be interpreted as being important to PD, that is PD-related.

ICA and PCA have the the following equation as their base:

$$Y = XA \tag{3.5}$$

where X are the observations, Y are the components and A is the unmixing matrix. This means mathematically, that the columns in X are projected into a new space Y with stretch and rotation factors A . The matrix notation in equation 3.5 can be replaced by sums as in equation 3.6, where the components y_{jk} can be calculated as linear¹ combination of the subjects x_{ji} and unmixing matrix a_{ik} . The indices are i (subjects), j (voxels) and k (components). The capital letters I , J and K are the numbers of subjects, voxels and components.

¹Non-linear transformations are also possible.

$$\begin{aligned}
y_{j1} &= a_{11}x_{j1} + a_{21}x_{j2} + \dots + a_{I1}x_{jI} \\
y_{j2} &= a_{12}x_{j1} + a_{22}x_{j2} + \dots + a_{I2}x_{jI} \\
&\vdots \\
y_{jK} &= a_{1K}x_{j1} + a_{2K}x_{j2} + \dots + a_{IK}x_{jI}
\end{aligned} \tag{3.6}$$

X can also be written on the left hand side as in equation 3.7. Then the columns in Y are projected into a new space X with stretch and rotation factors A^{-1} (called the mixing matrix).

$$X = YA^{-1} \tag{3.7}$$

The sums in equation 3.8 represent the matrix equation of 3.7, where the observations of each subjects x_{ji} can be calculated as linear combinations of the component y_{jk} and mixing matrix a_{ki}^{-1} .

$$\begin{aligned}
x_{j1} &= a_{11}^{-1}y_{j1} + a_{21}^{-1}y_{j1} + \dots + a_{K1}^{-1}y_{jK} \\
x_{j2} &= a_{12}^{-1}y_{j1} + a_{22}^{-1}y_{j2} + \dots + a_{K2}^{-1}y_{jK} \\
&\vdots \\
x_{jI} &= a_{1I}^{-1}y_{j1} + a_{2I}^{-1}y_{j2} + \dots + a_{KI}^{-1}y_{jK}
\end{aligned} \tag{3.8}$$

Projection methods can calculate Y and A when X is given. The difference between PCA and ICA is the criterion of how to extract the components Y and the matrix A . In principal component analysis the components represent the maximal amount of variance captured by the variables in X in decreasing order. This can be done by extracting the eigenvectors E of the covariance matrix of X . Y can then be calculated as in formula 3.5 by replacing A through E . Independent component analysis, on the other hand maximizes the statistical independence of the components.

Here is a point to mentioned: because the principal components are orthogonal to each other, PCA produces components which are always uncorrelated but not necessarily independent. Figure 3.2 illustrates that correlation between variables means dependence, but no correlation between variables does not necessarily mean independence. Only under certain circumstances are uncorrelated variables independent. If the variables in X have a normal distribution, and a principal component analysis is applied, then the Y will be uncorrelated and independent from each other.

Independence \Rightarrow Uncorrelatedness
but
 Independence $\not\Rightarrow$ Uncorrelatedness

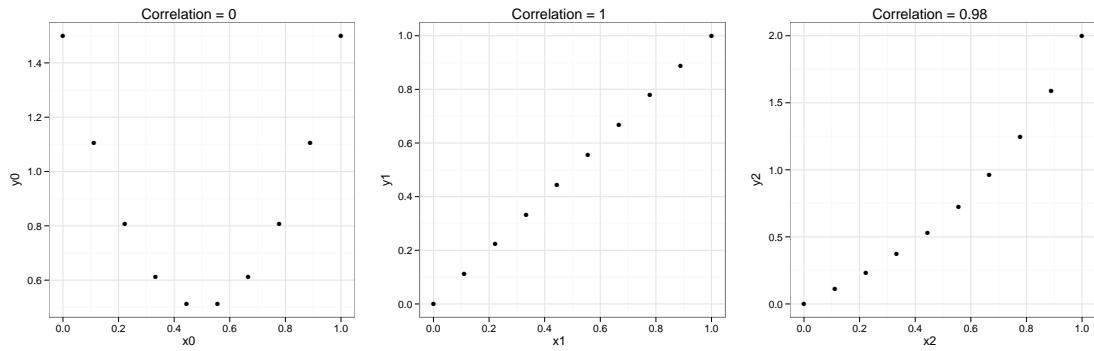


Figure 3.2.: Graph 1 (left) shows a correlation of 0, but there is a clear structure - y_0 is a polynomial of x_0 . Graph 2 (middle) shows a correlation of 1 and a clear linear relation between x_1 and y_2 . Graph 3 (right) represents a high correlation and dependent variables (but the relation is non-linear).

The PCA algorithm has a very strong relation to the normal distribution. If both variables have a normal distribution as in figure 3.3, principal component analysis will find two components which represent the two axes without any loss of information - the blue arrows are just rotated axes. The components are then in this case uncorrelated and independent from each other.

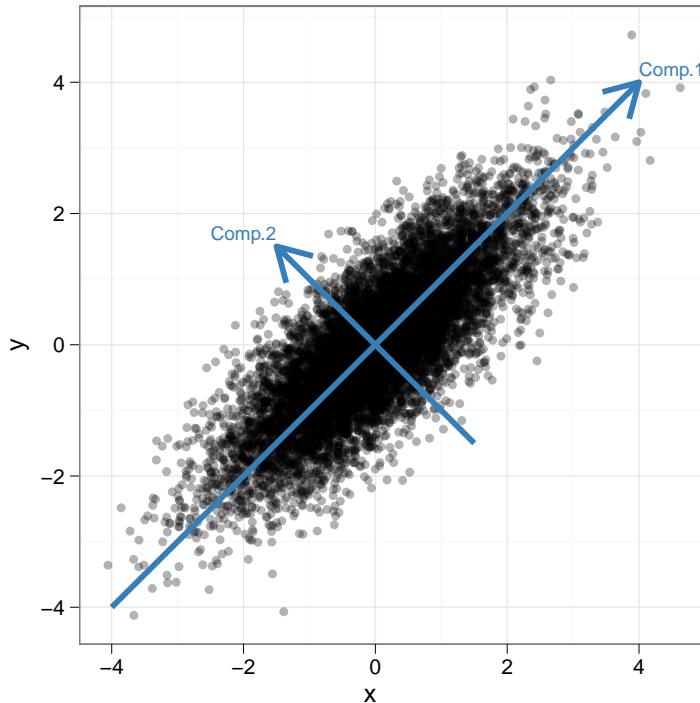


Figure 3.3.: Scatterplot of normal distributed data in x - and y -direction. The two blue arrows represents the two principal components.

PCA and ICA exhibit further differences in the characteristics of Y and A associated with each technique. Whereas the principal components are strictly uncorrelated and orthogonal, the independent components are maximal statistically independent from each other (therefore not orthogonal and not exactly uncorrelated). The mixing matrix A^{-1} is orthogonal in PCA but not in ICA.

In context of projection methods *second-order* and *higher-order* methods are sometimes mentioned. *Second-order methods* are used when having normally distributed variables. If X has an exact normal distribution then its distribution is completely determined by its covariance matrix. If X is not normal distributed, then the covariance matrix is just a faithful representation of the data. To perform principal component analysis *second-order methods* are used [23]. *Higher-order methods* use information of the distribution of X that is not contained in the covariance matrix. This means: the distribution of X must be Nongaussian (not normally distributed). The aim of the higher-order method ICA is to find *interesting* projections, and Gaussian distribution is the least interesting one.

The relation between ICA and PCA can be seen in equation 3.5. Both extract the components while maximizing a measure of *interestingness*. PCA defines the objective functions

with second-order statistics, while ICA uses higher-order statistics, (see Edwards & Oman (2003) [20] for an overview of projection methods applied in R).

3.2.1. Principal Component Analysis PCA

Principal component analysis is mainly used for dimension reduction in visualizing data (in 2D space). Representation by the first two components is only faithful if these components capture as much variability as possible.

Computation

As stated before, PCA maximizes the variance of the data in a set of orthogonal components, arranged by decreasing order of their eigenvalue (explained variance). In the data set displayed in figure 3.4 the line (rotated axis) shows an example of the 1st principal component, capturing as much of the variance in the points as possible. shows an example where the first component has a position to capture as much of the variance of the points as possible. The 2nd component is orthogonal to the 1st component and explains as much of the remaining variance as possible.

This approach could be implemented with an optimization algorithm, but this is not necessary because the unmixing matrix used to calculate the principal component can be computed by calculating the eigenvectors of the covariance matrix of x and y .

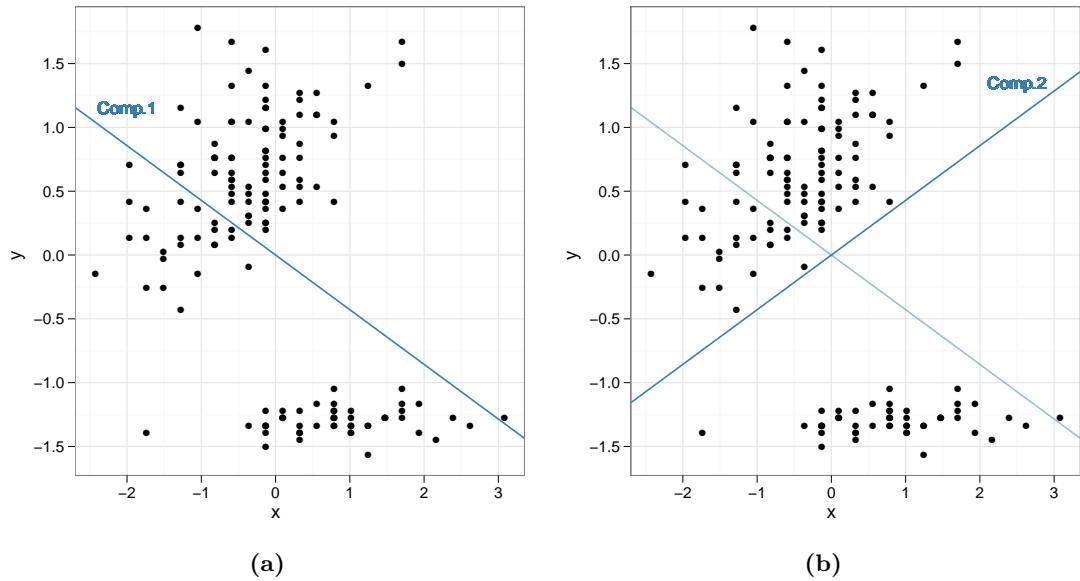


Figure 3.4.: Example with 'iris' data from R: x axis represents the demeaned variable 'sepal width', y axis the demeaned variable 'petal length'. Left figure shows component 1 capturing the maximal variance possible with one dimension. Right figure has component 2 perpendicular on component 1.

For a regular PCA the *recipe* would be to calculate the eigenvectors E and the eigenvalues v of $\text{cov}\langle X \rangle$. The components Y can be calculated as in equation 3.9, the observations X as in equation 3.10.

$$Y = XE \quad (3.9)$$

$$X = YE^{-1} = YE^T \quad (3.10)$$

$$(3.11)$$

In our analysis we used an adjusted PCA, where the principal components are calculated with adjusted eigenvectors; each eigenvector is divided by the square root of its associated eigenvalue. This creates components Y that all have a unit variance (compared to equation 3.9 where the variance corresponds to the eigenvalues).

The *recipe* then would be to calculate the unmixing matrix A as in equation 3.13; the mixing matrix A^{-1} can be calculated as in equation 3.14. The components Y can be calculated with the unmixing matrix as in equation 3.15, the observations X with the mixing matrix as in equation 3.16.

$$W = \text{diag}(1/v) \quad (3.12)$$

$$A = EW^{1/2} \quad (3.13)$$

$$A^{-1} = W^{-1/2}E^T \quad (3.14)$$

$$Y = XA \quad (3.15)$$

$$X = YA^{-1} \quad (3.16)$$

$$(3.17)$$

Characteristics

The components Y are orthogonal; the eigenvectors E are orthogonal, of unit length to a length and not symmetric, therefore $EE^T = I$; the unmixing matrix A is orthogonal and not scaled to length 1 (because of $W^{1/2}$), therefore $AA^T \neq I$, with $AA^{-1} = I$.

Order of Components

An order of the components is dictated by the amount of variance captured by each component, with the first principal component capturing the most variance and each subsequent principal component progressively less. Moreover, the eigenvalue of each eigenvectors represents the variance of the components: $v_k = \text{var}\langle y_{jk} \rangle$.

Because the components are orthogonal the variances of the components can be added.

$$\text{var}\langle x_{ji} \rangle = \sum_k^I \text{var}\langle y_{jk} \rangle$$

The relative proportion of variance is therefore for any chosen K

$$\frac{\sum_k^K \text{var}\langle y_{jk} \rangle}{\sum_k^I \text{var}\langle y_{jk} \rangle} \quad (3.18)$$

Number of Components

To extract the number of components for further investigation one can either choose a) a certain proportion - for example 80% - of the variance explained (see equation 3.18 or figure 3.5a) or b) create a screeplot of eigenvalues (figure 3.5b) and look for an "elbow" in the graph.

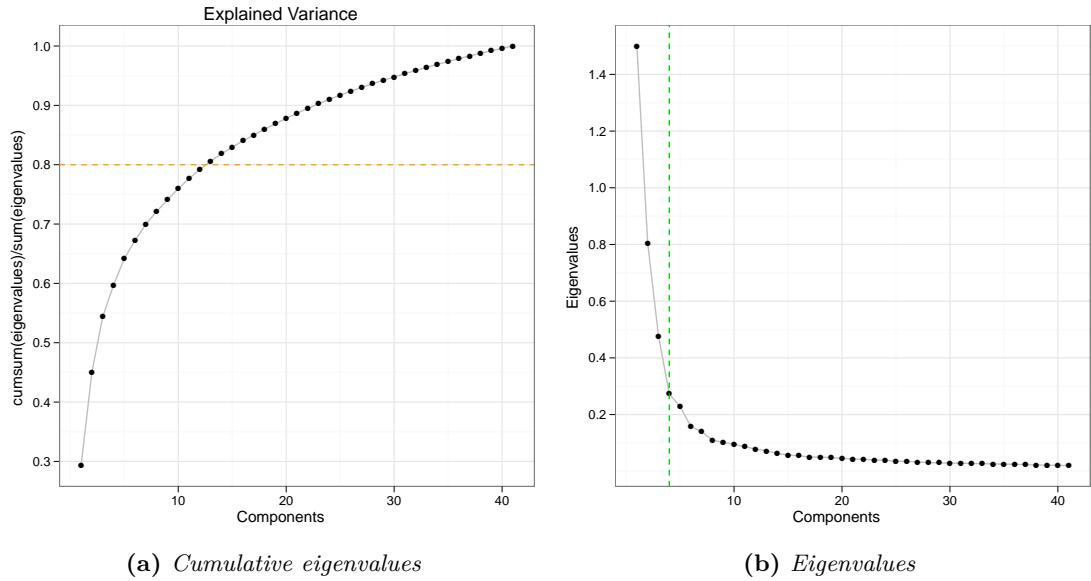


Figure 3.5.: Left figure shows the cumulative proportion of the eigenvalues; 13 components would be necessary to follow the 80% rule. Right figure has on the y axis the eigenvalues: just the first 4th components make sense. After the 4th component the eigenvalues are very small and do not contribute much more variance.

3.2.2. Independent component analysis ICA

The ICA model can be seen as a generative model, where the observations are generated by a process of linearly mixing the independent components Y . ICA is motivated by blind signal source theory and redundancy reduction. Bell & Sejnowski (1995) [21] give a good overview of blind separation, blind deconvolution and applications of ICA. Hyvärinen & Oja (1997) [24], Hyvärinen (1999) [23]² and Hyvärinen & Oja (2000) [22] explain the fastICA algorithm implemented in software.

A central problem of ICA is to define the independence of the components. Independence between two random variable is defined by equation 3.19, where the joint distribution of two random variables is the product of the distributions.

$$E\langle f_1(y_1)f_2(y_2)\rangle = E\langle f_1(y_1)\rangle E\langle f_2(y_2)\rangle \quad (3.19)$$

This traditional measurement of independence can not be used, because in practice the case of strictly independent variables never occurs. Hence, we must find another measurement to characterize independence. The key restriction in ICA is that the independent components must be as nongaussian as possible. If the variables in X are

²Of all Hyvärinen articles, this one seems to be the most understandable

gaussian, as in figure 3.6, then the ICA model can only be estimated up to an orthogonal transformation [22]. The mixing matrix would have the form of an identity matrix (1 on the diagonals) which means, that Y has exactly the same (normal) distribution as X . Although independent components have been found, it is not interesting at all in terms of pattern recognition because nothing *new* could be detected; it is just a rotation. Therefore it does not make any sense to have completely normally distributed variables in X or to achieve completely normally distributed variables in Y . Moreover in theory the estimation of a mixing matrix giving two normal distributed variables in X is not possible, but because in practice a perfect normal distributed variable will never occur, the computation is still possible; the mixing matrix will then contain values on the diagonal close to 1 and values close to zero elsewhere.

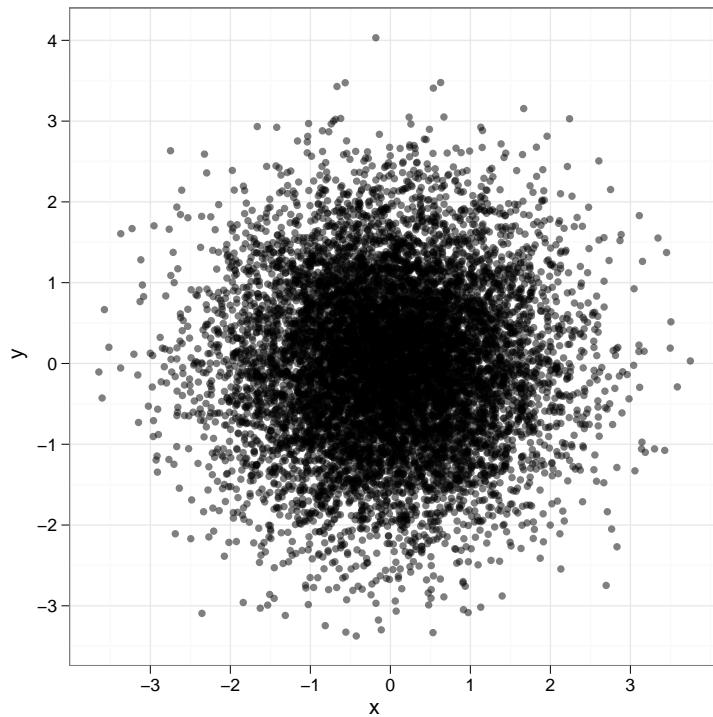


Figure 3.6.: Both x and y are normally distributed.

In a nutshell: If X contains approximately normally distributed variables, the solution of ICA for the mixing matrix will be close to the identity matrix, with independent and gaussian components. To get components which are independent and *interesting* (mixing matrix not equal to the identity matrix), the nongaussianity must be maximized.

Maximizing Independence \Rightarrow Maximizing Nongaussianity

Having established the criterion of maximizing the nongaussianity as a criterion we need a measure to quantify the nongaussianity. There are at least two possible measures of nongaussianity: kurtosis and negentropy[23].

Kurtosis is a normalized version of the 4th moment.

$$kurt\langle y \rangle = E\langle y^4 \rangle - 3(E\langle y^2 \rangle)^2$$

If y is normally distributed, then $E\langle y^4 \rangle = 3(E\langle y^2 \rangle)^2$ and $kurt\langle y \rangle$ is therefore zero. The kurtosis is always non negative and equal to zero for normally distributed variables.

Negentropy is a measurement used to quantify mutual information.

$$J\langle y \rangle = H\langle y_{gauss} \rangle - H\langle y \rangle$$

y_{gauss} is a gaussian variable with the same covariance as y . Negentropy is always non negative; gaussian variables have negentropy equal to zero. Unfortunately negentropy is very difficult to estimate, which leads to an approximation of the negentropy:

$$J\langle y \rangle \approx \{E\langle G(y) \rangle - E\langle G(y_{gauss}) \rangle\}^2 \quad (3.20)$$

where G is (α between 1 and 2)

$$G_{logcosh}(y) = \frac{1}{\alpha} \log \cosh \alpha y$$

or

$$G_{exp}(y) = -\exp\left(\frac{-y^2}{2}\right)$$

Kurtosis can be sensitive to outliers and is therefore not a robust measurement of nongaussianity. For this reason, negentropy is used for most applications.

Computation

There are various algorithms used to perform ICA. They can vary in terms of i) the function to maximize the nongaussianity and ii) the actual optimization. The iterative algorithms can contain *parallel* computations (all components are computed together) or with *successive* computations (one after the other).

In this report we used the **FastICA** algorithm (also called *fast fixed-point* algorithm) of Hyvärinen and Oja [24], [22] and [23] with the logcosh negentropy function $G_{\logcosh}(u)$ and successive computation algorithm. Thus the explanations to ICA below will only contain the algorithms and criteria of fastICA.

FastICA has two preprocessing steps:

- Centering: The mean of the variables are subtracted from the variables. (In our case this is done by using the SRP matrix instead of the raw CBF values.)
- Whitening: The variables in X are transformed to \tilde{X} so that the covariance matrix of \tilde{X} equals the identity matrix (the columns in \tilde{X} have all the same variance 1). This is done with PCA. Then $\tilde{X} = XA$, with $A = EW^{1/2}$ and $W = \text{diag}(1/v)$, where v are the eigenvalues and E the eigenvectors of $\text{cov}(X)$. The projection is made with a chosen number of components. It would make sense to use the number of whitening components explaining 80% of the variance. But this is computationally inefficient in our case. Instead we chose the same number of components as in our PCA analysis. The reason for this whitening step is to exclude *noise* (= white noise) of the data and to make it computationally faster. This step is also important to determine the number of independent components.

Implementation in R

There are a few independent component analysis implementations in R. Some are very specific to certain applications (e.g. spatial or temporal application) and others are very general. The implementation used in this work is the function **fastICA** of the package **fastICA**, containing the algorithm of Hyvärinen and Oja. The advantage of **fastICA** is that it contains a C code implementation which makes it very fast.

Characteristics

The components Y are: 1) maximally statistically independent from each other (this does not mean, that they *have to be* independent from each other!), 2) non-gaussian, 3) the variance of each component in Y is 1 and 4) the means of the components are 0. Neither are the components nor is the mixing matrix orthogonal.

Order of components

Negentropy can be used to order and measure the independence of the components.

- **Negentropy** of each component k as a measure of independence.
 $\widehat{J\langle Y_k \rangle} = \{\overline{G(Y_k)} - \overline{G(y_{gauss})}\}^2$ (y_{gauss} is a normally distributed random variable.)

In PCA, the eigenvalues associated with each eigenvector provide a measure of the variance explained by each component. In ICA, the norm of the mixing matrix A^{-1} provides a corresponding measure of component variance (the variance of the independent component are standardized to 1). Because A^{-1} is not orthogonal in ICA the variances can not be summed up; therefore a relative explained variance as seen in PCA (equation 3.18) can not be calculated (also because just a subset of independent components are calculated).

- **Norm** for each column k in the mixing matrix A^{-1} .
 $\|\kappa_k\|_2 = \sqrt{\sum_{i=1}^I (A_{ki}^{-1})^2}$. If the absolute values in the mixing matrix are high, then $\|\kappa_k\|_2$ is high, which means that its component k is frequently and therefore important.

In this report the order of the components is determined using the norm of the mixing matrix columns.

Number of components

It is suggested to use the following convention: the number of components is determined with PCA.

Goodness of Fit

The euclidean differences between X and \hat{X} was used to assess goodness of fit. Sum of squares ("Residuals") of the differences between the actual observations and values of the observations estimated with all k components or a subset k^* of them is shown below:

For each subject i :

$$\xi_i = \sum_j^J (x_{ji} - \hat{x}_{ji})^2, \quad \forall i$$

For the whole dataset:

$$\xi = \sum_i^I \sum_j^J (x_{ji} - \hat{x}_{ji})^2$$

Final advice

For those who assume that ICA can be applied to every dataset and will magically return a solution, Bell [21] has the following disclaimer (on his webpage³): "Let us remember, Independent Component Analysis (ICA) may not be achievable in general since (1) there may be no independent components, and (2) you might make fatal errors in estimating the component distributions. We only call it ICA because everyone else does."

³<http://cnl.salk.edu/~tony/ica.html>, accessed May 18, 2010.

3.3. Classification

When the dependent variable is categorial with two or more levels, classification methods are used.

If 2 levels Binomial Regression

2 or more levels Discriminant analysis, Multinomial Regression, (Ensemble) Decision Trees, k-nearest neighbour

The model type chosen depends on the purpose of the analysis: either predictive modeling or identifying important variables or both. Binomial and multinomial regression have interpretable coefficients and are used for predictive modeling, whereas discriminant analysis, decision trees and k-nearest neighbour are only used for predictive modeling and minimizing the prediction error.

The selected model should enable accurate prediction of participants and identification of the *importance* of a variable. An optimal choice is binomial regression. Binomial regression is one of several *families* in generalized linear models. Each family has different *link functions* which characterize the regression model. The binomial family has three of them: logit, probit and cloglog. We chose the logit link functions because the coefficients are more understandable and easier to interpret than the coefficients of the other two functions.

A good reference for logistic regression is *Regression Modeling Strategies* [26] from Harrell (2001).

The mixing matrix A^{-1} is used to distinguish PD's and controls. The mixing matrix A^{-1} is transposed in order to set the dimension to $I \times K$, where one row in the transposed mixing matrix corresponds to a single subject. Table 3.1 displays an example of the transposed mixing matrix where each row corresponds to the contribution of each component in a single subject.

	Group	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
1	PD	0.11	-0.01	0.01	-0.07	-0.04	0.05	-0.01
2	PD	-0.11	0.05	0.03	-0.01	0.02	0.04	-0.00
3	PD	-0.10	0.03	0.05	-0.05	0.04	-0.03	0.03
4	Control	0.08	0.01	0.07	0.04	-0.05	-0.07	0.03
5	PD	0.06	0.04	0.04	0.07	-0.18	0.01	0.07
6	Control	0.06	-0.06	0.02	-0.04	0.00	0.00	0.01

Table 3.1.: First 6 rows of an example data set used for logistic regression.

3.3.1. Theory

In logistic regression the dependent variable is treated as 0 and 1, but unlike the ordinary linear regression it is not possible to get values under 0 or over 1. In our case 0 would be *controls* and 1 *PD's*. 0 is sometimes associated with *failure* and 1 with *success*. Because the letters X and Y have already been used in the theory of projection methods, in this section we use other letters: L stands for the usually used X_i ($L_i^{(k)} = (A_{ki}^{-1})^T$), and G_i for the usually used Y .

The expected value of the dependent variable G_i is given in equation 3.21. The values in dependent variable G_i are independent from each other.

The expected value in logistic regression is in equation 3.21, the linear predictor in equation 3.22.

$$E\langle G_i | L_i^{(k)} \rangle = \mu_i = h\langle \eta_i \rangle = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (3.21)$$

$$\eta_i = \beta_0 + \sum_k^K \beta_k L_i^{(k)} = h^{-1}\langle \mu_i \rangle = g\langle \mu_i \rangle = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \log(odds) \quad (3.22)$$

The *odds* are representing the ratio between the probability of success versus the probability of failure. The link function $g\langle \rangle$ is also known as *logit*. It transforms the expected value μ to a linear additive term. Predictions can be made for μ (also called response-prediction) or η (also called link-prediction).

Interpretation of coefficients

The estimated coefficients β_k can be interpreted in terms of *log odds ratio* η : β_k is the change in log odds ratio estimates per unit change in the covariate $L^{(k)}$ ⁴. Unfortunately it is nearly impossible to make a statement related to the predicted probabilities μ , but there is a possibility to use a *nomogram* as seen in [26] (p. 256).

3.3.2. Model assessment & selection

Model selection

Model selection is used to exclude unimportant variables from the model. Usually a large model with K and a small model with K^* parameters are compared, where $K > K^*$ and $K^* \subsetneq K$. The difference between the two models is calculated and assessed. This can be

⁴ $\exp \beta_k$ is the change in odds ratio estimates per unit change in the covariate $L^{(k)}$.

done in *stepwise* order, where models are compared until there is no difference. There are two criteria used to define the difference between two models:

(a) Testing the difference in residual deviance as in equation 3.23 where $D\langle \cdot \rangle$ is the residual deviance for a certain model and $\hat{\phi}$ is the dispersion parameter, which is assumed to be 1 in case of logistic regression. The T value is approximately $\chi^2_{df-df^*}$ distributed ($df = I - K$ are the degrees of freedom for the large model and $df^* = I - K^*$ are the degrees of freedom for the small model). The null hypothesis is that all extra coefficients are zero.

$$T = \frac{D\langle G_i; \beta_{k^*} \rangle - D\langle G_i; \beta_k \rangle}{\hat{\phi}} \quad (3.23)$$

(b) Another possibility uses the *Akaikes information criterion (AIC)* to define the difference between two models. AIC is defined as in equation 3.24, with the "penalty-term" $2K$ to include the number of variables.

$$AIC = -2(\text{maximized log likelihood}) + 2K \quad (3.24)$$

If (b) is stepwise applied, the direction can be 'forward' (start with a model containing few variables), 'backward' (start with a model with a large number of variables) or 'both' (start in the middle, then forward and backward direction). If the smaller model fits the data as well as a bigger model, then the smaller model is preferred. In this analysis, we used stepwise backward AIC.

Residual analysis: Examining the Assumptions

Residual analysis is used to discover observations with a large influence and any other systematic deviations from the *correct* model. The following graphs were used for residual analysis:

- Tukey-Anscombe plot: $\hat{\mu}$ on x-axis, deviance residuals $R_i^{(D)}$ ⁵ on y-axis. If transformations are required to fit the model, than this can be seen in this plot. The expected value of the residuals should be 0.
- Variance plot: $\hat{\mu}$ on x-axis, square root of the absolute deviance residuals $\sqrt{|R_i^{(D)}|}$ on y-axis. The expected value of the y axis should be constant.
- Leverages: diagonal elements of the hat matrix on x-axis, $R_i^{(D)}$ on y-axis. If points have a diagonal element bigger than $2\frac{K}{I}$ and if the value of the residual is high, this could be an observation with a high influence. There are other criteria to discover these observations (e.g. the criterion of Huber).

⁵In generalized linear models are more than just type of residual existing: response residuals, Pearson residuals, working residuals and deviance residuals.

- Half normal plot: extreme values on the top right area.

Cross-validation

Cross-validation is a measure of quality [27] and used to assess the predictive power of the logistic regression model if new data is predicted. In the case of cross-validation, the model is fitted with only a fraction of the data set. The left out data (data not used to fit the model) is then predicted using the fitted model. In general the data set can be divided into H parts, this is called H -fold cross validation. If $H = I$ the cross-validation is also called *leave-one-out*.

The H -fold cross-validation used in this study was as follows:

1. Split the data in H equal-sized parts
2. For the h^{th} part: fit a model to the other $H - 1$ parts of the data, apply stepwise AIC, and predict the h^{th} part of the data with the fitted model. Do this for $h = 1, \dots, H$.

We have chosen $H = I$ because this represents a real world application of our model (having an existing model and predicting an unknown person).

The data set after the cross-validation would look as in table 3.2 and the confusion matrix for a certain cutoff point as in table 3.3. The prediction values are continuous between 0 and 1. To get categorial values (*PD* or *control*), an *threshold* for the prediction has to be determined. We are choosing 0.5, the middle between 0 and 1. The threshold can also be determined by using a receiver operating characteristics (ROC) curve. ROC curves can visualize and compare the performance [25] of a classifiers by taking the threshold into account.

Subject	Group	Prediction
1	PD	0.8
2	PD	0.1
3	PD	0.9
4	Control	0.3
5	PD	0.6
6	Control	0.2

Table 3.2.: First 6 rows of the data set.

Prediction			
		positive	class.error
negative		# false positive	$\frac{\#false\ positive}{\#negative}$
negative	# true negative	# true positive	$\frac{\#false\ negative}{\#positive}$
positive	# false negative		

Table 3.3.: Confusion matrix for 2 groups: positive (PD) and negative (control), ('#' means 'number of').

We are using ROC curves with *false positive rate* on the x-axis and *true positive rate* on the y-axis, see figure 3.7. The area under the curve (AUC) is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [25].

A few informative measures are as follows:

- false positive rate = $\frac{\#false\ positive}{\#positive}$
- true positive rate = $\frac{\#true\ positive}{\#positive}$
- sensitivity = true positive rate = probability of predicting disease given true state is disease
- specificity = 1 - false positive rate = probability of predicting non-disease given true state is non-disease
- accuracy = $\frac{\#true\ positive+\#true\ negative}{\#positive+\#negative}$

In ROC analysis the cutoff point is varied. For every cutoff the *false positive rate* and the *true positive rate* are calculated and plotted. The area under the curve (AUC) is also calculated and can be used to compare the performance between models.

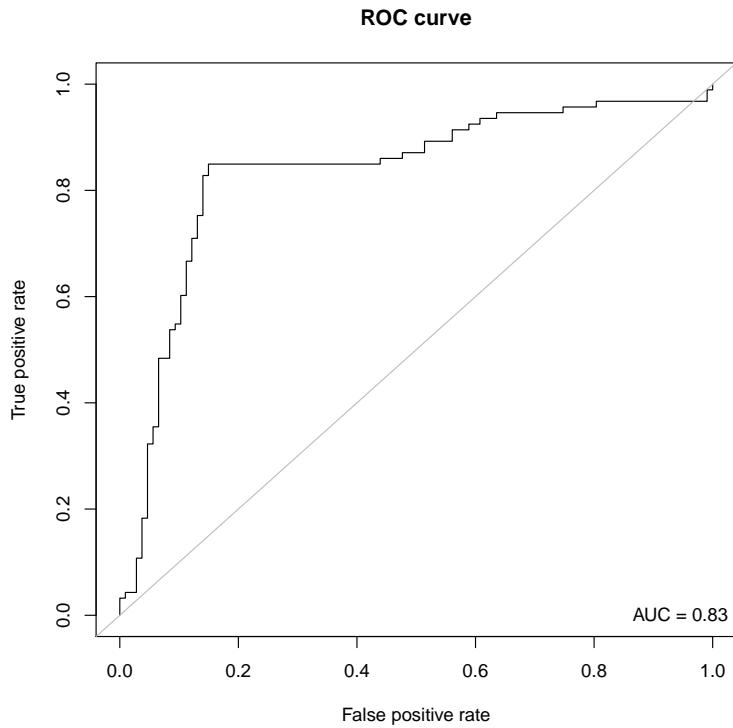


Figure 3.7.: ROC curve: false positive rate on x-axis, true positive rate on y-axis. A black line close to the gray line would indicate a random process as well as an area under the curve lower than 0.5. The point (0,1), left top edge would indicate a perfect classification.

3.4. Building Pattern

The model with k^* parameters determined with stepwise AIC tells us which variables are necessary to optimally distinguish controls from PD. A single PD-pattern constructed as a linear combination of disease-related variables could provide an easily accessible single imaging biomarker of PD. As we have seen in chapter 3.3.1 the meaning of one coefficient β_{k^*} of the regression in equation 3.22 is *the change in log odds that $G_i = 1$ per unit change in $L^{(k)}$* .

If coefficient β_{k^*} ...

- ... is negative, then the $\log(odds)$ decrease with increasing change in variable $L^{(k^*)}$.
- ... is positive, then the $\log(odds)$ increase with increasing change in variable $L^{(k^*)}$.
- ... has a high absolute value, then the change in $\log(odds)$ is stronger than with a small absolute value.

The *sign* of the coefficient indicates whether the high values of a variable are more related to controls or PD and the *absolute value* of the coefficient quantifies the importance of the variable. The linear combination of the components with the coefficients (equation 3.25) therefore leads to an image which is associated with the highest probability of PD be classified as PD. This image is also known as **PD-related pattern**. This pattern is standardized by its mean and standard deviation (equation 3.26) to achieve values which deviate from 0 - as in the SRP matrix which was input for the model.

$$X_{PD} = \beta_0 + \beta_{k^*} Y^{(k^*)} \quad (3.25)$$

$$X_{PD}^Z = \frac{(X_{PD} - \bar{X}_{PD})}{sd(X_{PD})} \quad (3.26)$$

X_{PD} and X_{PD}^Z both have the J elements.

3.5. Assessment of Model

The entire model can be assessed and checked with cross-validation to obtain the predictive power for new subjects. It is also important to assess the variance of each voxel estimation in the PD-related pattern with bootstrap. If we assume that each voxel has a normal distribution, and if we know the variance of each voxel, then the voxel estimation can be tested with null hypothesis that voxel j is equal to zero.

3.5.1. Cross-validation

As mentioned in section 3.3.2 cross-validation is a measure of quality and used to assess the predictive power of a model if new data is predicted. Again we employed leave-one-out cross-validation for the entire model is as follows:

1. Calculate the subject residual profile SRP without subject i
2. Perform a projection method (PCA or ICA) with the SRP matrix as input
3. Train a the logistic regression with output of 2)
4. Predict subject i with the trained classifier

3.5.2. Bootstrap

Bootstrap is a tool for assessing statistical accuracy [27], [26] and is useful for making statistical inference without any knowledge about the distribution of an estimator. In our case the estimation of a voxel is known, but not the variability of the estimation. Because the variability of the estimation can not be calculated theoretically it has to be assessed with simulation, here bootstrapping, outlined below.

Bootstrapping repeats the model fit $g(\cdot)$ B -times with a different data set (X_1^*, \dots, X_I^*) . In each case, a new data set is created by resampling the original data with replacement. A new model is created by i) generating the SRP-matrix, ii) computing the projection method, iii) fitting a logistic regression with stepwise AIC, iv) building pattern by linear combination of coefficients and components and v) standardizing the pattern by subtracting the mean and dividing by its standard deviation. Under the assumption that the data set is a representative sample we obtain an approximative distribution for each voxel j . Knowing the estimation value and the bootstrapped variability for voxel j we can test the null hypothesis that voxel j equals $\mu_0 = 0$. Eidelberg (2009) [16] assumes that each voxel has a nomal distribution. We follow this idea.

Some notation: The observations for each subject are named as X_i . $\hat{\Theta}_j$ is the estimator for $g(X_1, \dots, X_I)$. A * indicates a sample with replacement.

Expected value of the estimator $\hat{\Theta}_j$ is

$$E[\hat{\Theta}_j^*] = B^{(-1)} \sum_{b=1}^B \hat{\Theta}_j^{*,b}$$

Variance of the estimator $\hat{\Theta}_j$ is

$$V[\hat{\Theta}_j^*] = \frac{1}{B-1} \sum_{b=1}^B (\hat{\Theta}_j^{*,b} - E[\hat{\Theta}_j^*])^2$$

Bias of estimator $\hat{\Theta}_j$ is

$$Bias[\hat{\Theta}_j^*] = \hat{\Theta}_j - E[\hat{\Theta}_j^*]$$

Some pseudo code

(1) Sample the data with replacement: X_1^*, \dots, X_I^*

(2) Compute the estimation for each voxel j : $\hat{\Theta}_j^{*,b} = g\langle X^* \rangle$

Estimation involves

- building SRP matrix with bootstrap sample
- applying projection method to SRP matrix
- performing a logistic regression with stepwise AIC to obtain important components
- building PD related pattern and standardizing it (subtracting mean and dividing by standard deviation)

(3) Repeat 1. and 2. B=1000 times

(4) Compute the expected value E and the variance V and the bias of $\hat{\Theta}_j^*$

(5) The z-scored voxel value j is then: $\hat{\Theta}_j^Z = \frac{\hat{\Theta}_j - 0}{\sqrt{V(\hat{\Theta}_j^*)}}$

(6) If the j^{th} element in $\hat{\Theta}_j^Z$ is smaller than $z_{\alpha/2}$ or bigger than $z_{1-\alpha/2}$, then the value of voxel j is statistically different than 0 in either increasing (positive) or decreasing (negative) direction.

3.6. Comparing PCA and ICA

Recent work has compared different projection methods in a variety of contexts. Baek et al. (2002) [14] compared PCA and ICA applying in face recognition with the FERRET data set. FERRET is an established standard data set with over 2,000 face images and often used in face recognition. In this article the classification accuracy was used to compare ICA and PCA. McKeown, Sejnowski, T.J. & et al. (1998) [17] was comparing PCA and ICA introducing a probability of observing the data under the null hypothesis that the ICA assumptions are valid.

The generative model of the independent component analysis assumes that the data is a mixture of different *independent* components whereas principal component analysis assumes that the data is a mixture of components which represents the data the best.

To compare PCA and ICA we measure the **performance** of each entire model by performing a cross-validation and obtain the classification accuracy. What might be also help is to measure the **independence** of the components. With the previous mentioned negentropy we can make an approximate estimation for nongaussianity of the components.

Negentropy

Negentropy $\hat{J}\langle u \rangle$ is a measure for the nongaussianity. ICA maximizes the nongaussianity of components and therefore maximizes the independence. PCA should have lower values for negentropy, but how much independence do we "gain" with independent component analysis? The expected value in formula 3.27⁶ is estimated with the sample mean. If ICA produces components which are clearly more independent than PCA, ICA might be preferred.

$$\hat{J}\langle u \rangle = \{E\langle \log \cosh u \rangle - E\langle \log \cosh y_{gauss} \rangle\}^2 \quad (3.27)$$

Performance of entire model

The performance of the entire model is assessed with cross-validation, a method to measure of how accurate the model estimation is.

⁶This formula is from the function *ica.R.def* in the R-package *fastICA*.

4. Results

The results were produced using the statistical open source software R [30], version 2.11.0 (2010-04-22). For R pseudo code see appendix B.

4.1. Data Preparation

The data was preprocessed (chapter 2) yielding smoother normalized perfusion images. The further data preparation was done as in equation 3.3 and 3.4 described in chapter 3.1. For details and graphs see appendix C.

4.2. Projection Methods

The projections methods are applied to the subject residual profile (SRP) as mentioned in chapter 3.2.

4.2.1. Principal Component Analysis (PCA)

Principal component analysis performed with X (SRP matrix) gives principal components Y (equation 3.15) and mixing matrix A^{-1} (equation 3.14).

Number of Components

Figure 4.1 illustrates two methods used to select the number of components retained for further analysis. To follow the 80%-variance-explained rule (figure 4.1a) 30 components would be necessary. In the scree plot of the eigenvalues (figure 4.1b) the eigenvalues are very small after the 6th (or 5th) component. 30 components would overfit the logistic model, so the first 6 components were chosen.

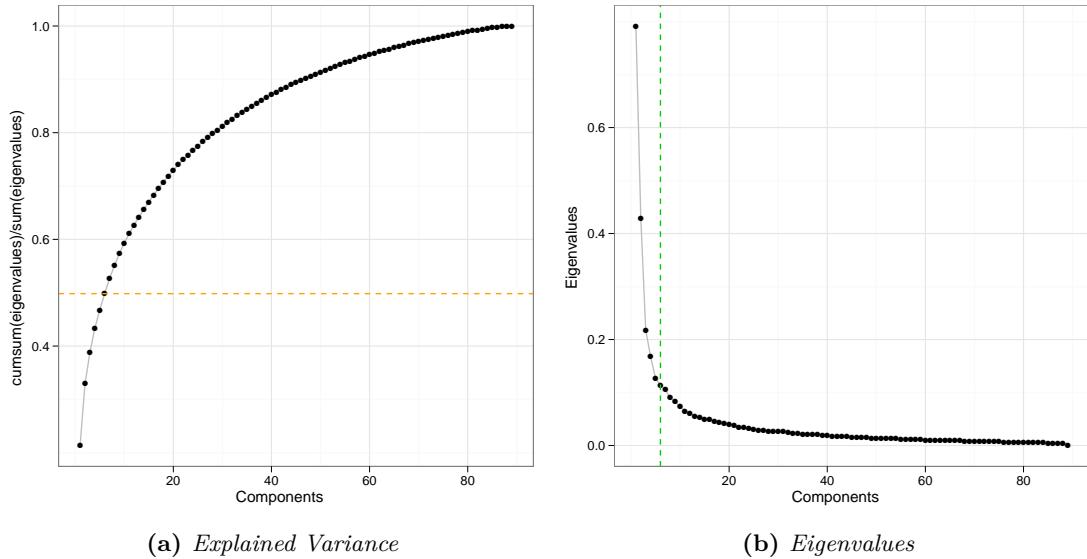


Figure 4.1.: Left, scree plot of eigenvalues versus components. After the 6th component the eigenvalues are very small and therefore contribute very little. **Right,** the cumulative proportion of variance explained by each component; 30 components would be necessary to follow the 80 % rule, with 6 components about 50 % of the cumulative variance is achieved.

More details and graphs to the mixing matrix and the components can be found in appendix C.2.

4.2.2. Independent component analysis (ICA)

Independent component analysis performed with X (SRP matrix) gives - corresponding to formula 3.7 ($X = YA^{-1}$) - independent components Y and the mixing matrix A^{-1} . The number of components is determined with PCA (therefore **6** components, see figure 4.1) and only 6 components are calculated (PCA always calculates as many components as given variables).

Order of Components

Six independent components were calculated. The components were ordered by the norm of the mixing matrix in decreasing order (a corresponding measure to the eigenvalues used in PCA). Table 4.1 and Figure 4.2 show that only component 3 has a high negentropy (measure for nongaussianity; approximation, formula 3.20; estimation, formula 3.27), while the other components have a negentropy close to zero. Therefore an ordering by negentropy is possible but does not make sense.

	Negentropy	Norm of Mixing Matrix
Comp.1	0.0000	0.4422
Comp.2	0.0004	0.4080
Comp.3	0.0012	0.3271
Comp.4	0.0002	0.2359
Comp.5	0.0003	0.2217
Comp.6	0.0001	0.2120

Table 4.1.: Negentropy and norm of the independent components.

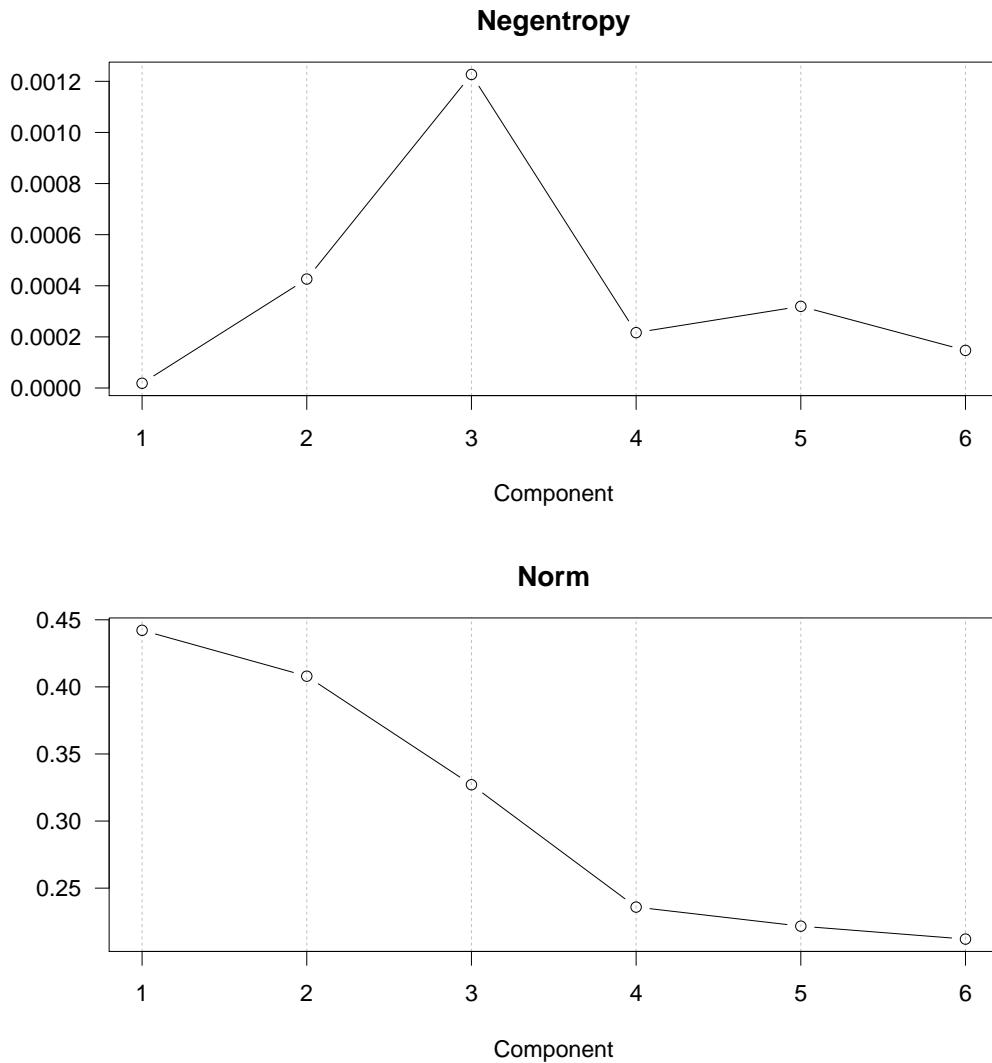


Figure 4.2.: Negentropy (top) for each component and norm (bottom) for each column in the mixing matrix. Component 3 has the highest negentropy and components are ordered in decreasing 'norm-order'.

More details and graphs to the mixing matrix and the components can be found in appendix C.3.

4.3. Classification

Classification of *PD* and *control* was done by a logistic regression as explained in chapter 3.3 with the mixing matrix as covariates. The mixing matrix A^{-1} was transposed so that

every row of $(A^{-1})^T$ was representing a subject.

4.3.1. Principal Component Analysis PCA

Model output for the logistic regression with all 6 variables is shown in table 4.2, the output for the logistic regression after stepwise AIC model selection in table 4.3. Coefficients for all components 1, 4 and 5 have a negative sign, component 5 has the highest absolute value, component 4 the lowest absolute value. Therefore component 5 decreases the *log odds* more than component 4 with one unit change in the variables. The results further down are all corresponding to the logistic regression *after* stepwise AIC model selection.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.383	0.382	3.622	0.000
Comp.1	-13.917	4.443	-3.132	0.002
Comp.2	-0.221	4.138	-0.053	0.957
Comp.3	6.769	6.706	1.009	0.313
Comp.4	-12.952	7.065	-1.833	0.067
Comp.5	-39.609	12.151	-3.260	0.001
Comp.6	-6.260	7.876	-0.795	0.427

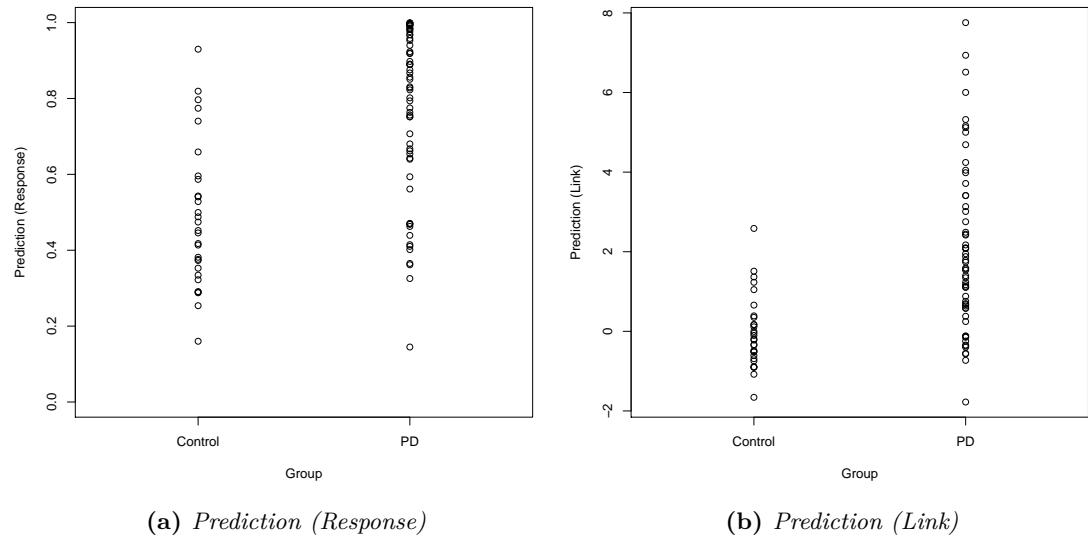
Table 4.2.: Model output for logistic regression.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.347	0.373	3.607	0.000
Comp.1	-13.987	4.389	-3.187	0.001
Comp.4	-11.908	6.650	-1.791	0.073
Comp.5	-38.405	11.851	-3.241	0.001

Table 4.3.: Model output for logistic regression after stepwise AIC model selection.

Prediction

The two figures in 4.3 are both showing the prediction of the logistic regression. Figure 4.3a shows the response (μ , see equation 3.21) and figure 4.3b the *log odds* (η , see equation 3.22). The prediction 0.93 of control is subject 11548, the prediction 0.14 of PD is subject 6417 (lower than any subject of the control group!).

**Figure 4.3.:** Prediction for logistic regression.

Analysis of Residuals

The graphical residual analysis shown in figure 4.4 does not show any violation of the assumptions of a logistic regression. Two rather extreme values - subject 6417 (nbr. 20) and subject 11548 (nbr. 73) have high absolute residuals, table 4.4 shows the data for both observations.

Subject	Cognition	Comp.1	Comp.4	Comp.5	$\hat{\mu}$	$\hat{\eta}$
6417	PDU	0.064	0.029	0.049	-1.775	0.145
11548	Control	-0.113	-0.021	0.016	2.585	0.930

Table 4.4.: Data for subject 6417 and 11548 including the prediction for μ (between 0 and 1) and η (linear predictor).

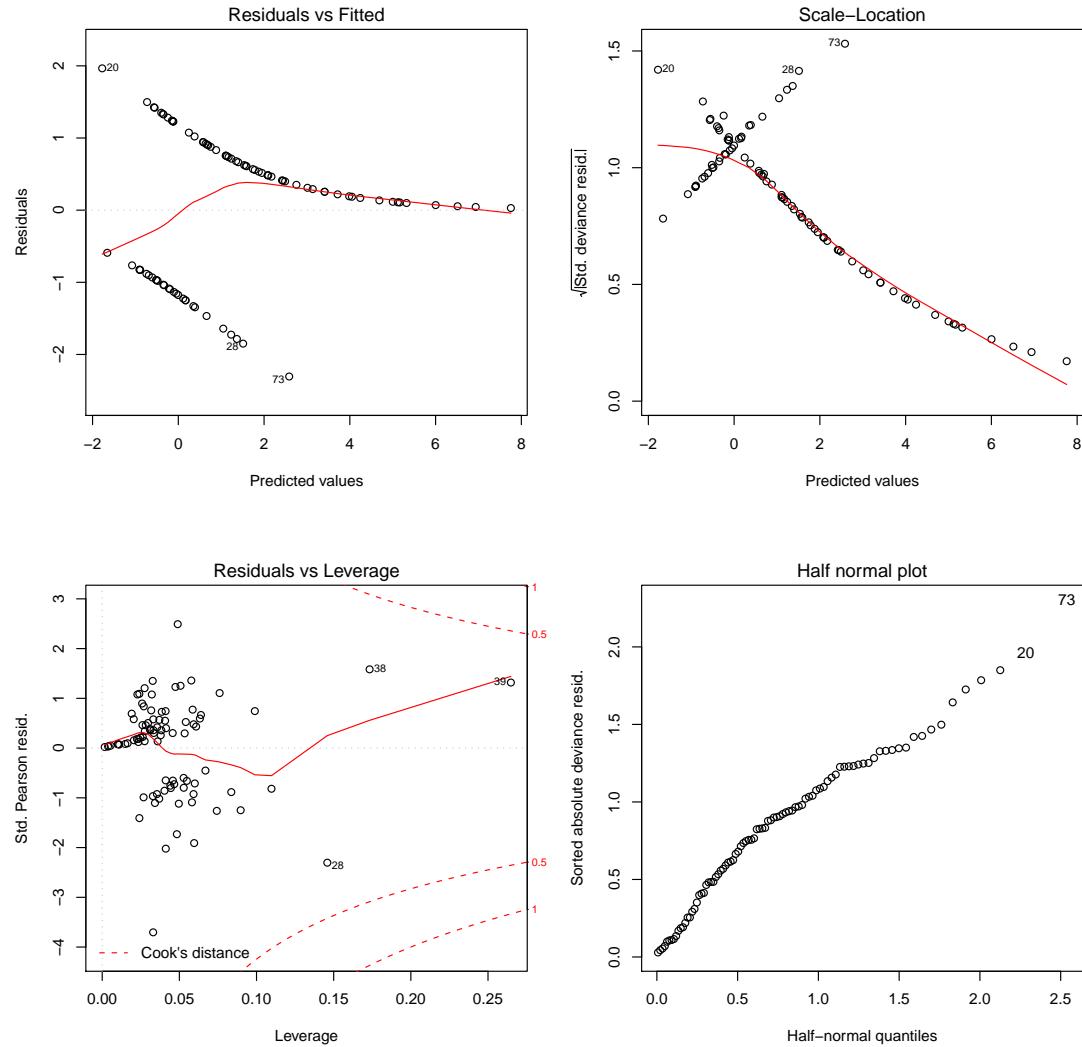


Figure 4.4.: Residual plot for logistic regression. The assumptions seem not to be violated. PD's have positive residuals, controls negative residuals. To emphasize a few interesting things: The Tukey-Anscombe-plot (top left) shows that negative residuals occur only in predictions between (-2) and (3), whereas positive residuals can vary between (-2) and 8. The square root of absolute standardized deviance residuals below 0.5 occur only in PD's (top right). This fact might makes the variance unstable. The leverage plot (bottom left) shows no observation with a large influence. Finally the halfnormal plot (bottom right) shows no fat tail or extreme value.

Mixing matrix and components

The significant variables in the logistic regression are shown in figure 4.5, where each subject represents a line. The shape of the lines of controls is different from the structure

of PDD, but not so much from the structure of the PDU's. Figure 4.6 and 4.7 are representing boxplots for component 1, 4 and 5 where the difference between certain groups - control/PD or control/PDU/MCI/PDD - are visible.

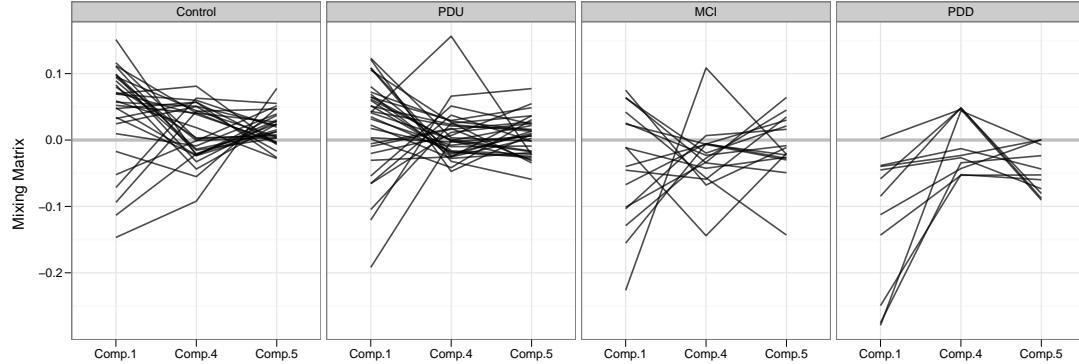


Figure 4.5.: Spaghetti plot of mixing values of component 1, 4 and 5. Each subject is represented by one line and each box represents one cognition level (control, PDU, MCI, PDD). The PDU subject with a high mixing value of component 4 is subject 7681 and the MCI subject with a high mixing value of component 4 is subject 12209.

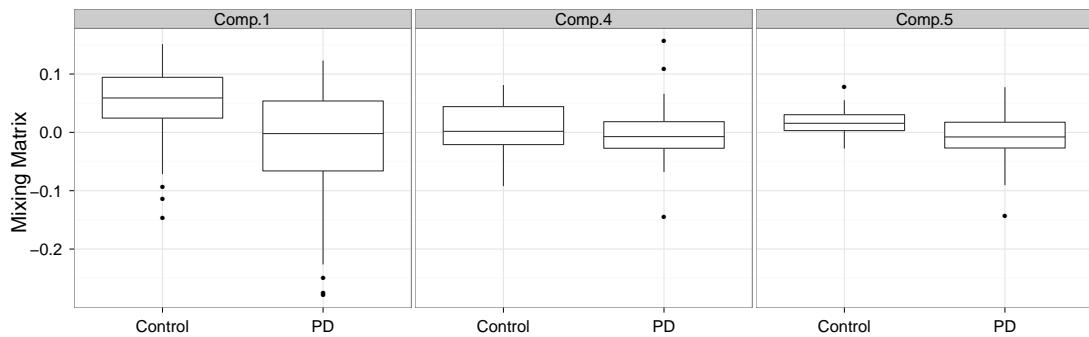


Figure 4.6.: Boxplots for component 1, 4 and 5 for controls and PD's.

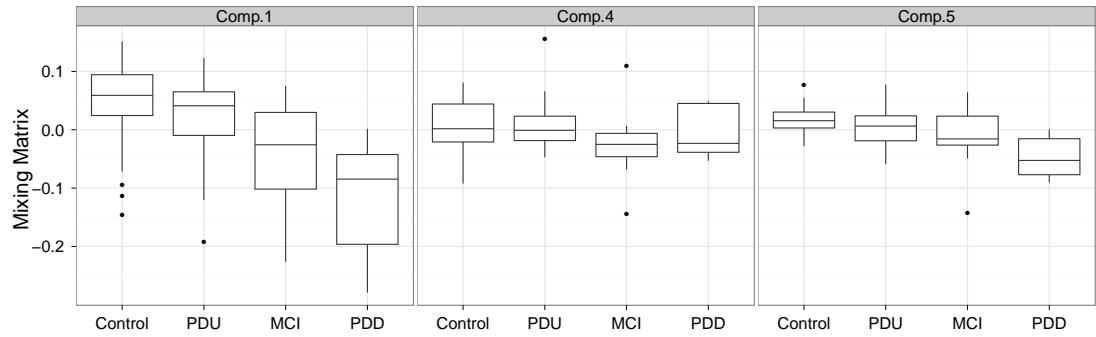


Figure 4.7.: Boxplots for component 1, 4 and 5 for control, PDU, MCI, PDD.

The corresponding principal components are visualized in a pairs plot in figure 4.8 and the image plots in 4.9. The pairs plot shows the correlation between each component and possible clusters.

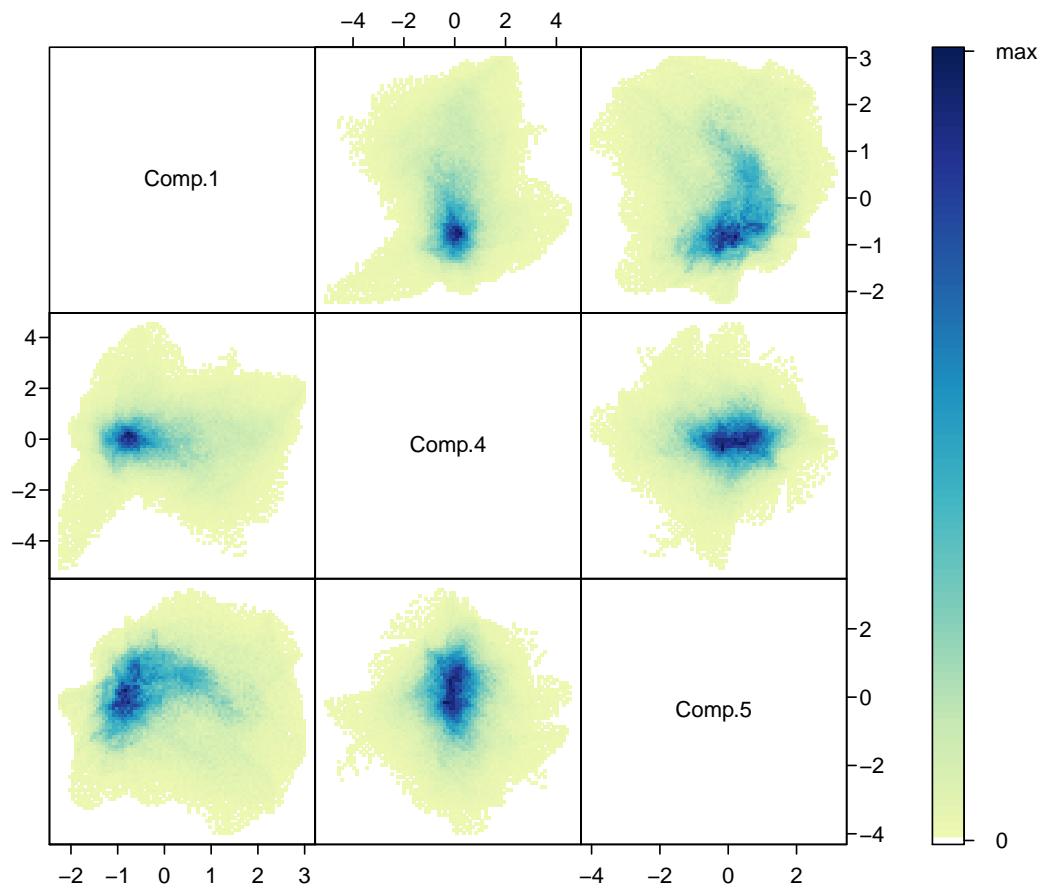


Figure 4.8.: Pairs plot of component 1, 4 and 5. The density goes from white (0) to dark blue (maximum).

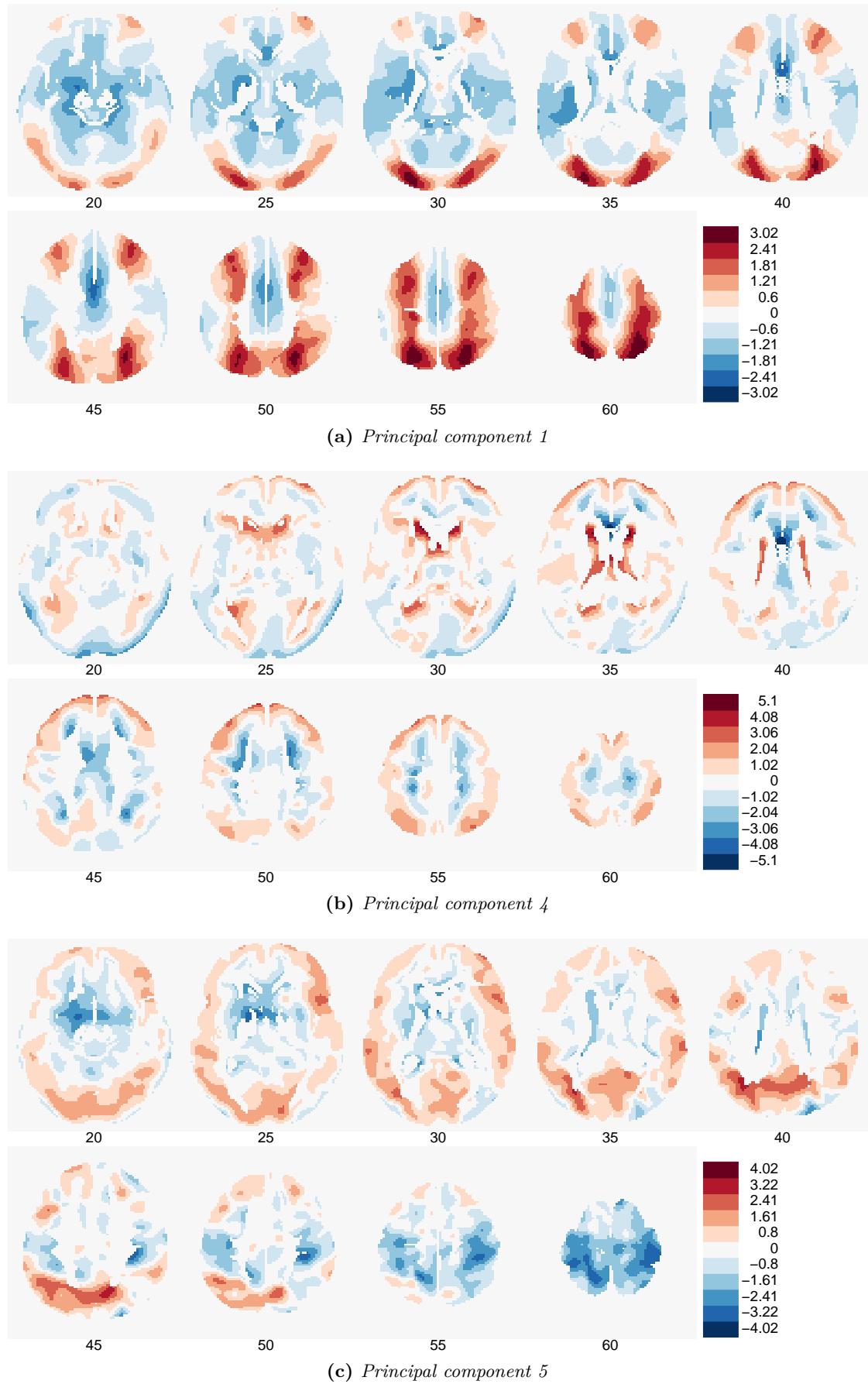


Figure 4.9.: Images of all principal components 1, 4 and 5 - red indicates increased, blue decreased relative cerebral blood flow. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

Cross-validation

The ROC curve of the cross-validation is presented in figure 4.10 (bottom left). Confusion matrix in table 4.5 shows an overall error rate of 0.28. For the confusion matrix in table a cutoff of 0.5 was used. The mosaicplot in figure 4.10 (top left) is explained in Hofmann (2003) [29]. PD's are better classified than controls and unfortunately both groups have a wide range in prediction (0 to 1). The corresponding

	Control (pred)	PD (pred)	class.error
Control	16	13	0.448
PD	12	48	0.200

Table 4.5.: *Performance of logistic regression with PCA (cross-validation with leave-one-out).*

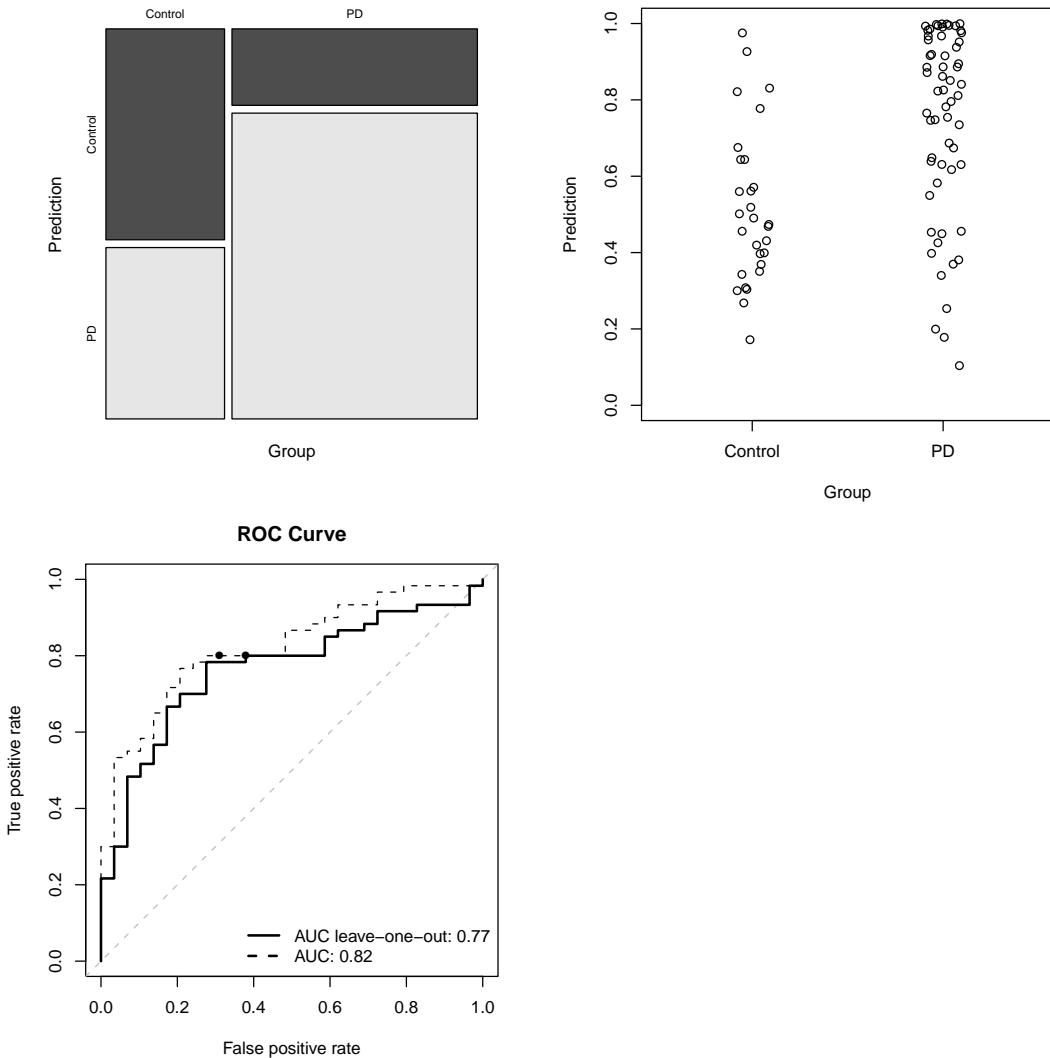


Figure 4.10.: Performance of logistic regression with PCA (cross-validation with leave-one-out). The mosaicplot (top left) shows the classification of PD's and controls with a cutoff of 0.5. The proportion of misclassification in PD's is smaller than in controls. The prediction for PD's and controls (top right) have a wide range between 0 and 1. The ROC curve (bottom left) shows the curve for the leave-one-out cross-validation (—) and for the regular prediction (---). The two points are indicating the cutoff of 0.5.

4.3.2. Independent Component Analysis ICA

Model output for the logistic regression with all 6 variables is shown in table 4.6, the output for the logistic regression after stepwise AIC model selection in table 4.7. Coefficients for component 1 has a positive sign, component 6 a negative sign. The *log odds* changes β_1

with one unit change in variable component 1 and β_5 with one unit change in variable component 5. The results further down are all corresponding to the logistic regression *after* stepwise AIC model selection.

Model Output

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.383	0.382	3.622	0.000
Comp.1	34.073	9.013	3.781	0.000
Comp.2	3.113	4.748	0.656	0.512
Comp.3	-0.693	5.605	-0.124	0.902
Comp.4	-5.882	7.100	-0.828	0.407
Comp.5	-1.344	7.393	-0.182	0.856
Comp.6	-28.422	10.058	-2.826	0.005

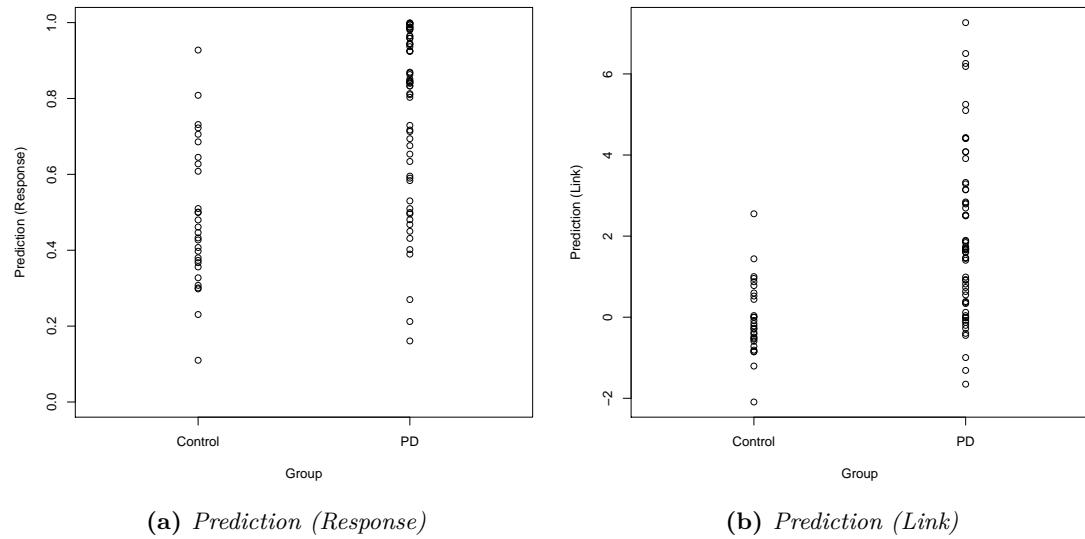
Table 4.6.: Model output for logistic regression (ICA).

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.312	0.362	3.626	0.000
Comp.1	32.605	8.380	3.891	0.000
Comp.6	-25.549	9.223	-2.770	0.006

Table 4.7.: Model output for logistic regression after stepwise AIC model selection (ICA).

Prediction

The two figures in 4.11 are both showing the prediction of the logistic regression. Figure 4.11a shows the response (μ , see equation 3.21) and figure 4.11b the *log odds* (η , see equation 3.22). The prediction of 0.93 for control is subject 11548, the prediction of 0.16 for PD is subject 6417.

**Figure 4.11.:** Prediction for logistic regression (ICA).

Analysis of Residuals

The graphical residual analysis shown in figure 4.12 does not show any violation of the assumptions of logistic regression. Two rather extreme values subject 6417 (nbr. 20) and subject 11548 (nbr. 73) have high absolute residuals, table 4.8 shows the data for the both subjects.

Subject	Cognition	Comp.1	Comp.6	$\hat{\mu}$	$\hat{\eta}$
6417	PDU	-0.085	0.008	-1.651	0.161
11548	Control	0.104	0.084	2.551	0.928

Table 4.8.: Data for subject 6417 and 11548 including the prediction for μ (between 0 and 1) and η (linear predictor).

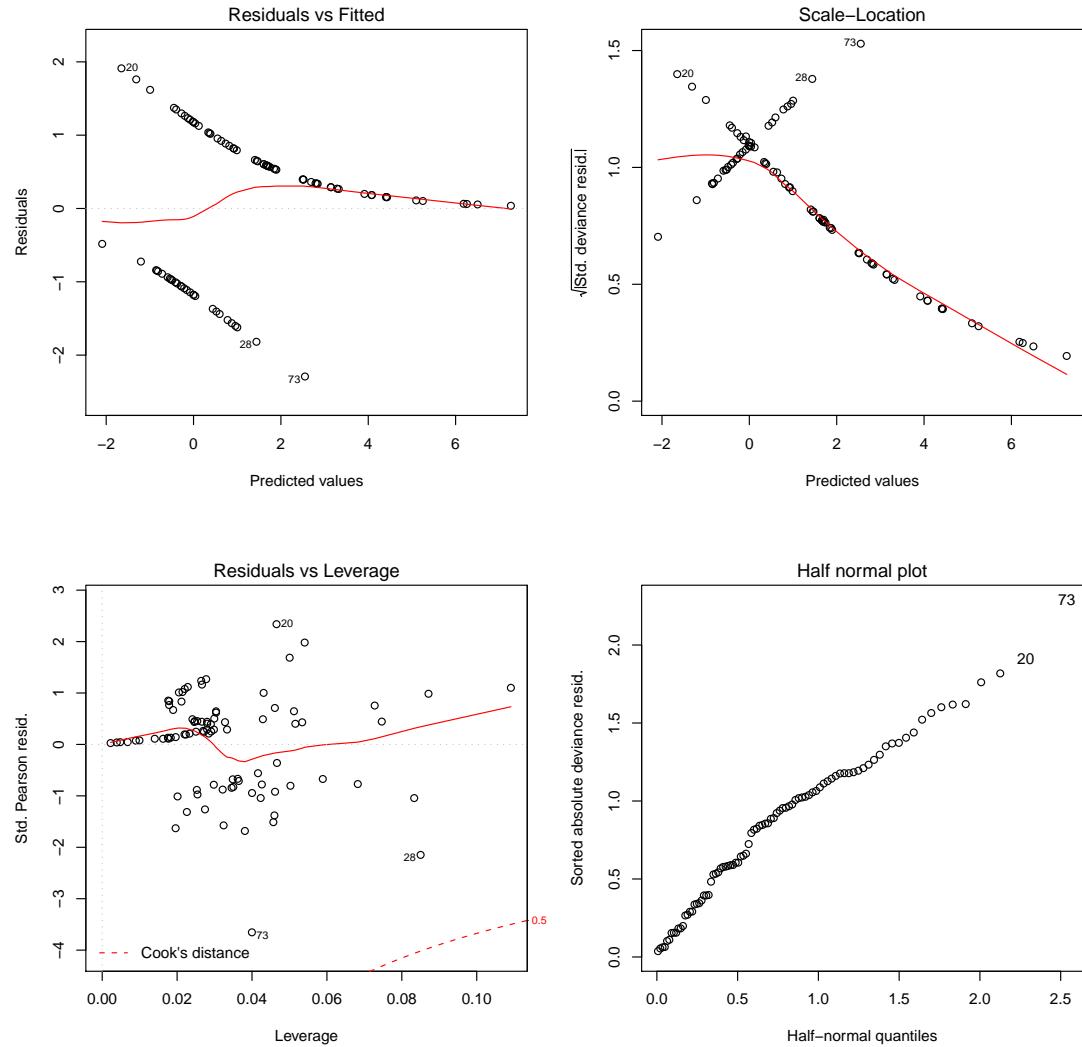


Figure 4.12.: Residual plot for logistic regression (ICA). The assumptions seem not to be violated. PD's have positive residuals, controls negative residuals. To emphasize a few interesting things: The Tukey-Anscombe-plot (top left) shows that negative residuals occur only in predictions between (-2) and (3), whereas positive residuals can vary between (-2) and 7. Also the square root of absolute standardized deviance residuals below 0.5 occur only in PD's (top right). This fact might makes the variance unstable. The leverage plot (bottom left) shows no observation with a large influence. Finally the halfnormal plot (bottom right) shows no fat tail or extreme value.

Mixing Matrix and Components

The significant variables in the logistic regression are shown in figure 4.13, where each subject represents a line. The slope of the lines of controls is rather positive, for PDU's

positive or negative, for MCI's rather negative and for PDD's always negative with high values in mixing values of component 1 and lower values in component 6. Figure 4.15 and 4.16 are representing boxplots for component 1 and 6 where the difference between certain groups - control/PD or control/PDU/MCI/PDD - are visible. Because only 2 variables are significant a two dimensional scatterplot in figure 4.14 will help to understand the relation between component 1 and 6. One can also see a ribbon for controls, then a little bit more right, a ribbon for PDU's, then MCI's and PDD's.

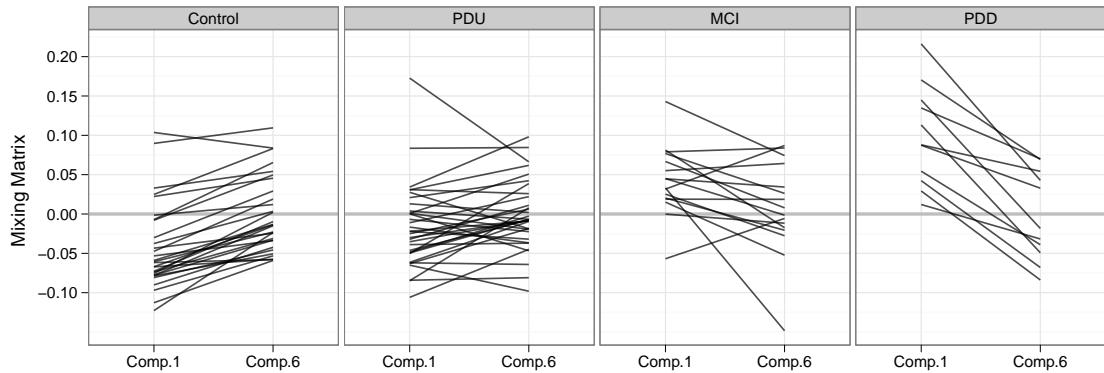


Figure 4.13.: Spaghetti plot of variables discovered with logistic regression. Each subject is represented by one line and each box represents one cognition level (control, PDU, MCI, PDD). The PDU subject with a high mixing value of component 1 is subject 11365, the MCI subject with a low mixing value of component 6 is subject 7108.

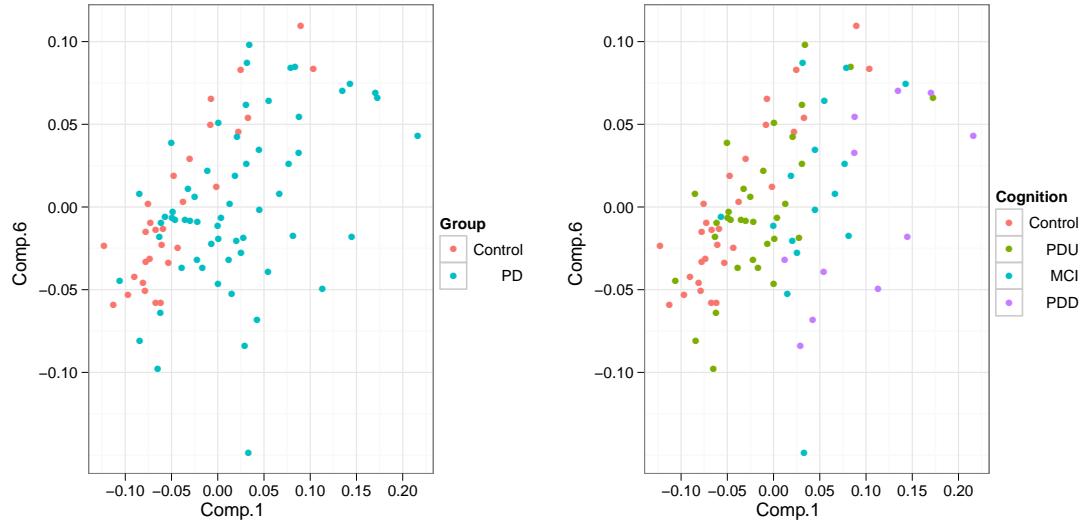


Figure 4.14.: Scatterplot for mixing values of component 1 and 6, colors are group levels (left) and cognition levels (right).

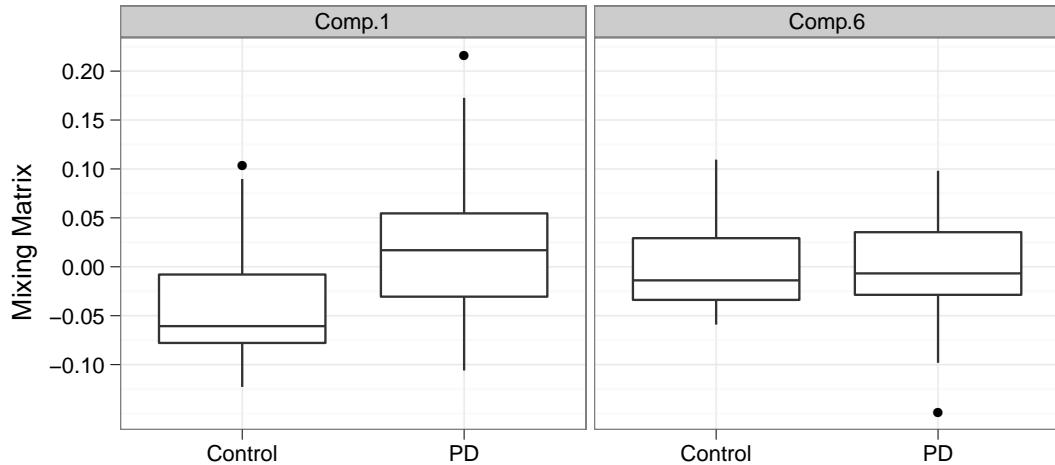


Figure 4.15.: Boxplots for component 1 and 6 for controls and PD's.

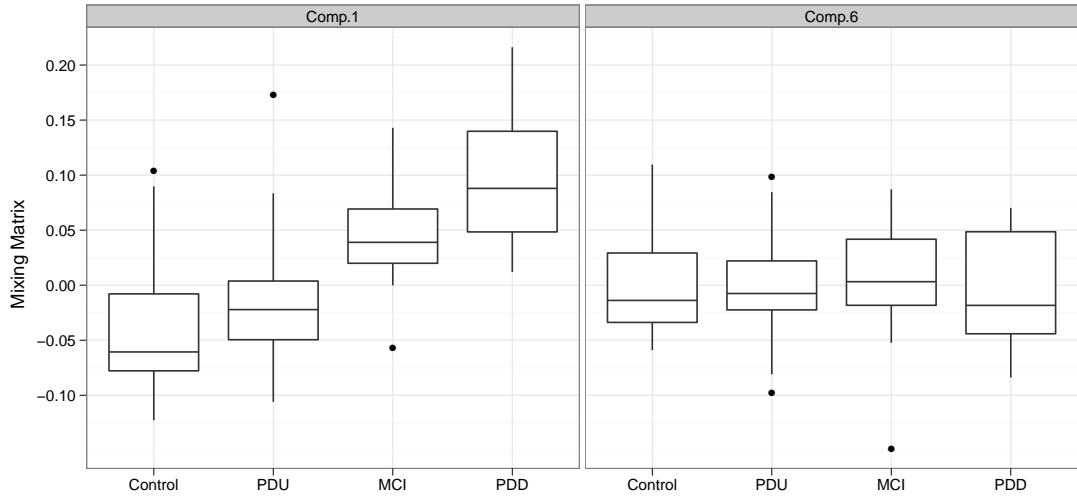


Figure 4.16.: Boxplots for component 1 and 6 for control, PDU, MCI, PDD.

The corresponding independent components are visualized in a pairs plot in figure 4.17 and the image plots in 4.18. The pairs plot shows the correlation between each component and possible clusters.

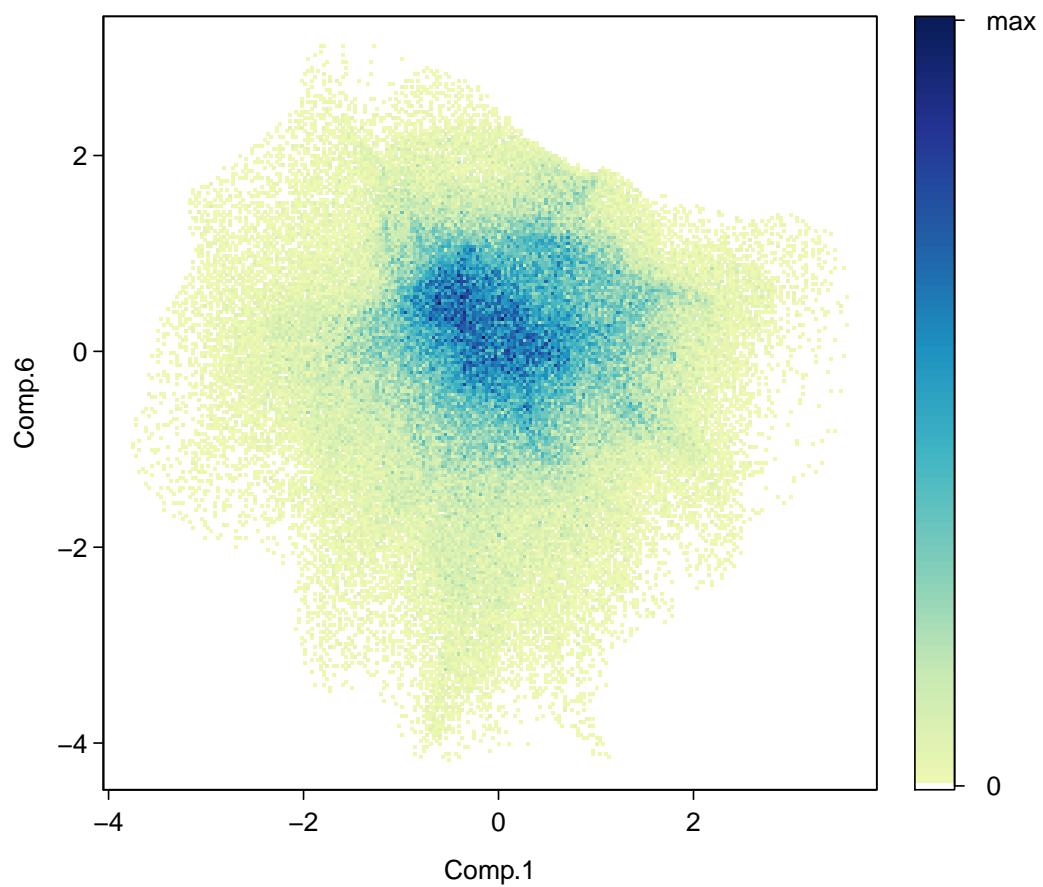


Figure 4.17.: Scatterplot of component 1 and 6. The density goes from white (0) to dark blue (maximum).

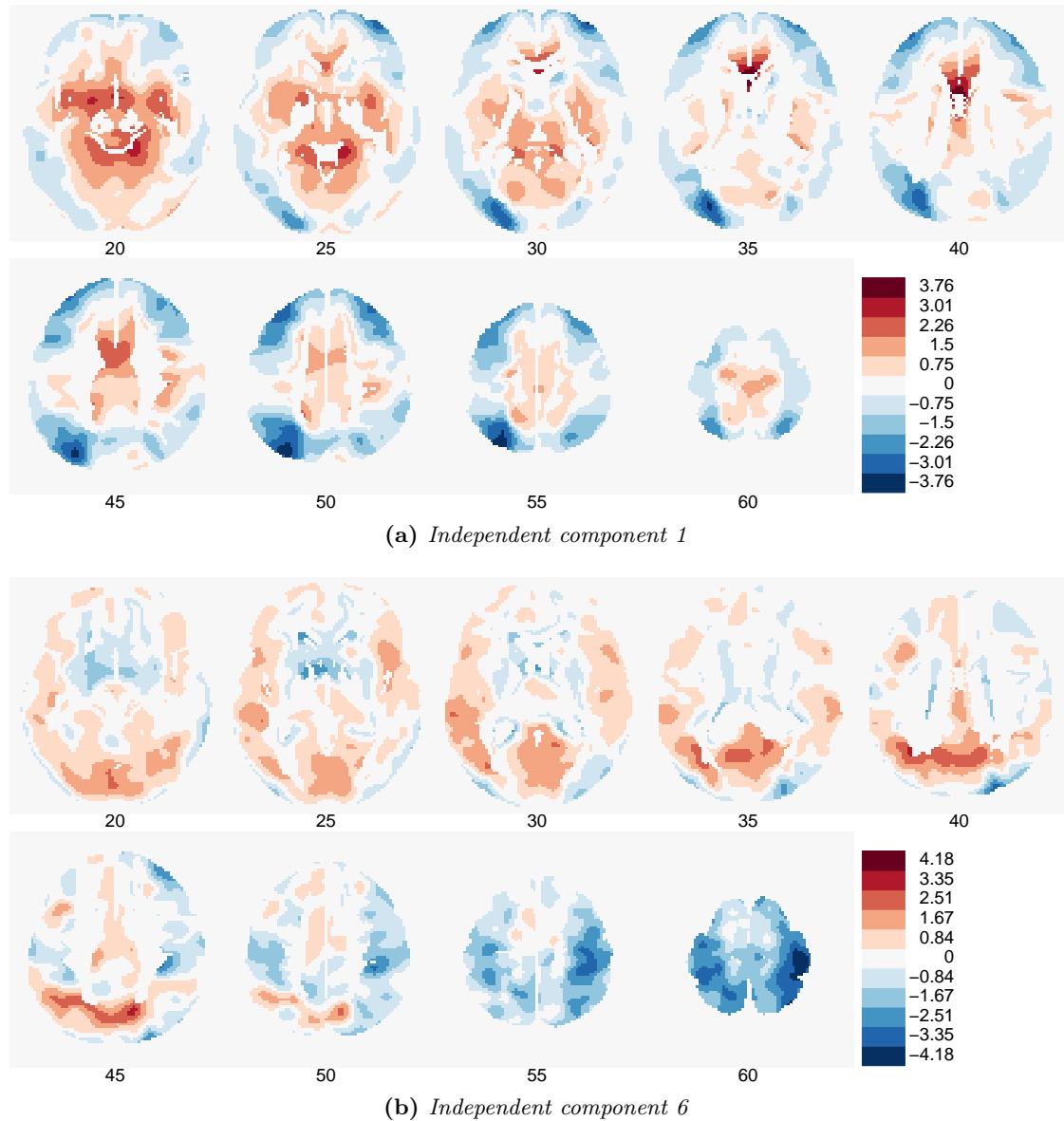


Figure 4.18.: Images of independent components 1 and 6 - red indicates increased, blue decreased relative cerebral blood flow. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

Cross-validation

The ROC curve of the cross-validation is presented in figure 4.19 (bottom left). The confusion matrix in table 4.9 shows a relative error rate of 0.28, while a cutoff of 0.5 was used. The mosaicplot in figure 4.19 (top left) shows that PD's are better classified as

controls. Plot on the top right shows the prediction for PD's and controls with a range from 0 to 1.

	Control (pred)	PD (pred)	class.error
Control	17	12	0.414
PD	13	47	0.217

Table 4.9.: *Performance of logistic regression with ICA (cross-validation with leave-one-out).*

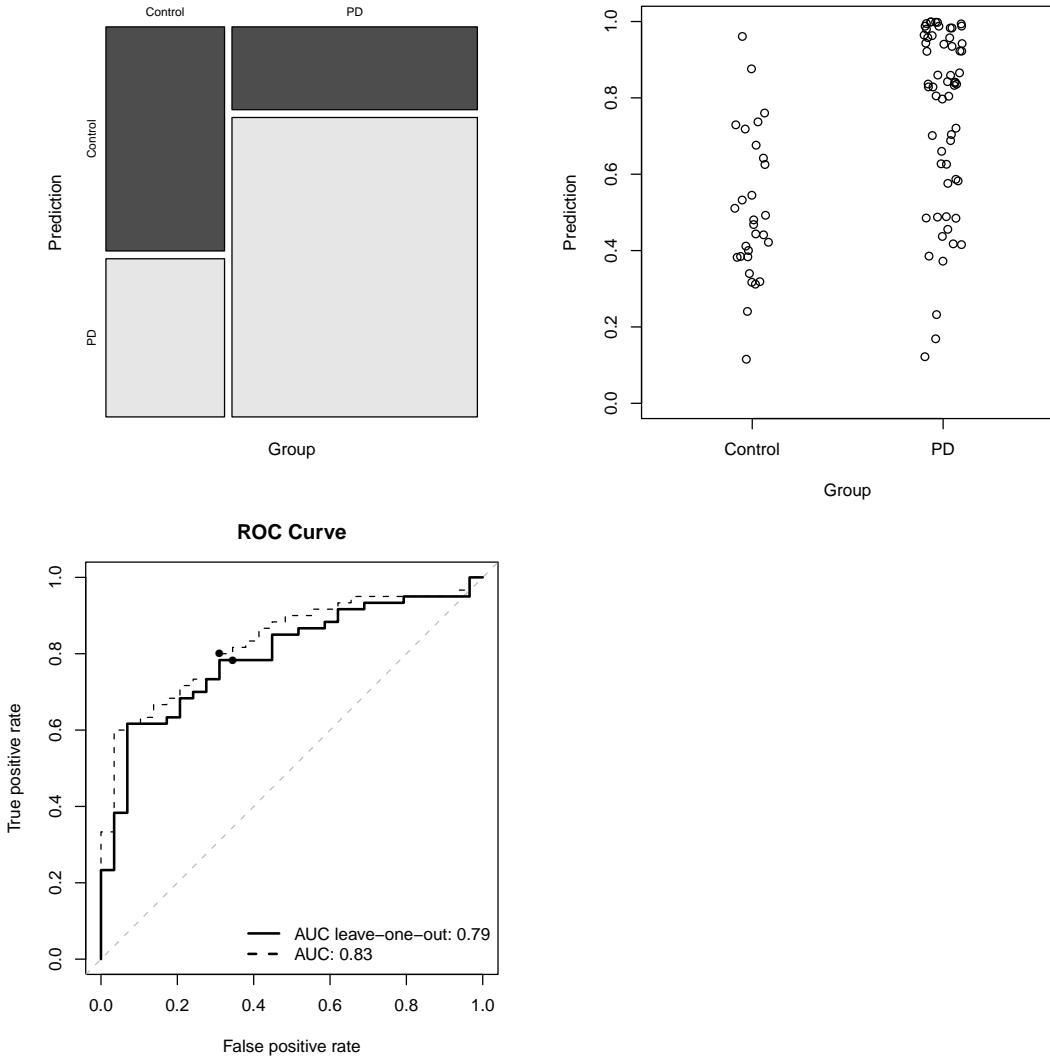


Figure 4.19.: Performance of logistic regression with PCA (cross-validation with leave-one-out). Top left is a mosaicplot which shows the classification of PD's and controls with a cutoff of 0.5. The proportion of misclassification in PD's is smaller than in controls. Top right shows the prediction for PD's and controls with a wide range from 0 to 1. The plot on bottom left is a ROC curve for the leave-one-out cross-validation (—) and for the regular prediction (---). The two points are indicating the cutoff of 0.5.

4.4. Building Pattern

The pattern are built as mentioned in chapter 3.4 (linear combination of the regression coefficients and the components). Figure 4.20 shows a density scatterplot of the PD-

related pattern built with ICA and PCA. Both pattern have similar voxel values. Note that in this is the 'raw' pattern without consideration of the voxels variance. Thus a statement about 'extreme low' oder 'extreme high' areas is not possible (the difference of zero can not be tested).

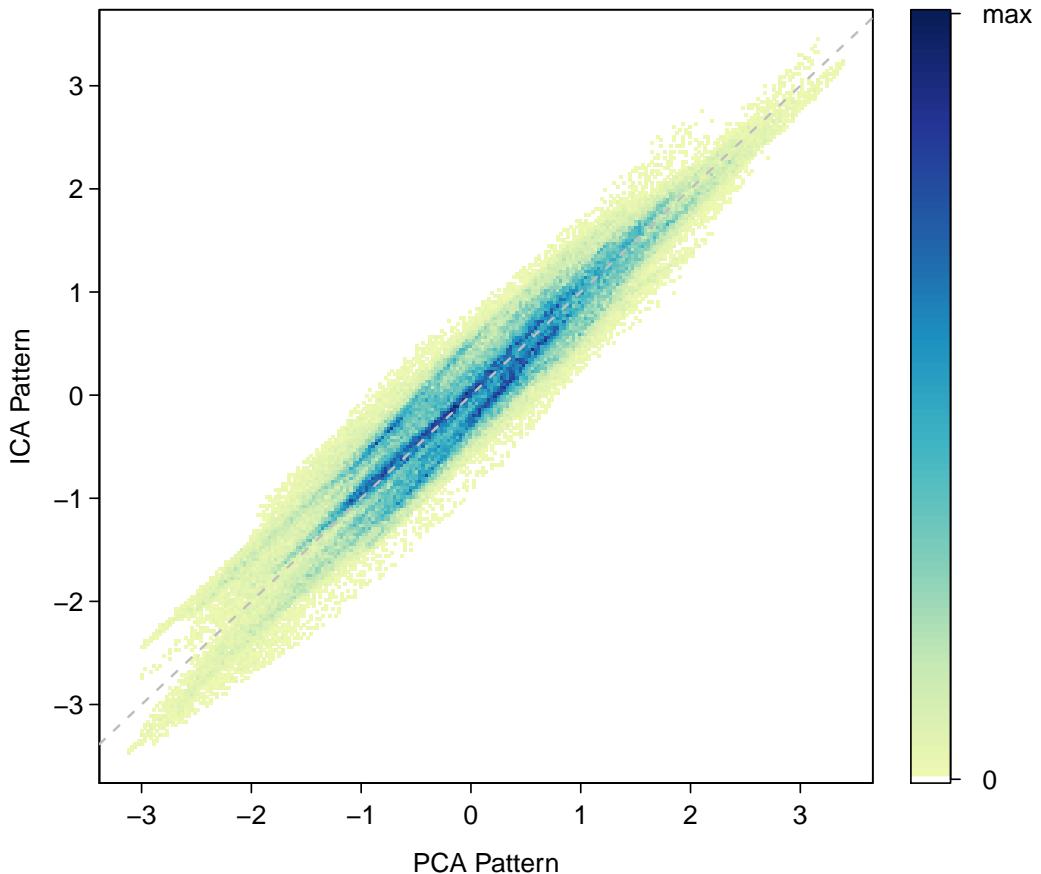


Figure 4.20.: Scatter-density-plot for PCA (x-axis) and ICA (y-axis) pattern. This plot shows that both pattern seems to be almost identical.

4.4.1. PCA

Figure 4.21 shows the histogram of the PD-related pattern and figure 4.22 the corresponding image in two dimensional axial view.

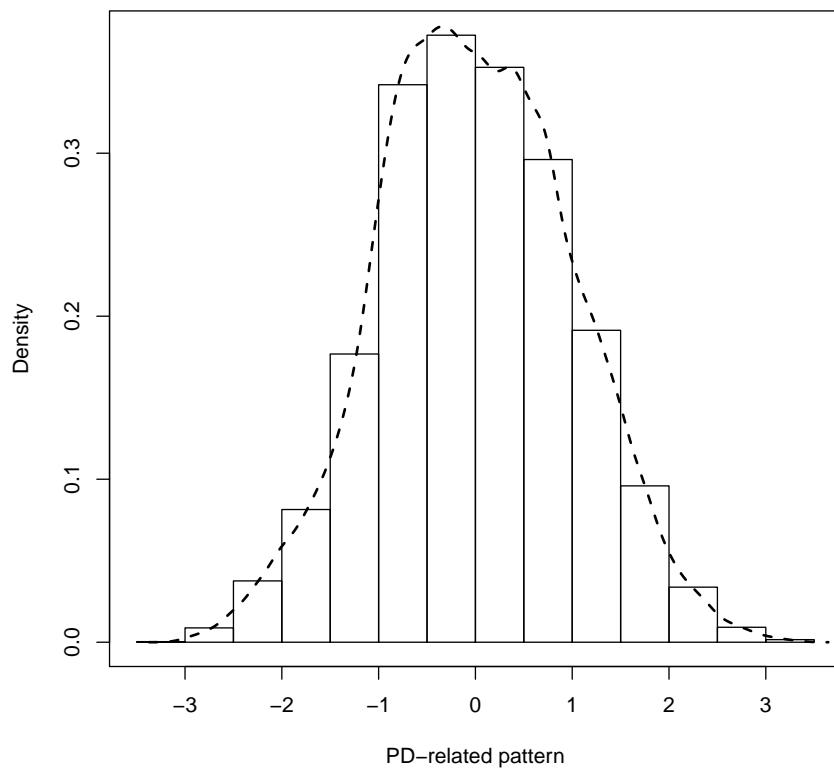


Figure 4.21.: Histogram for PD-related pattern (PCA).

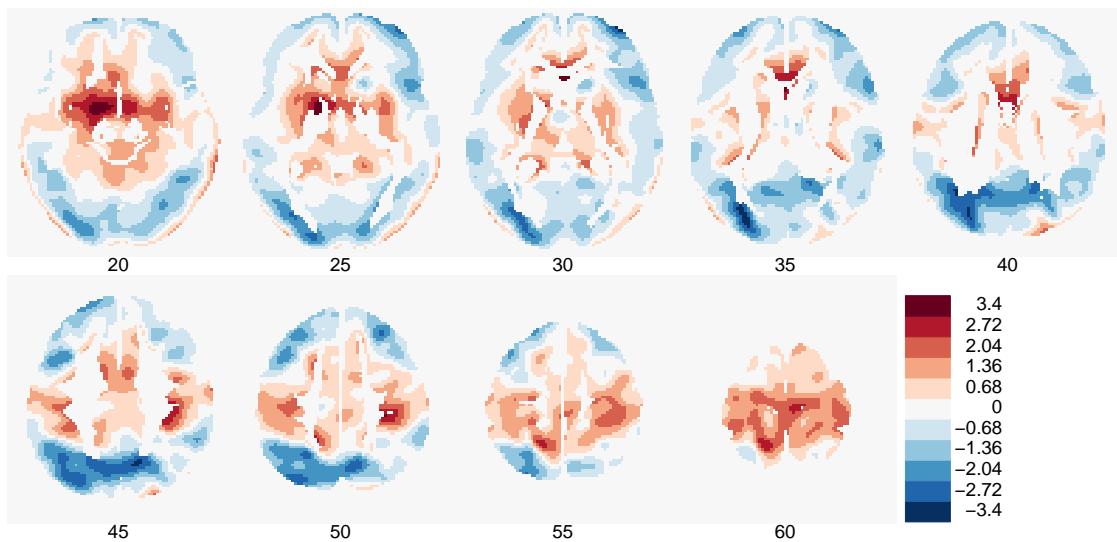


Figure 4.22.: Imageplot for PD-related pattern (PCA) - blue indicates relative increased and red relative decreased cerebral blood flow values. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

4.4.2. Independent component analysis (ICA)

Figure 4.23 shows the histogram of the PD-related pattern and figure 4.24 the corresponding image in two dimensional axial view..

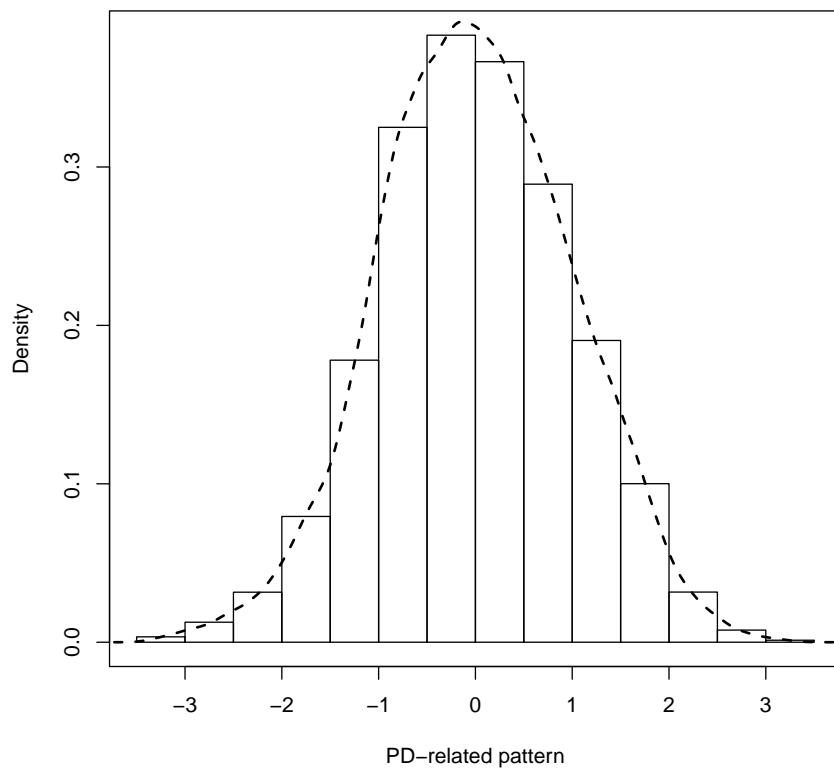


Figure 4.23.: Histogram for PD-related pattern (ICA).

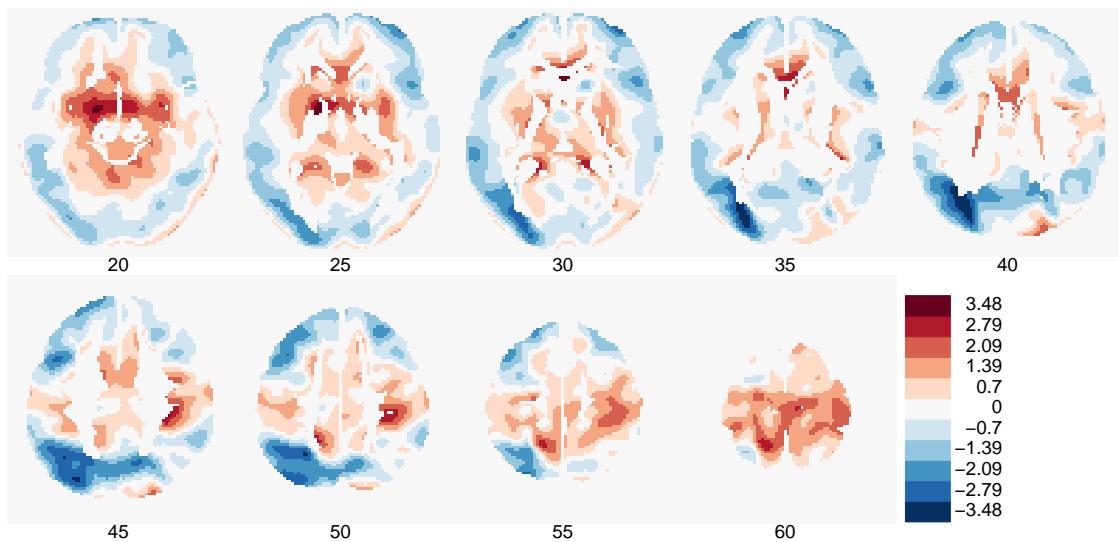


Figure 4.24.: Imageplot for PD related pattern (ICA) - blue indicates relative increased and red relative decreased cerebral blood flow values. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

4.5. Assessment of Model

The method can be assessed and checked with cross-validation to see how the method is performing if a new subject has to be classified by the entire model. It is also important to include the variation (=precision) of each voxel estimation for testing purposes.

4.5.1. Cross-validation of Model

Cross-validation of the entire model with leave-one-out was done as shown in section 3.5.1.

Principal component analysis (PCA)

PCA showed an overall error rate of **0.31** (cutoff=0.5). 45% of controls and 80% of PD's were correct classified (table 4.10). There was no misclassification for PDD's and only 1 misclassification for MCI's (table 4.11). ROC curve and predictions are in figure 4.25.

	Control (pred)	PD (pred)	class.error
Control	13	16	0.552
PD	12	48	0.200

Table 4.10.: Performance of whole model with PCA (cross-validation with leave-one-out).

	Control (pred)	PD (pred)
Control	13	16
PDU	11	22
MCI	1	15
PDD	0	11

Table 4.11.: Performance of whole model with PCA for cognition groups (cross-validation with leave-one-out).

Misclassified subjects:

Parkinson's disease subjects 5801, 6257, 6417, 7595, 7681, 8374, 8860, 12210, 12391, 12892, 13196, 13198.

Control subjects 5486, 5703, 5704, 5802, 6766, 6770, 6849, 7107, 7683, 8618, 8858, 8862, 8948, 9117, 11548, 11890.

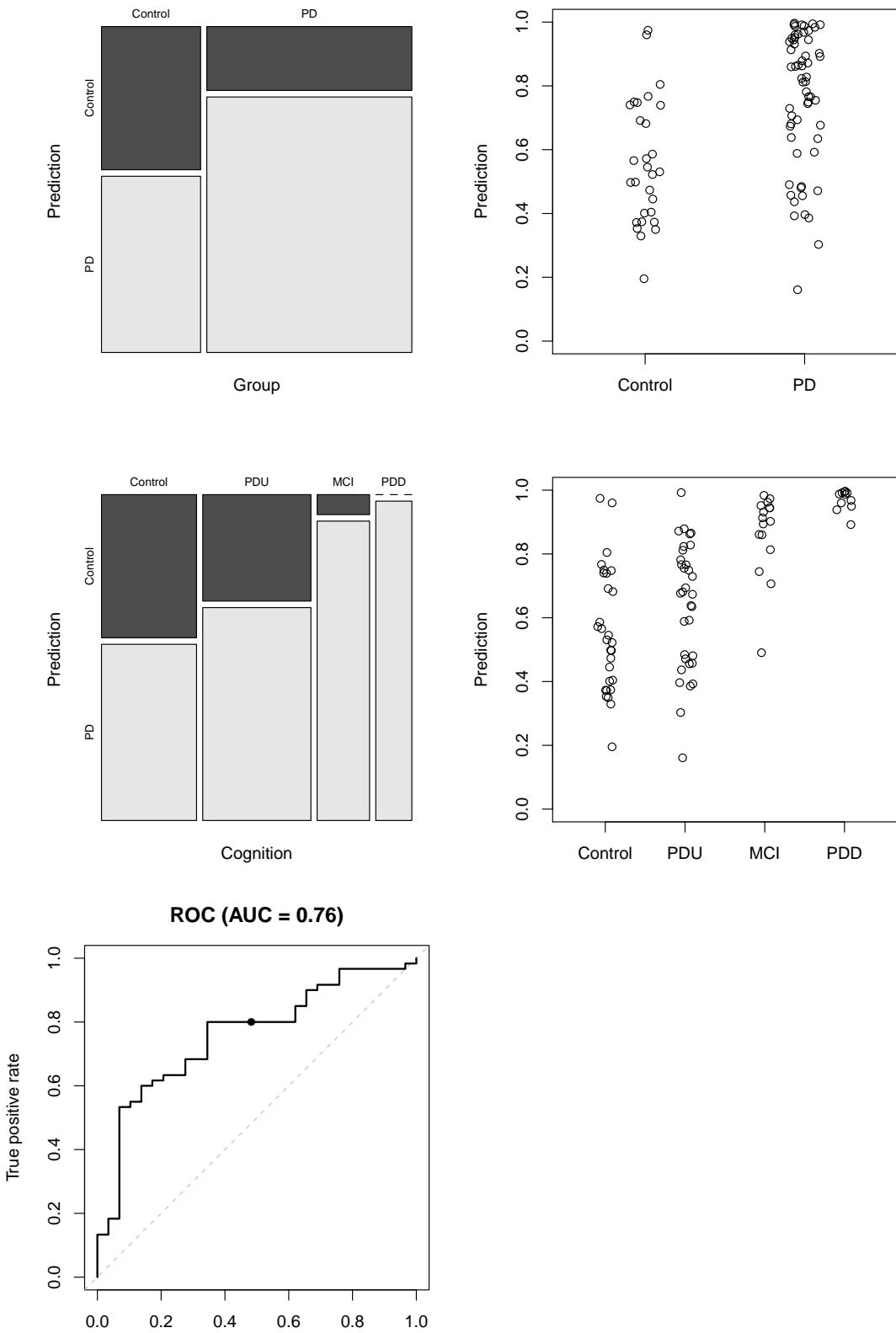


Figure 4.25.: Performance of model (cross-validation with leave-one-out).

Independent component analysis (ICA)

Independent component analysis showed an overall error rate of **0.24** (cutoff=0.5). Less than 59% of controls and 85% of PD's were correct classified (table 4.12). There was no misclassification for PDD's and only 1 misclassification for MCI's (table 4.13). ROC curve and predictions are in figure 4.26.

	Control (pred)	PD (pred)	class.error
Control	17	12	0.414
PD	9	51	0.150

Table 4.12.: Performance of whole model with ICA (cross-validation with leave-one-out).

	Control (pred)	PD (pred)
Control	17	12
PDU	8	25
MCI	1	15
PDD	0	11

Table 4.13.: Performance of whole model with ICA for cognition groups (cross-validation with leave-one-out).

Misclassified subjects:

Parkinson's disease subjects 5801, 6257, 6417, 8374, 8860, 12210, 12391, 12892, 13196.

Control subjects 5703, 6766, 6770, 7107, 7683, 8618, 8619, 8858, 8862, 9117, 11548, 11890.

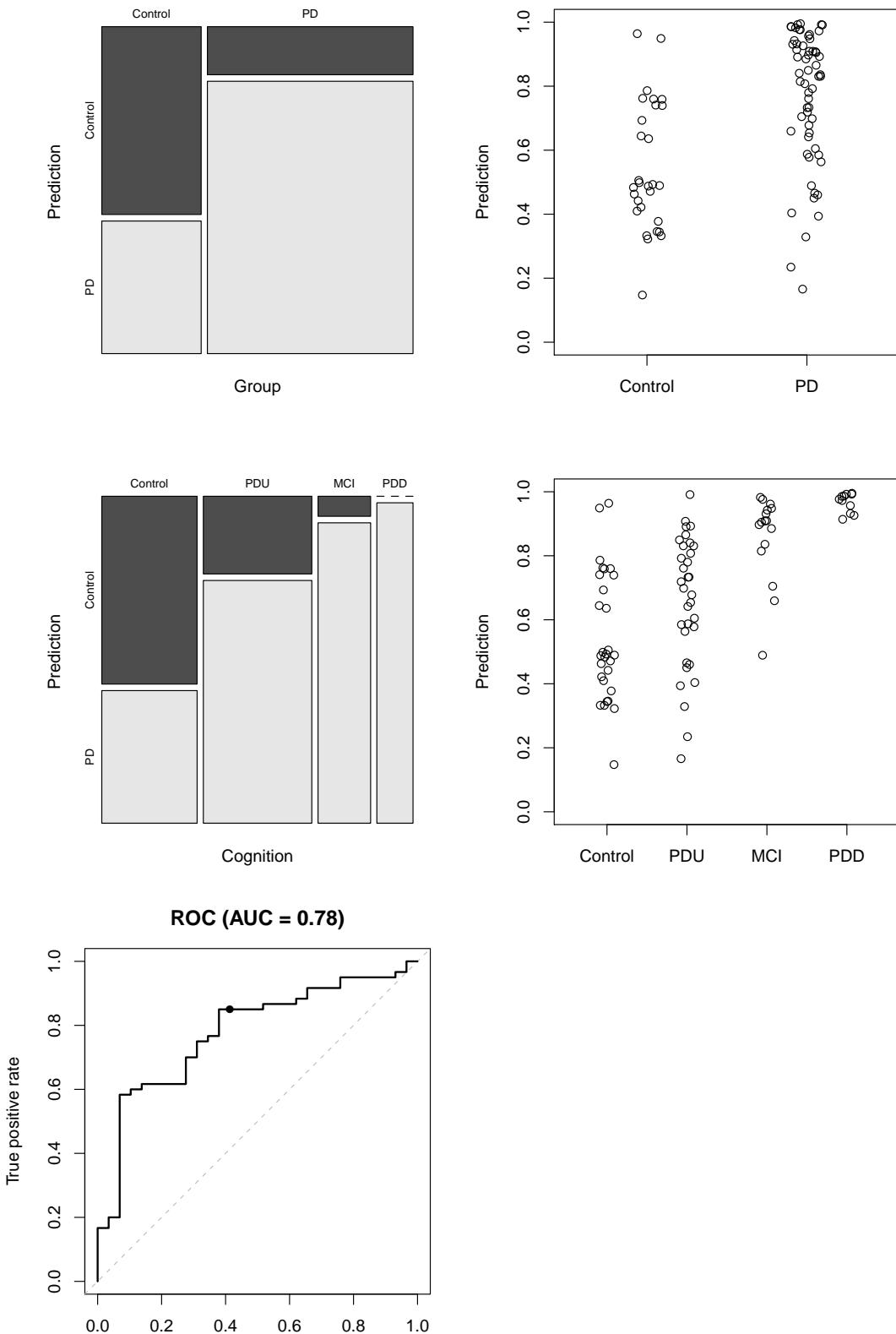


Figure 4.26.: Performance of model (cross-validation with leave-one-out).

4.5.2. Bootstrap

Bootstrap was done as shown in section 3.5.2.

Each voxel of the 'raw' PD-related pattern was divided through its standard deviation computed with bootstrap. Including the variance of the voxel and standardizing the estimation of each voxel with its variance might change the pattern but if the PCA and the ICA pattern are compared (figure 4.28) they are very similar.

Figure 4.27 shows the so called *bootstrapped PD-related pattern* for ICA and PCA. The interpretation of the bootstrapped images is difficult. Axial images in figure 4.27 show blue, white and red colored parts. We would tend to say that blue means decreased, white means unchanged and red increased cerebral blood flow. Broghammer (2009) [28] claims that because control and Parkinson's disease cerebral blood flow is statistically different an overall shift occurs. The statement for an image therefore changes: decreased (blue) areas are in reality strong decreased areas, unchanged (white) areas are in reality slight decreased areas and increased areas (red) are in reality unchanged areas.

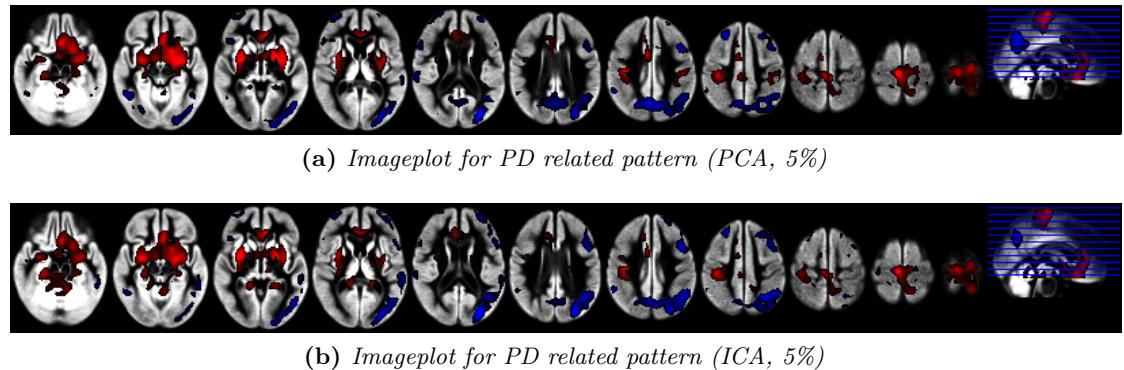


Figure 4.27.: Imageplot for PD related pattern: ICA and PCA and Z-Scores - blue indicates decreased relative cerebral blood flow and red increased relative cerebral blood flow.

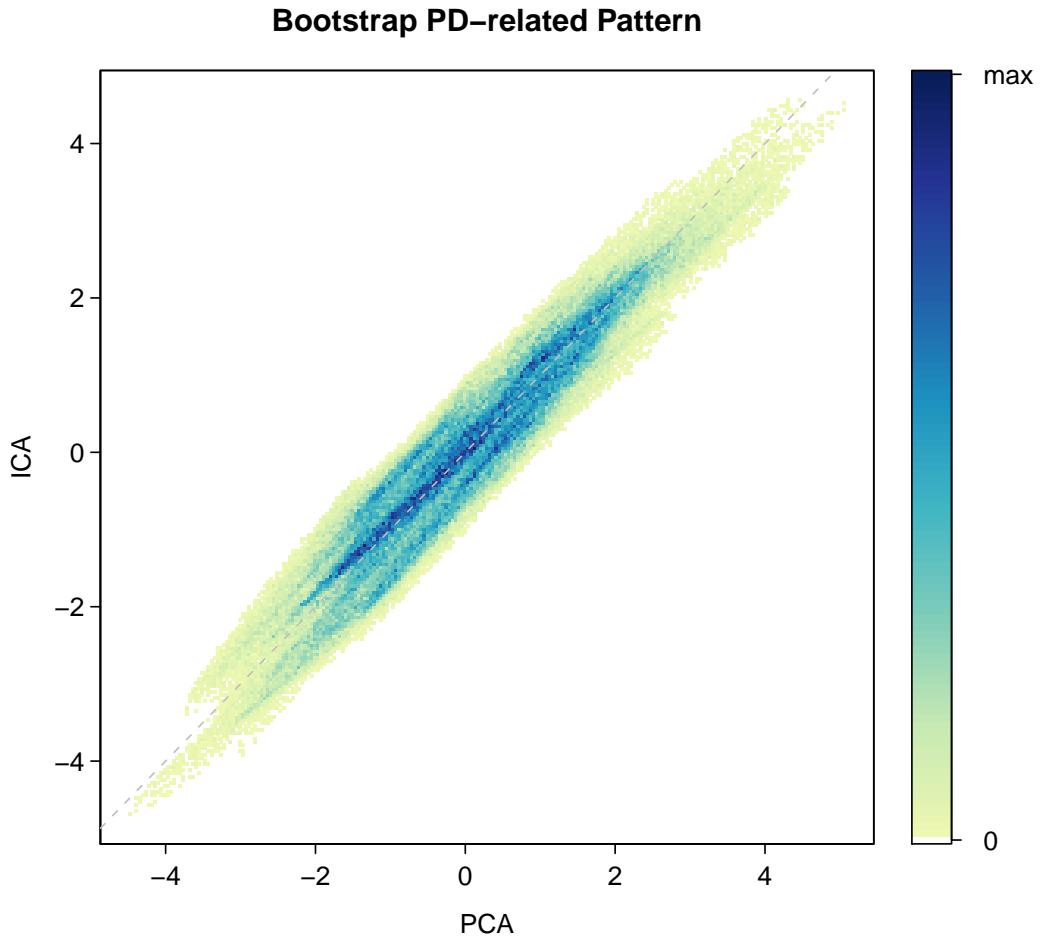


Figure 4.28.: Scatter-density-plot for PCA (x-axis) and ICA (y-axis) pattern. This plot shows the similarity of the ICA and the PCA pattern.

Figure 4.30 (PCA) and 4.32 (ICA) show the results of the bootstrapping. In (a) is the relation between the 'raw' pattern and the bootstrapped pattern, in (b) the 'raw' pattern versus the mean of the bootstrap values, in (c) the 'raw' pattern versus the variance of the bootstrap values and in (d) the 'raw' pattern versus the bias (expected value - estimator) of the bootstrap values. In both models (PCA and ICA) bootstrap shows irregularities in the estimation of the model. First, there is as systematic bias (see graph (d)). Negative values are to low estimated and positive values to high. Means - the model is overestimated and voxels which seem significant ($|voxel| > 1.96$) might be not significant. We did not apply a correction of the estimation. Graph (c) shows that some

voxels have a high variance. These voxels are in ICA and PCA localized in the occipital region (around $30 \times 7 \times 15$ ($x \times y \times z$)).

The histograms in figure 4.29 (PCA) and 4.31 (ICA) shows a symmetric bootstrapped pattern with a maximal variance of 5 and a maximal absolute bias of 1.5.

Principal component analysis (PCA)

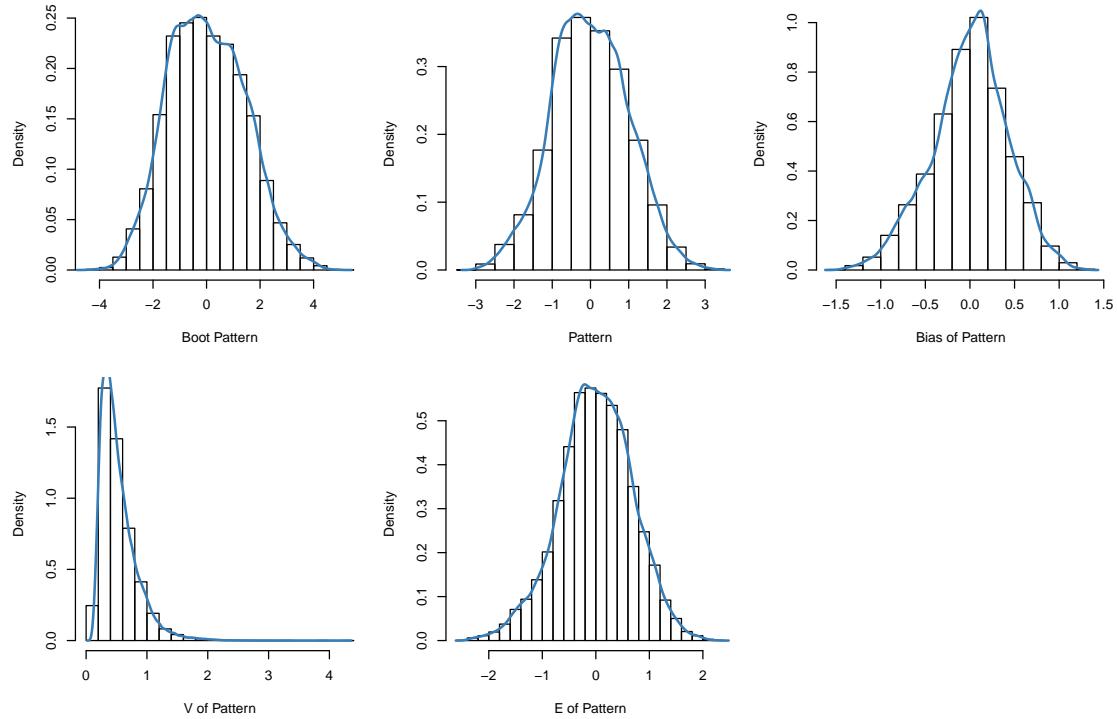
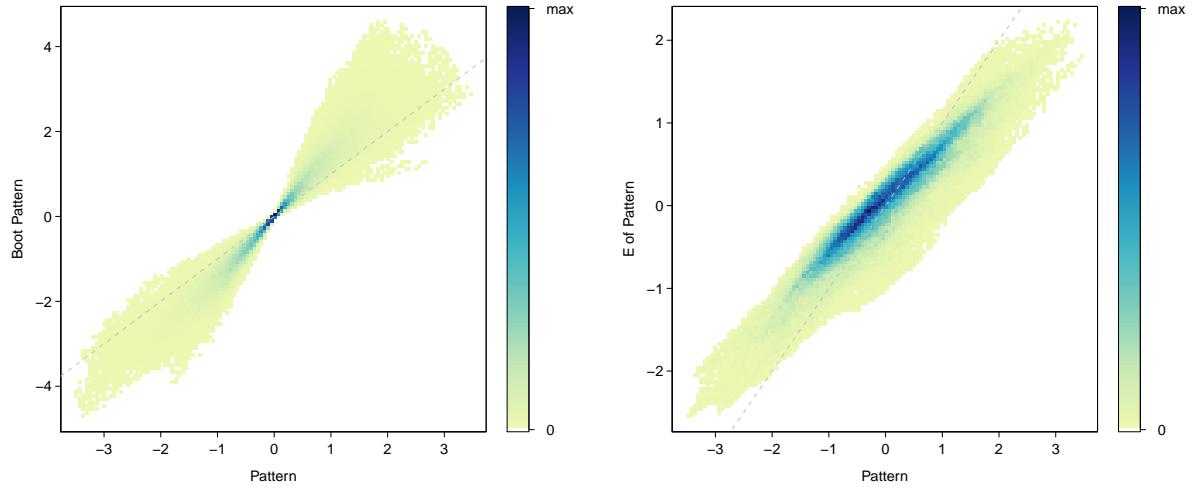
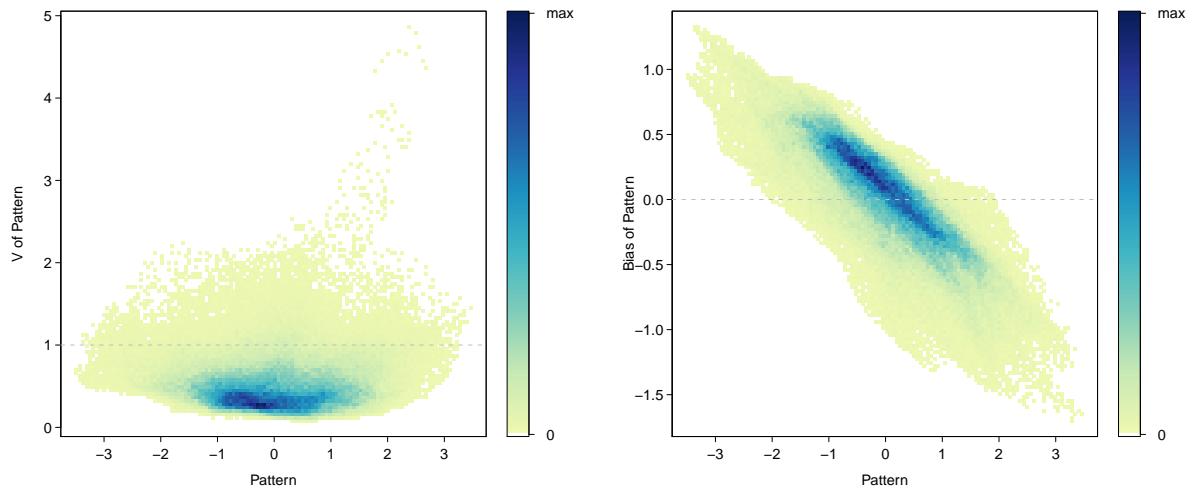


Figure 4.29.: Histogram of bootstrapped PD-related pattern, 'raw' pattern, bias, variance and mean of pattern.



(a) Y-axis: estimation for each voxel divided by its standard deviation.

(b) Y-axis: mean for each voxel.



(c) Y-axis: variance for each voxel.

(d) Y-axis: bias for each voxel.

Figure 4.30.: X-axis is representing the 'raw' PD-related pattern, y-axis is the bootstrapped pattern (top left), the mean for each voxel (top right), the variance for each voxel (bottom left) and the bias for each voxel (bottom right).

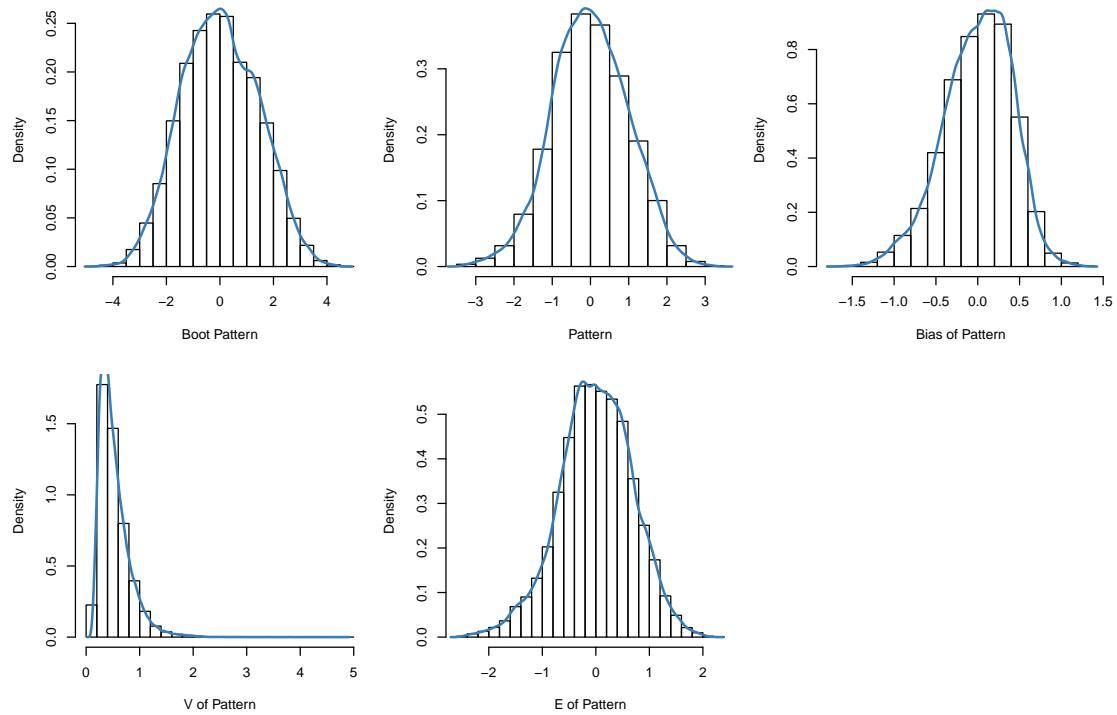
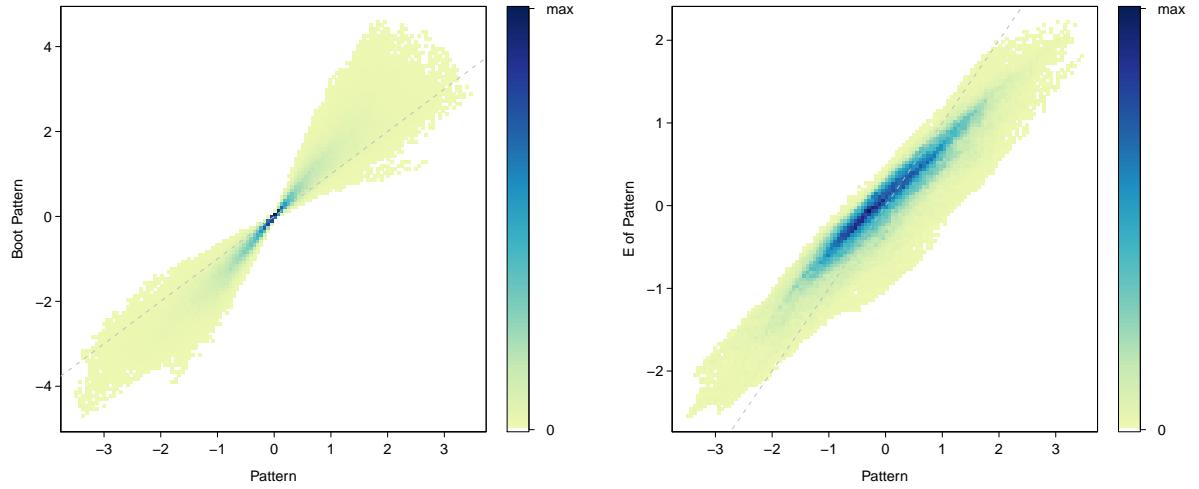
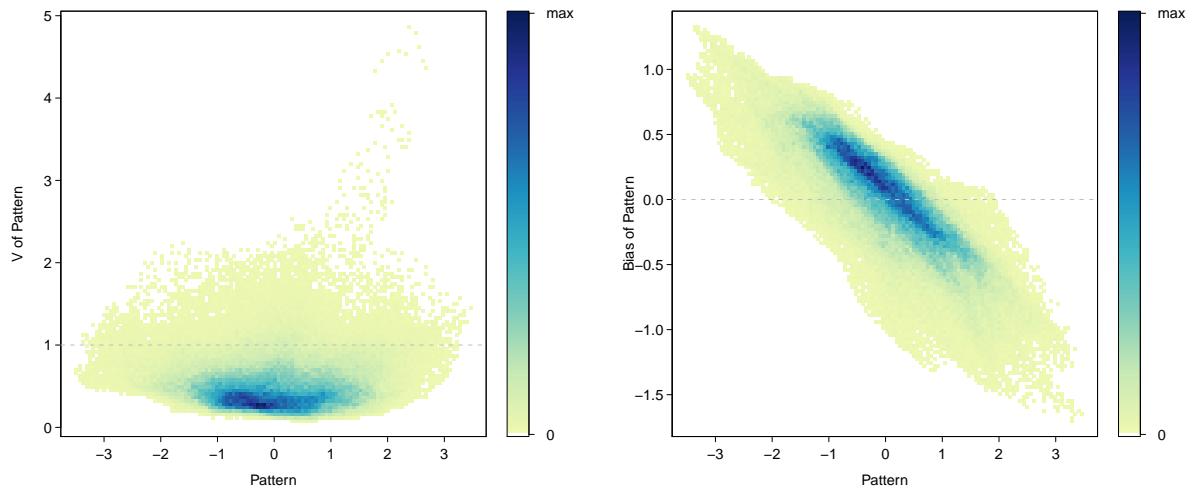
Independent component analysis (ICA)

Figure 4.31.: Histogram of bootstrapped PD-related pattern, 'raw' pattern, bias, variance and mean of pattern.



(a) Y-axis: estimation for each voxel divided by its standard deviation.

(b) Y-axis: mean for each voxel.



(c) Y-axis: variance for each voxel.

(d) Y-axis: bias for each voxel.

Figure 4.32.: X-axis is representing the 'raw' PD-related pattern, y-axis is the bootstrapped pattern (top left), the mean for each voxel (top right), the variance for each voxel (bottom left) and the bias for each voxel (bottom right).

4.6. Comparing PCA and ICA

Whether PCA or ICA is recommended can be decided with the following questions.

1. How **independent** are the components? With the previous mentioned measure of *negentropy* we can make an approximate estimation for nongaussianity of components.
2. What is the **performance** of the entire model? With cross-validation we can obtain the model accuracy.

4.6.1. Negentropy or Kurtosis

The negentropy should be in independent components higher than in principal components and the norm vince versa. Remember: significant components in PCA have been 1, 4 and 5; in ICA component 1 and 6. Figure 4.33 illustrates that principal component 4 has the highest negentropy overall significant components and independent component 1 has the lowest one. In addition is norm of the 3 significant components in PCA much higher than the 2 significant components in ICA. This can also be seen if a prediction of X is made. The sum of square of the difference between X and \hat{X} (predicted with components 1, 4 and 5) is for PCA 413441.9 and for ICA (predicted with components 1 and 6) 481591.77.

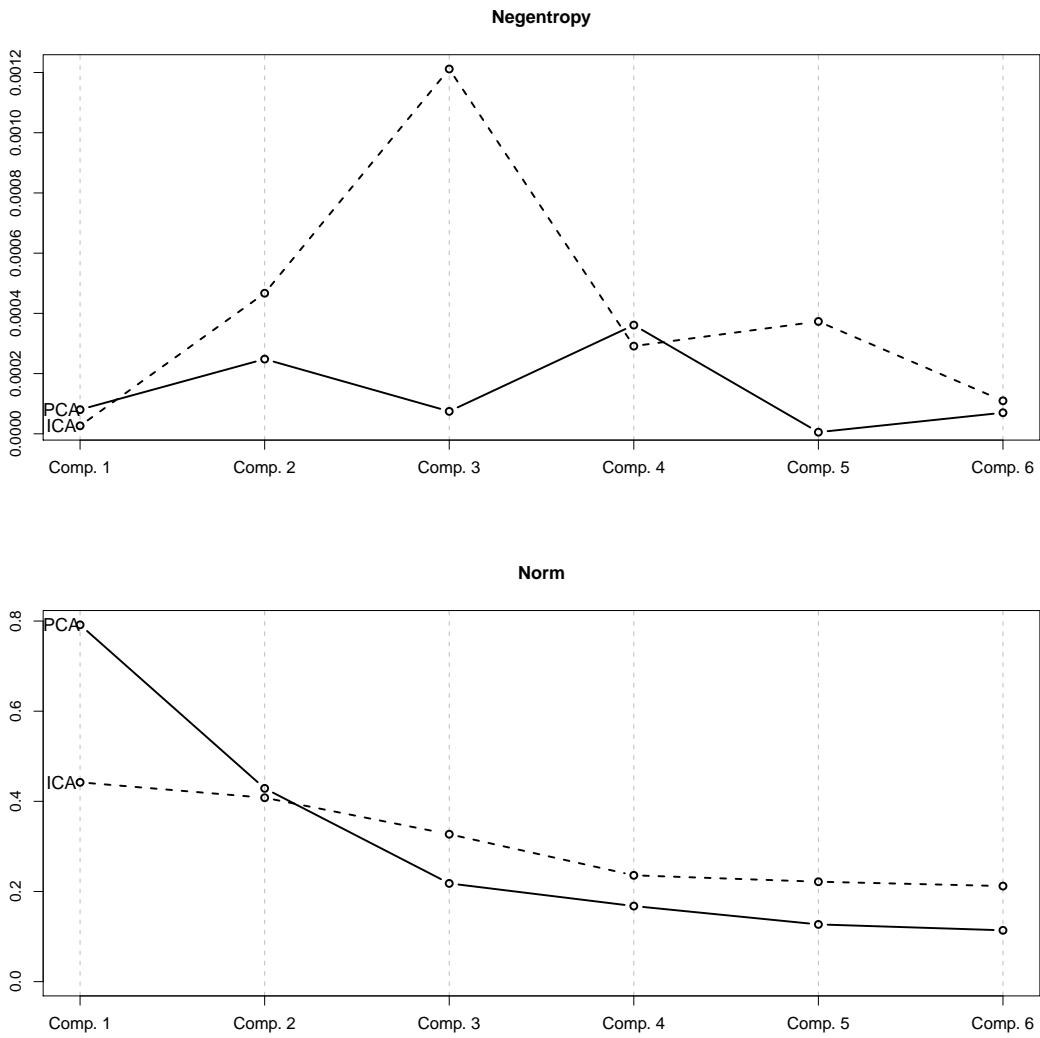


Figure 4.33.: Negentropy (top) of all components 1 to 6 and norm (bottom) for each columns in the mixing matrix. Significant components in PCA were 1, 4 and 5, in ICA 1 and 6.

4.6.2. Performance of Model

Regarding the misclassificationrate of **0.24** in ICA is performing better than PCA which has a misclassificationrate of **0.31**, area under the curve in illustrated in figure 4.34. PCA (table 4.14) shows 16 misclassified controls and 12 misclassified PD's (11 PDU, 1 MCI), ICA (table 4.15) has only 12 misclassified controls and 9 misclassified PD's (8 PDU, 1 MCI). Figure 4.34 has a bigger area under the curve for ICA than for PCA. The scatterplots in figure 4.35 and 4.36 show no structure difference in prediction for

ICA or PCA. The mosaicplots are representing the confusion matrix in a graphical manner.

Confusion matrix of leave one out PCA (Errorrate: 0.31):

	Control (pred)	PD (pred)	class.error
Control	13	16	0.552
PD	12	48	0.200

	Control (pred)	PD (pred)
Control	13	16
PDU	11	22
MCI	1	15
PDD	0	11

Table 4.14.: Performance of entire model with PCA for (cross-validation with leave-one-out).

Confusion matrix of leave one out ICA (Errorrate: 0.24):

	Control (pred)	PD (pred)	class.error
Control	17	12	0.414
PD	9	51	0.150

	Control (pred)	PD (pred)
Control	17	12
PDU	8	25
MCI	1	15
PDD	0	11

Table 4.15.: Performance of entire model with ICA for (cross-validation with leave-one-out).

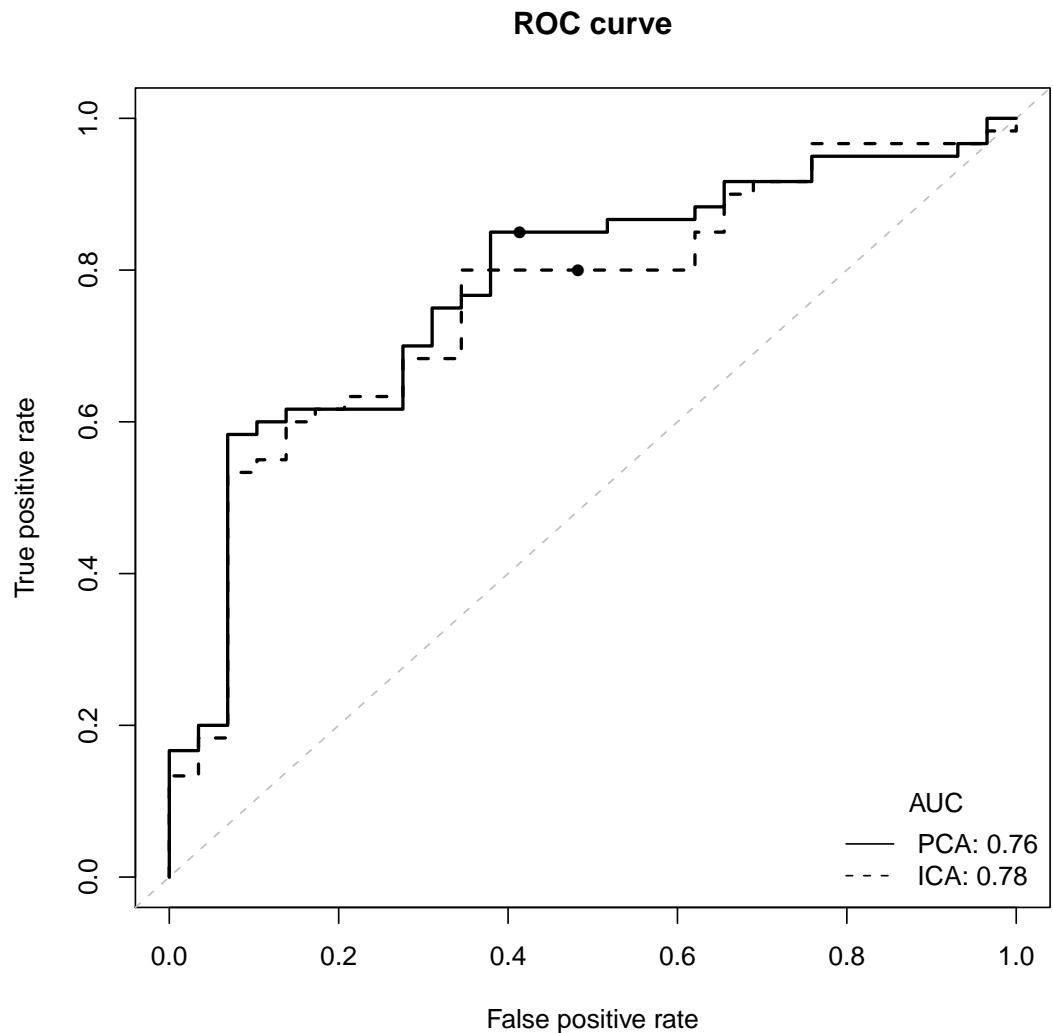


Figure 4.34.: ROC curve for entire model. PCA (—) and ICA (- - -) have a similar area under the curve.

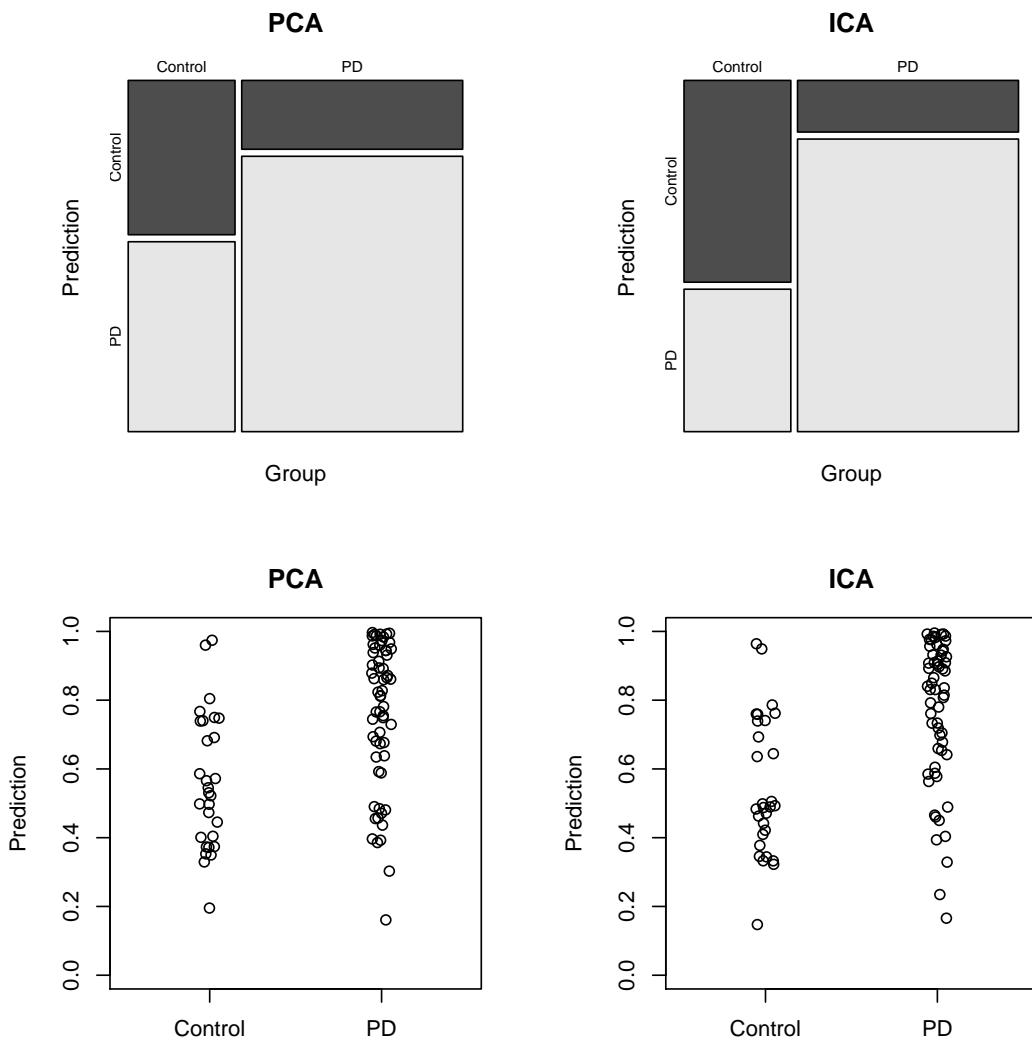


Figure 4.35.: Performance of entire model. The mosaicplot (representing the confusion matrix) shows that ICA classifies better than PCA (bigger areas in the diagonal). The scatterplots (with jitter in x direction) are showing no structural difference between ICA and PCA.

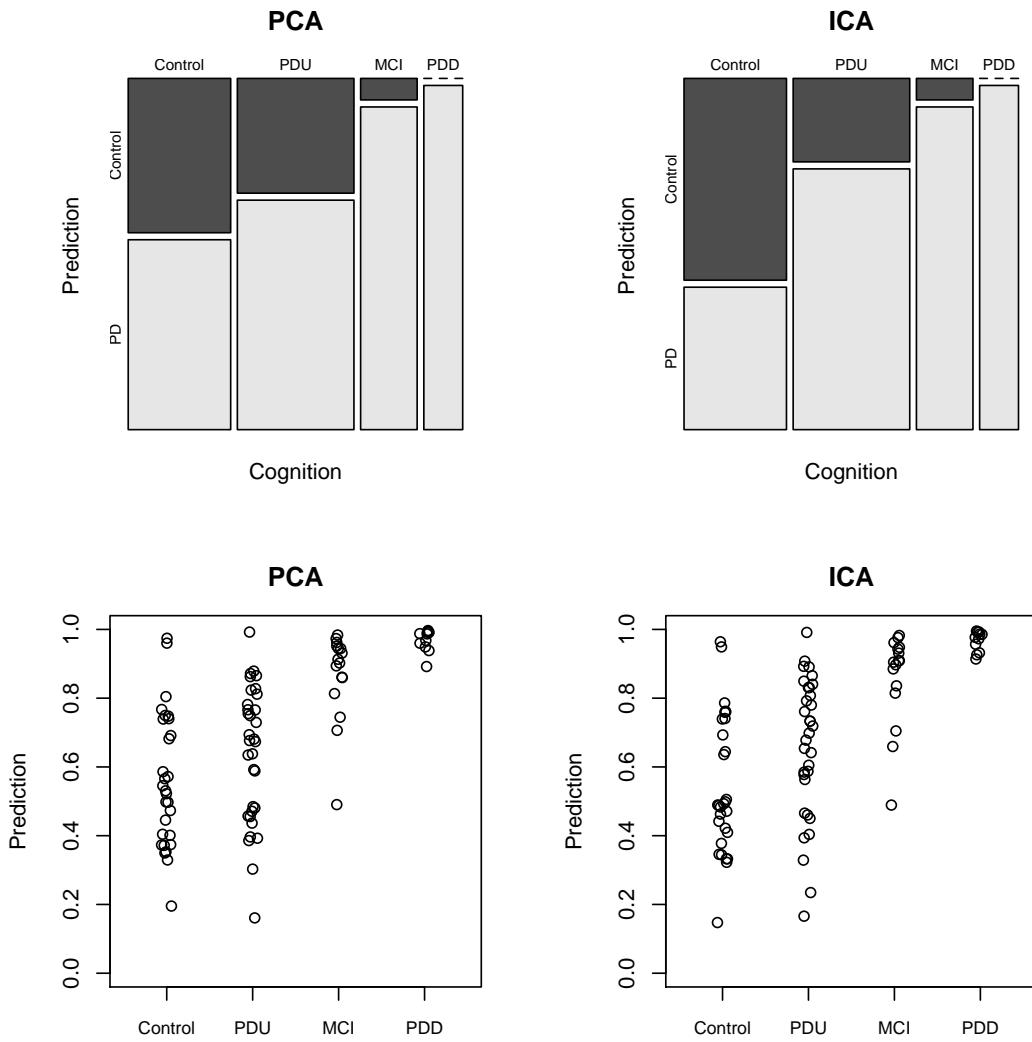


Figure 4.36.: Performance of entire model for cognition groups. The mosaicplot are showing that most of the misclassifications in PD is caused by PDU. The two controls with a prediction close to 1 are subject 6770 and 11548. The MCI with a prediction under 0.5 is subject 12391.

4.7. Subset: Exclusion of Controls

This chapter reports the results of the analysis of only a subset of the subjects. As seen in tables 4.11 (PCA) and 4.13 (ICA) the distinction between controls and cognitive unimpaired PD's (PDU) is not very good. That might be not a weakness of the methods ICA or PCA rather the fact that there is no difference in cerebral blood flow between

healthy people and cognitive unimpaired PD's. In fact, the differences between these two groups are not of particular interest at this early stage of research. What is more important is to make the distinction between cognitive unimpaired PD's (PDU) and mild impaired PD's (MCI). To classify a person within the cognitive groups PDU, MCI and PDD is difficult. Even more because a fraction of the people are taking drugs can distort the results of the cognition test assessments (e.g. MoCA). These test results are used to classify the patients as PDU, MCI and PDD. Therefore one is particularly interested in developing a biomarker which allows a reliable prediction of a subject as cognitive unimpaired or cognitive impaired (mild or severe).

That is the reason why we excluded controls and built new group labels for PD's:

unimpaired cognitive unimpaired: PDU

impaired cognitive impaired: MCI, PDD

4.7.1. PCA: Projection methods & Classification

PCA was performed with a data set without controls. Then the logistic regression was applied where response variable was the group label impaired or unimpaired and the explanatory variables were the mixing values for each component. Table 4.16 shows the logistic regression output after a stepwise model selection with AIC. Component 1, 2, 4 and 5 were statistically significant. Interesting is the fact that the intercept is no longer significant as in the previous models.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.058	0.403	-0.144	0.885
Comp.1	24.591	7.239	3.397	0.001
Comp.2	9.193	6.016	1.528	0.126
Comp.4	57.893	18.262	3.170	0.002
Comp.5	-25.506	11.366	-2.244	0.025

Table 4.16.: Model output for logistic regression after stepwise AIC model selection (PCA).

Figure 4.37 show the prediction where 3 subjects are interesting. The unimpaired subject with a prediction close to 1 is subject 11365, the two misclassified impaired subjects are MCI's: 12391 (0.03), 6255 (0.31).

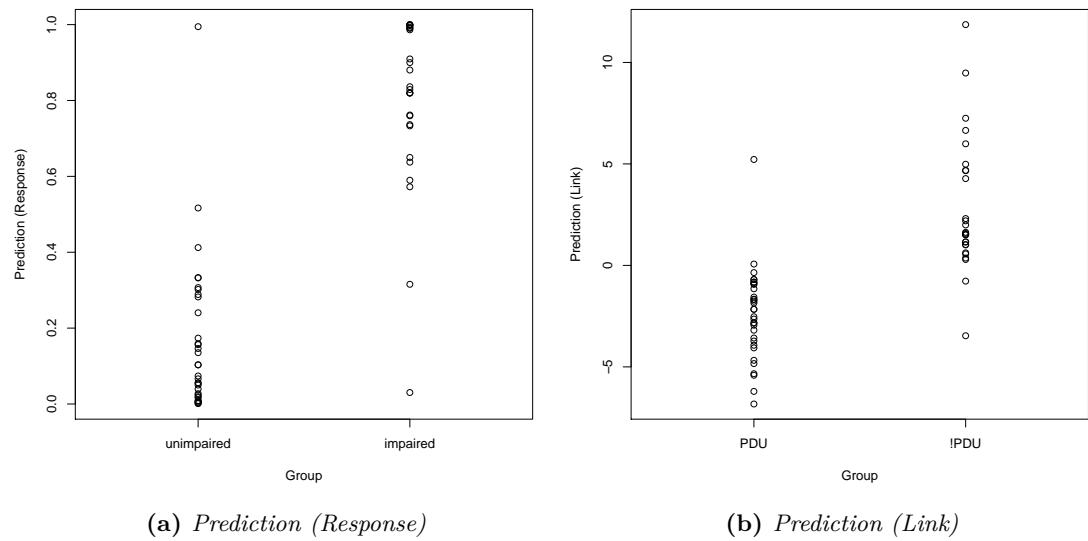


Figure 4.37.: *Prediction for logistic regression (PCA).*

The graphical residual analysis shown in 4.38 gives evidence that the assumptions of a logistic regression are violated. Subject 11365 (number 54) and 12391 (number 43) have extreme residuals. This is illustrated in every of the 4 graphs. Nevertheless the logistic regression is used.

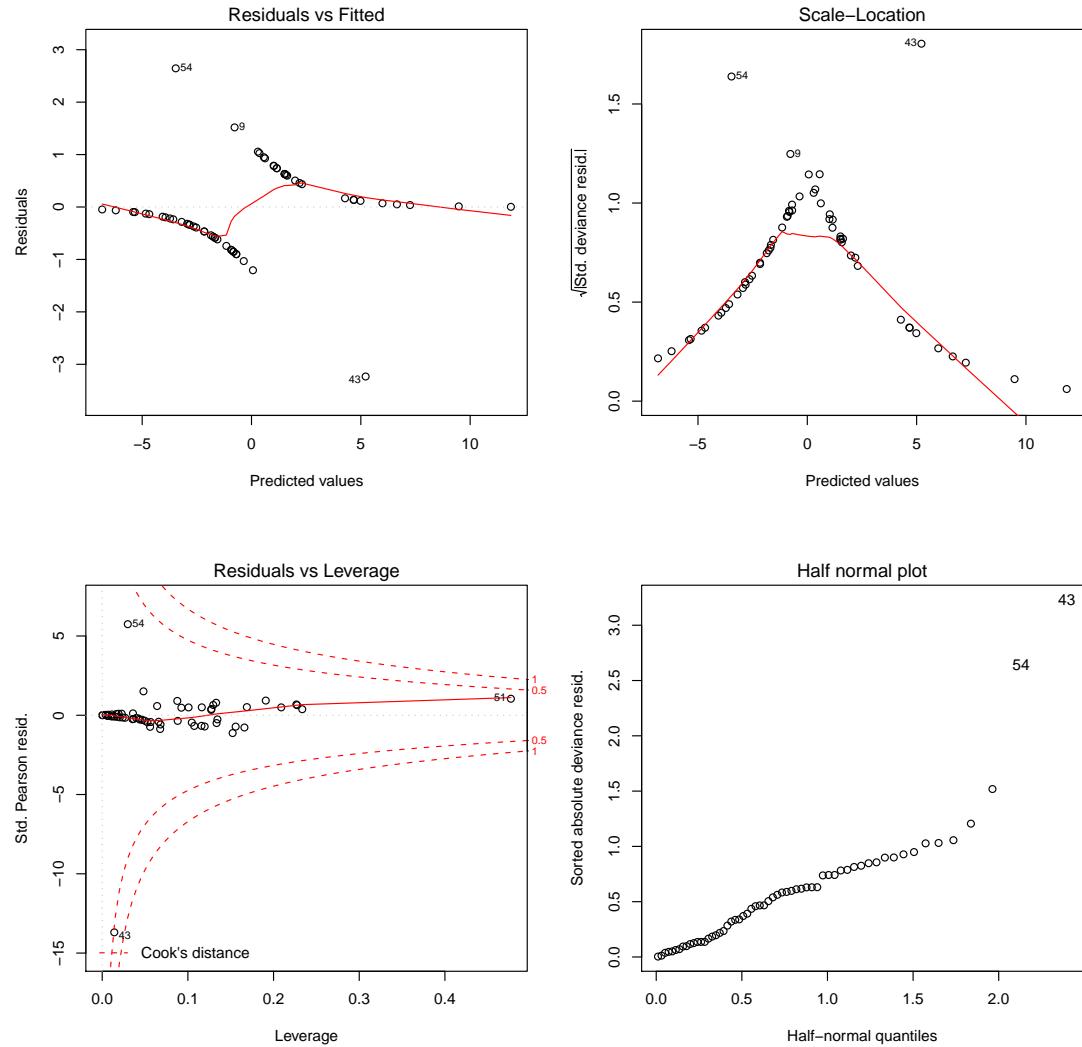


Figure 4.38.: Residual plot for logistic regression (PCA). The scale location plot shows a violence assumption that the variance is constant. Note the extreme residuals of observations 54 and 43

The significant variables in the logistic regression are shown in figure 4.39, where each subject represents a line.

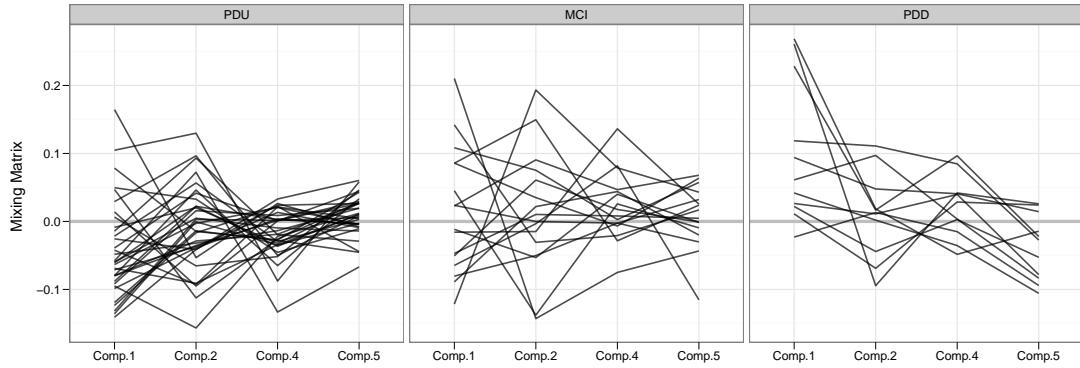


Figure 4.39.: Spaghetti plot of mixing values of component 1, 2, 4 and 5. Each subject is represented by one line and each box represents one cognition level (PDU, MCI, PDD). PDU's have a range between (-0.15) and 0.15 for component 1 and 2, but a smaller range for component 4 and 5, with in particular negative values in component 4. MCI's values in component 1, 2 and 4 have a wide range, but not for component 5. The values can be positive or negative. PDD's have with one exception only positive values for component 1, and negative values for component 5. Values for component 2 or 4 can be negative or positive.

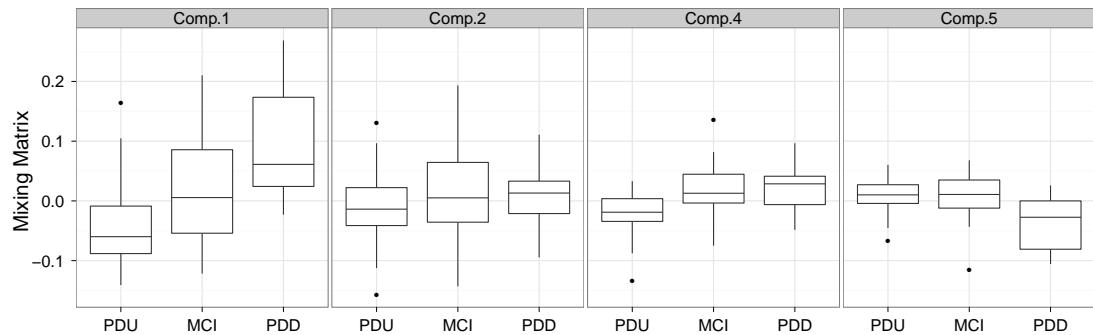


Figure 4.40.: Boxplots for component 1, 2, 4 and 5 for PDU, MCI and PDD. Component 1 discriminates PDU's from PDD's. Component 4 shows lower values for PDU's and a higher level for both MCI's and PDD's. PDU's and MCI's seem to have the same level in Component 5, while PDD's have a lower values.

The corresponding principal components are visualized in a pairs plot in figure 4.41 and the image plots in 4.42 and 4.43.

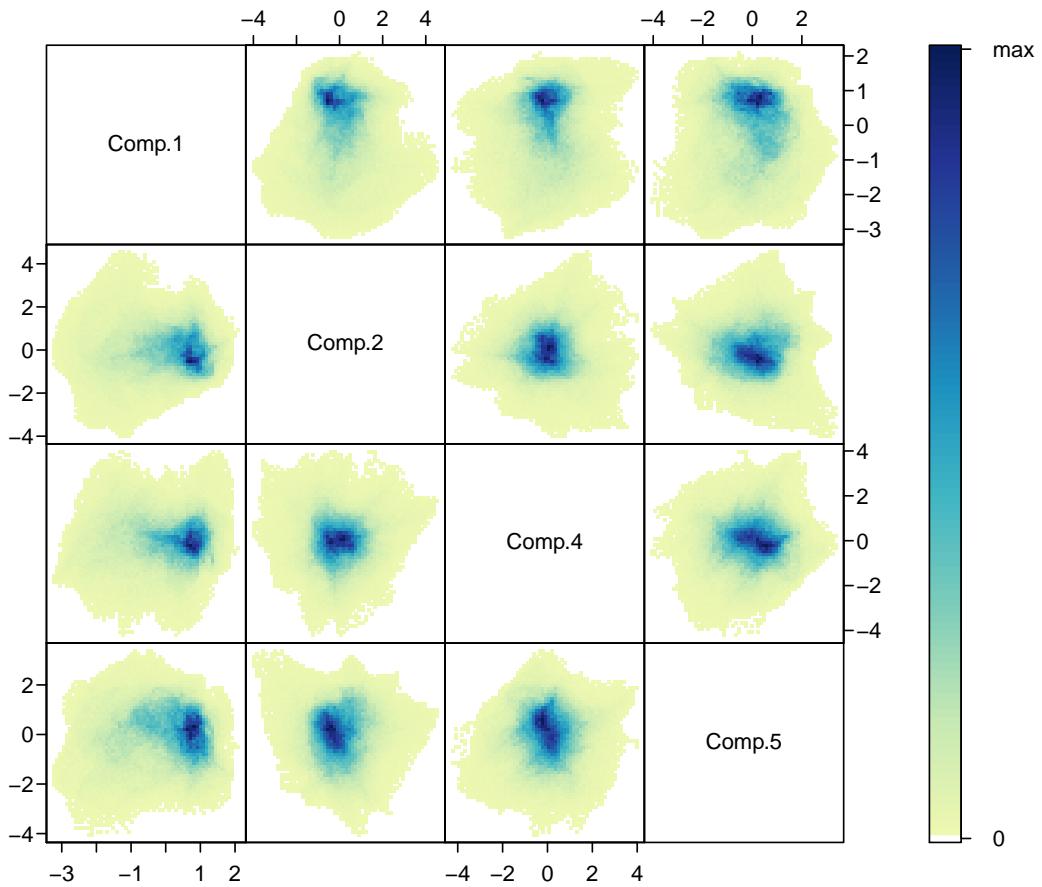
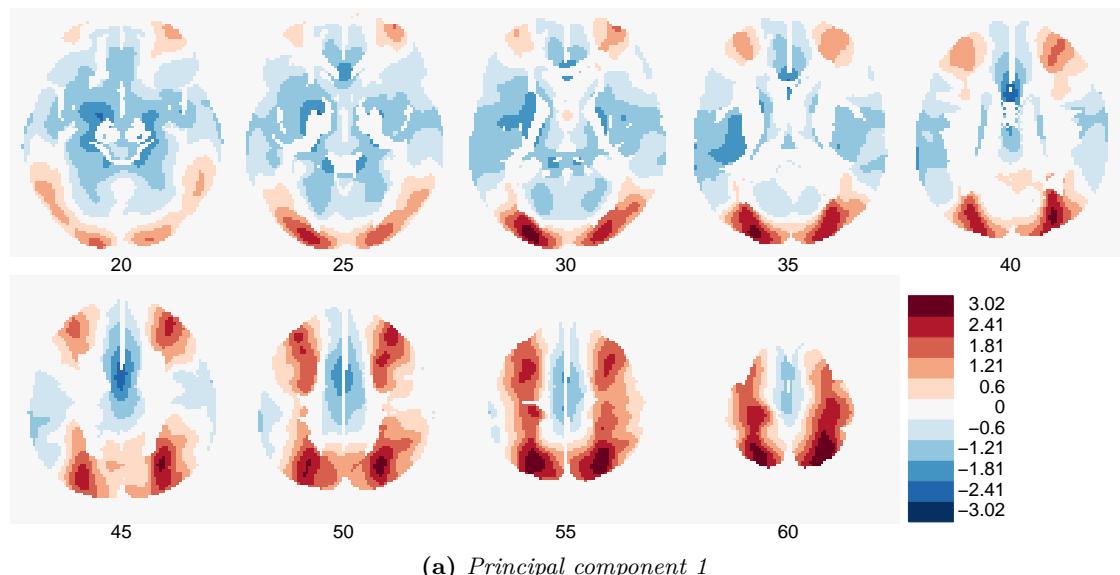
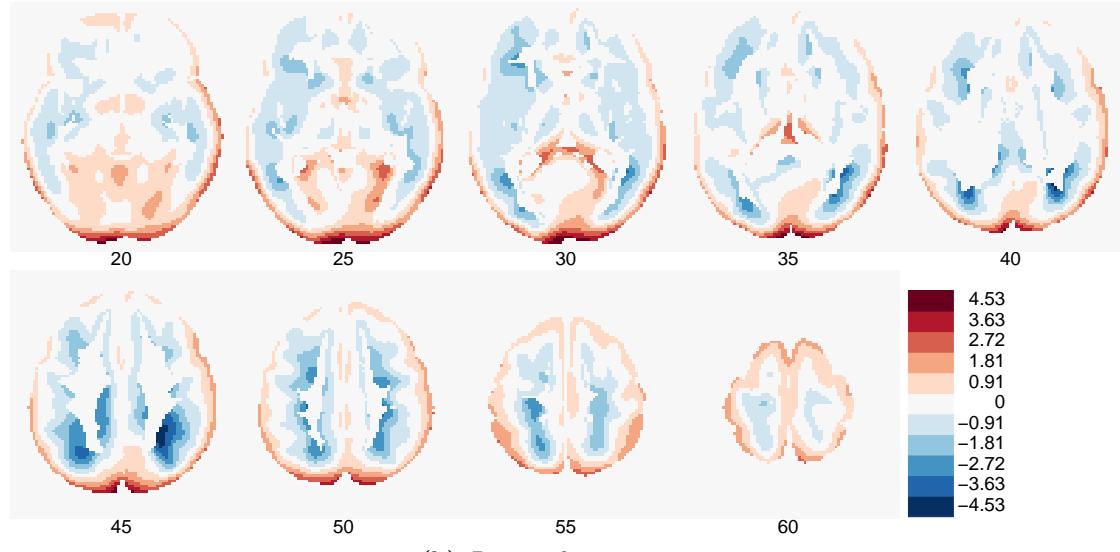


Figure 4.41.: Pairs plot of component 1, 2, 4 and 5. The density goes from white (0) to dark blue (maximum). The components seem not symmetrical and no cluster is visible.



(a) Principal component 1



(b) Principal component 2

Figure 4.42.: Images of all principal components 1 and 2 - red indicates increased, blue decreased relative cerebral blood flow. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

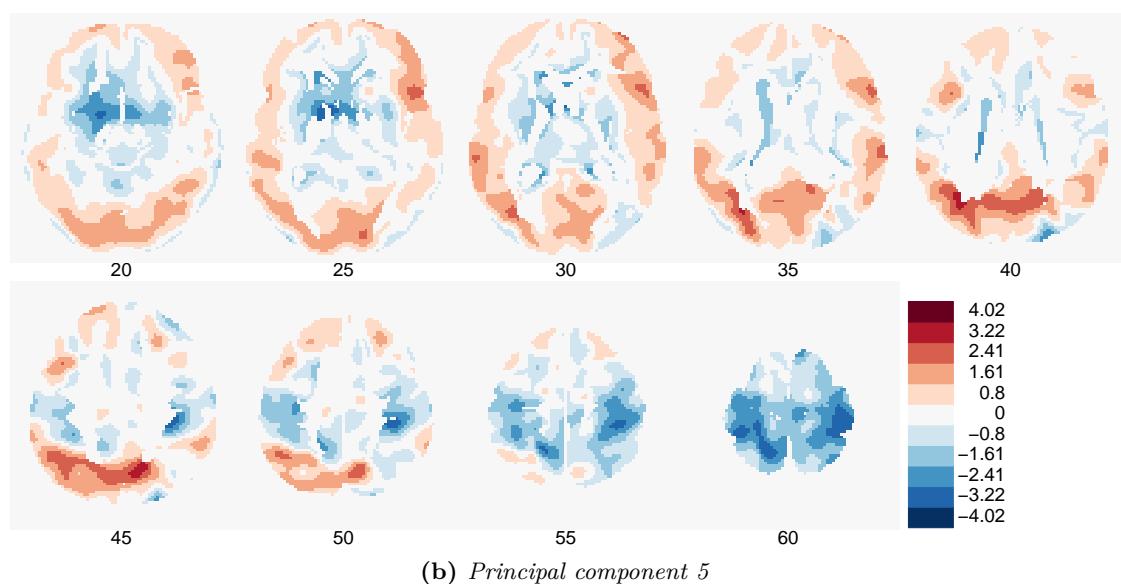
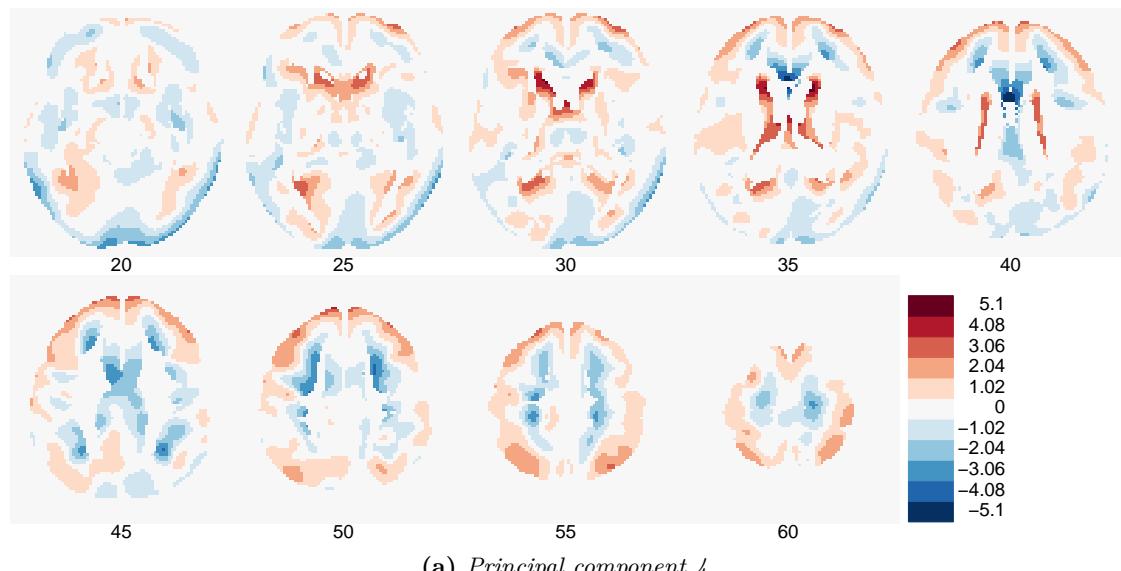


Figure 4.43.: Images of all principal components 4 and 5 - red indicates increased, blue decreased relative cerebral blood flow. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

4.7.2. ICA: Projection methods & Classification

PCA was performed with a data set without controls. Then the logistic regression was applied where response variable was the group label impaired or unimpaired and the explanatory variables were the mixing values for each component. Table 4.17 shows the

logistic regression output after a stepwise model selection with AIC. Component 1, 3, 4 and 5 were statistically significant. Interesting is the fact that the intercept is not longer significant as in the previous models.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.045	0.415	-0.109	0.913
Comp.1	-26.982	8.034	-3.359	0.001
Comp.3	-19.597	9.564	-2.049	0.040
Comp.4	34.133	10.849	3.146	0.002
Comp.5	-50.109	14.835	-3.378	0.001

Table 4.17.: Model output for logistic regression after stepwise AIC model selection (ICA).

Figure 4.44 shows the predictions of the logistic regression. The unimpaired person with a prediction close to 1 is 11365, the impaired subject with a prediction close to 0 is 12391 (MCI).

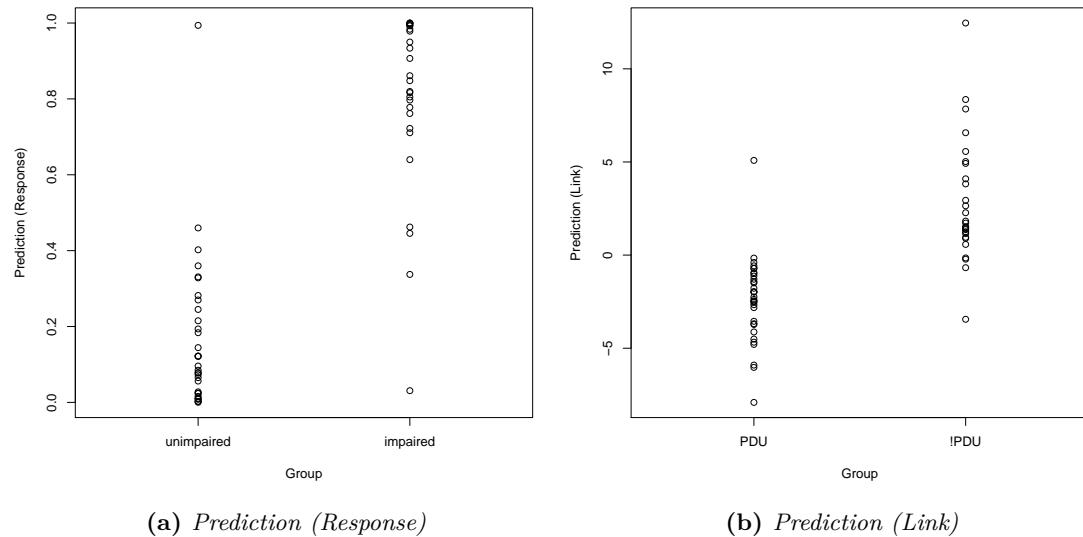


Figure 4.44.: Prediction for logistic regression (PCA).

The graphical residual analysis in figure 4.45 shows two subjects 12391 (nbr. 54, impaired) and 11365 (nbr. 43, unimpaired) which have high absolute residuals. The graph 'Residuals vs Leverage' even considers that these two observations are outliers. Nevertheless the logistic regression is used (but see 5.1 for suggestions if the same analysis is done again).

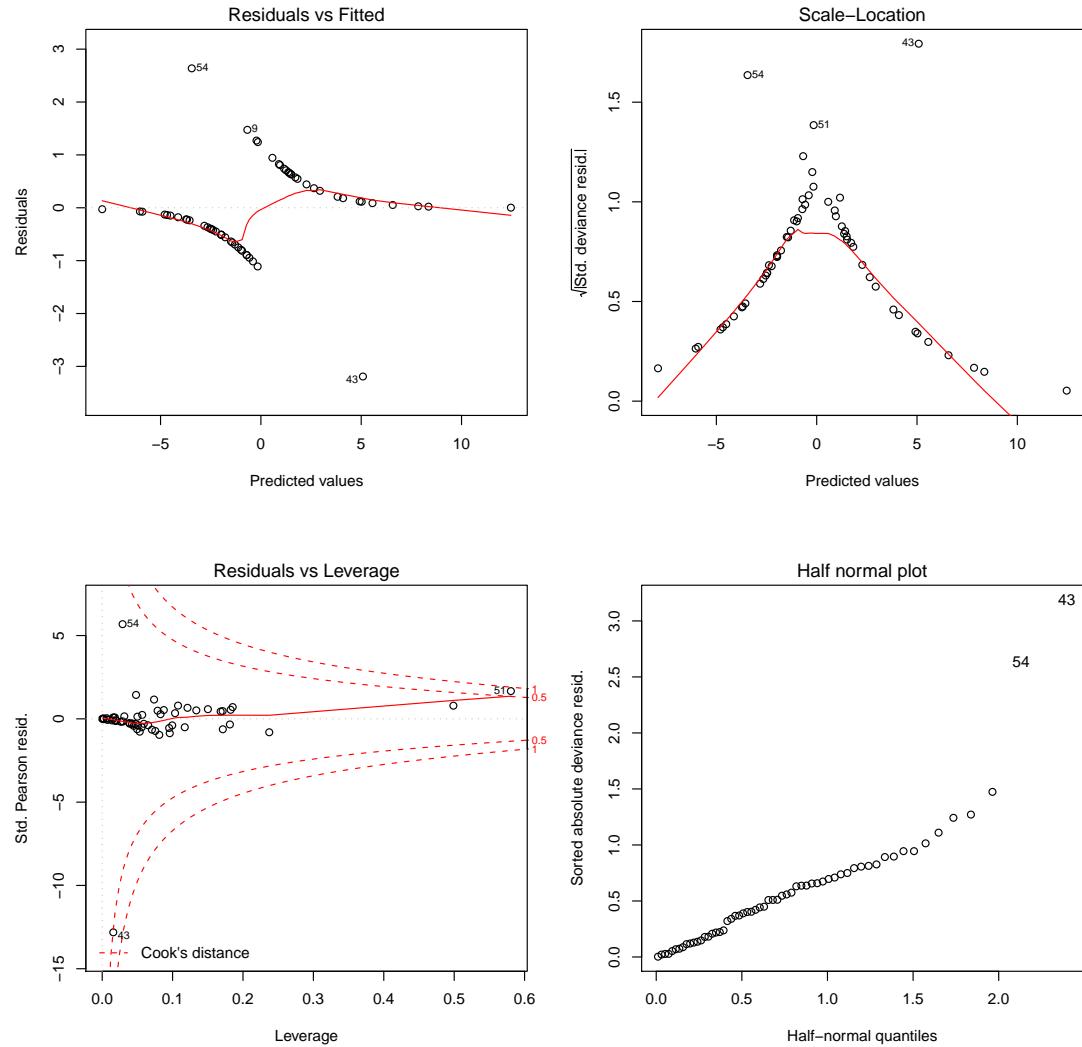


Figure 4.45.: Residual plot for logistic regression (ICA).

The significant variables in the logistic regression are shown in figure 4.46, where each subject represents a line.

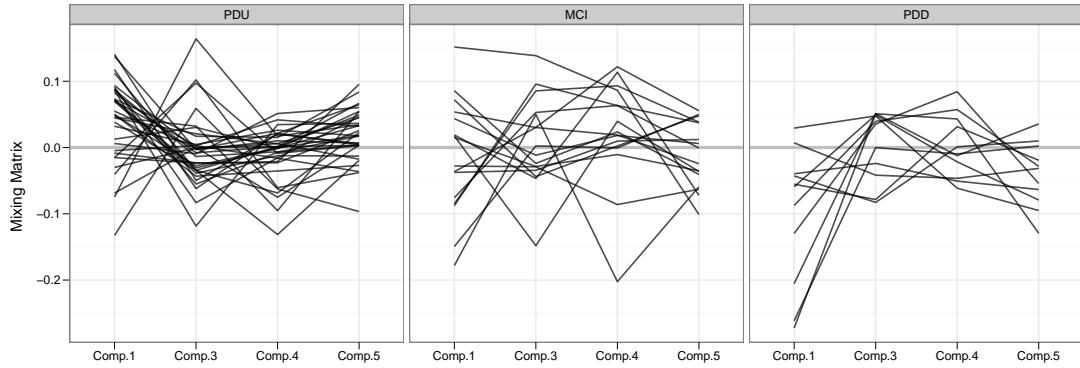


Figure 4.46.: Spaghetti plot of mixing values of component 1, 3, 4 and 5. Each subject is represented by one line and each box represents one cognition level (PDU, MCI, PDD). The shape of the lines in PDU's are forming more an 'u' where the PDD's have more a 'n'. In general, PDU's have positive values in component 1 and 5, where PDD's have negative values. MCI's have in all components either positive or negative values.

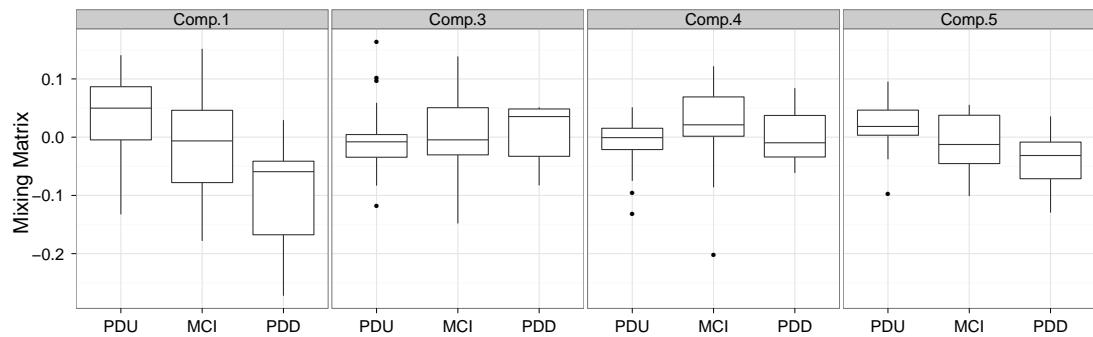


Figure 4.47.: Boxplots for component 1, 3, 4 and 5 for PDU, MCI and PDD.

The corresponding independent components are visualized in a pairs plot in figure 4.48 and the image plots in 4.49 and 4.50.

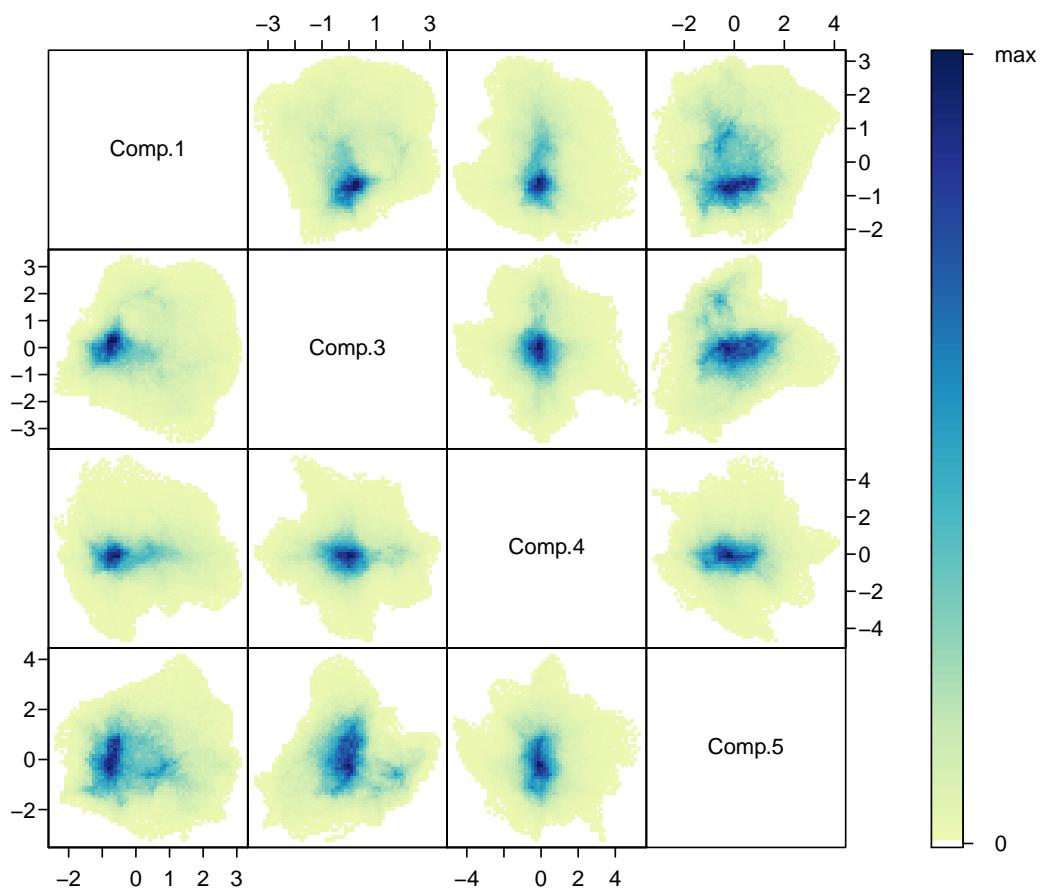
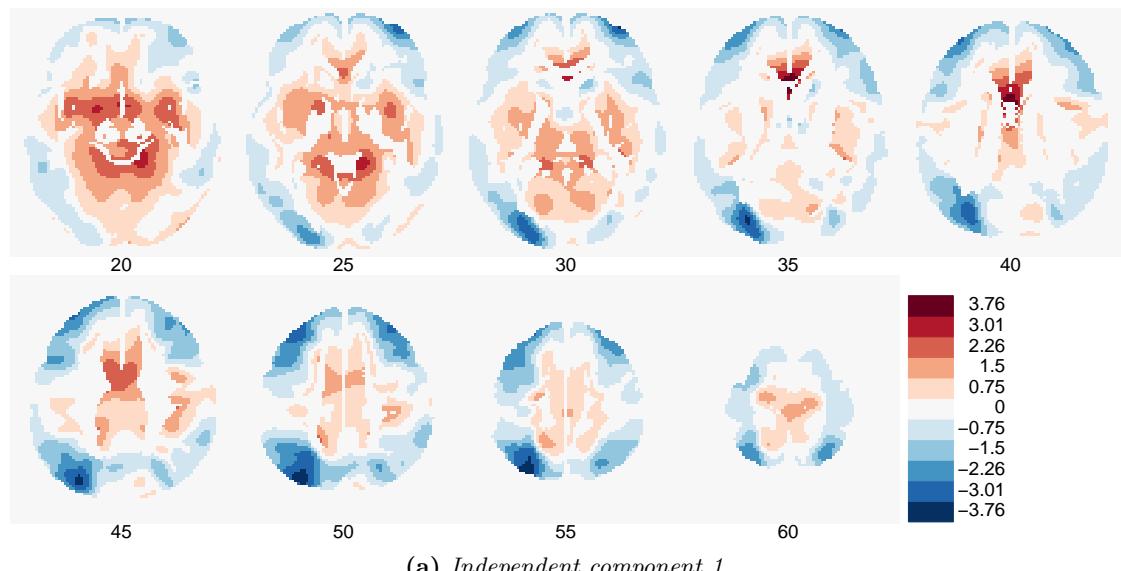
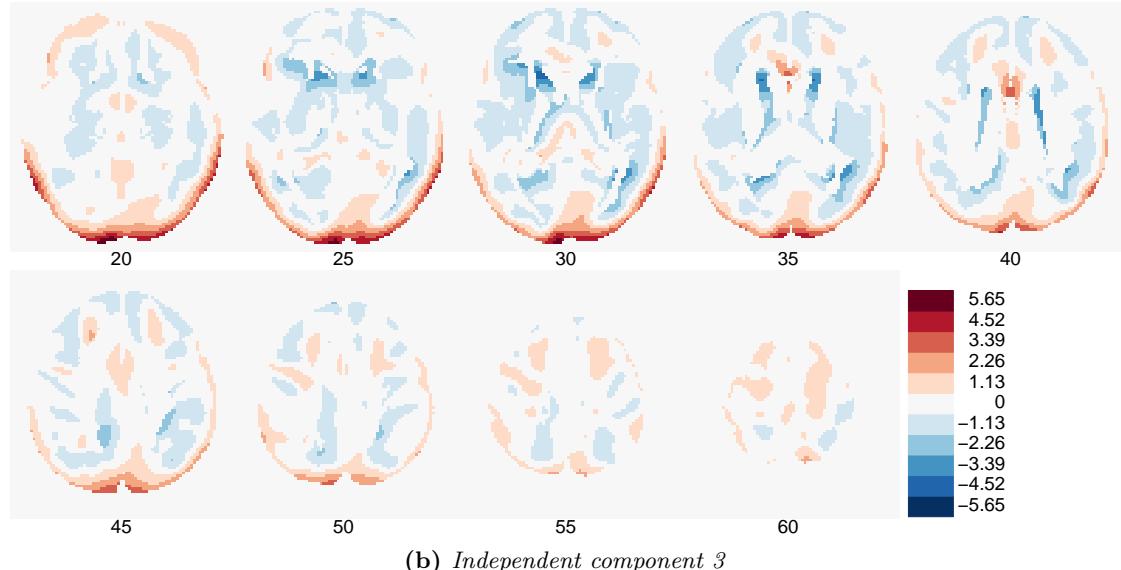


Figure 4.48.: Pairs plot of independent component 1, 3, 4 and 5. The density goes from white (0) to dark blue (maximum). Clusters are visible between component 1 and 5 and 3 and 5.



(a) Independent component 1



(b) Independent component 3

Figure 4.49.: Images of all Independent components 1 and 3 - red indicates increased, blue decreased relative cerebral blood flow. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

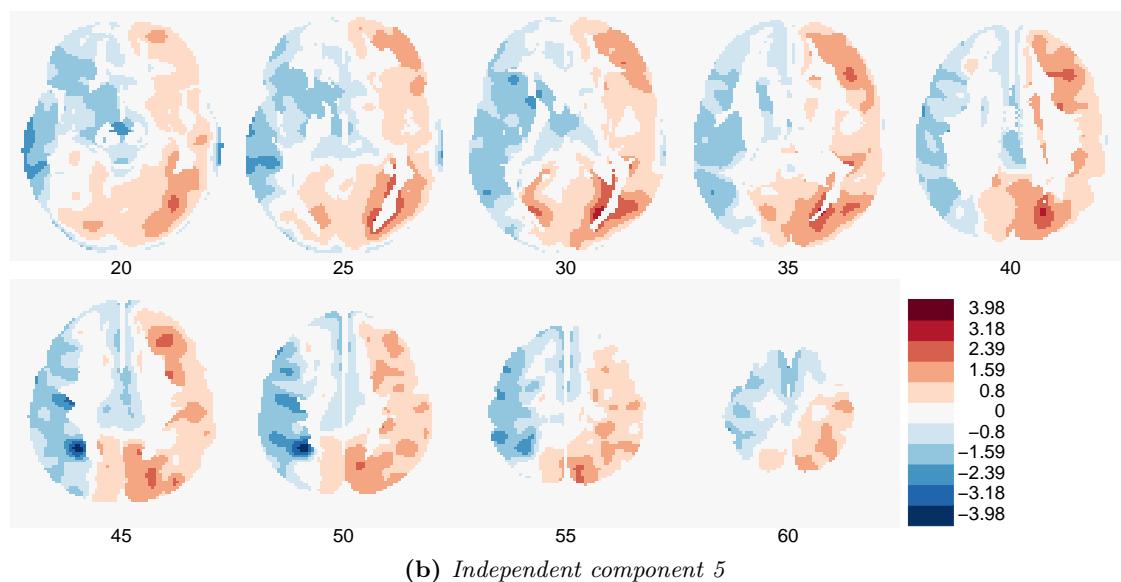
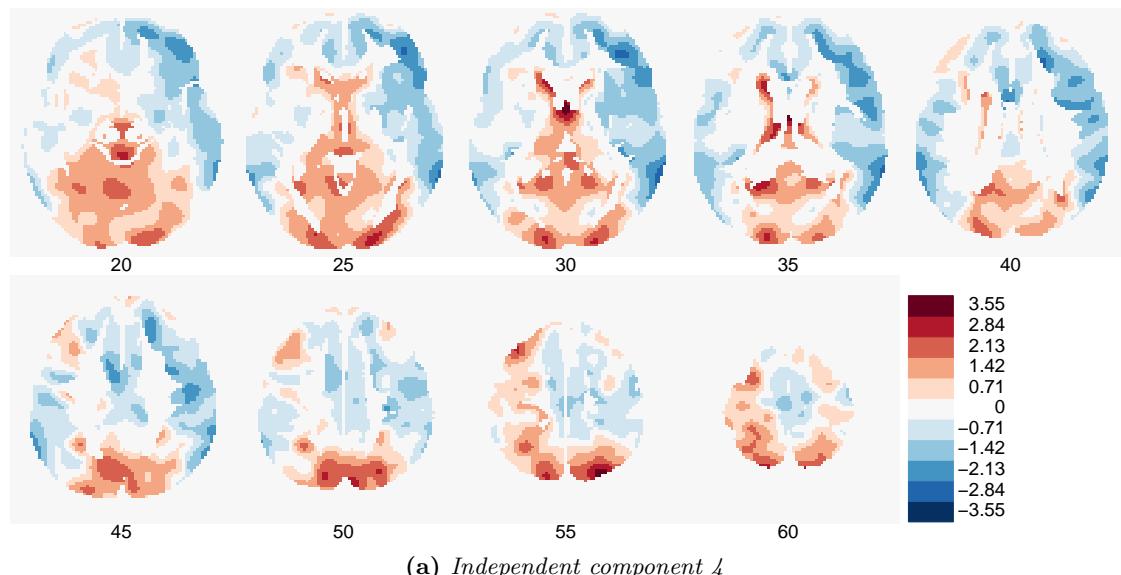


Figure 4.50.: Images of all Independent components 4 and 5 - red indicates increased, blue decreased relative cerebral blood flow. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side, right hemisphere is on the right hand side.

4.7.3. Negentropy or Kurtosis

Component 4 (which is significant in ICA) has a high negentropy. The negentropy of ICA is higher compared to PCA. The norm of PCA and ICA is about the same.

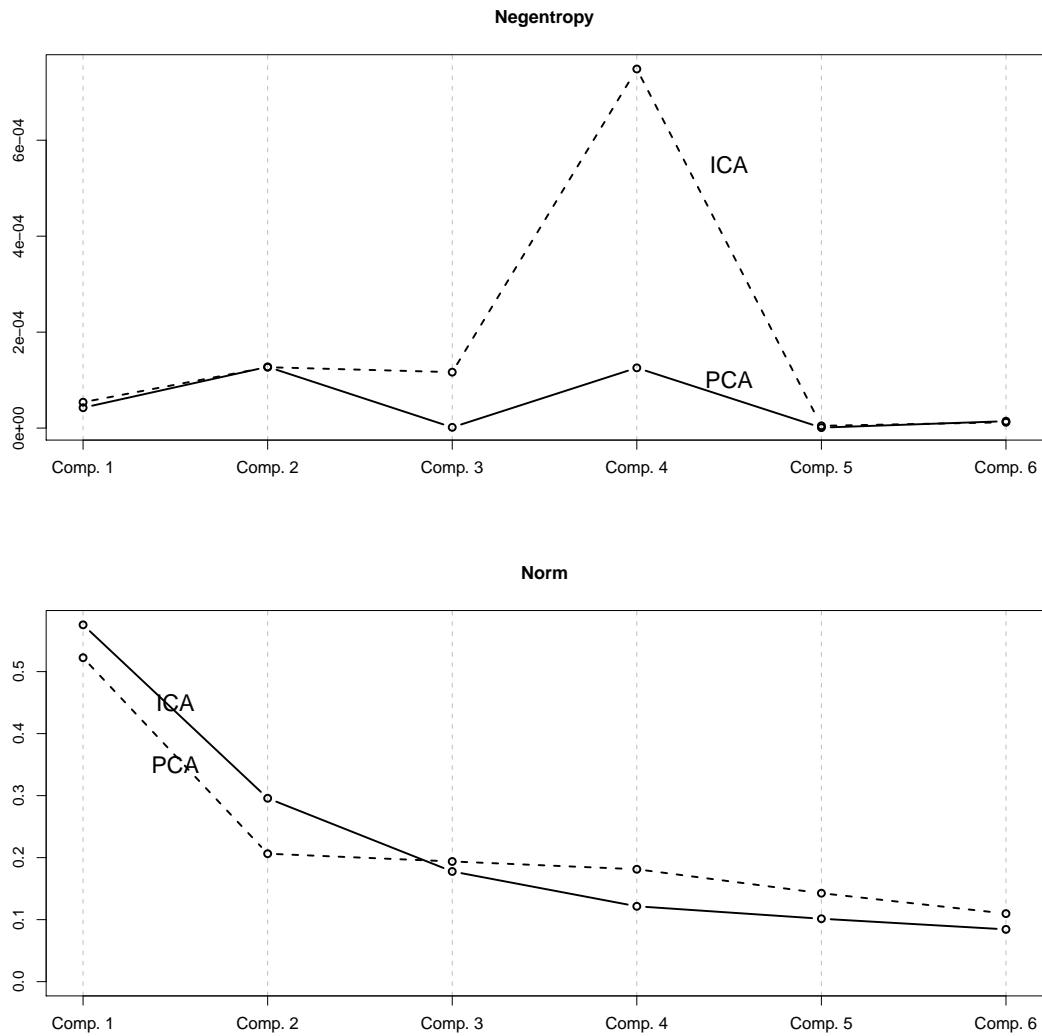


Figure 4.51.: Negentropy (top) of all components 1 to 6 and norm (bottom) for each columns in the mixing matrix. Significant components in PCA were 1, 2, 4 and 5, in ICA 1, 2, 3 and 5.

4.7.4. Performance of Model

The performance of the model can be measured with a leave-one-out cross-validation of the entire model. Table 4.18 and 4.19 are showing that PCA is slightly outperforming ICA. The difference between the performance of PCA and ICA is, that PCA classifies 1 PDU, 1 MCI and 1 PDD more. The error rate for PCA is 0.1, for ICA 0.15.

Figure 4.52 shows the ROC curves with an area under the curve of 0.9 for PCA, and 0.89 for ICA.

Confusion matrix of leave one out PCA (Error rate: 0.1):

	unimpaired (pred)	impaired (pred)	class.error
unimpaired	31	2	0.061
impaired	4	23	0.148

	unimpaired (pred)	impaired (pred)
PDU	31	2
MCI	4	12
PDD	0	11

Table 4.18.: Performance of entire model with PCA for cognition groups (cross-validation with leave-one-out).

Confusion matrix of leave one out ICA (Error rate: 0.15):

	unimpaired (pred)	impaired (pred)	class.error
unimpaired	30	3	0.091
impaired	6	21	0.222

	unimpaired (pred)	impaired (pred)
PDU	30	3
MCI	5	11
PDD	1	10

Table 4.19.: Performance of entire model with ICA for cognition groups (cross-validation with leave-one-out).

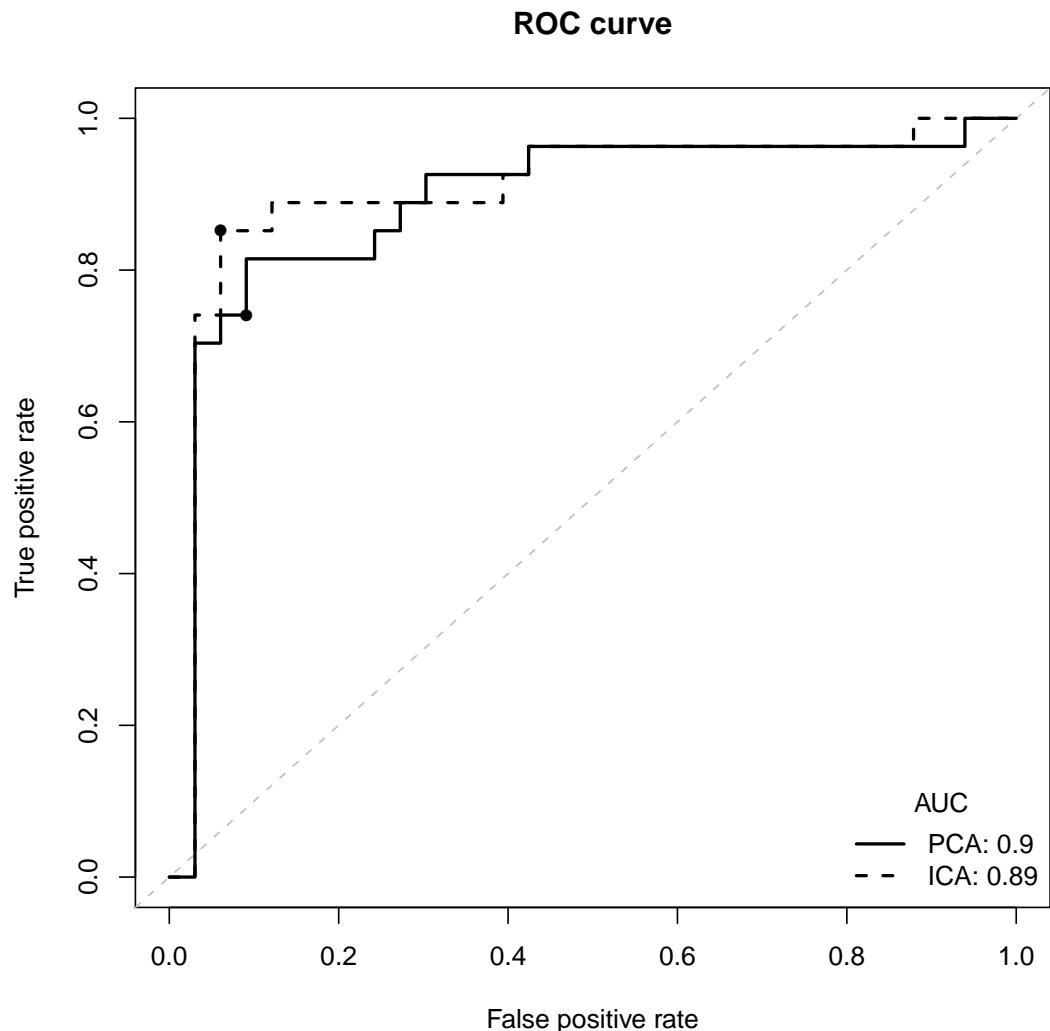


Figure 4.52.: ROC curve for entire model.

Figure 4.53 shows the prediction for the group labels unimpaired and impaired. The unimpaired subject with a prediction close to 1 is 11365 in both PCA and ICA. The 4 impaired subjects in PCA with a probability lower than 0.5 are 12391 (0.01), 8779 (0.23), 6255 (0.28), 8015 (0.41). ICA has 6 impaired subject with values lower than 0.5: 6255 (0.29), 8015 (0.50), 8779 (0.36), 11368 (0.38), 12209 (0.25) and 12391 (0.01). Figure 4.54 shows the predictions corresponding to the cognition levels PDU, MCI and PDD. PDD's lower than the cluster on the top are subjects 8014, 9118, 9554, 11368.

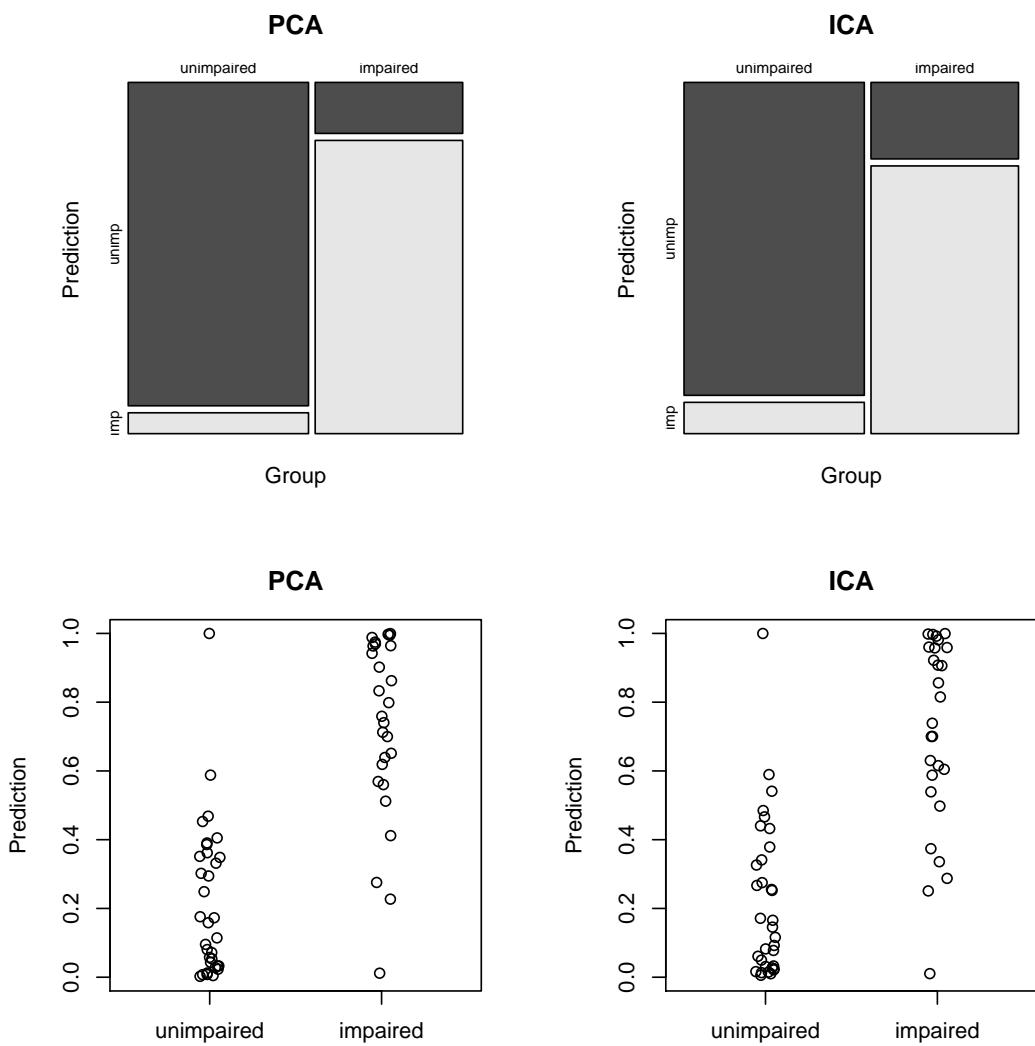


Figure 4.53.: Performance of entire model.

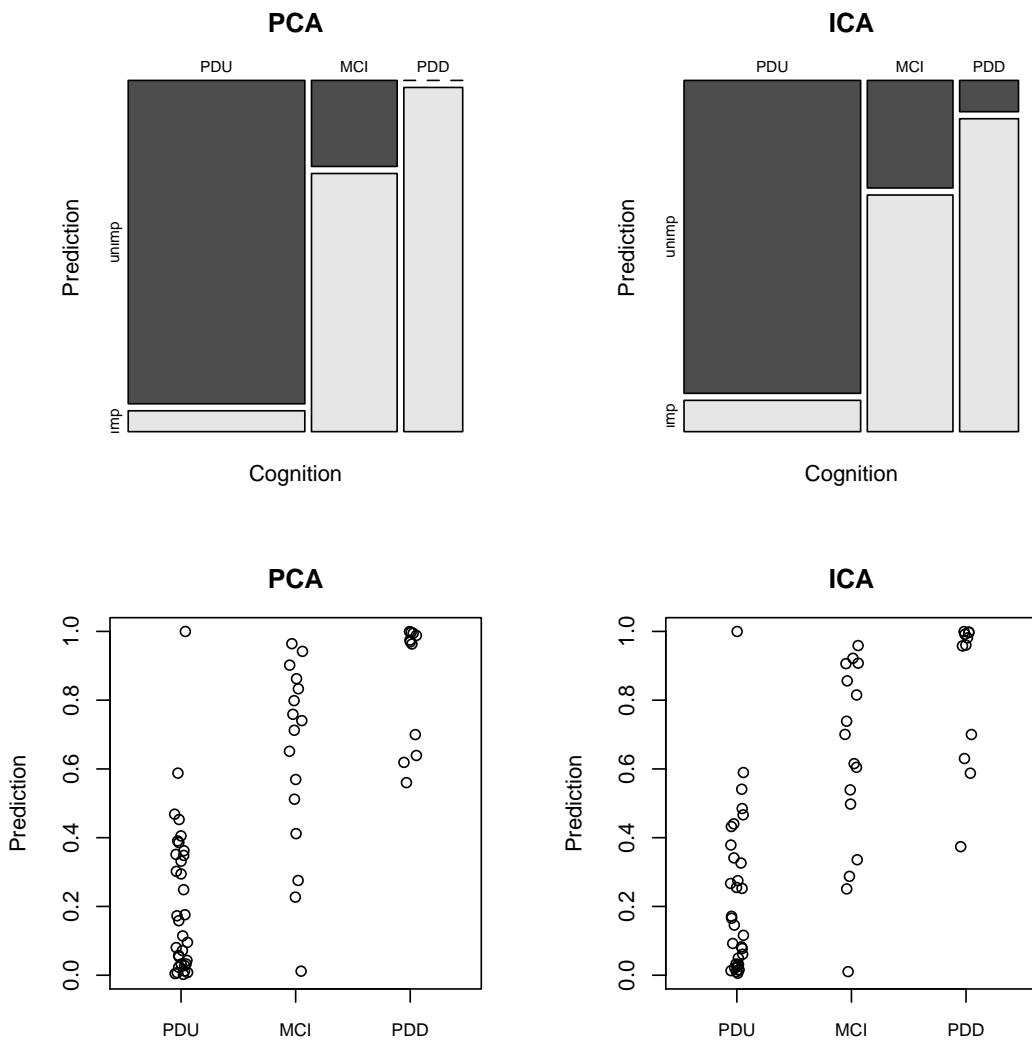


Figure 4.54.: Performance of entire model for cognition groups.

Misclassified subjects in PCA:

Impaired subjects 6255, 8015, 8779, 12391.

Unimpaired subjects 8863, 11365.

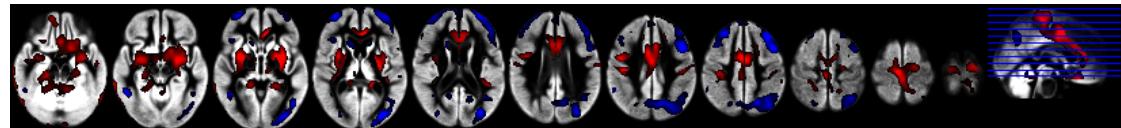
Misclassified subjects in ICA:

Impaired subjects 6255, 8015, 8779, 11368, 12209, 12391.

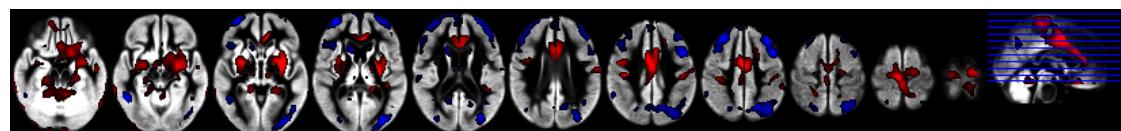
Unimpaired subjects 7595, 8863, 11365.

4.7.5. Bootstrap

Bootstrap was applied to measure the precision of the estimation. The estimation was then divided through its standard deviation. Figure 4.27 shows the image plots with highlighted significant areas, figure 4.55 shows a scatterplot for PCA and ICA.



(a) *Imageplot for PD related pattern (PCA, 5%)*



(b) *Imageplot for PD related pattern (ICA, 5%)*

Figure 4.55.: *Imageplot for PD related pattern: ICA and PCA and Z-Scores - blue indicates decreased relative cerebral blood flow and red increased relative cerebral blood flow.*

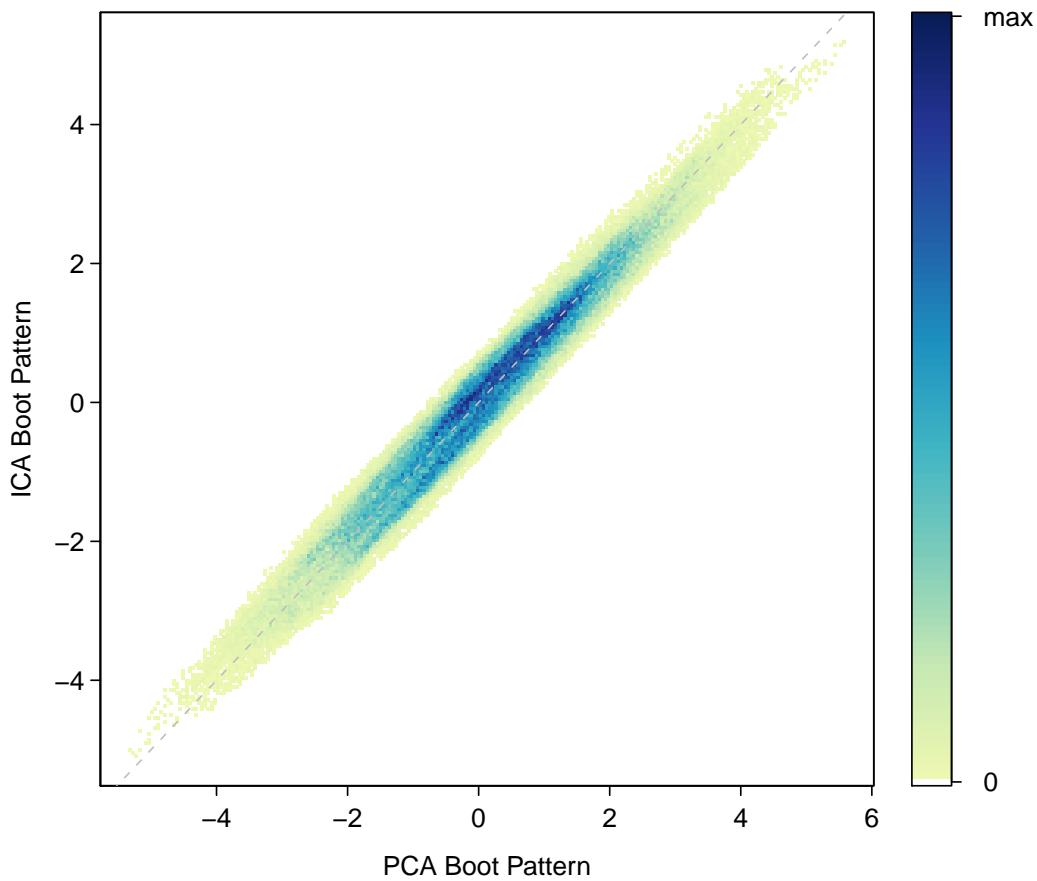
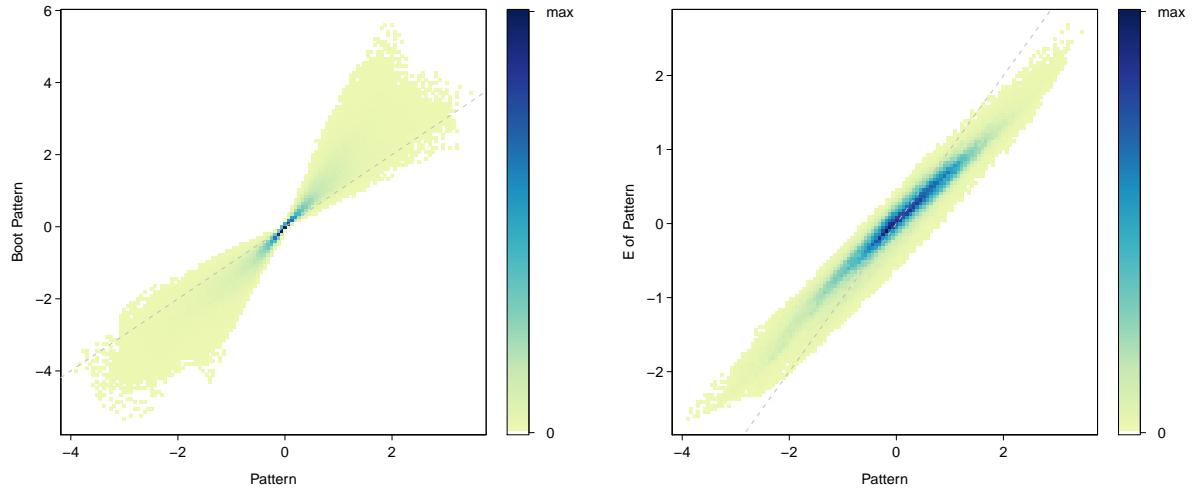


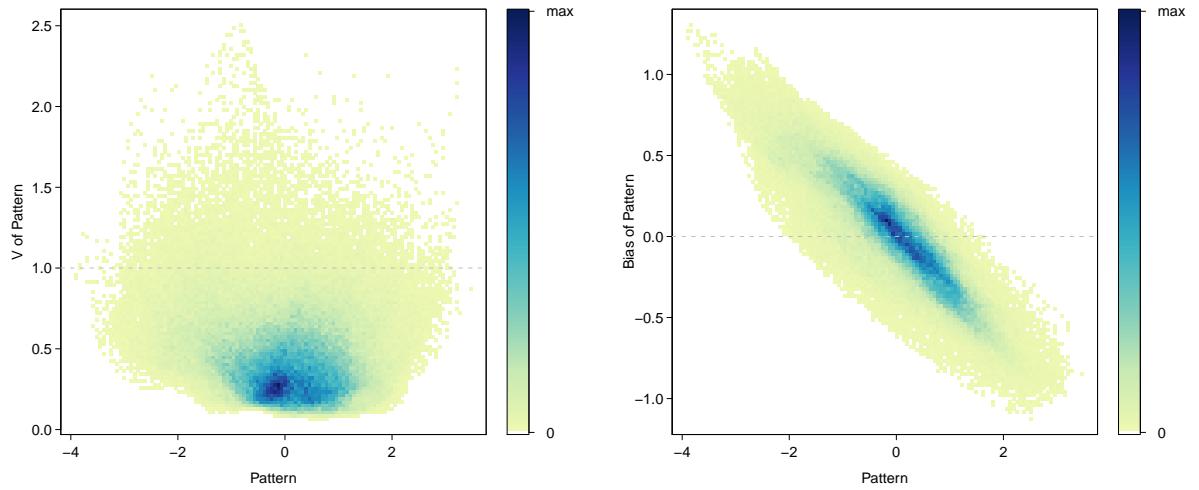
Figure 4.56.: Scatter-density-plot for PCA (x-axis) and ICA (y-axis) pattern. This plot shows that both pattern are almost the same.

The same problem of bias (see figure 4.57 and 4.57) as with the complete data set occur also with the subset of the data set. It is interesting that the high variance in the occipital region disappeared - which would indicate that some controls might have strange values in the occipital region.



(a) Y-axis: estimation for each voxel divided by its standard deviation.

(b) Y-axis: mean for each voxel.



(c) Y-axis: variance for each voxel.

(d) Y-axis: bias for each voxel.

Figure 4.57.: PCA: X-axis is representing the 'raw' PD-related pattern, y-axis is the bootstrapped PCA pattern (top left), the mean for each voxel (top right), the variance for each voxel (bottom left) and the bias for each voxel (bottom right).

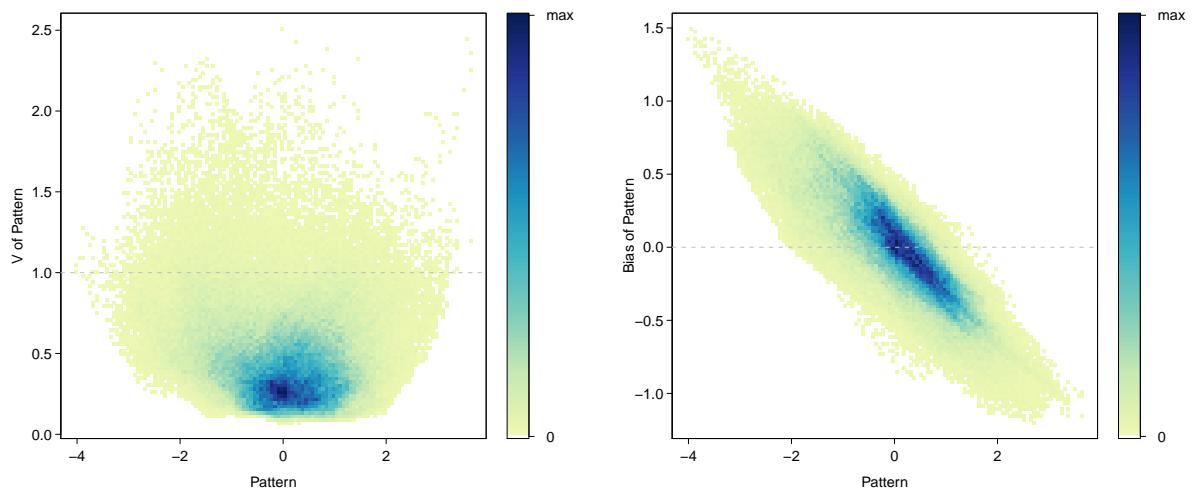
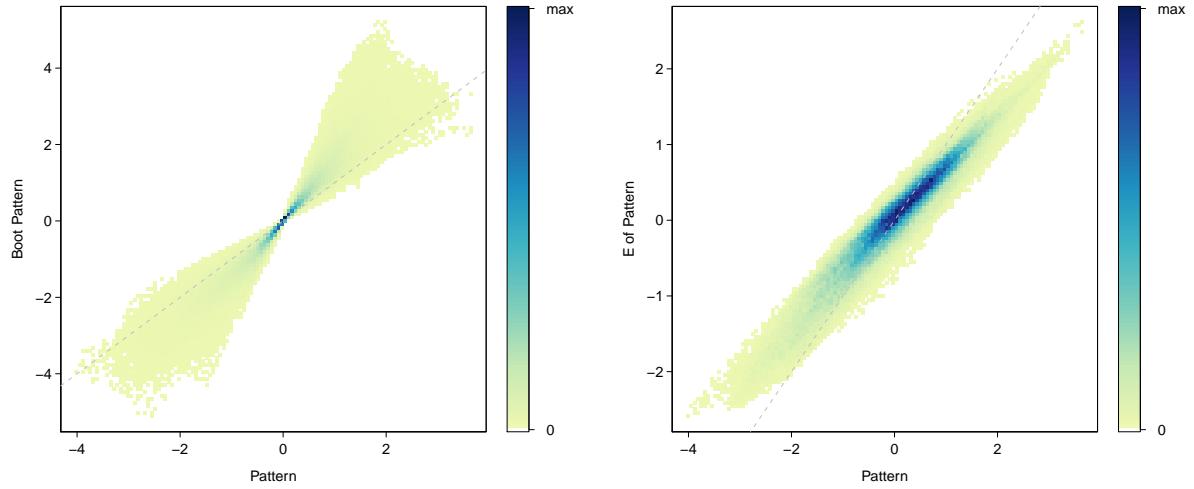


Figure 4.58.: ICA: X-axis is representing the 'raw' PD-related pattern, y-axis is the bootstrapped ICA pattern (top left), the mean for each voxel (top right), the variance for each voxel (bottom left) and the bias for each voxel (bottom right).

5. Discussion

The goals of this report were - as mentioned in the introduction:

1. Identify perfusion patterns which optimally distinguish Parkinson's disease participants from healthy controls.
2. Combine the patterns to define a typical Parkinson's disease brain-perfusion-pattern.
3. Describe, employ and compare the two possible methods for this type of pattern recognition analysis: principal component analysis (PCA) and independent component analysis (ICA).
4. Decide whether PCA or ICA is recommended for this type of analysis.

The perfusion pattern which optimally distinguished Parkinson's disease participants from healthy controls (1.) was identified in chapter 4.3. The Parkinson's disease brain-perfusion-pattern (2.) was built in chapter 4.4 and then combined with results of bootstrap in chapter 4.5. ICA and PCA (3.) were compared in chapter 4.6.

We also tried to identify components which are related to co-information of subject such as cognitive or motor scores, gender, age and duration of the disease. The results of single linear regressions are in appendix C.4.

This chapter should give the answer to item 4, whether PCA or ICA is recommended for this type of analysis. Let us remember - we performed PCA and ICA on two data sets. Once on a full data set with healthy and Parkinson's disease participants and once a subset with only Parkinson's diseased persons.

In our case ICA showed at least one high nongaussian component (see figures 4.33 and 4.51). This component turned out to be significant in the case of the subset data set. In the case of the full data set independent components showed similar negentropy than principal components.

With the entire data set the classification error¹ of a model with PCA is 0.31 and with ICA 0.24. Thus ICA performs clearly better than PCA. With the subset data set PCA has a classification error of 0.1 and ICA or 0.15, which would indicate the reverse - PCA outperforms ICA. But while looking on absolute misclassification - PCA has 6 and ICA 9 misclassified subjects - the difference of 5% are only 2 subjects. With the full data set the relative difference is 7% and the absolute 7 subjects.

¹with leave-one-out cross-validation

PCA is a standard method and applied since more than 100 years. Adjustments can be made in preparing the data set or choosing the correlation instead of the covariance. ICA was developed in the 90s and has adjustment parameters as well as many different implementations which will produce different results. Choosing the parameters carefully and selecting an appropriate algorithm can make the analysis with ICA challenging, or, on the other hand, much more adaptable. The computation time of ICA can be - depending how many components are chosen - much longer than PCA.

The leave-one-out cross-validation of the classification of the subset data set is very good with only 6 and 9 misclassifications respectively. This means that an unknown person with Parkinson's disease, is correct classified as cognitive unimpaired or impaired with 10% chance of a misclassification.

It is irritating that in one case ICA has a lower error rate and in once case PCA. In the case of the full data set (where ICA was better performing), the model was sufficient (not considering the bias in the estimation). Whereas in the case of the subset of data (where PCA was better performing), the logistic regression showed weaknesses in the graphical residual analysis and at least two outliers. We are assuming that with an appropriate classification model (e.g. robust estimation) PCA might be not outperforming ICA anymore.

We are recommending for this type of analysis a model with ICA, because it showed a lower misclassification and is in general a flexible method and - at least in R - fast and easy to handle.

5.1. Outlook

The work in this report was done during 3 months. Unfortunately these 3 months were not enough time to develop a perfect model. Some parts of the work are not satisfying enough done or could consider alternatives for future work.

Data preparation 5 subjects were excluded of the analysis due to missing values. Instead of excluding the subjects consider data imputation (might not work).

ICA Independent component analysis was applied with the function fastICA in R. Recently the ProDenICA algorithm was recently (May 2010) released in R, which might be useful to compare different ICA algorithms. ProDenICA is an abbreviation for *Product Density Estimation* and described in *Elements of Statistical Learning* [27]. Furthermore ICA is also available for non-linear generative processes.

Comparison PCA and ICA are producing not equal, but rather similar results. It would be good to know *why* - under which conditions - the results of PCA and ICA are similar (except the case when the data is normally distributed). For this the ICA algorithm has to be further explored. Also the comparison itself needs to

be improved. A possibility is to test whether PCA or ICA is more accurate, e.g. testing two area under under ROC curves (Wilcoxon-Test).

Robust regression For the logistic and the linear regression one should rather use robust regression or transform variables (e.g. to achieve linear relations). Robust regression is useful to detect outliers (which might not be detected by the graphical residual analysis of a linear regression) and to estimate the coefficients in a *robust* manner, so that the influence of the outliers on the estimation is marginal.

In particular the single linear regressions in appendix C.4 were not sufficient enough to capture the relations between the mixing values and the co-informations (that is the reason why the results are in the appendix). First the relation might be non-linear, thus transformation or alternatives to linear regression should be used. Secondly, existing outliers are a problem. Outliers are defined as observations which have a strong influence on a estimation. If such influencing observations appear one has to ask 2 questions: why they occur (may be errors in data preparation) and how to treat them. Outliers can be excluded from the analysis because they do not represent the population (if so) and distort the results (the location of the regression line). Another option is to apply *robust* regression. More informations to robust regression can be found in Heritier, S. et al., 2009. *Robust methods in biostatistics*, J. Wiley. In R package **robust** or **robustbase** are recommended.

A. Appendix A: Misc

A.1. Details to Cognitionlevels

Those with PDD (dementia) were diagnosed on the basis of the Movement Disorders Task Force criteria [5]. Those with MCI (mild cognitive impaired) did not show significant impairment in functional activities of daily living but showed performance at 1.5 standard deviation or more below norms on at least two standardized neuropsychological tests, in at least one of the Task Force cognitive domains (1. Executive function; 2. Memory; 3. Attention, working memory and speed of processing; and 4. Visuospatial/visoerceptual function). Other PD's were classed as PDU (unimpaired). Global cognitive status was assessed using the Montreal Cognitive Assessment.

A.2. Details to Mosaicplot

A mosaicplot is a possibility to visualize two or more categorical variables. A advantage to confusion matrices is, that the relative amount of a certain level is showed very clearly. Article *Constructing and reading mosaicplots* from Hoffmann [29] offers more informations and possibilities about mosaicplots. The following 3 examples show respectively a mosaicplot and a corresponding table:

Control	31
PD	64

Table A.1.: Number of Parkinson's disease and healthy controls.

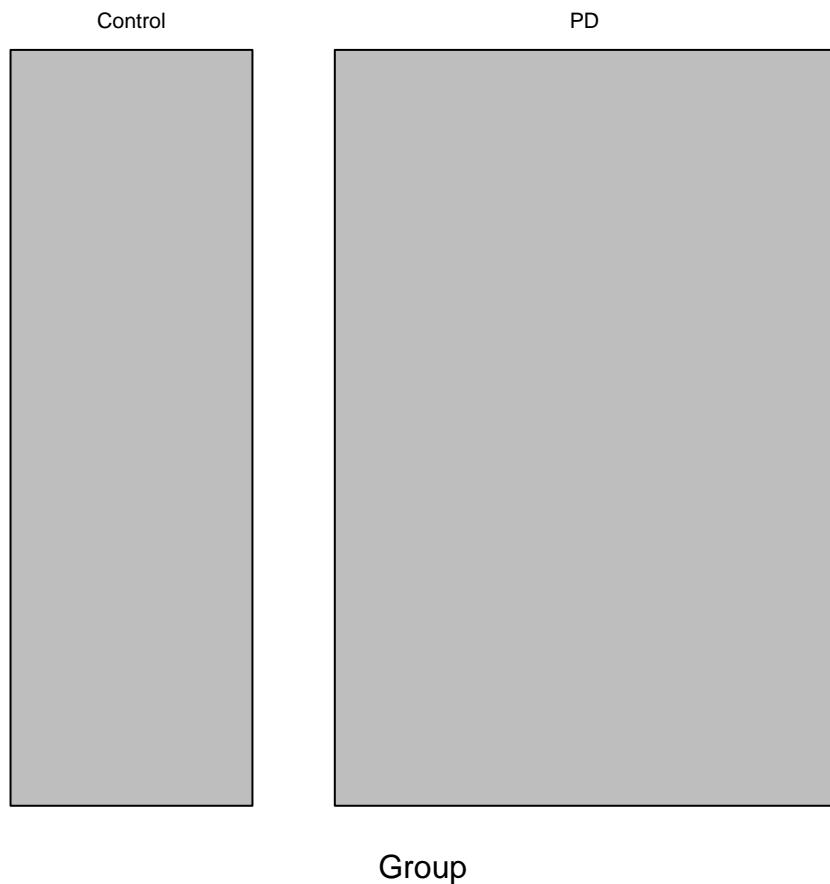


Figure A.1.: Mosaicplot: the length of the boxes in x-direction shows the proportion of controls (about 1/3) and PD's (about 2/3).

	Control	PD
male	21	46
female	10	18

Table A.2.: Number of Parkinson's disease and healthy controls splitted up by Gender.

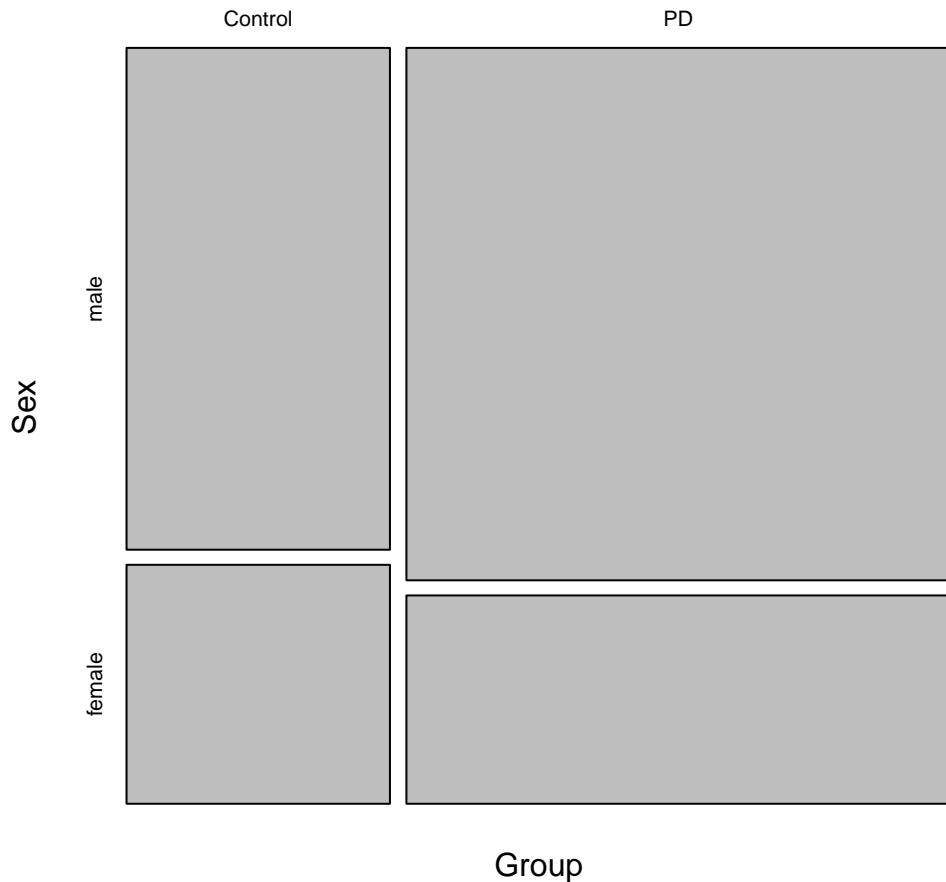


Figure A.2.: Mosaicplot: length of boxes in x-direction shows again the proportion of controls and PD's, whereas the length of boxes in y-direction is indicating the proportion of men and women within controls or PD. This mosaicplot shows that controls and PD's are splitted up in 1/3 women and 2/3 men, but the proportion of women is in PD's slightly higher than in the control group.

	Control	PDU	MCI	PDD
male	21	24	13	9
female	9	12	5	2

Table A.3.: Number of Cognition level groups and healthy controls splitted up in men and women.

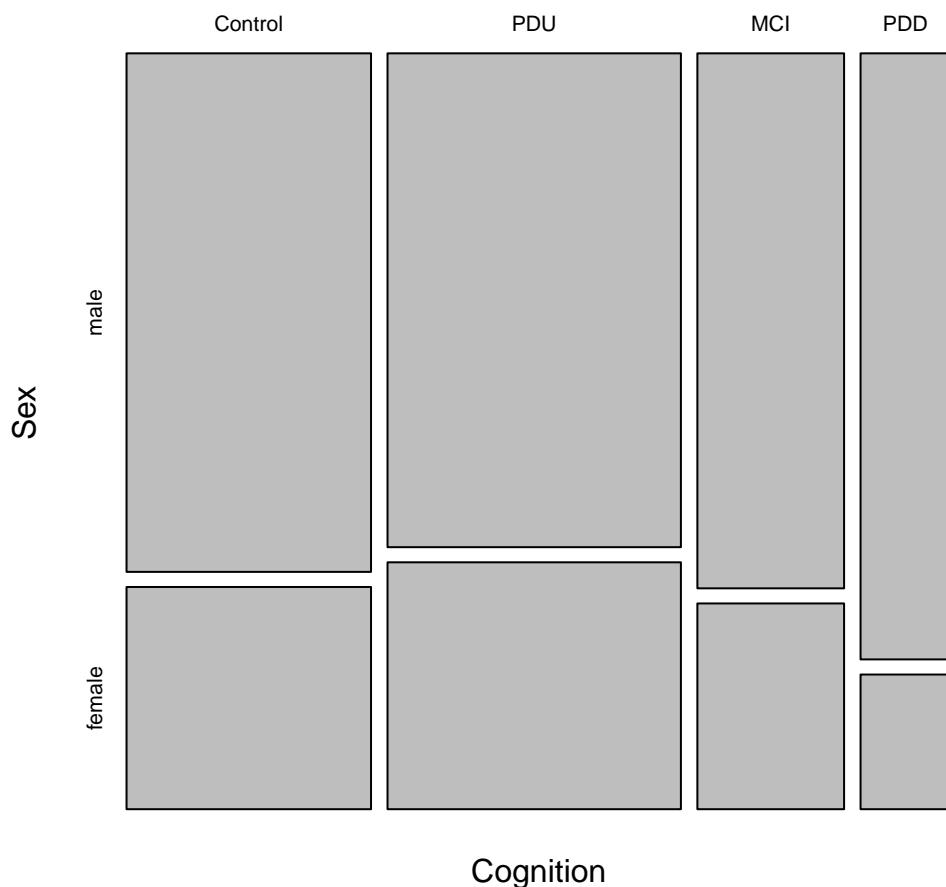


Figure A.3.: Mosaicplot shows the relative size of cognition levels (small to big: PDD, MCI, control, PDU) and that PDD's have the smallest proportion of women and PDU's the biggest.

B. Appendix B: Pseudo Code

R pseudo code of each model step is listed in the gray box below.

B.1. Data preparation

```
# Packages
# -----
library(dcemriS4)          # Read NIfTI Files

# Parameter
# -----
thresh <- 0.10

# Import Gray Matter (GM) Mask File
# -----
gmMask <- readNIIfTI('raverage_GM_mwcids_30cont_10PDU_10MCI_10PDD.nii')

# Trim the GM mask so that it is only runs from [,,15:65]
# (necessary for early ASL scans which had a massive spiral
# artefact at the level of brain stem/inferior cerebellum)
# -----
dim.gmMask <- dim(gmMask)
gmMask[1:dim.gmMask[1], 1:dim.gmMask[2], c(1:14, 66:69)] <- 0

# Distinguish white matter (out), gray matter (in)
# -----
ind_out <- which(gmMask < thresh)      # White Matter
ind_in <- which(gmMask >= thresh)       # Gray Matter

# Import Cerebral Blood Flow Data
# (each person has an own image file)
# -----

inputfile <- dir()          # names of files

volSize <- length(inputfile)
cbf.temp <- matrix(0, nrow = volSize, ncol = prod(dim.gmMask))

for (i in 1:volSize)
{
    dat.temp <- readNIIfTI(inputfile[i])
    dat.temp.shift <- dat.temp@scl_slope*dat.temp + dat.temp@scl_inter
    # read below

    ## The data of some images are not correctly scaled.
    ## Therefore the output 'scl.slope' and 'scl.inter' are used.
    ## For more information read in the informations in the package dcemriS4.

    cbf.temp[i, ] <- c(dat.temp.shift, recursive = TRUE)
```

```

}

# Set White Matter to zero
# -----
cbf <- cbf.temp
cbf[,ind_out] <- 0

# Exclude the White Matter
# -----
cbf.temp <- cbf[,ind_in]

# Take the logarithm of the CBF values
# -----
cbf.log <- log(cbf.temp)

# Create Subject Residual Profile (SRP)
# ----

# a) Row Averages
row.av <- apply(cbf.log, 1, mean)

# b) Subtract Row Averages
srp.temp <- cbf.log - row.av

# c) Column Averages
col.av <- apply(srp.temp, 2, mean)

# d) Subtract Column Averages
srp <- t(t(srp.temp) - col.av)

```

B.1.1. Projection Methods: PCA

```

# Parameters
# -----
n.comp <- 6

# Calculating Covariance matrix of SRP values
# -----
cov.mat <- cov(srp)                                     # SRP Matrix has dimension (J x I)

# Calculating Eigenvectors and Eigenvalues
# -----
v <- eigen(cov.mat)$values                            # Eigenvalues
E <- eigen(cov.mat)$vectors                          # Eigenvectors

# Set negative eigenvalues to 0
# -----
v[v<0] <- abs(v[v<0])

# Calculate Loadings
# -----
W <- diag(1/v)
W.h <- diag(1/v^(1/2))
W.mh <- diag(1/v^(-1/2))
loadings.div <- E %*% W.h                         ## A = E %*% W^(1/2), dimension (I x K)
loadings.mult <- t(W.mh %*% t(E)) ## (A^(-1))^T = (W.mh %*% E^T)^T, dimension (I x K)

# Calculate Principal Components
# -----

```

```
components <- srp%*%loadings.div
```

B.2. Projection Methods: ICA

```
# Set Random Seed (takes always the same random seed)
# -----
set.seed(33)

# Package
# -----
library(fastICA)

# Parameters
# -----
n.comp <- 6

Perform ICA
# -----
ica.output <- fastICA(srP, n.comp, method = 'C', row.norm = FALSE, alg.typ =
  'deflation', maxit = 500, fun = 'logcosh', alpha = 1, tol = 0.0001, verbose =
  TRUE)
# SRP has dimension (J x I)

# Mixing Matrix
# -----
mixMatrix <- t(x$A)

# Order Components
# -----
norm.MM <- (apply((mixMatrix)^2, 2, sum))
order.MM <- order(norm.MM, decreasing=TRUE)

mixMatrix <- mixMatrix[,order.MM]

# Components
# -----
ic <- x$S[,order.MM]
```

B.3. Classification

```
# Packages
# -----
library(MASS)          # for stepAIC()

# Parameters
# -----
n.comp <- 6

# Import 2nd Information of Subjects (MoCA, etc. )
# -----
load('dat_subj.rda')

# Import Mixing Matrix
# -----
load('mixMatrix.rda')
mixMatrix <- data.frame(mixMatrix[,1:n.comp])    # Choose Number of Components
```

```

# Merge 2nd Information of Subjects and Mixing Matrix
# -----
dat <- merge(dat.subj, mixMatrix, by='Subject_scan_number')

# Perform Logistic Regression
# -----
mod <- glm(Group ~ . - Subject_scan_number, family = binomial(link = 'logit'),
           data = dat)

# stepwise AIC in backward direction
# -----
mod.step <- stepAIC(mod, direction = 'backward')

```

B.4. Assessment of Model: Cross-validation

```

# Parameters
# -----
n.comp <- 6

# Import Gray Matter (GM) Mask File
# -----
gmMask <- readNIfTI('raverage_GM_mwcids_30cont_10PDU_10MCI_10PDD.nii')

# Import Rdata Files
# -----
# Vector with TRUE or FALSE for Prob>0.1 that a voxel is Gray Matter
load('ind_in.rda')

# CBF log data
load('cbf_log.rda')

# Subject Numbers
load('inputfile.rda')

# 2nd Information (e.g. Age, Sex, ...)
load('dat_subj.rda')

# Start Cross-Validation
# -----
for (k in 1:nrow(cbf.log))
{
  # =====
  # I) Data preparation
  # =====

  # 'In Bag' observations k
  cbf.log.ib <- cbf.log[-k,]

  # 'Out of Bag' observation k
  cbf.log.oob <- cbf.log[k,]

  # Create Subject Residual Profile
  # for 'In Bag' Observations
  row.av <- apply(cbf.log.ib, 1, mean)
  srp.ib.temp <- cbf.log.ib-row.av

```

```

col.av <- apply(srps.ip.temp, 2, mean)
srps.ip <- t(t(srps.ip.temp)-col.av)

# Create Subject Residual Profile
# for 'Out of Bag' Observations
srps.oob.temp <- cbf.log.oob-mean(cbf.log.oob)
srps.oob <- srps.oob.temp-col.av

# =====
# II) Perform PCA
# =====

# a) cov matrix
cov.mat <- cov(t(srps.ip))

# b) Calculating Eigenvalues and Eigenvectors
v <- eigen(cov.mat)$values
E <- eigen(cov.mat)$vectors

# Set negative eigenvalues to its absolute value
v[v < 0] <- abs(v[v < 0])

# c) compute mixing and un-mixing amtrix
W <- diag(1/v)
W.h <- diag(1/v^(1/2))
W.mh <- diag(1/v^(-1/2))

# Un-mixing Matrix A
unmixMatrix <- E %*% W.h

# Mixing Matrix A^{-1}
# mixMatrix <- t(W.mh %*% t(E))

# d) principal components
pc <- t(srps.id) %*% unmixMatrix

# e) calculate mixing matrix for 'out of bag' observation
# A^{-1} = Y^{-1} %*% X
mixMatrix.oob <- t(ginv(pc) %*% srps.oob)

# =====
# III) Logistic Regression
# =====

# a) Merge Grouplevels and Mixing Matrix
# -----
dat.id <- merge(dat.subj$Group, t(mixMatrix))
dat.oob <- data.frame(dat.subj$Group, t(mixMatrix.oob))

# a) Model + stepAIC with 'In Bag' Data
# -----
mod <- glm(Group ~ .-Subject_scan_number, family = binomial(link='logit'),
           data = dat.id)
mod.step <- stepAIC(mod, direction='backward')

# b) Prediction for 'Out of Bag' Data
# -----
pred.loo[k, ] <- data.frame(dat.oob, pred.response = predict(mod.step, type
                  = 'response', newdata = dat.oob),
                  pred.link = predict(mod.step, type = 'link', newdata = dat.oob))

}

```

B.5. Assessment of Model: Bootstrap

```

# Parameters
# =====
n.comp <- 6
m <- 500

# Import Rdata
# =====
# Vector with TRUE or FALSE for Prob>0.1 that a voxel is Gray Matter
load('ind_in.rda')

# SRP data
load('srp.rda')

# Subject Numbers
load('inputfile.rda')

# 2nd Information (e.g. Age, Sex, ...)
load('dat_subj.rda')

# load PD-related pattern after classification
load('network_pat_temp.rda')

# Empty matrix for storing results
# row (voxels) and columns (m runs)
# =====
nw.pat.boot <- matrix(NA, nrow=dim(srp)[2], ncol=m)

# =====
# Bootstrap
# =====

for (i in 1:m)
{

  # i) Boot subjects: Sample with replacement
  # -----
  subj.boot <- sample(1:length(inputfile), length(inputfile), replace=TRUE)

  # ii) Boot subjects: Sample with replacement
  # -----
  cbf.log.boot <- cbf.log[subj.boot,]

  # iii) Creating SRP matrix with boot subjects
  # -----
  row.av <- apply(cbf.log.boot, 1, mean)
  srp.boot.temp <- cbf.log.boot-row.av
  col.av <- apply(srp.boot.temp, 2, mean)
  srp.boot <- t(t(srp.boot.temp)-col.av)

  # iv) Performing PCA or ICA
  # -----
  x <- fastICA(t(srp.boot), n.comp, method = 'C', row.norm = FALSE, alg.typ
    = 'deflation', ...)

  # v) Extracting Mixing Matrix and Components
  # -----
  mixMatrix <- t(x$A)
}

```

```
components <- x$S

# vi) Order components (only for ICA)
norm.MM <- (apply((mixMatrix)^2, 2, sum))
order.MM <- order(norm.MM, decreasing = TRUE)
mixMatrix <- mixMatrix[,order.MM]
components <- components[,order.MM]

# vii) Logistic Regression
# -----
dat.boot <- merge(dat.subj$Group, t(mixMatrix))
mod <- glm(Group ~ .-Subject_scan_number, family = binomial(link = 'logit',
  ), data = dat.boot)
mod.step <- stepAIC(mod, direction = 'backward')

# viii) Build PD-related Pattern
# -----
coef <- mod.step$coef
k <- which(names(mixMatrix) %in% names(coef[-1]))
nw.pat.boot[,i] <- coef[1]+apply(t(components[,k])*coef[-1]), 1, sum)

}

# store the different measures (mean, variance, bias, bootpattern)
# in a separate data frame
# =====
dat.boot <- data.frame(pattern = nw.pat.temp, pattern_mean_boot = apply(nw.pat.
  boot
+ , 1, mean), pattern_var_boot = apply(nw.pat.boot, 1, var))
dat.boot$pattern_boot <- dat.boot$pattern/sqrt(dat.boot$pattern_var_boot)
dat.boot$pattern_bias_boot <- dat.boot$pattern_mean_boot-dat.boot$pattern
}
```

C. Appendix C: Results

C.1. Data preparation

Figure C.1, C.2 and C.3 give an overview of the *natural logarithmic cerebral blood flow data*, $\log\langle CBF \rangle$. The mean $\log\langle CBF \rangle$ for PD is 5.86 and for controls 5.97. Difference between the means is statistically significant with p-value $< 2.2e - 16$ (nonparametric Wilcoxon-Mann-Whitney test for 2 independent samples with $H_0 : \mu_{control} = \mu_{PD}$, assumption: the distributions of controls and PD are the same with a shift).

The logarithmic cerebral blood flow data is then transformed to a *subject residual profile (SRP)*. Figure C.4, C.5 and C.6 give an overview of the SRP matrix. The mean SRP for all subject is 0.

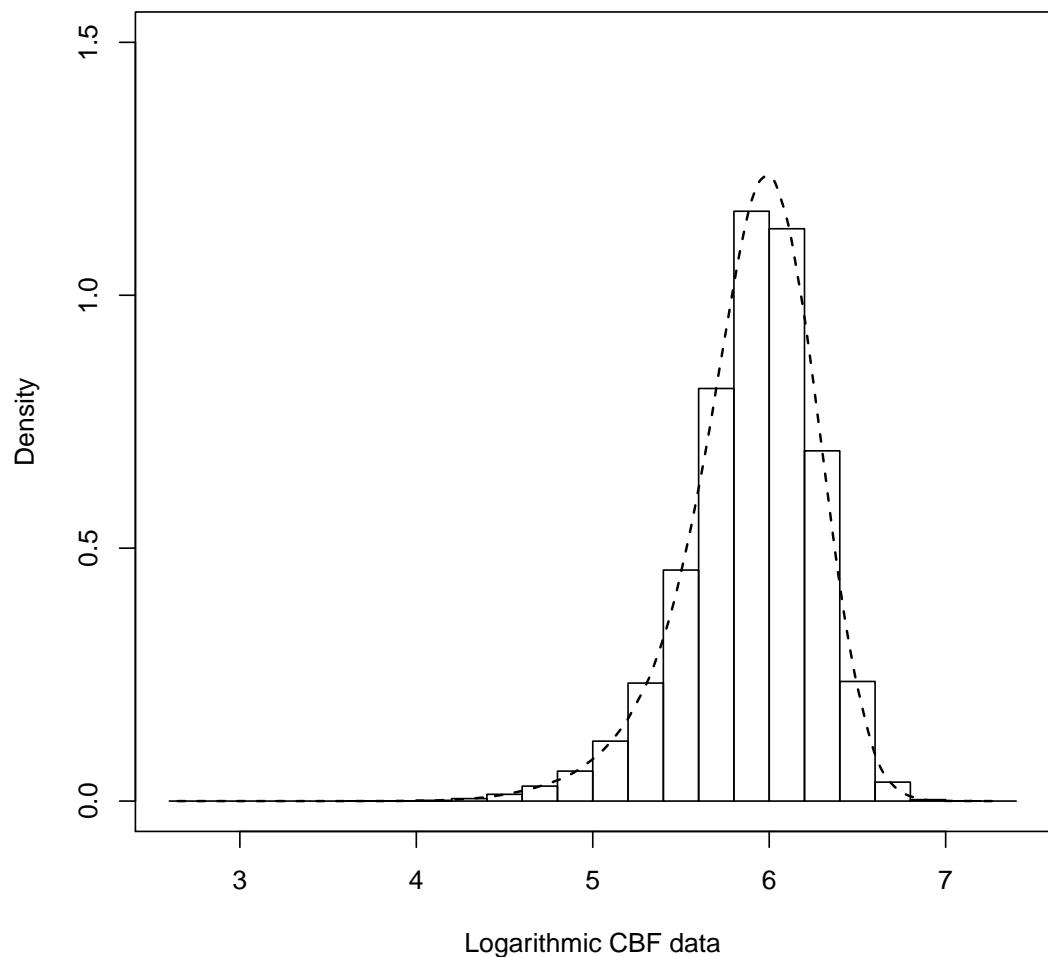


Figure C.1.: Histogram of cerebral blood flow (CBF) data.

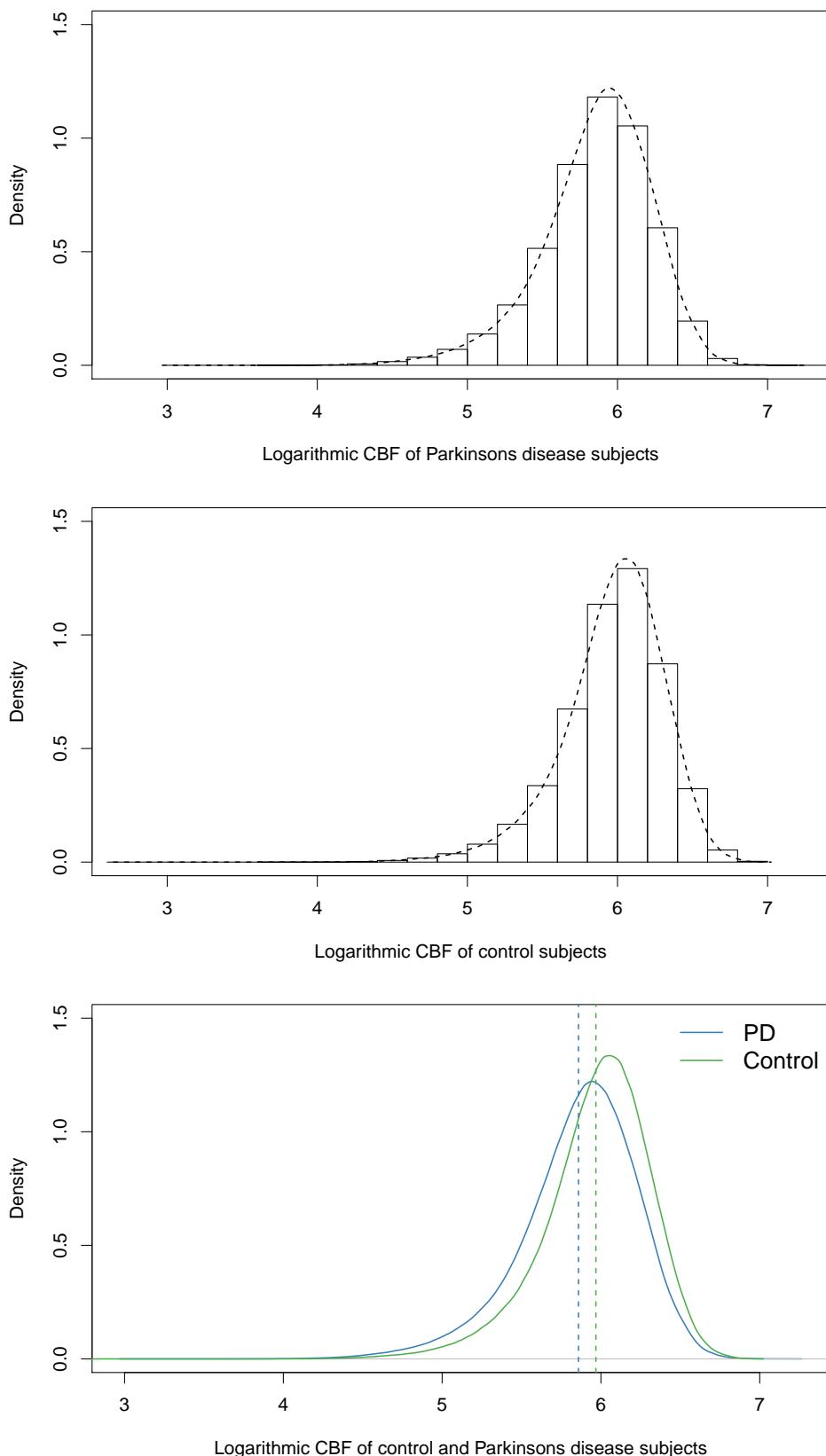


Figure C.2.: Histogram and density for Parkinson's disease (top) and control participants (middle) and together (bottom). Difference between the means (indicated by the green and blue dotted line) is statistically significant.

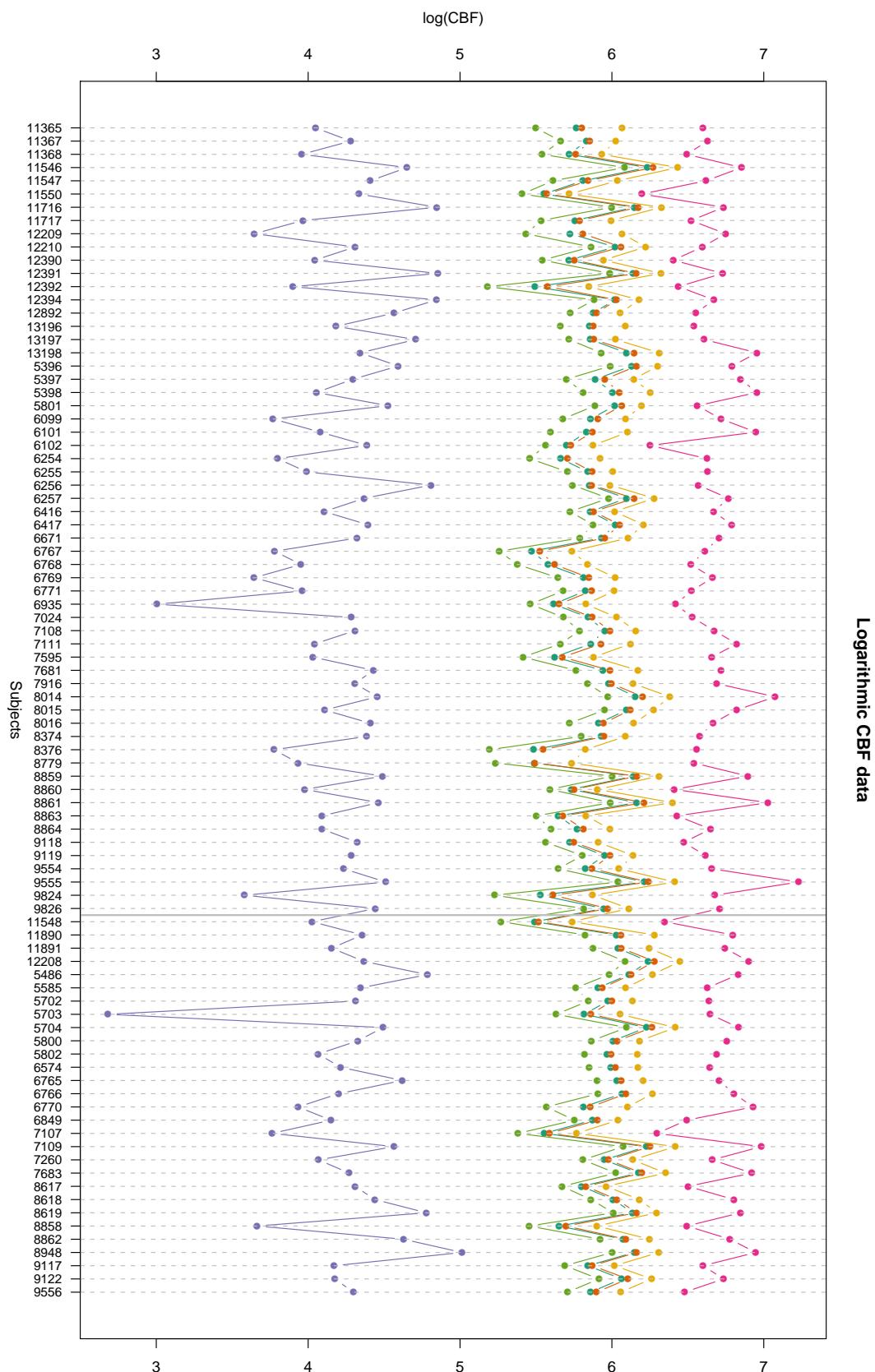


Figure C.3.: Minimum, 1st quartile, mean, median, 3rd quartile and maximum of logarithmic cerebral blood flow (CBF) data. Subjects 11365 to 9826 are Parkinson's disease participants, 11548 to 9556 are controls. This graph was used to detect extreme low values.

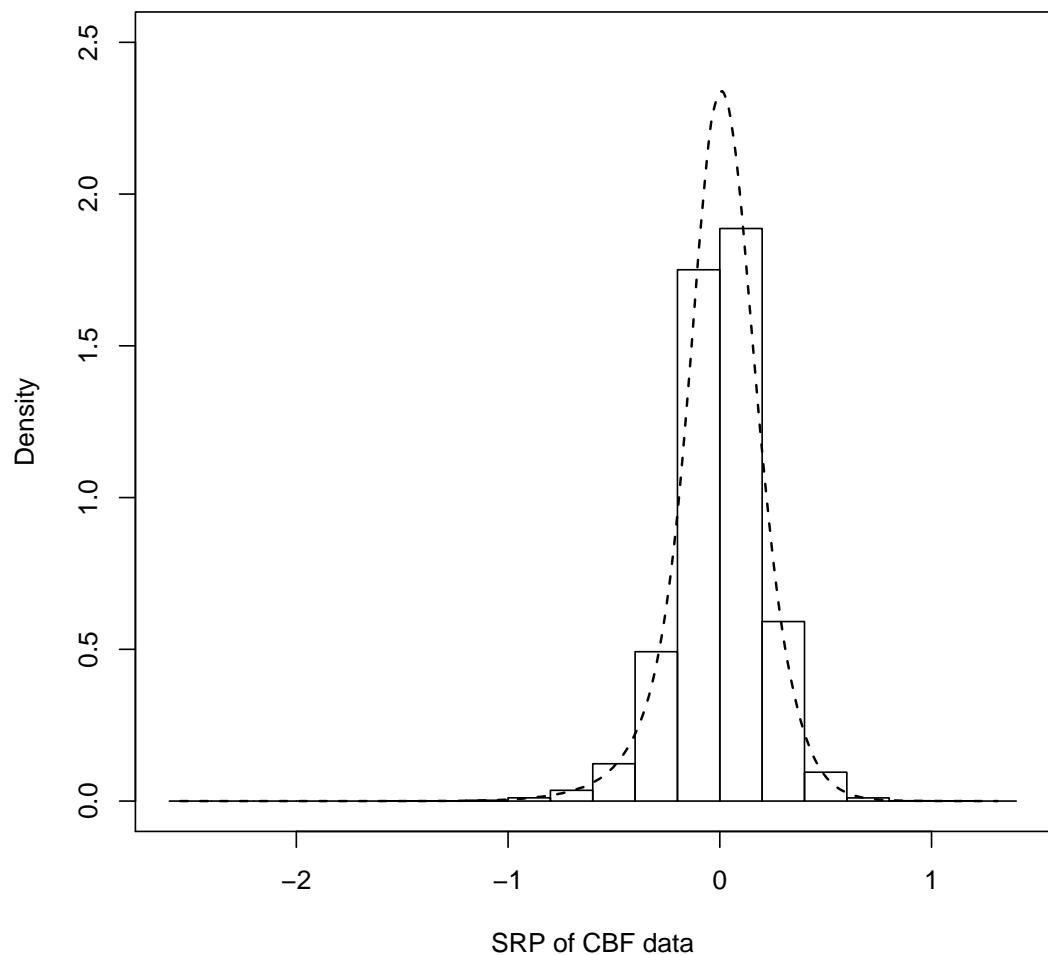
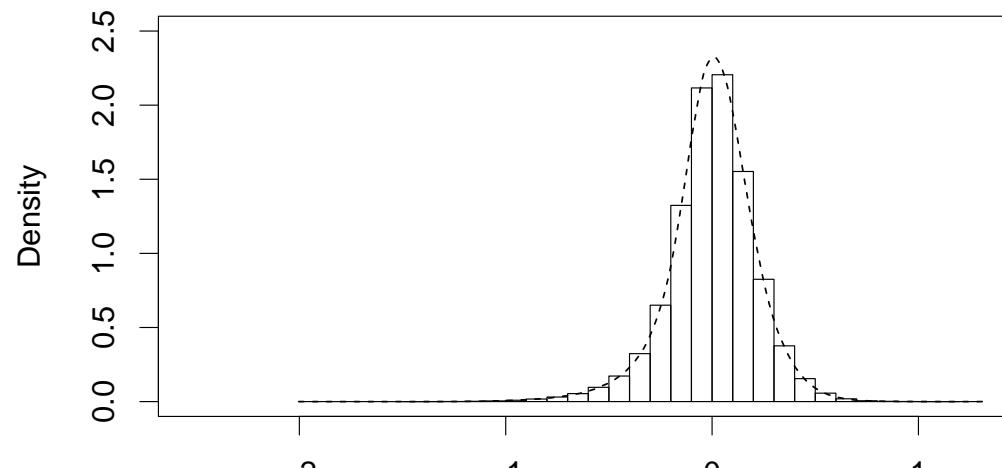
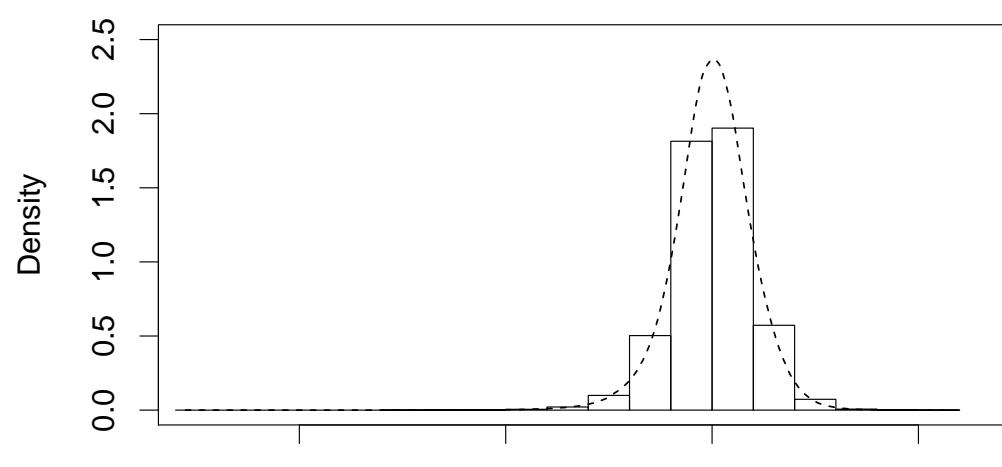


Figure C.4.: Histogram of the subject residual profile (SRP) of logarithmic cerebral blood flow (CBF) data.



SRP of Parkinson's disease subjects



SRP of control subjects

Figure C.5.: Histogram and density of the subject residual profile (SRP) of logarithmic cerebral blood flow (CBF) for Parkinson's disease (top) and control participants (bottom). The mean is for both 0.

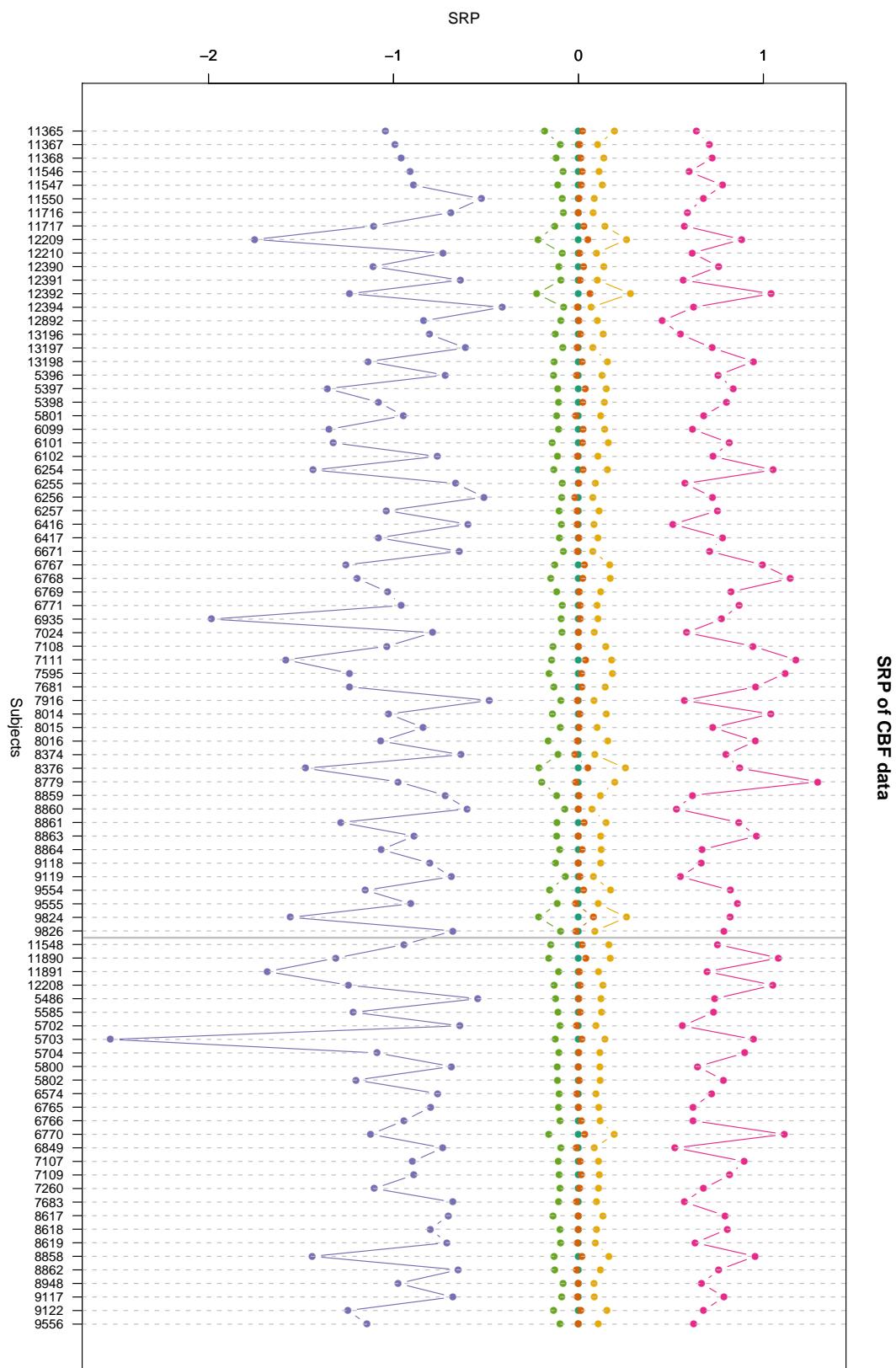
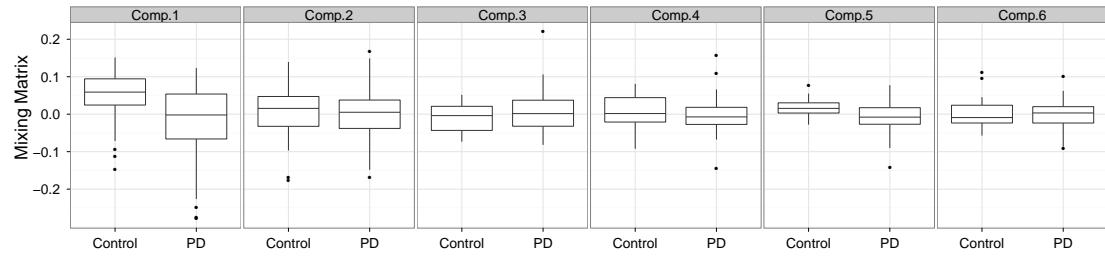


Figure C.6.: Minimum, 1st quartile, mean, median, 3rd quartile and maximum of the subject residual profile (SRP) of cerebral blood flow (CBF) data. Subjects 11365 to 9826 are Parkinson's disease participants, 11548 to 9556 are controls. This graph was used to detect extreme low values, then these observations were checked if the low values were the actual values or created by data preprocessing.

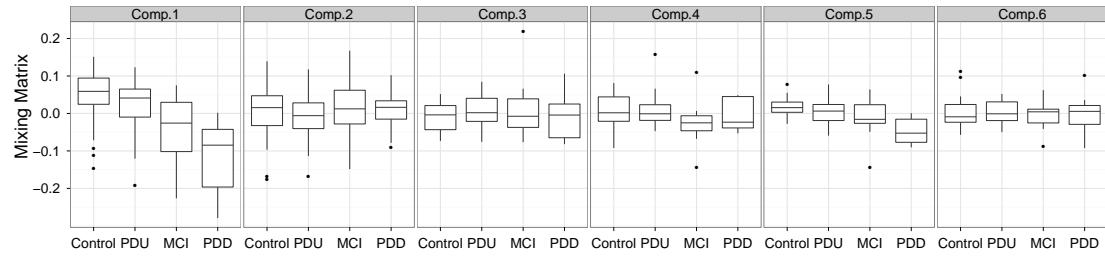
C.2. Projection Methods: PCA

Mixing Matrix A^{-1}

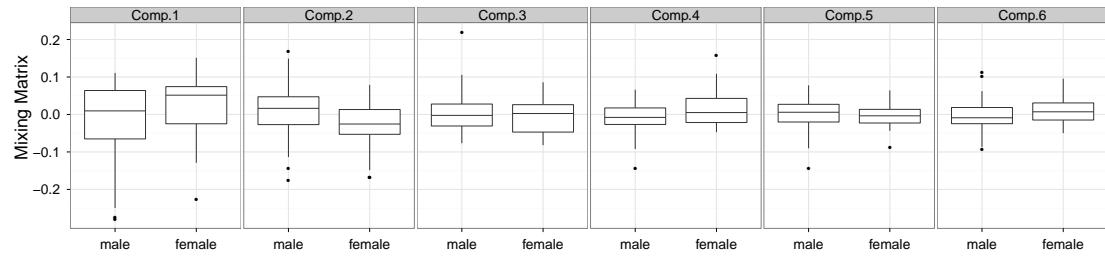
Figure C.7 and C.8 show the relation between the columns in the mixing matrix (expression of a particular component) and variables such as group, cognition, sex, age, MoCA, UPDRS, Hoehn & Yahr and Duration.



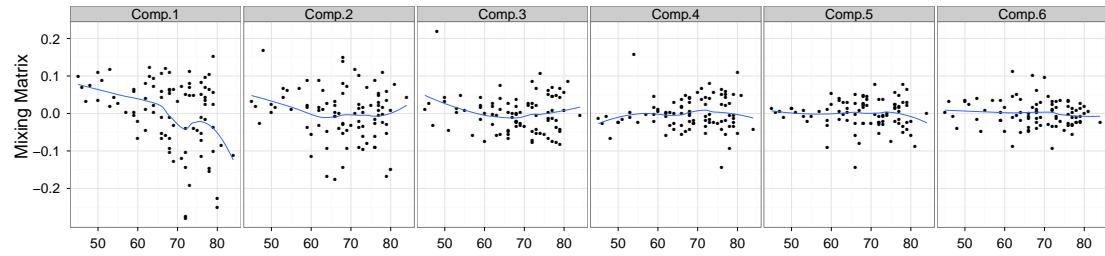
(a) Group (controls, PD): Controls seem to visually have higher values in the mixing values of component 1 and 5 than PD's.



(b) Cognition (controls, PDU, MCI, PDD): Negative trend in mixing values of component 1 and 5.

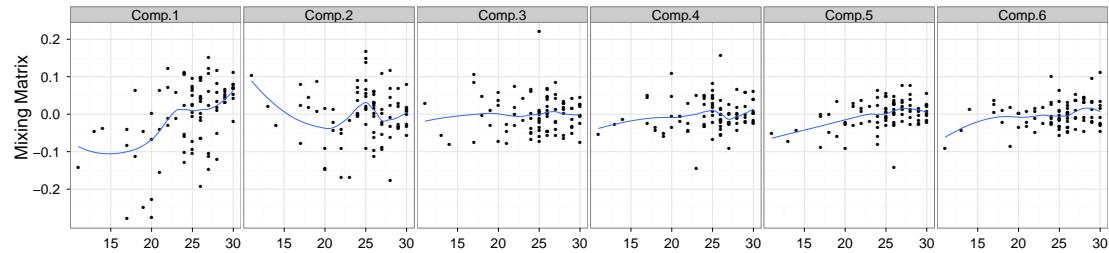


(c) Sex (male, female): Component 1 and 2 seem to distinguish male and female.

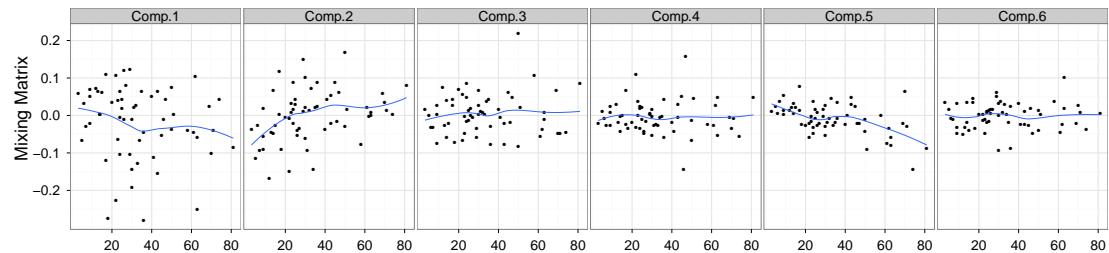


(d) Age: The blue line is a lowess smoother to emphasize the relation. Subjects around 50 seem to have positive mixing values in component 1 whereas older subjects can have the full range of possible values.

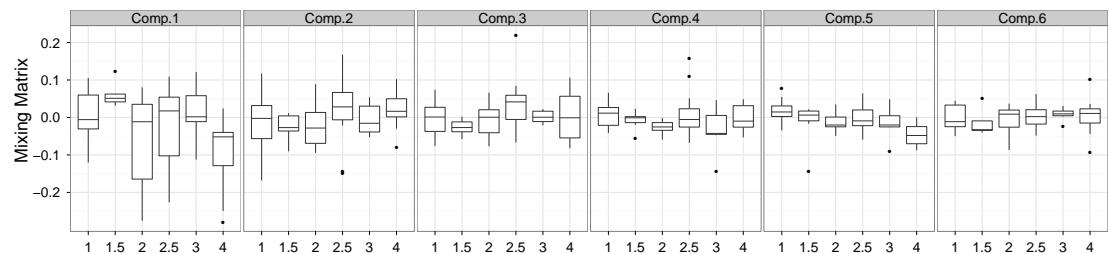
Figure C.7.: Y-axis is representing the columns of the mixing matrix, x-axis is representing group, cognition, sex and age. Component 1 seems to be important for all variables.



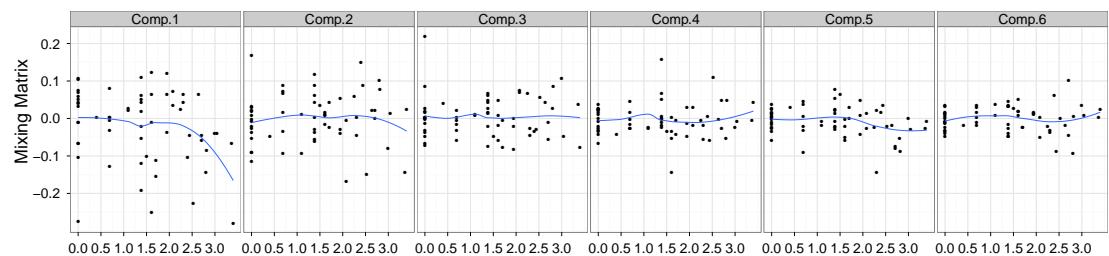
(a) MoCA: Subjects with high MoCA values tend to have positive mixing values in component 1 whereas low MoCA values imply negative mixing values of component 1.



(b) UPDRS: Mixing values of component 2 seem to have a slight trend.



(c) Hoehn & Yahr (1, 1.5, 2, 2.5, 3, 4): no clear trend visible.



(d) Logarithmic Duration: The duration was log transformed (due to just a few PD's with a long duration); there is no trend visible.

Figure C.8.: Y-axis is representing the columns of the mixing matrix, x-axis is representing MoCA, UPDRS, Hoehn & Yahr and duration. The blue line is a lowess smoother to emphasize the relation. Component 1 seems to be important for variable MoCA.

Components Y

Figure C.9 shows the histogram and the density of the components 1 to 6 and C.10 shows the relation between the first six components in a density plot. Images in figures C.11 and C.12 represent the classical two dimensional axial view.

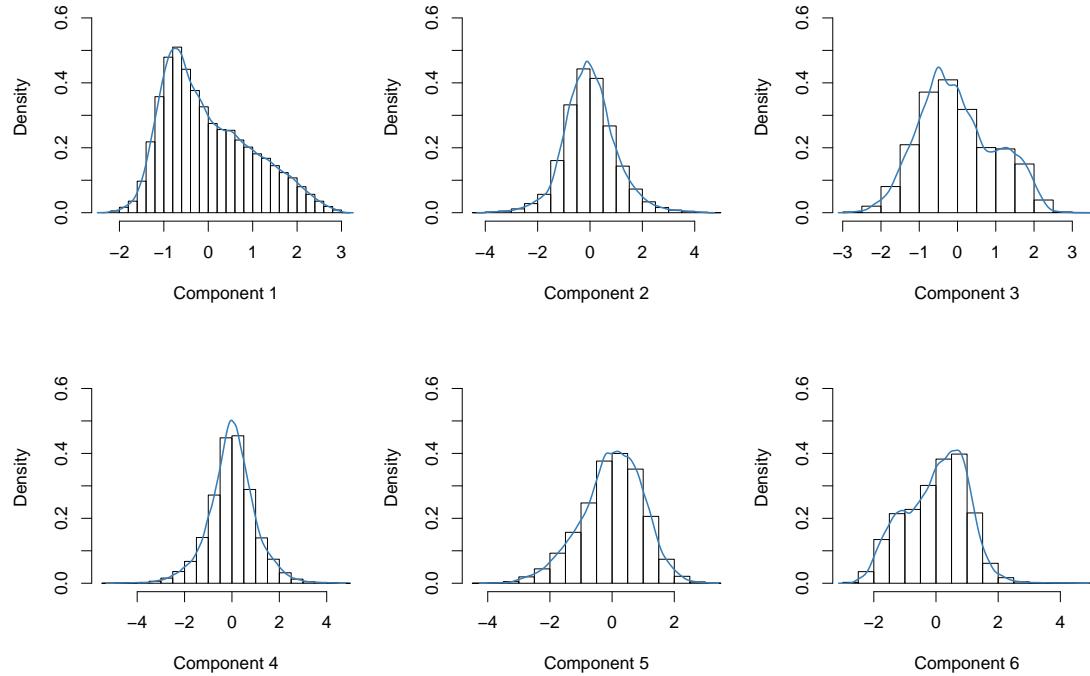


Figure C.9.: Histogram and density of component 1 to 6: component 2 and 4 seem to have a symmetric distribution.

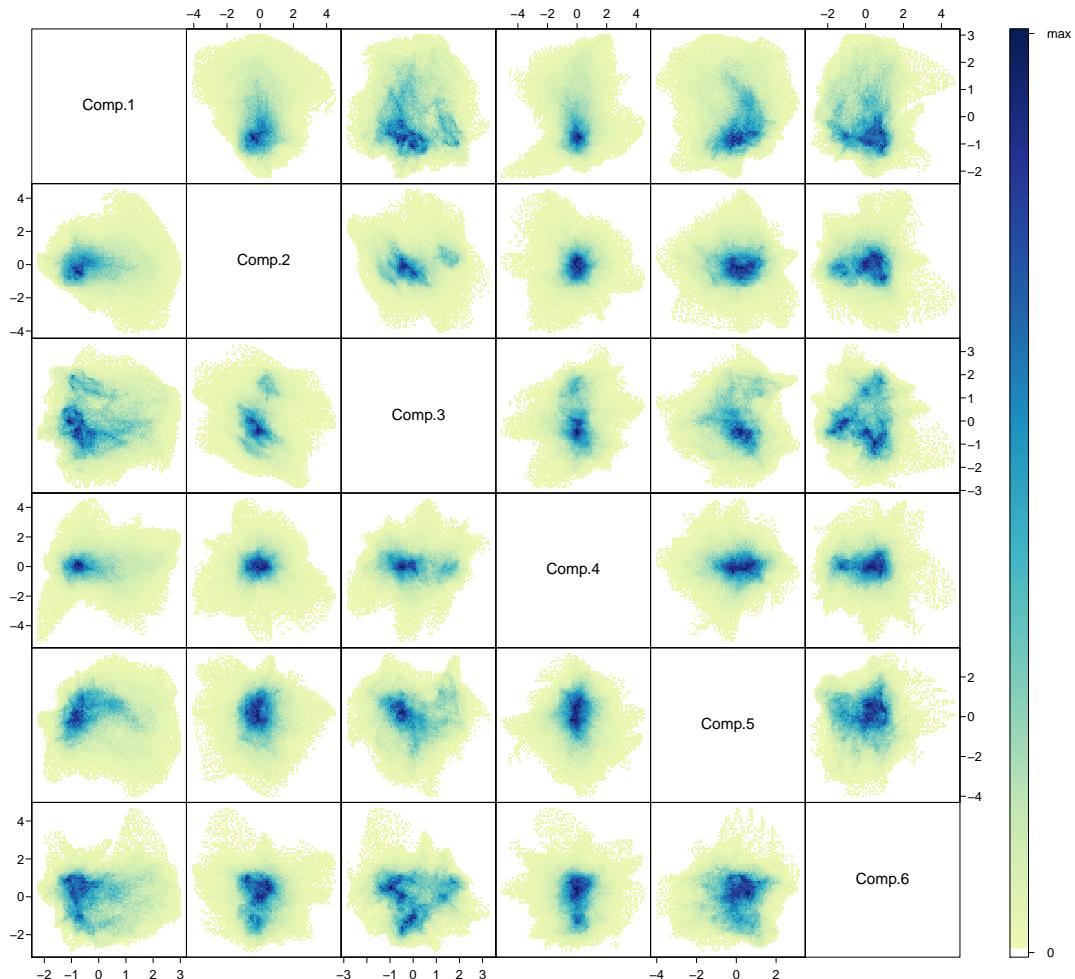


Figure C.10.: Pairs plot for principal components 1 to 6. The shape of the darkblue areas is of interest. Component 3 and 6 is interesting in terms of clusters. Note that the range in x-axis and y-axis is not always the same.

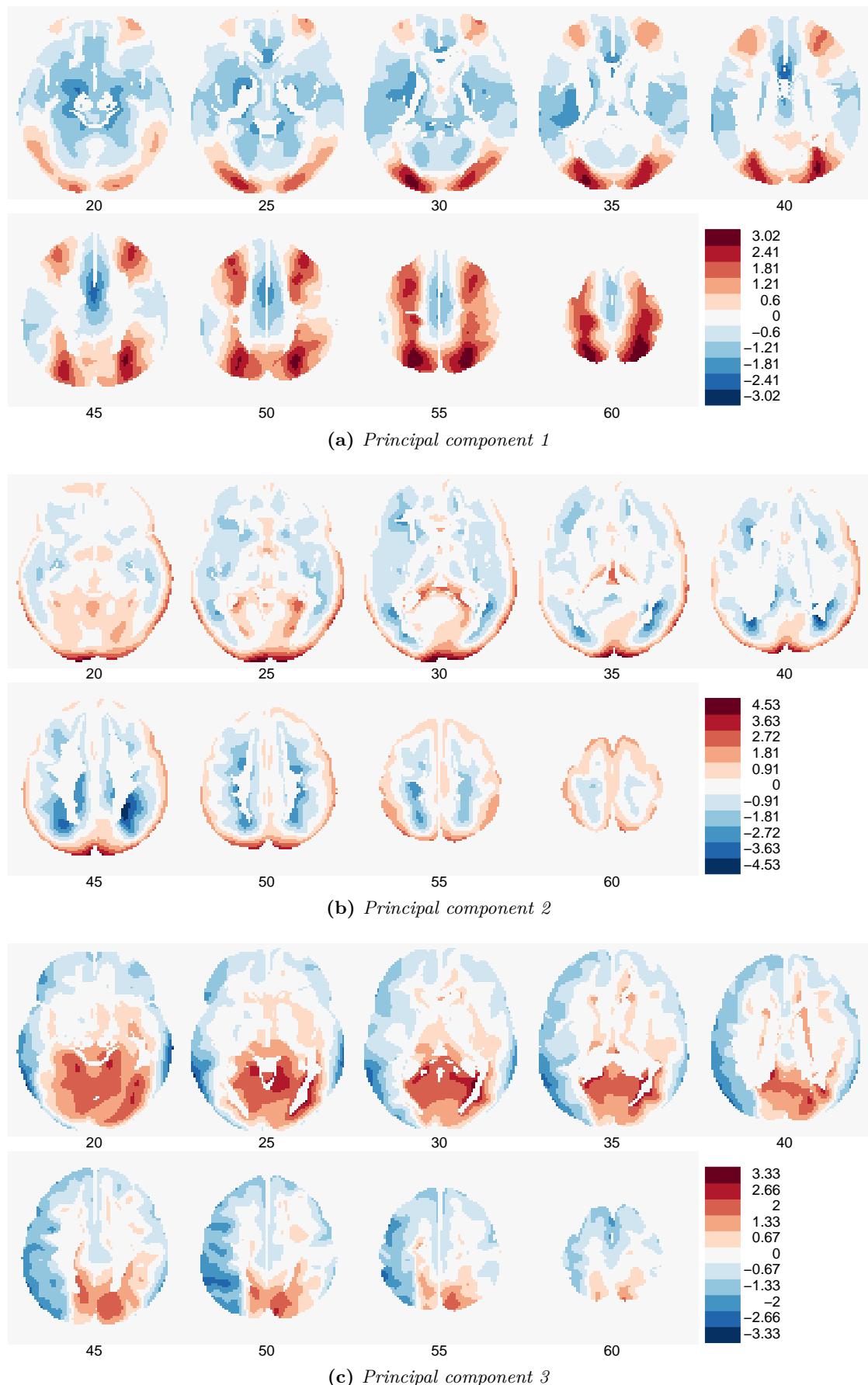


Figure C.11.: Images of principal components 1, 2 and 3 - red indicates relative increased, blue relative decreased cerebral blood flow values. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere is on the left hand side of the image, right hemisphere is on the right. (This is the 'neurological' view of the images. The 'radiology' view would have the left hemisphere on the right hand side and vice versa.)

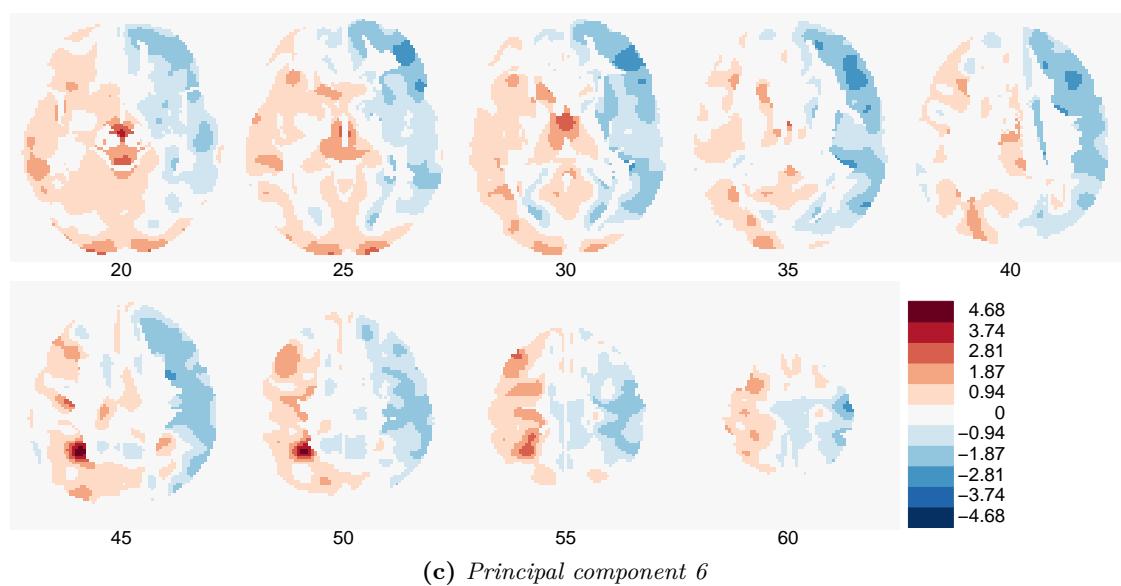
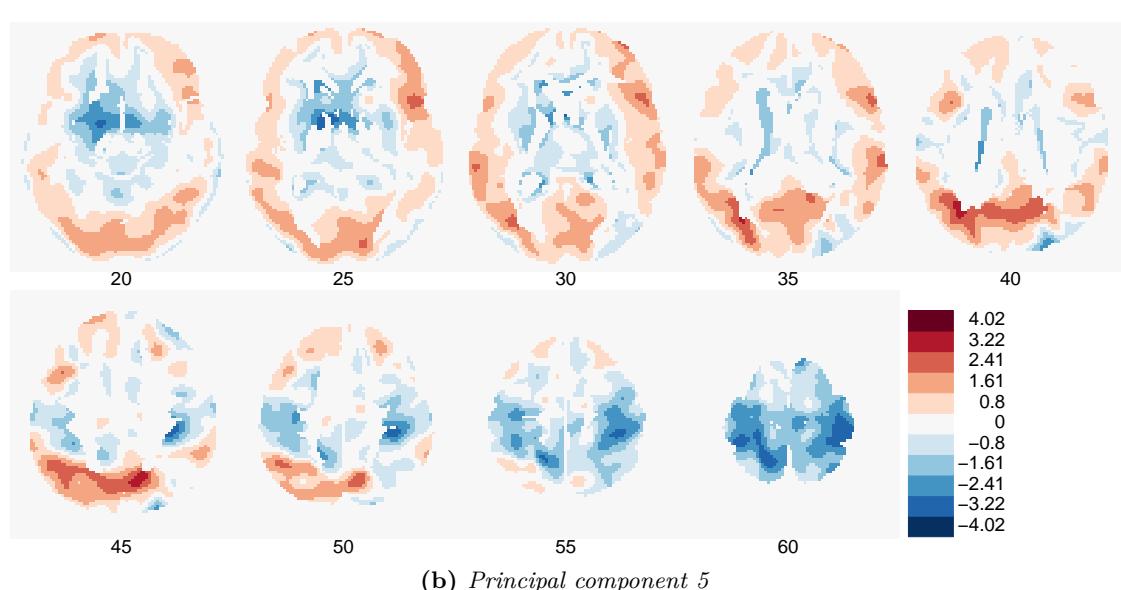
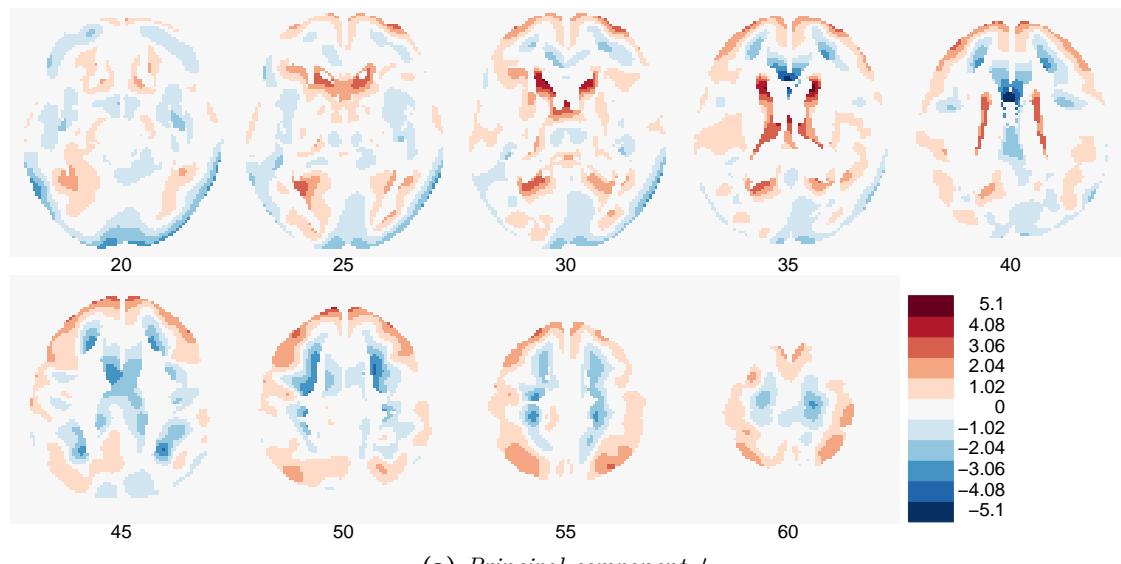
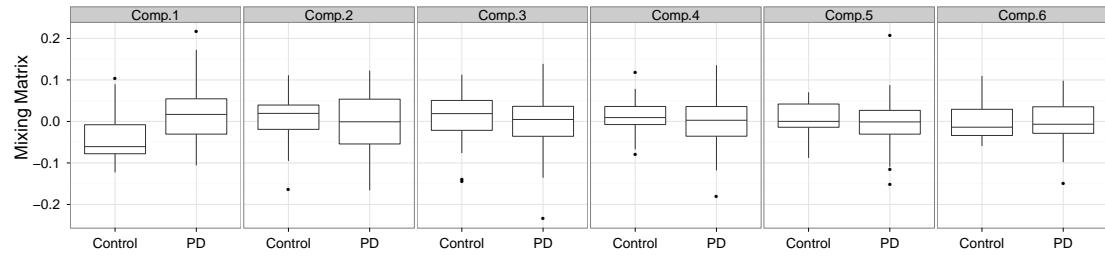


Figure C.12.: Images of principal components 4, 5 and 6 - red indicates relative increased, blue relative decreased cerebral blood flow values. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side of the image, right hemisphere is on the right.

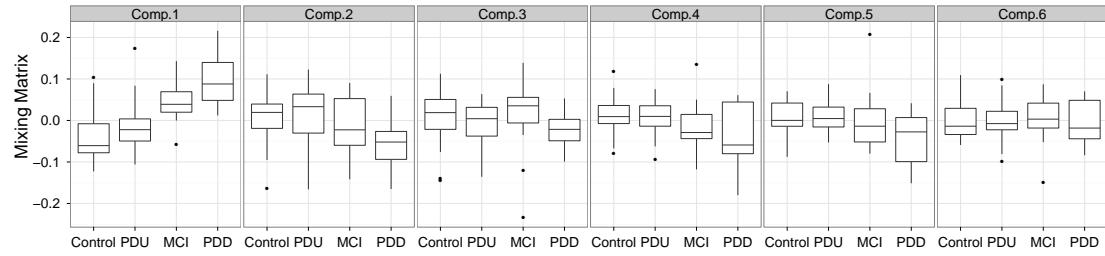
C.3. Projection Methods: ICA

Mixing Matrix A^{-1}

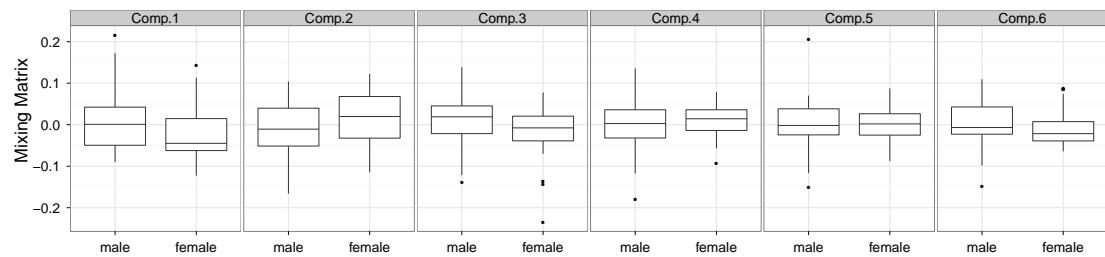
Figure C.13 and C.14 show the relation between the columns in the mixing matrix and the variables as group, cognition, sex, age, MoCA, UPDRS, Hoehn & Yahr and Duration.



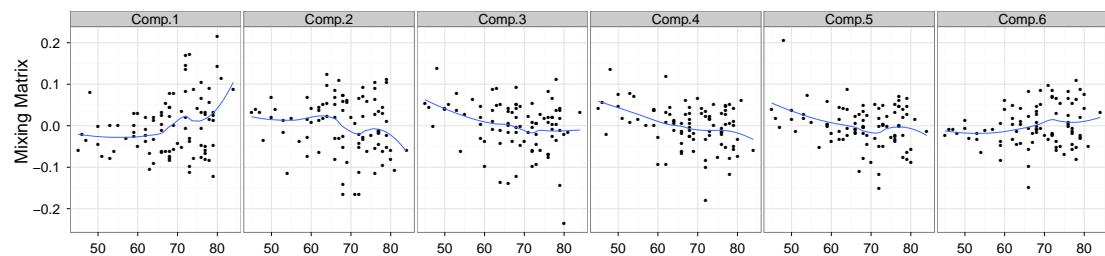
(a) Group (controls, PD): Component 1 seems to distinguish controls and PD's.



(b) Cognition (controls, PDU, MCI, PDD): Component 1 seems to have a strong trend between cognition and the mixing values.

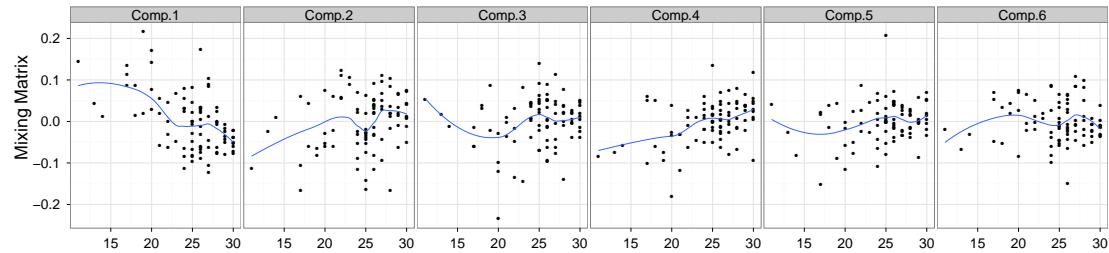


(c) Sex (male, female): Component 1, 2 and 3 seem to distinguish male and female.

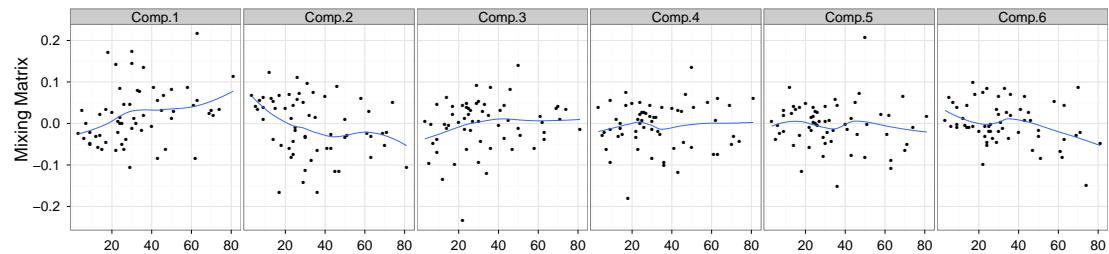


(d) Age: The blue line is a lowess smoother to emphasize the relation. Every component seems to have a slight trend, but not necessarily a linear one.

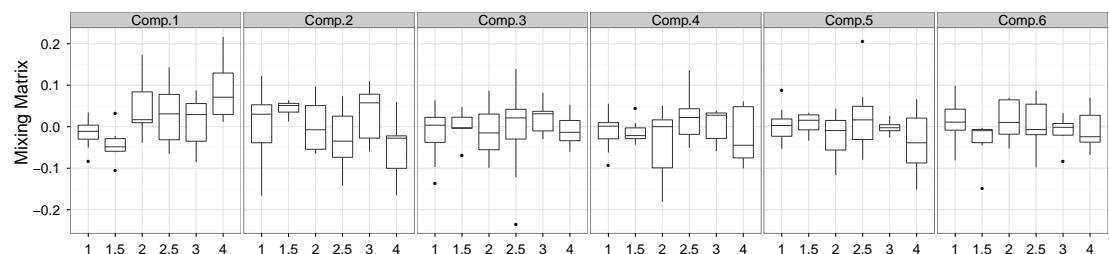
Figure C.13.: Y-axis is representing the columns of the mixing matrix, x-axis is representing group, cognition, sex and age.



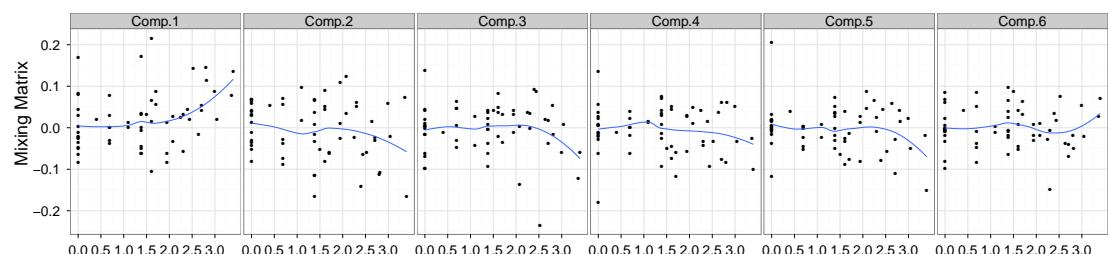
(a) MoCA: Component 1, 2 and 4 show a negative trend between MoCA and the mixing values.



(b) UPDRS: Component 1 shows a positive trend and component 2 a negative trend between UPDRS and the mixing values.



(c) Hoehn & Yahr (1, 1.5, 2, 2.5, 3, 4): no clear relation visible.



(d) Logarithmic Duration: The duration was log transformed (due to just a few PD's with a long duration), there is no trend visible.

Figure C.14.: Y-axis is representing the columns of the mixing matrix, x-axis is representing MoCA, UPDRS, Hoehn & Yahr and duration. The blue line is a lowess smoother to emphasize the relation. Component 1 and 2 seem to be important for MoCA & UPDRS.

Components Y

Figure C.15 shows the histogram and the density of the components 1 to 6 and C.16 shows the relation between them in a density plot. Images in figures C.17 and C.18 then represent the classical two dimensional axial view.

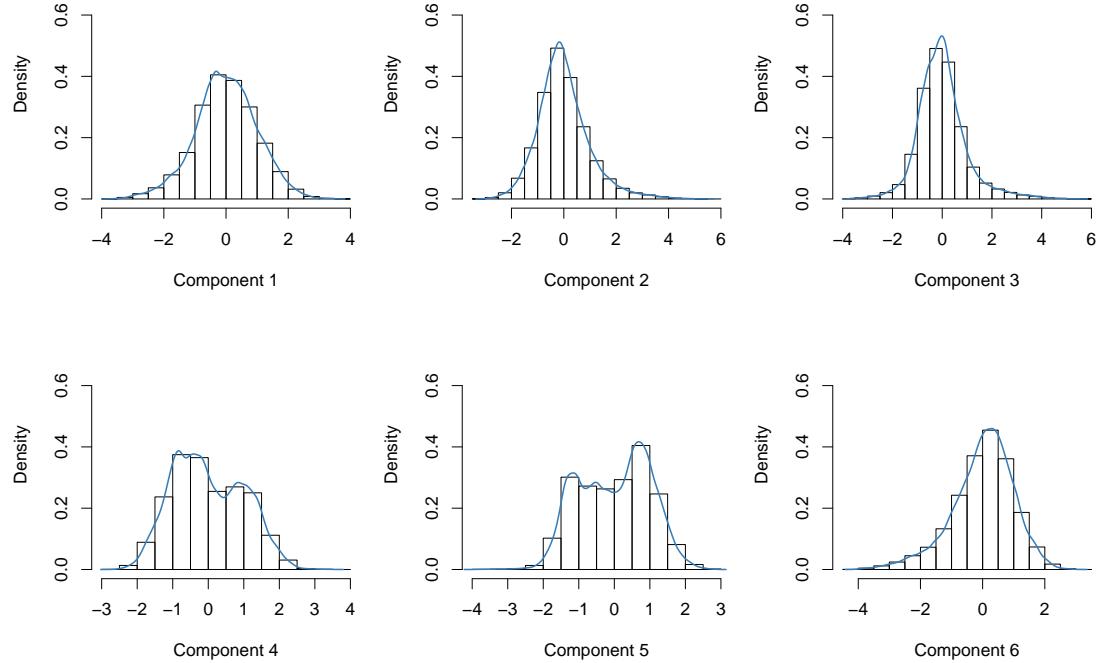


Figure C.15.: Histogram and density of independent component 1 to 6: component 1 seems to have a symmetric distribution.

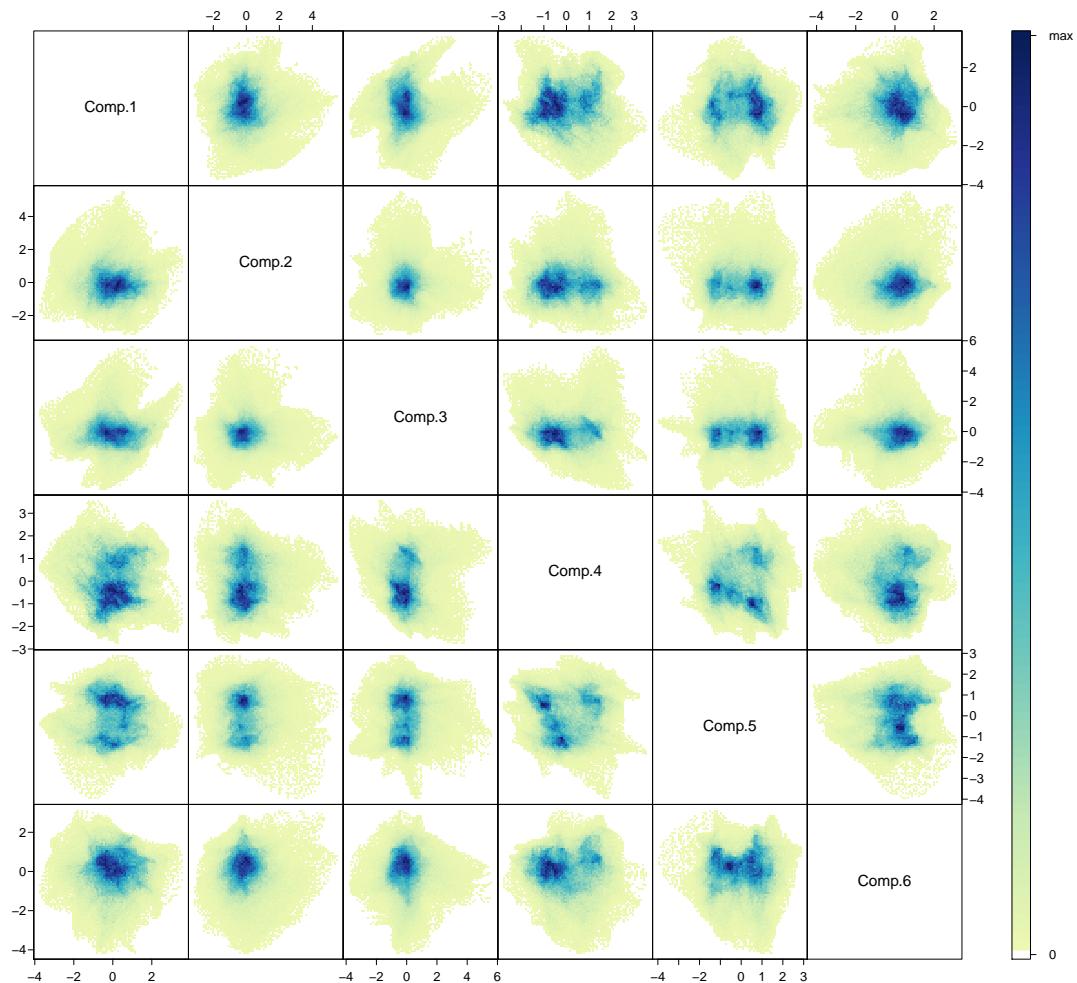


Figure C.16.: Pairs plot of independent component 1 to 6. Component 4 and 5 seems to be interesting in terms of clustering.

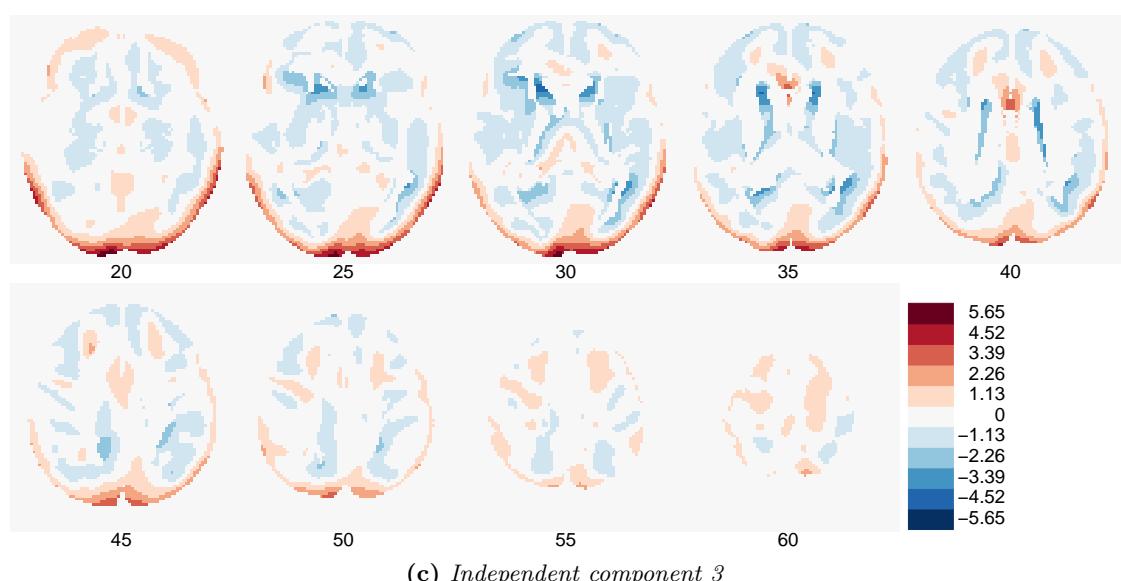
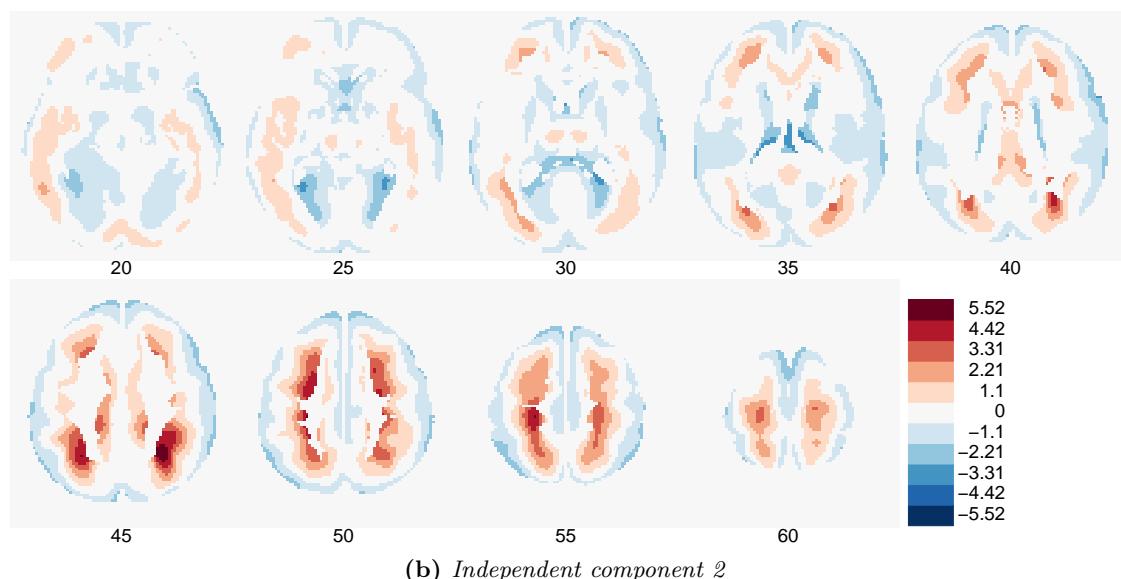
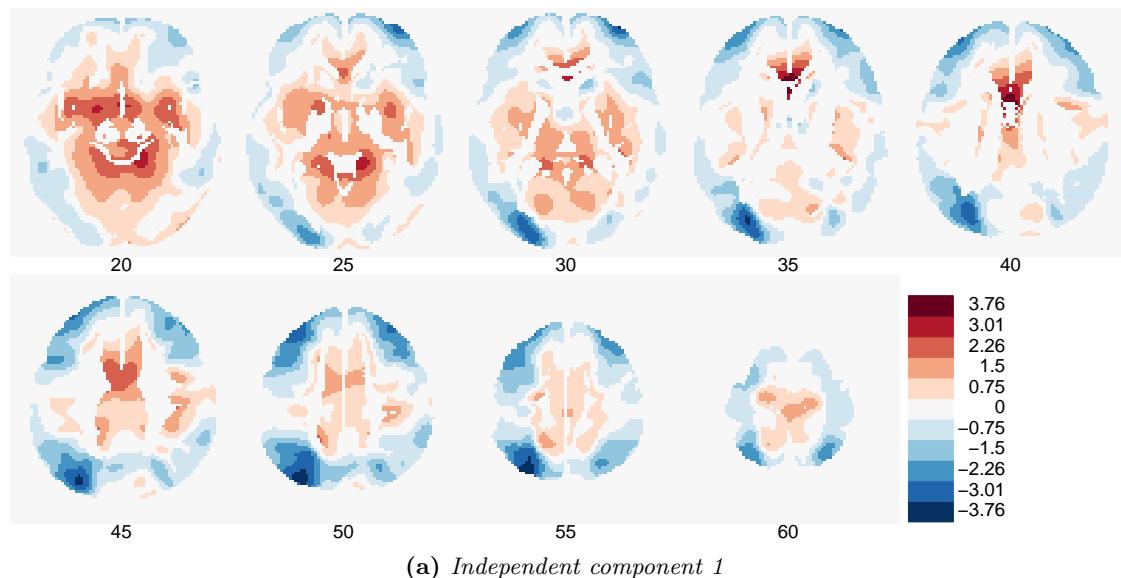
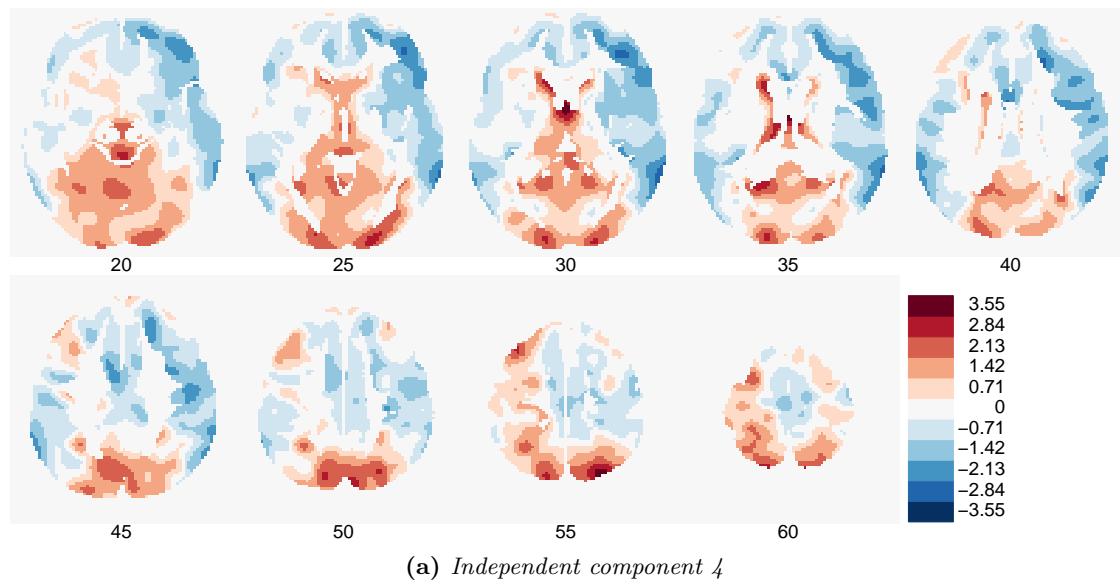
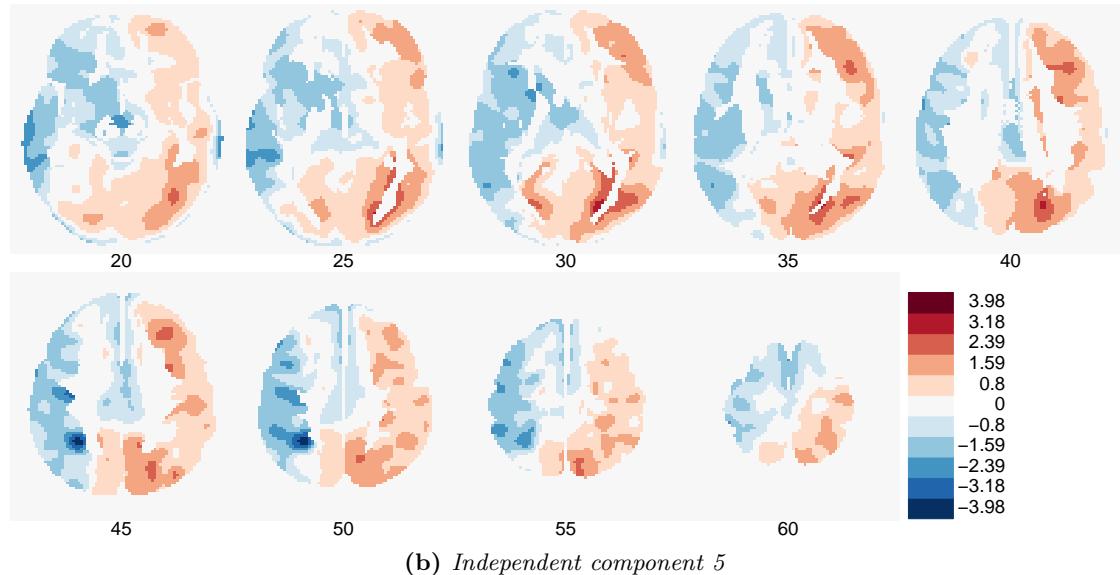


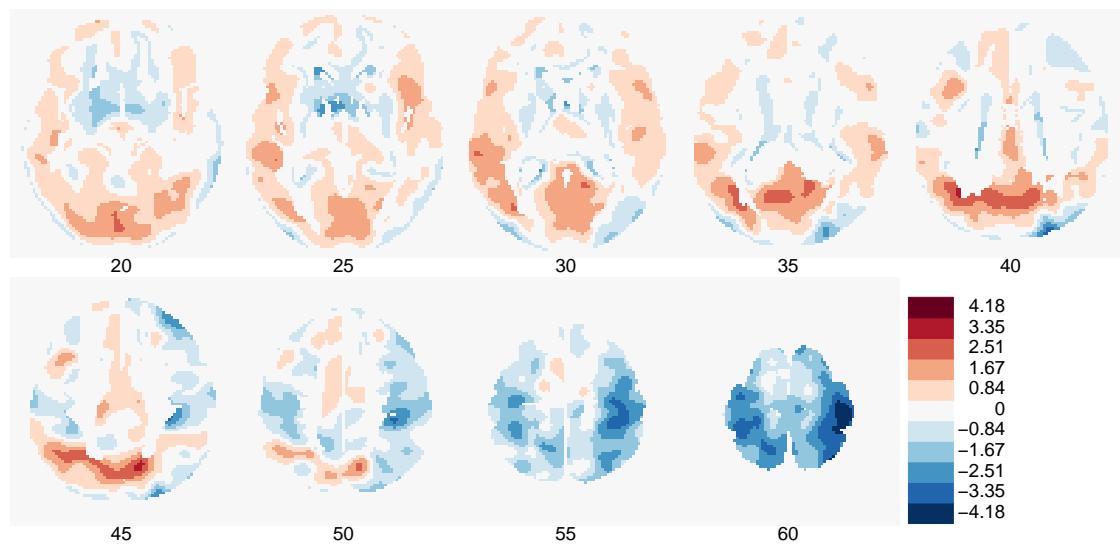
Figure C.17.: Images of independent components 1, 2 and 3 - red indicates relative increased, blue relative decreased cerebral blood flow values. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side of the image, right hemisphere is on the right.



(a) Independent component 4



(b) Independent component 5



(c) Independent component 6

Figure C.18.: Images of independent components 4, 5 and 6- red indicates relative increased, blue relative decreased cerebral blood flow values. The slices are in z-direction from 20, 25, ..., 60. The view of the image is from the above: left hemisphere on the left hand side of the image, right hemisphere is on the right.

C.4. Linear Regression of Components and Co-Information

In the classification we were only analyzing if the groups can be distinguished by certain mixing values. Here we clarify if other relations exist.

The mixing values from each component were regressed with co-information of subjects (MoCA, UPDRS, Hoehn & Yahr, disease duration, age, gender). This tells us which components (images) are related to a certain measure. For example if the mixing values of component 1 and MoCA have a high correlation, it is assumed that the image of component 1 has a relation to MoCA and therefore to the cognition. If this is the case, the sign of the estimation is important to make any statements. It has to be mentioned that one does not know the variance per voxel (compared to the bootstrapped pattern).

A multiple regression with the mixing values of one component as the response variable and all co-information variables as explanatory variables was not used due to correlation between the explanatory variables (see figures 2.2 and 2.3 in the beginning). Instead linear regression models with only one intercept and one slope were used.

The section is splitted up in 2 graphs which show the relations graphically and in tables with R^2 , p-values, and the estimation for the intercept α and the slope β . If the covariate has categories, then an anova is applied. The R^2 then refers to the sum of squares of the covariate divided trough the total sum of squares.

When the data was analyzed a residual analysis was made. It showed that in some cases at the assumptions of a linear regression - expected value of 0 for residuals, constant variance of residuals, normal distribution, no leverages - were violated. These facts are can also be seen in the graphs in figure C.19, C.20, C.21 and C.22. For a further analysis I would suggest a) transformations of variables to avoid violations and b) applying robust regression.

C.4.1. Principal component analysis (PCA)

Linear regression models with statistically significant coefficients have been:

PC 1 versus MoCA, Age, Hoehn & Yahr and Duration

PC 2 versus Sex and UPDRS

PC 4 versus Sex

PC 5 versus MoCA, UPDRS, Hoehn & Yahr and Duration

PC 6 versus MoCA

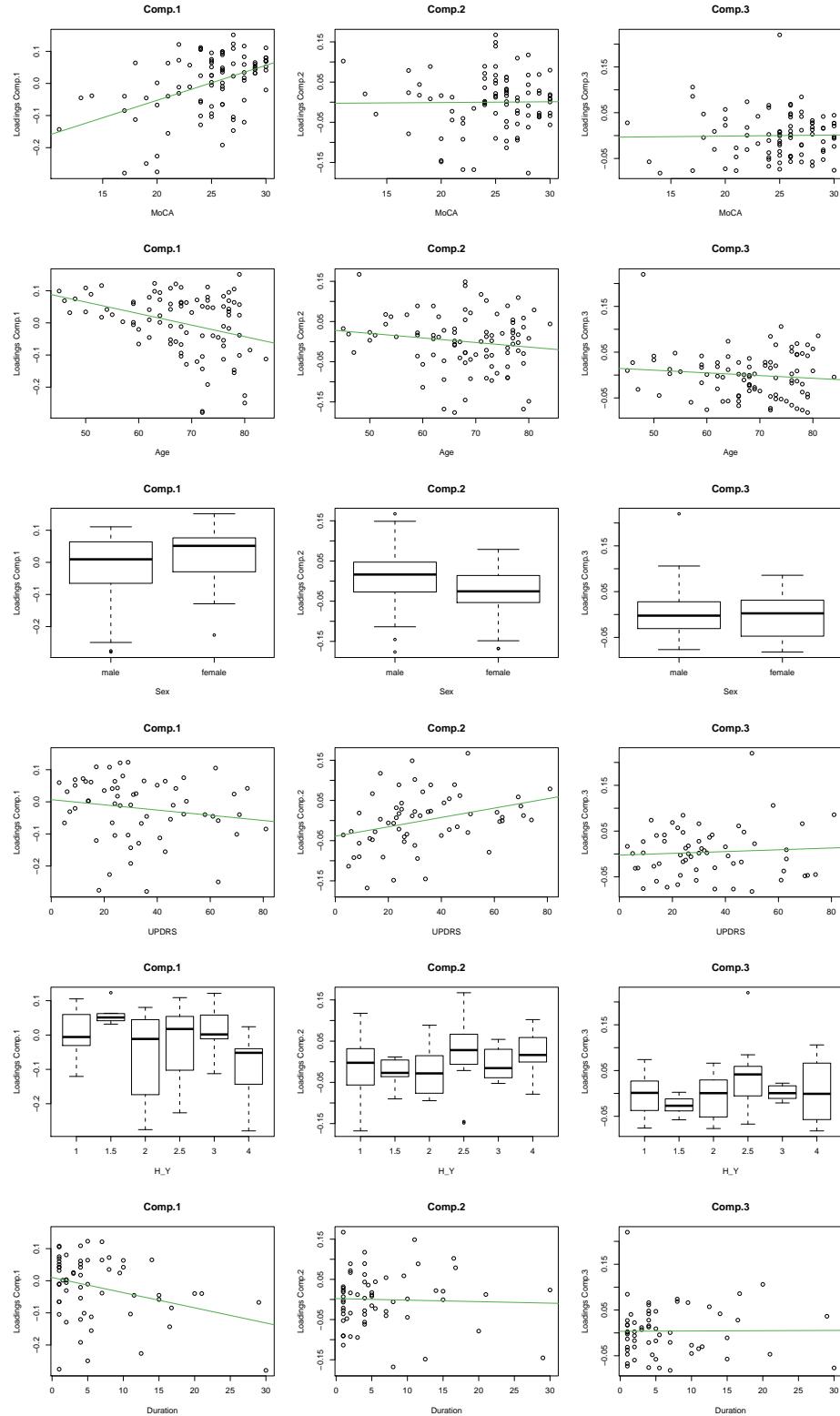


Figure C.19.: Component 1, 2 and 3 versus variables MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration. Each component has one column, each variable one row. The green lines are representing the intercept and the slope of the regression. Note that the assumption of a linear regression are violated in component 1 versus age and component 1 versus duration.

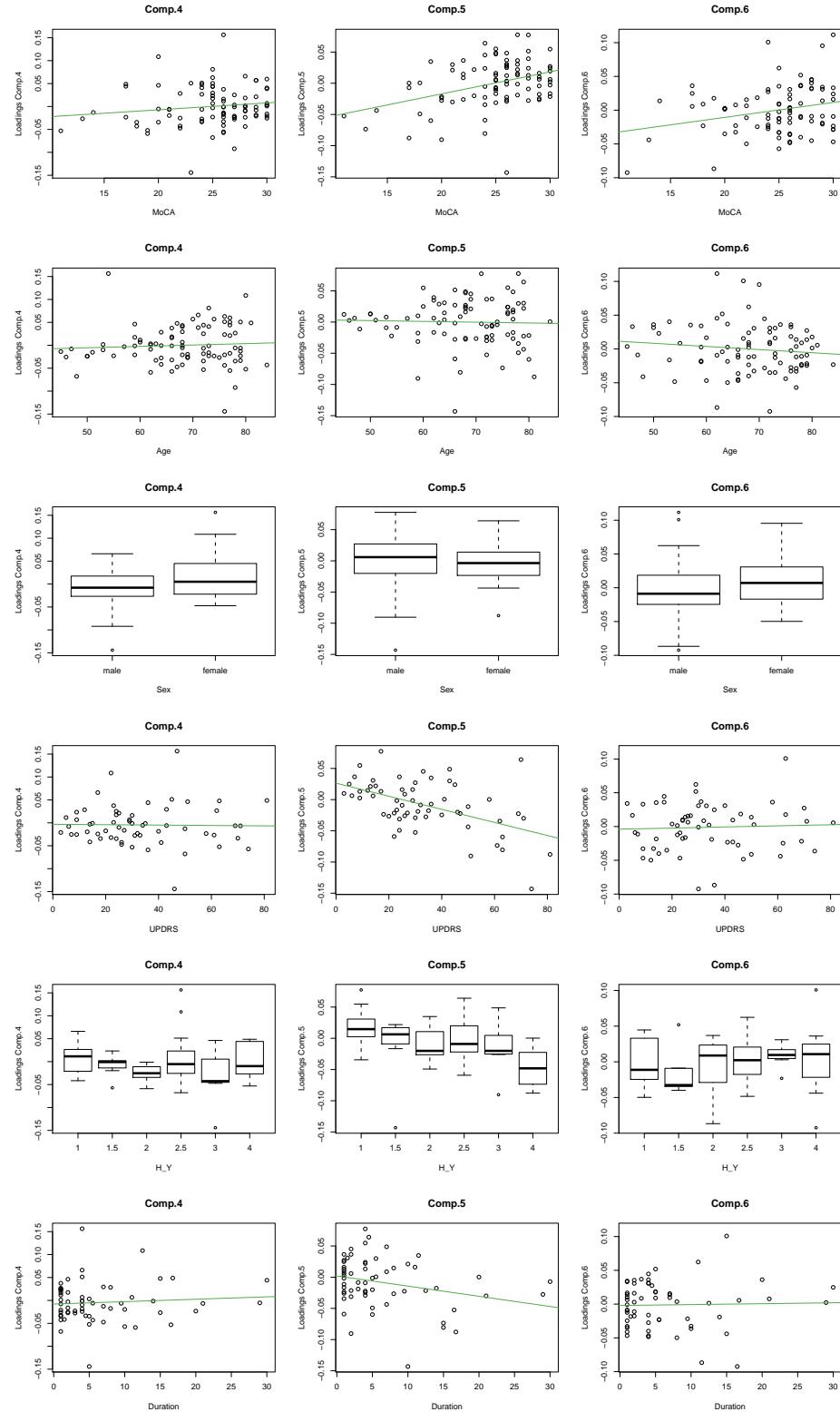
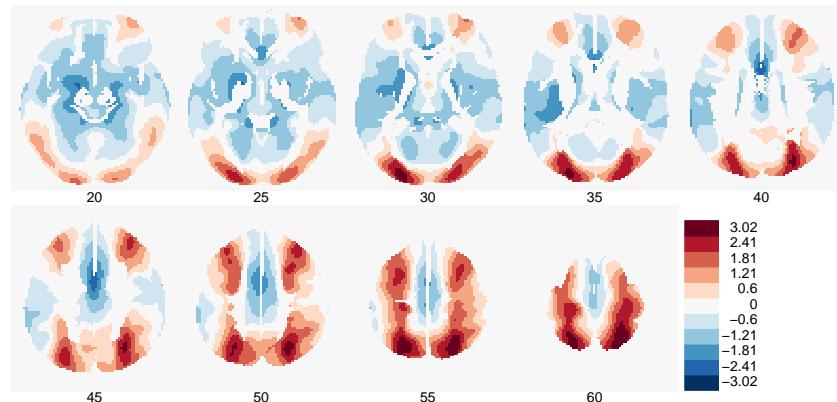
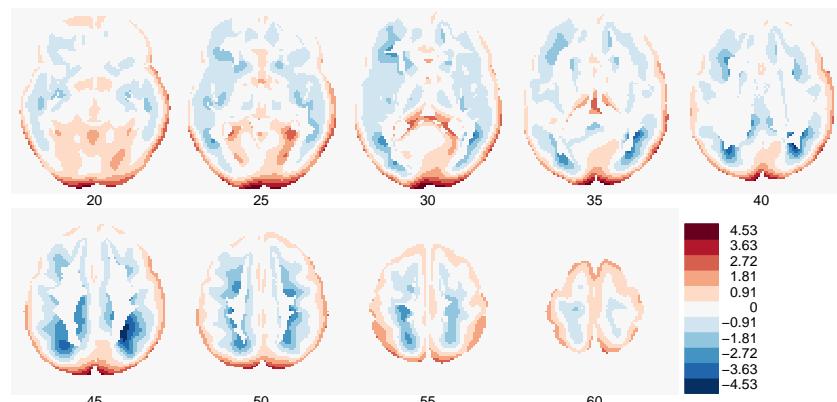


Figure C.20.: Component 4, 5 and 6 versus variables MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration. Each component has one column, each variable one row. The green lines are representing the intercept and the slope of the regression.

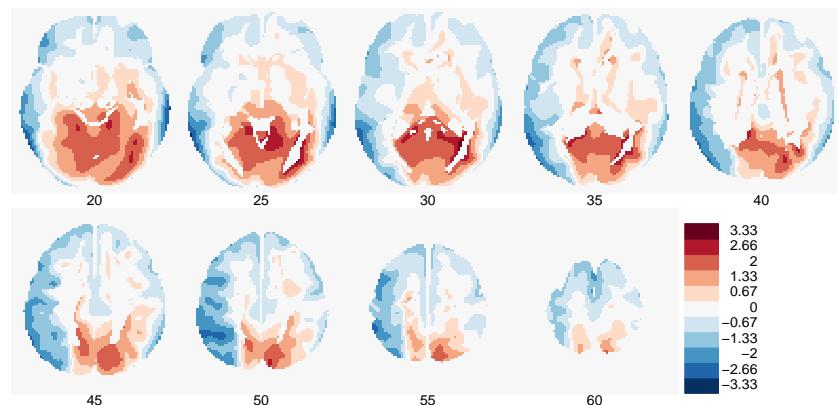


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.270	0.245	-	0.007	-	0.010
$\hat{\beta}$	0.011	-0.004	-	-0.001	-	-0.005
P-value for $\hat{\beta}$	0.000	0.001	0.095	0.216	0.015	0.011
R^2	0.221	0.121	0.032	0.026	0.224	0.106

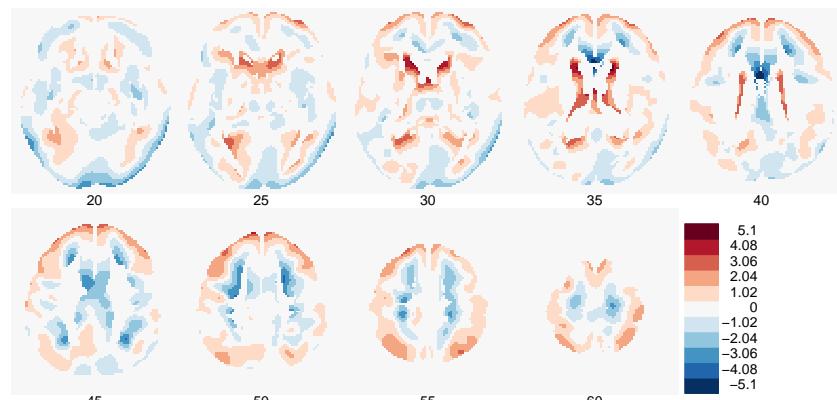
Table C.1.: Image and output for linear regression models for principal component 1

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.005	0.077	-	-0.039	-	0.002
$\hat{\beta}$	0.000	-0.001	-	0.001	-	-0.000
P-value for $\hat{\beta}$	0.919	0.162	0.035	0.010	0.465	0.781
R^2	0.000	0.022	0.050	0.108	0.080	0.001

Table C.2.: Image and output for linear regression models for principal component 2

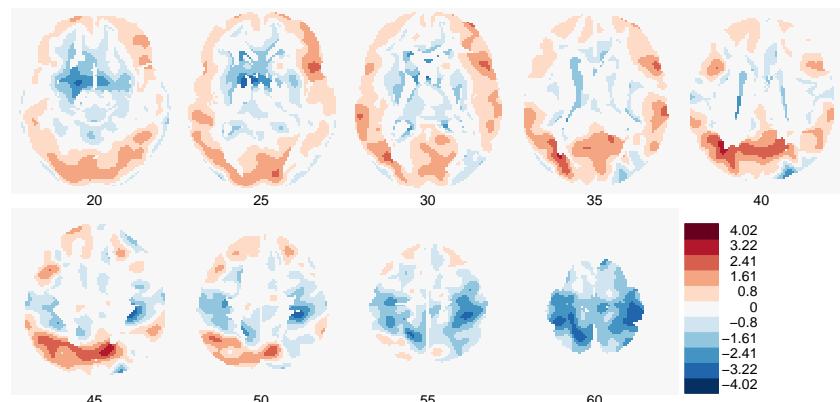


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.006	0.040	-	-0.003	-	0.004
$\hat{\beta}$	0.000	-0.001	-	0.000	-	0.000
P-value for $\hat{\beta}$	0.858	0.311	0.738	0.597	0.185	0.953
R^2	0.000	0.012	0.001	0.005	0.127	0.000

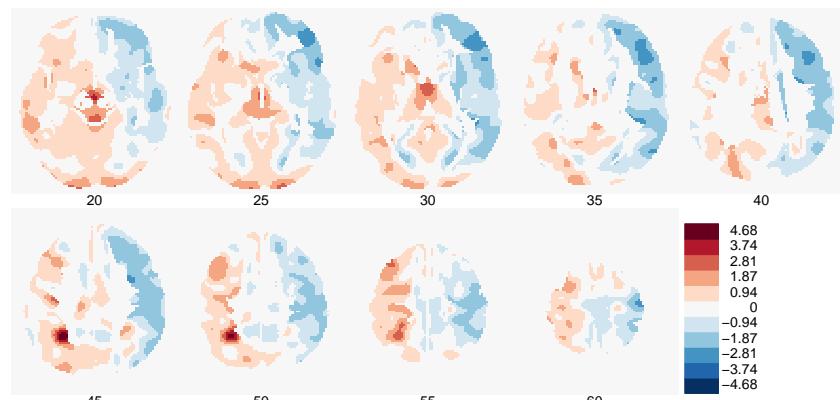
Table C.3.: Image and output for linear regression models for principal component 3

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.037	-0.021	-	-0.003	-	-0.008
$\hat{\beta}$	0.002	0.000	-	-0.000	-	0.001
P-value for $\hat{\beta}$	0.188	0.553	0.019	0.886	0.219	0.540
R^2	0.020	0.004	0.062	0.000	0.119	0.006

Table C.4.: Image and output for linear regression models for principal component 4



	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.088	0.009	-	0.026	-	0.002
$\hat{\beta}$	0.004	-0.000	-	-0.001	-	-0.002
P-value for $\hat{\beta}$	0.000	0.775	0.764	0.000	0.007	0.036
R^2	0.145	0.001	0.001	0.257	0.250	0.074

Table C.5.: Image and output for linear regression models for principal component 5

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.055	0.032	-	-0.004	-	-0.002
$\hat{\beta}$	0.002	-0.000	-	0.000	-	0.000
P-value for $\hat{\beta}$	0.017	0.268	0.280	0.728	0.820	0.847
R^2	0.063	0.014	0.013	0.002	0.039	0.001

Table C.6.: Image and output for linear regression models for principal component 6

C.4.2. Independent component analysis (ICA)

Linear regression models with statistically significant coefficients have been:

IC 1 versus MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration

IC 2 versus MoCA, UPDRS and Hoehn & Yahr

IC 4 versus Age, Sex

IC 5 versus Age

IC 6 versus UPDRS

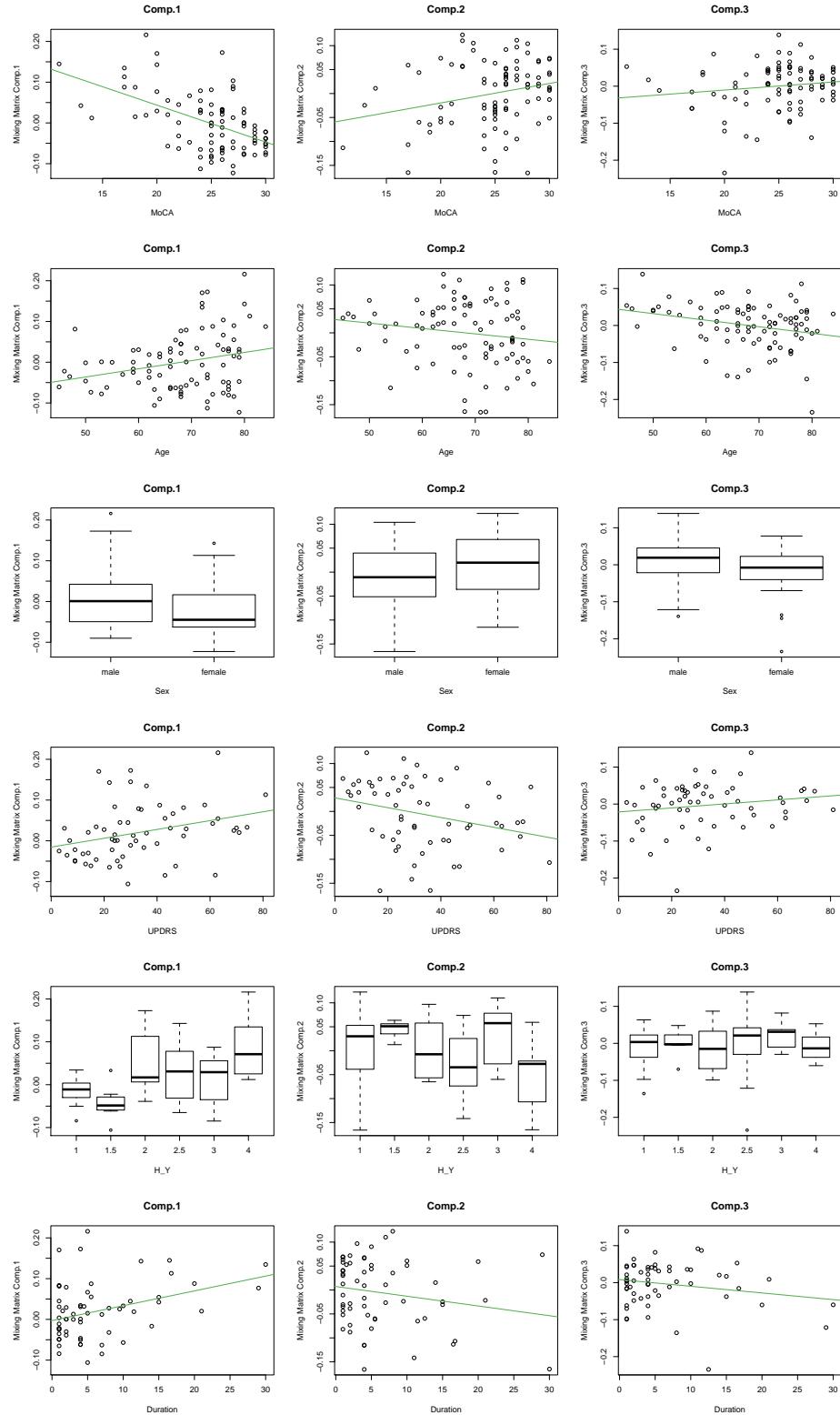


Figure C.21.: Independent component 1, 2 and 3 versus variables MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration. Each component has one column, each variable one row. The green lines are representing the intercept and the slope of the regression.

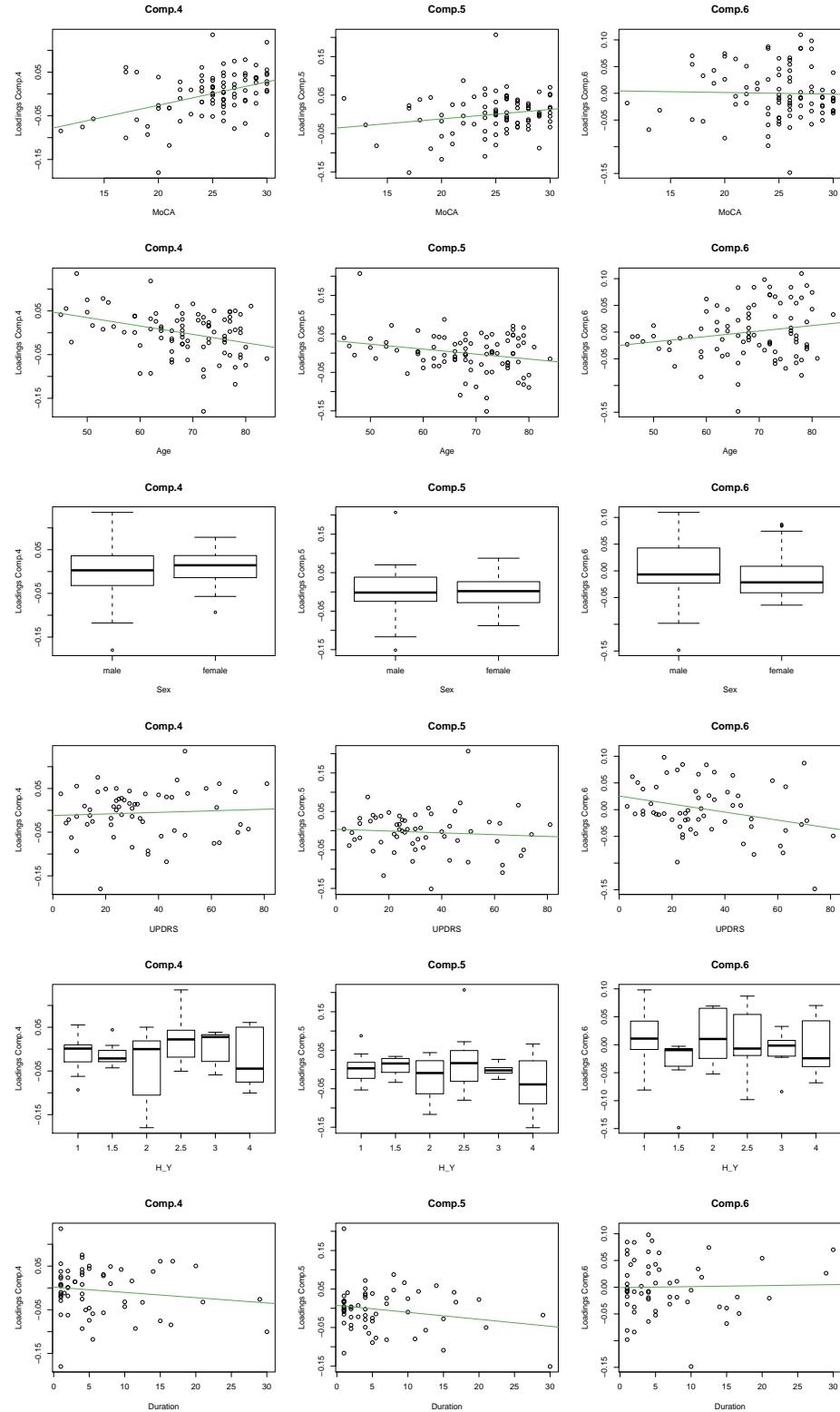
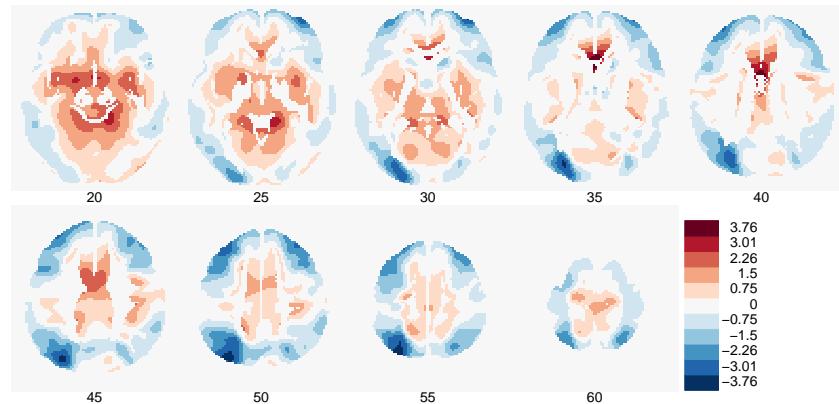
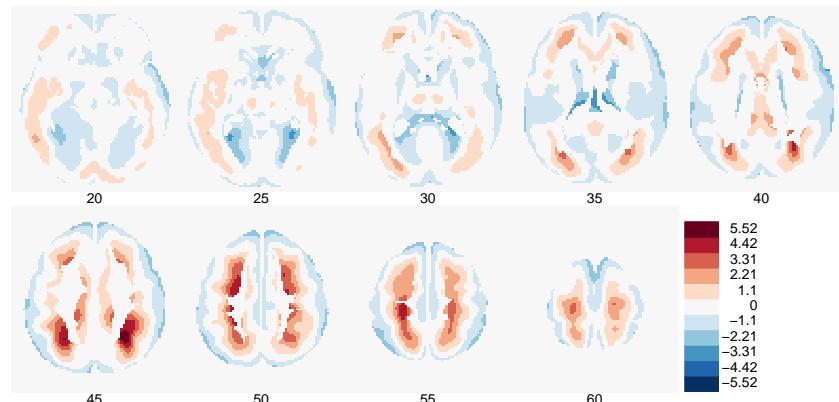


Figure C.22.: independent component 4, 5 and 6 versus variables MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration. Each component has one column, each variable one row. The green lines are representing the intercept and the slope of the regression.

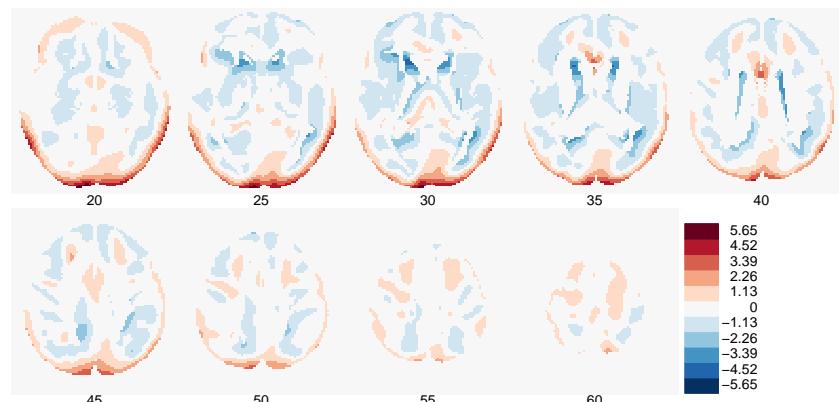


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.225	-0.137	-	-0.015	-	-0.003
$\hat{\beta}$	-0.009	0.002	-	0.001	-	0.004
P-value for $\hat{\beta}$	0.000	0.014	0.041	0.017	0.001	0.005
R^2	0.273	0.068	0.047	0.094	0.328	0.126

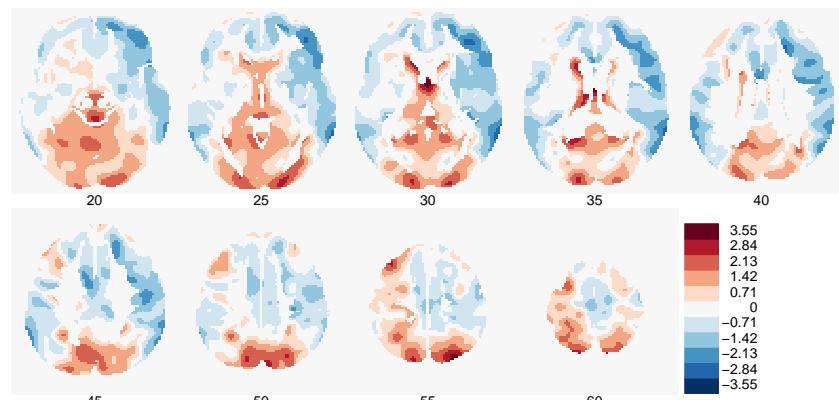
Table C.7.: Image and output for linear regression models for independent component 1

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.101	0.077	-	0.028	-	0.007
$\hat{\beta}$	0.004	-0.001	-	-0.001	-	-0.002
P-value for $\hat{\beta}$	0.021	0.154	0.083	0.028	0.044	0.136
R^2	0.060	0.023	0.034	0.081	0.186	0.038

Table C.8.: Image and output for linear regression models for independent component 2

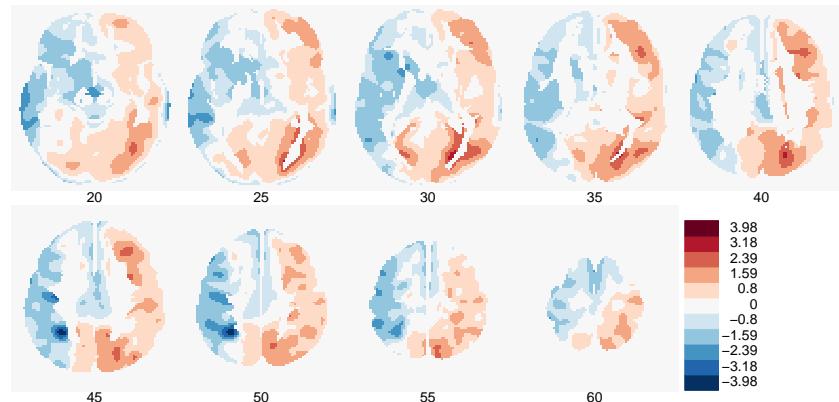


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.054	0.121	-	-0.021	-	0.008
$\hat{\beta}$	0.002	-0.002	-	0.001	-	-0.002
P-value for $\hat{\beta}$	0.170	0.011	0.037	0.186	0.899	0.128
R^2	0.022	0.071	0.049	0.030	0.029	0.040

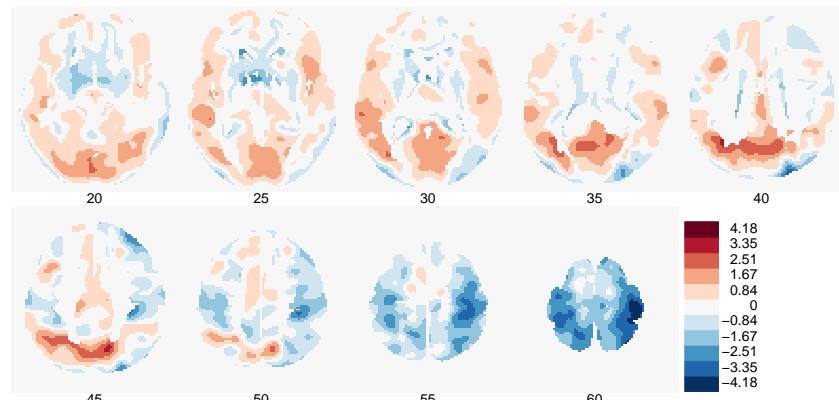
Table C.9.: Image and output for linear regression models for independent component 3

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.133	0.131	-	-0.012	-	0.002
$\hat{\beta}$	0.005	-0.002	-	0.000	-	-0.001
P-value for $\hat{\beta}$	0.000	0.001	0.167	0.631	0.196	0.263
R^2	0.180	0.116	0.022	0.004	0.124	0.022

Table C.10.: Image and output for linear regression models for independent component 4



	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.061	0.088	-	0.004	-	0.008
$\hat{\beta}$	0.002	-0.001	-	-0.000	-	-0.002
P-value for $\hat{\beta}$	0.060	0.026	0.851	0.536	0.253	0.091
R^2	0.040	0.056	0.000	0.007	0.112	0.048

Table C.11.: Image and output for linear regression models for independent component 5

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.008	-0.069	-	0.025	-	-0.001
$\hat{\beta}$	-0.000	0.001	-	-0.001	-	0.000
P-value for $\hat{\beta}$	0.811	0.074	0.289	0.023	0.262	0.852
R^2	0.001	0.036	0.013	0.086	0.110	0.001

Table C.12.: Image and output for linear regression models for independent component 6

C.4.3. Subset of data set: Principal component analysis (PCA)

Here only a subset as mentioned in section 4.7 are used.

Graphs in figure C.23 and C.24 show the mixing values of the principal component 1 to 6 on the y-axis and the 2nd information on the x-axis illustrate the relation between the response variables and the explanatory variables graphically. The green line is the regression line. Tables C.13 to C.18 show the relations with p-values.

Graphs and tables indicates following relations (no assumptions violated and p-value of coefficient < 0.05):

PC 1 and MoCA are related (Age and Duration have not a constant variance of the residuals, although Hoehn & Yahr is significant, the graph shows no meaningful relation.)

PC 2 is related to Sex and UPDRS.

PC 4 is related to MoCA

PC 5 is related to Hoehn & Yahr and Duration (MoCA and UPDRS show not a constant variance)

Linear regression of component 1 and MoCA has the highest R^2 .

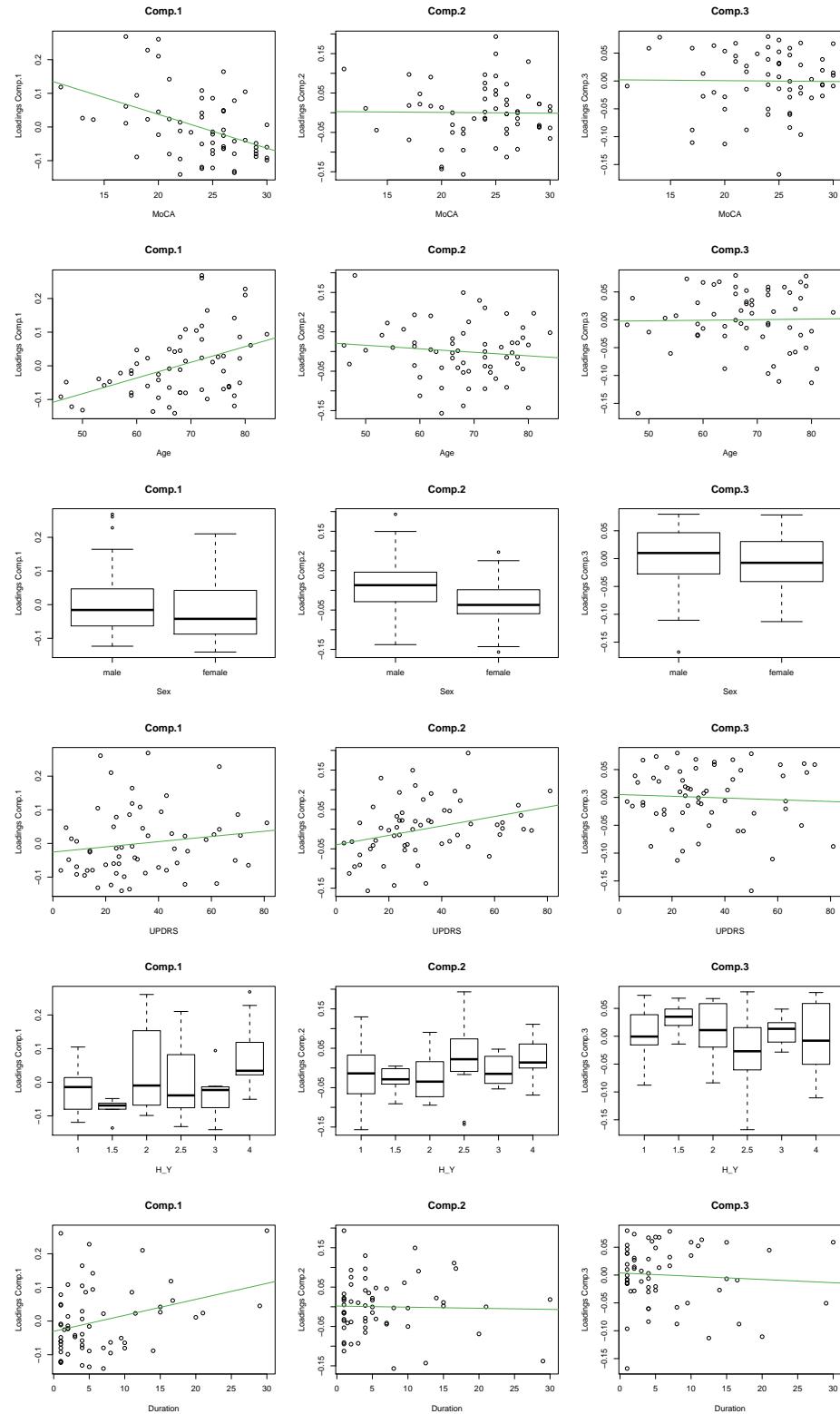


Figure C.23.: Graphs of principal component 1, 2, 3 (columns) and MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration (rows).

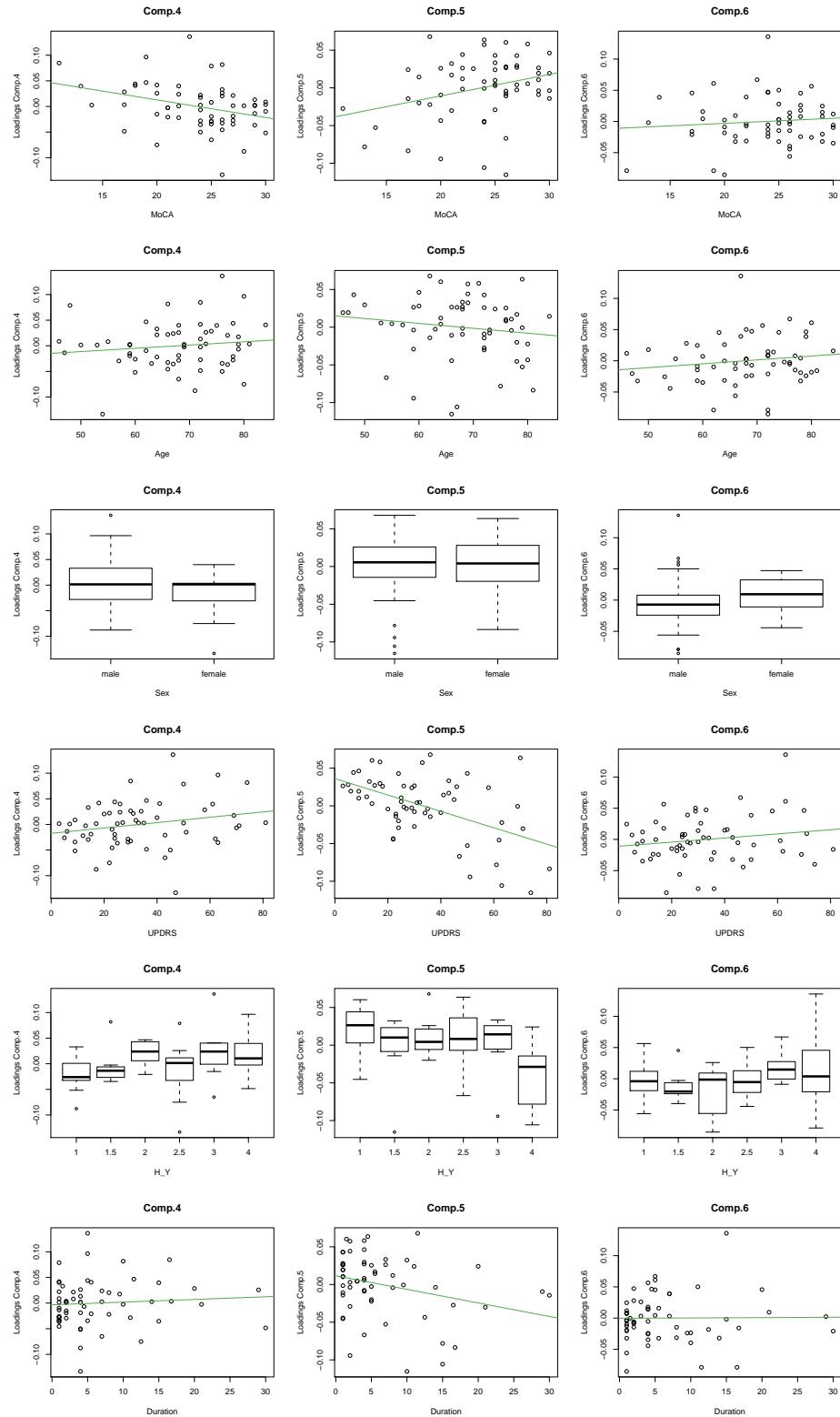
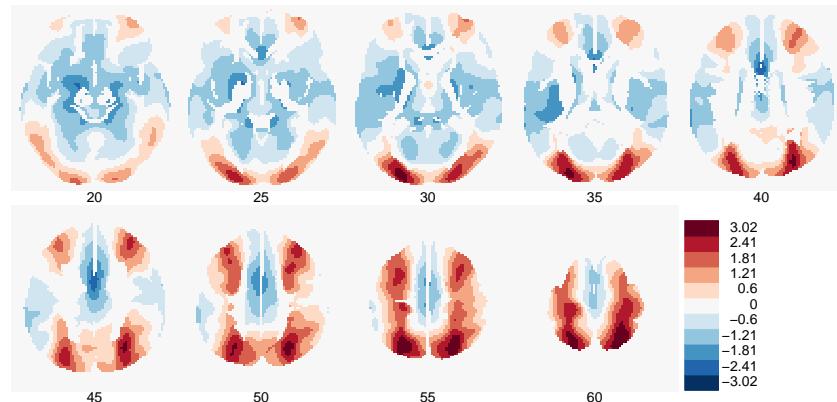
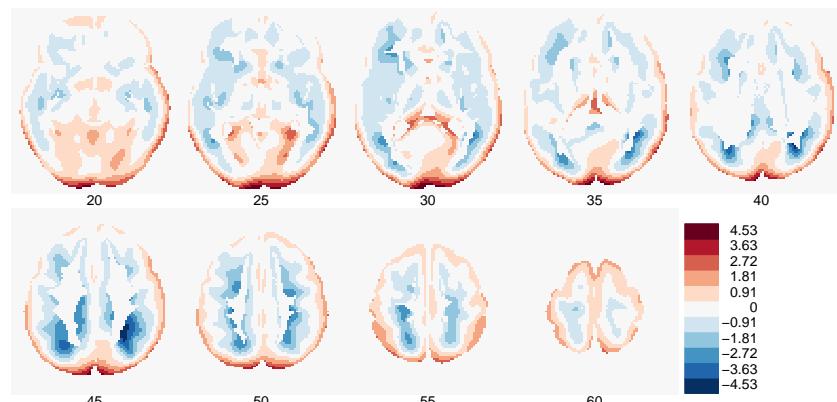


Figure C.24.: Graphs of principal component 4, 5, 6 (columns) and MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration (rows).

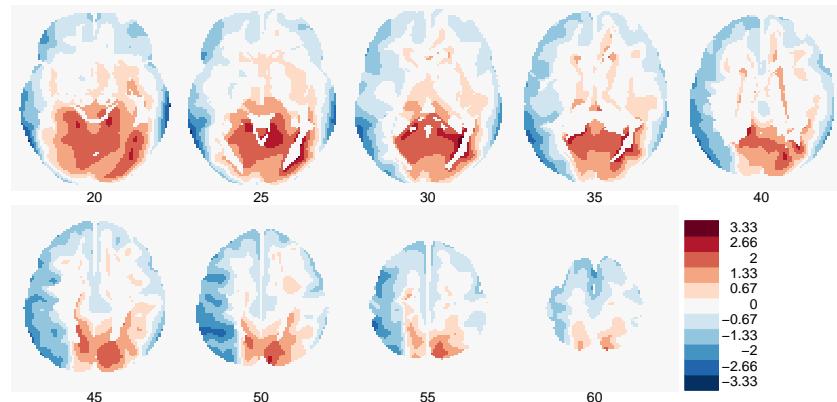


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.238	-0.316	-	-0.026	-	-0.030
$\hat{\beta}$	-0.010	0.005	-	0.001	-	0.005
P-value for $\hat{\beta}$	0.000	0.001	0.346	0.243	0.020	0.012
R^2	0.201	0.187	0.015	0.023	0.214	0.105

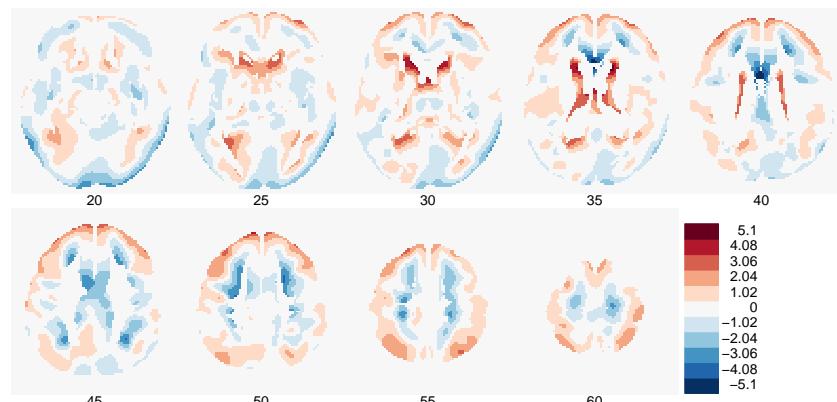
Table C.13.: Image and output for linear regression models for principal component 1

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.005	0.061	-	-0.040	-	0.002
$\hat{\beta}$	-0.000	-0.001	-	0.001	-	-0.000
P-value for $\hat{\beta}$	0.922	0.378	0.036	0.010	0.373	0.840
R^2	0.000	0.013	0.073	0.109	0.092	0.001

Table C.14.: Image and output for linear regression models for principal component 2

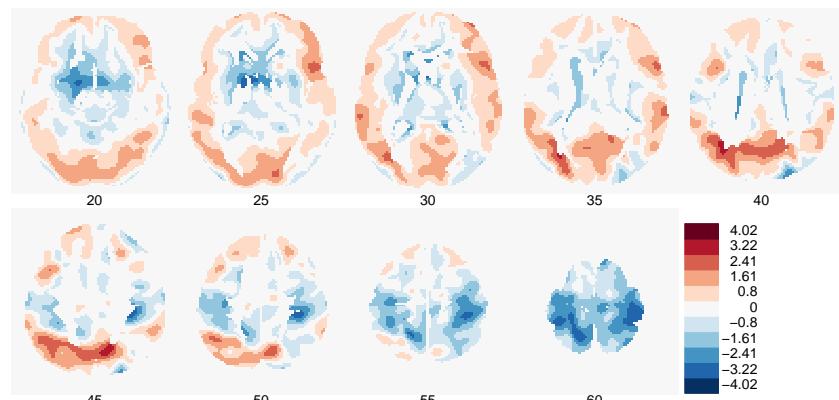


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.004	-0.007	-	0.005	-	0.004
$\hat{\beta}$	-0.000	0.000	-	-0.000	-	-0.001
P-value for $\hat{\beta}$	0.925	0.897	0.423	0.672	0.198	0.588
R^2	0.000	0.000	0.011	0.003	0.124	0.005

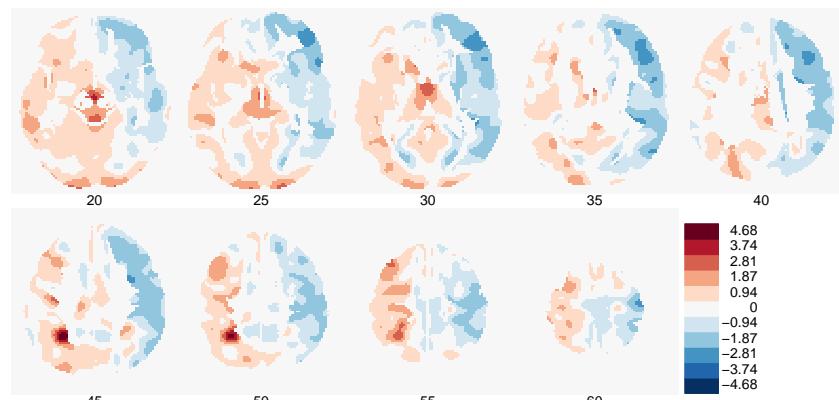
Table C.15.: Image and output for linear regression models for principal component 3

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.082	-0.043	-	-0.017	-	-0.003
$\hat{\beta}$	-0.003	0.001	-	0.001	-	0.001
P-value for $\hat{\beta}$	0.009	0.330	0.095	0.084	0.060	0.558
R^2	0.112	0.016	0.047	0.050	0.174	0.006

Table C.16.: Image and output for linear regression models for principal component 4



	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.068	0.044	-	0.036	-	0.012
$\hat{\beta}$	0.003	-0.001	-	-0.001	-	-0.002
P-value for $\hat{\beta}$	0.018	0.274	0.870	0.000	0.020	0.023
R^2	0.092	0.021	0.000	0.262	0.214	0.085

Table C.17.: Image and output for linear regression models for principal component 5

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.019	-0.042	-	-0.011	-	-0.000
$\hat{\beta}$	0.001	0.001	-	0.000	-	0.000
P-value for $\hat{\beta}$	0.472	0.251	0.244	0.192	0.319	0.940
R^2	0.009	0.023	0.023	0.029	0.100	0.000

Table C.18.: Image and output for linear regression models for principal component 6

C.4.4. Subset of data set: Independent component analysis (ICA)

Here only a subset as mentioned in section 4.7 are used.

Graphs in figure C.25 and C.26 with the mixing values of the principal component 1 to 6 on the y-axis and the 2nd information on the x-axis illustrate the relation between the response variables and the explanatory variables graphically. The green line is the regression line. Tables C.19 to C.24 show the relations with p-values.

Graphs and tables indicates following relations (no assumptions violated and p-value of coefficient < 0.05):

IC 1 and MoCA / Age are related (Duration shows not a constant variance of the residuals and although Hoehn & Yahr is significant, the graph shows no meaningful relation.)

IC 2 is related UPDRS.

IC 4 is related to Sex.

IC 5 is related MoCA, UPDRS, Hoehn & Yahr and Duration (Age shows a non-constant variance). Hoehn & Yahr levels 1 and 1.5 have positive values in component 5 and levels 2 to 4 have rather negative values.

IC 6 is related to UDPRS.

Linear regression of component 1 and MoCA has the highest R^2 .

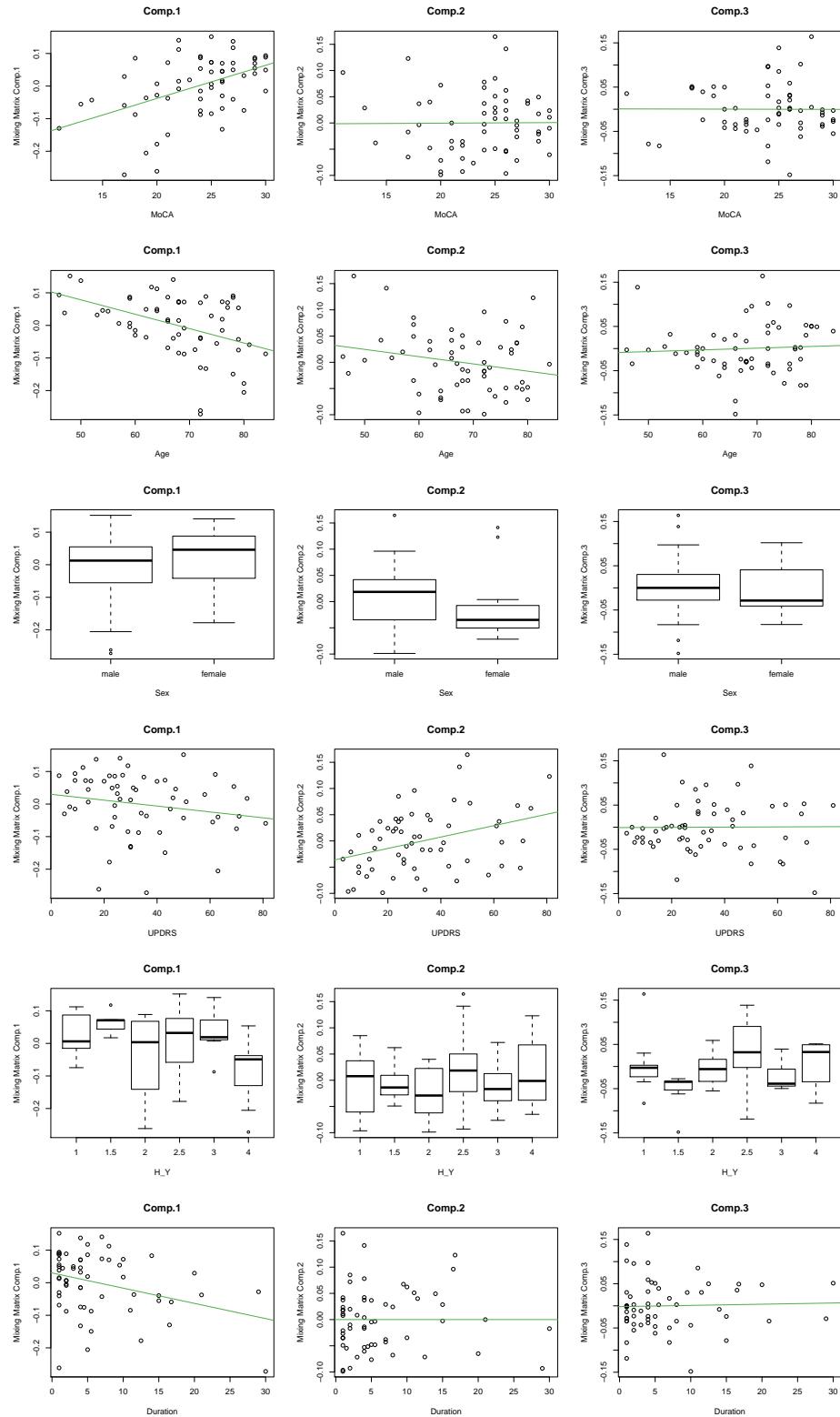


Figure C.25.: Graphs of independent component 1, 2, 3 (columns) and MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration (rows).

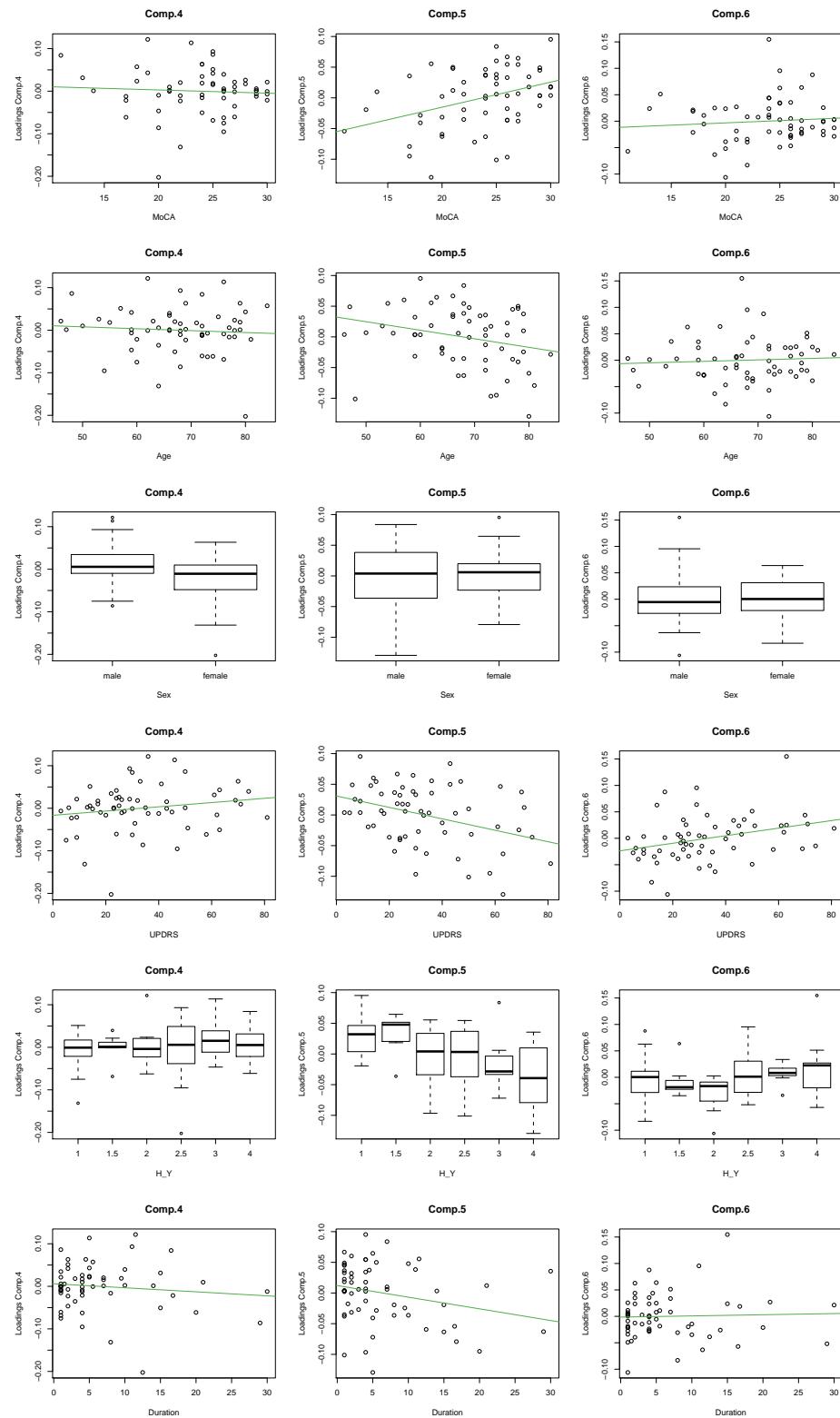
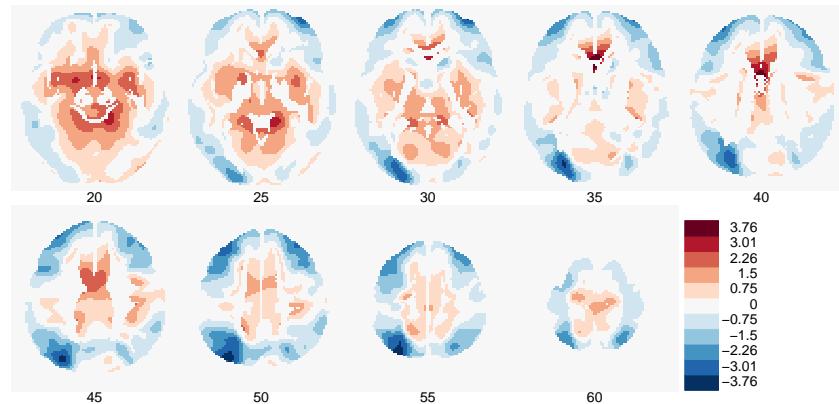
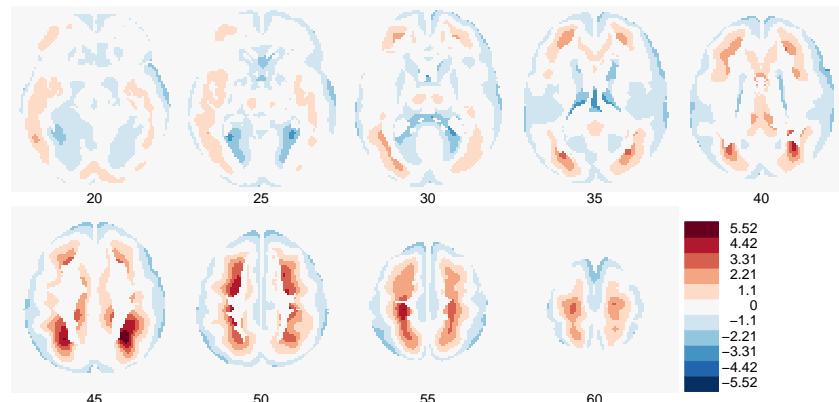


Figure C.26.: Graphs of independent component 4, 5, 6 (columns) and MoCA, Age, Sex, UPDRS, Hoehn & Yahr and Duration (rows).

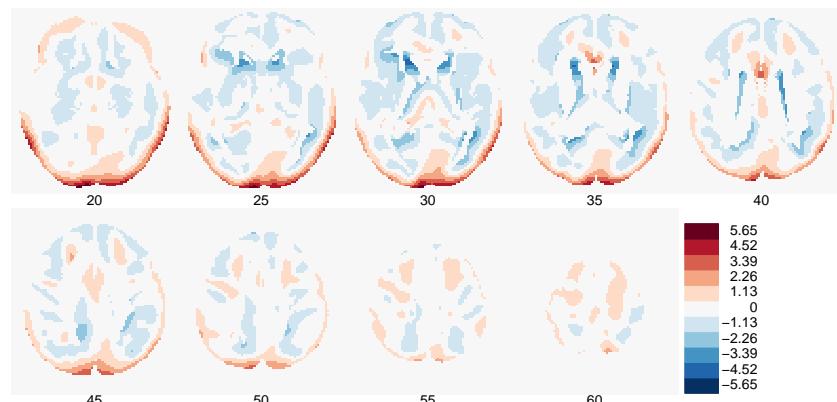


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.242	0.299	-	0.030	-	0.030
$\hat{\beta}$	0.010	-0.004	-	-0.001	-	-0.005
P-value for $\hat{\beta}$	0.000	0.001	0.242	0.152	0.021	0.009
R^2	0.228	0.185	0.024	0.035	0.213	0.112

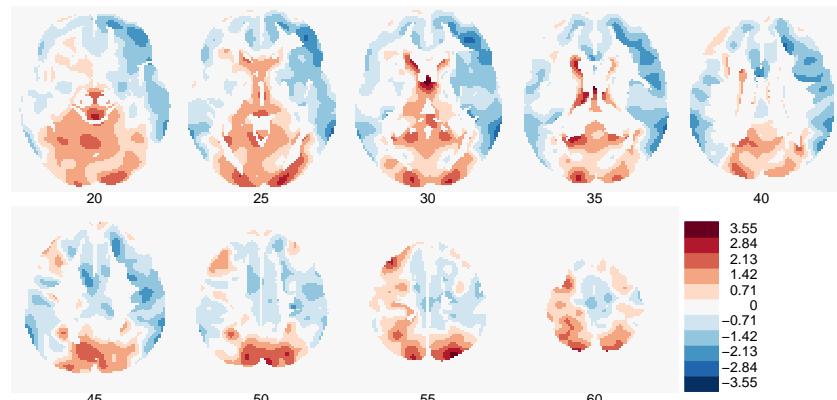
Table C.19.: Image and output for linear regression models for independent component 1

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.003	0.094	-	-0.036	-	0.000
$\hat{\beta}$	0.000	-0.001	-	0.001	-	-0.000
P-value for $\hat{\beta}$	0.951	0.100	0.123	0.005	0.431	1.000
R^2	0.000	0.046	0.041	0.129	0.084	0.000

Table C.20.: Image and output for linear regression models for independent component 2

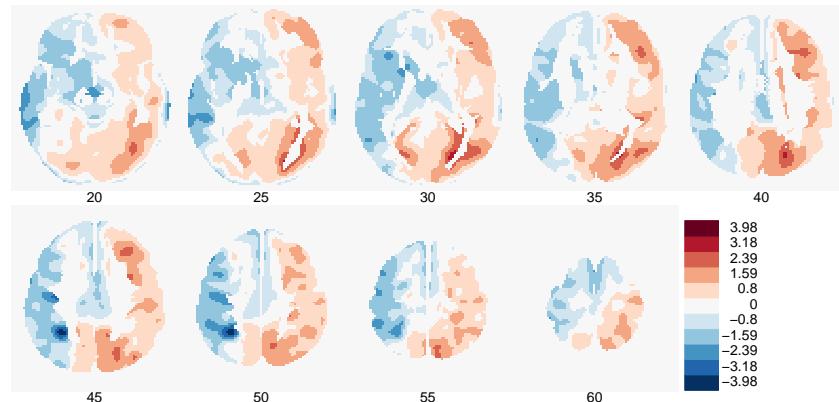


	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.001	-0.026	-	-0.001	-	-0.002
$\hat{\beta}$	-0.000	0.000	-	0.000	-	0.000
P-value for $\hat{\beta}$	0.975	0.638	0.557	0.947	0.021	0.809
R^2	0.000	0.004	0.006	0.000	0.212	0.001

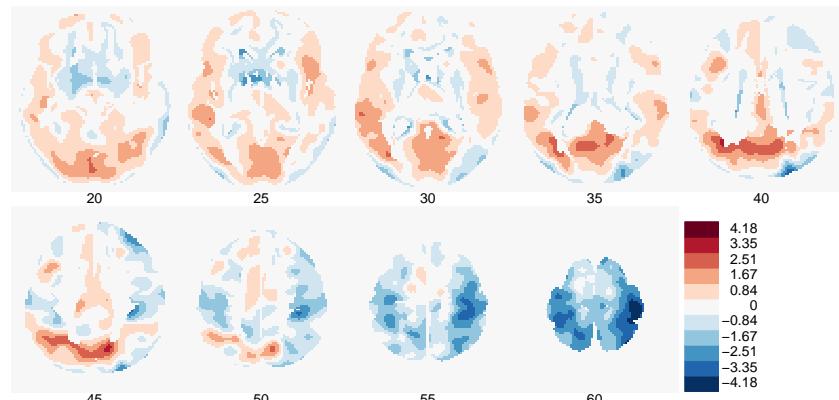
Table C.21.: Image and output for linear regression models for independent component 3

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	0.018	0.030	-	-0.017	-	0.006
$\hat{\beta}$	-0.001	-0.000	-	0.000	-	-0.001
P-value for $\hat{\beta}$	0.649	0.576	0.014	0.180	0.905	0.380
R^2	0.004	0.005	0.100	0.031	0.028	0.013

Table C.22.: Image and output for linear regression models for independent component 4



	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.097	0.094	-	0.031	-	0.012
$\hat{\beta}$	0.004	-0.001	-	-0.001	-	-0.002
P-value for $\hat{\beta}$	0.004	0.046	0.745	0.004	0.007	0.045
R^2	0.134	0.067	0.002	0.137	0.249	0.068

Table C.23.: Image and output for linear regression models for independent component 5

	MoCA	Age	Sex	UPDRS	H & Y	Duration
$\hat{\alpha}$	-0.020	-0.019	-	-0.024	-	-0.001
$\hat{\beta}$	0.001	0.000	-	0.001	-	0.000
P-value for $\hat{\beta}$	0.502	0.652	0.808	0.012	0.206	0.793
R^2	0.008	0.004	0.001	0.105	0.122	0.001

Table C.24.: Image and output for linear regression models for independent component 6

Bibliography

Epidemiology

- [1] de Lau, L.M. & Breteler, M.M., 2006. Epidemiology of Parkinson's disease. *The Lancet Neurology*, 5(6), 525-535.
- [2] Lang, A.E. & Lozano, A.M., 1998. Parkinson's disease-First of two parts. *The New England journal of medicine*, 339(15), 1044.

Data

- [3] Goetz, C.G. et al., 2003. The Unified Parkinsons Disease Rating Scale (UPDRS): status and recommendations. *Mov Disord*, 18(7), 73850.
- [4] Hoehn, M.M. & Yahr, M.D., 1967. Parkinsonism: onset, progression, and mortality. *Neurology*, 57(10 Suppl 3), S11-26.
- [5] Litvan, I. et al., 2003. Movement Disorders Society Scientific Issues Committee report: SIC Task Force appraisal of clinical diagnostic criteria for Parkinsonian disorders. *Movement Disorders: Official Journal of the Movement Disorder Society*, 18(5), 467-486.
- [6] Montreal Cognitive Assessment online. <http://www.mocatest.org/>, accessed March 31, 2010.
- [7] Nasreddine, Z.S. et al., 2005. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699.
- [8] Neuroimaging Informatics Technology Initiative (2010). NIFTI Data Format. <http://nifti.nimh.nih.gov/nifti-1>, accessed March 31, 2010.
- [9] Pooley, R.A., 2005. Fundamental Physics of MR Imaging1. *Radiographics*, 25(4), 1087-1099. (<http://radiographics.rsna.org/content/25/4/1087.full>, accessed May 25, 2010.)
- [10] Wellcome Department of Cognitive Neurology, University College London, UK. <http://www.fil.ion.ucl.ac.uk/spm>, accessed May 19, 2010.

Previous work

- [11] Alexander, G.E. & Moeller, J.R., 1994. Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: A principal component approach to modeling brain function in disease. *Human Brain Mapping*, 2(1-2), 79-94.
- [12] Ashburner, J. & Friston, K.J., 2000. Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6), 805-821.
- [13] Ashburner, J. & Friston, K.J., 2001. Why voxel-based morphometry should be used. *NeuroImage*, 14(6), 1238-1243.
- [14] Baek, K. et al., 2002. PCA vs. ICA: A comparison on the FERET data set. In *Joint Conference on Information Sciences*, Durham, NC. pp. 824827.
- [15] Draper, B.A. et al. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, 91(1-2), 115-137.
- [16] Eidelberg, D., 2009. Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. *Trends in Neurosciences*, 32(10), 548-557.
- [17] McKeown, M.J., Sejnowski, T.J. & others, 1998. Independent component analysis of fMRI data: examining the assumptions. *Human Brain Mapping*, 6(5-6), 368-372.
- [18] McKeown, M.J., Hansen, L.K. & Sejnowski, T.J., 2003. Independent component analysis of functional MRI: What is signal and what is noise? *Current Opinion in Neurobiology*, 13(5), 620-629.
- [19] Turk, M. & Pentland, A., 1991. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1), 71-86.

Methods

PCA

- [20] Edwards, J. & Oman, P., 2003. Dimensional reduction for data mapping. *R News*, 3(3). (p. 2-7)

ICA

- [21] Bell, A.J. & Sejnowski, T.J., 1995. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6), 1129-1159.
- [22] Oja, E. & Hyvärinen, A., 2000. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5), 411-430.

- [23] Hyvärinen, A., 1999. Survey on independent component analysis. *Neural Computing Surveys*, 2(4), 94128.
- [24] Hyvärinen, A. & Oja, E., 1997. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7), 1483-1492.

Logistic Regression & Bootstrap & Cross-Validation

- [25] Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- [26] Harrell, F.E., 2001. Regression modeling strategies, Springer. (p. 87: Bootstrap, p. 215: Logistic Regression)
- [27] Hastie, T., Tibshirani, R. & Friedman, J.H., 2009. The elements of statistical learning, Springer. (p. 241 - 249: Cross-validation, p. 249 - 254: Bootstrap, p. 314 - 317: ROC curve)

Results

- [28] Borghammer, P. et al., 2009. Subcortical elevation of metabolism in Parkinson's disease—a critical reappraisal in the context of global mean normalization. *NeuroImage*, 47(4), 1514-1521.
- [29] Hofmann, H., 2003. Constructing and reading mosaicplots. *Computational Statistics & Data Analysis*, 43(4), 565-580.
- [30] R Project (2010). Download of R. <http://www.r-project.org/>, accessed April 1, 2010. Used packages have been: dcmriS4 (for importing, exporting and visualizing medical images files with format NIfTI), ggplot2, IDPmisc (visualizing large data sets), RColorBrewer (color sets), fastICA (independent component analysis), ROCR (ROC analysis), faraway (halfnormal plot). See <http://cran.r-project.org/web/views/MedicalImaging.html> (accessed April 30, 2010) for more informations about possibilities of medical imaging files in R.