We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Principal Component Analysis Applied to SPECT and PET Data of Dementia Patients – A Review

Elisabeth Stühler and Dorit Merhof
University of Konstanz
Germany

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia, followed by vascular and frontotemporal dementia. Approximatly 8% of the population in developed countries is impaired by AD at the age of 65, with the risk expanding to 30% for individuals over the age of 85 years (Petrella et al. (2003)). Due to the increasing life expectancy, the spread of AD is estimated to triple over the next 50 years (Petrella et al. (2003)). If AD remained untreated, the economic impact on society would increase dramatically (Carr et al. (1997); Mueller et al. (2005)), but it is even more important to alleviate the psychological strain on patients and their relatives. Normally, a patient affected by AD has an anticipated average life expectancy of 8-10 more years, divided into several stages of the disease. The neuropathological stages of AD are described in detail in Braak & Braak (1991), where the development of amyloid deposition and neurofibrillary changes within the brain are explained. These changes can already be observed in the preclinical phase, i.e., before clinical symptoms occur. Clinical symptoms usually begin (in early stages) with memory and learning impairment, followed by alterations in judgement, display of social behavioral problems and reduced faculty of speech. In late stages of AD, motoric and sensory functions are affected as well (Selkoe (2001)).

First pharmaceuticals for treatment of AD symptoms were recently developed, and there are several more under clinical trials at the moment, which in turn require the early detection of AD (Petrella et al. (2003)). Cases with early-onset AD are usually diagnosed with mild cognitive impairment (MCI). According to Tabert et al. (2006), about 10% of cases with amnestic MCI (i.e., patients with memory deficits) and about 50% of MCI cases with further cognitive domain deficits will convert to AD within three years.

In early stages of AD, structural changes within the brain are difficult to detect, as they are restrained to very specific areas (e.g., hippocampal atrophy) until AD is advanced to a middle or later stage. Petrella et al. (2003) advise therefore to resort to nuclear medicine imaging which captures more subtle pathological changes, rather than to magnetic resonance imaging (MRI) or X-ray computed tomography (CT) as they are less capable for early detection of dementia. Prevalent in clinical assessment of AD are positron emission tomography (PET) and single-photon emission computed tomography (SPECT), where PET is observed to perform superior to SPECT for distinguishing between AD and a control group (CTR), e.g., in Herholz et al. (2002b).

In nuclear medicine, the biomarkers used for detection of AD include increased β-amyloid deposition, decreased glucose metabolism and reduced blood flow in the brain, which are

among many indicators for AD. Furthermore, AD can be correlated to several risk factors, such as the genetic inheritance of the ϵ4 allele of the apolipoprotein E (APOE) or the increased accumulation of tau proteins in the cerebrospinal fluid (CSF).

SPECT or PET images are typically evaluated by clinical reading, but this procedure requires expert knowledge, is time-consuming and rater-dependent. Therefore, statistical analyses for automated detection or prediction of AD progression in MCI have been subject to recent research.

As SPECT or PET datasets contain a large amount of information, i.e., more than 10^5 voxel-values within the whole-brain region, and as usually up to 100 subjects are considered in a study, statistical analysis of the 3D-images is very challenging. It includes univariate analysis where a voxel-wise comparison is performed to differentiate between AD and normal controls (CTR), e.g., in Dukart et al. (2010) and Habeck et al. (2008). More recently also multivariate analysis, such as principal component analysis (PCA), has been applied to enable statistical evaluation of all voxel-values at the same time, thereby accounting not only for differences in single intensity values but also correlations between regions. This usually outperforms univariate analysis in the early identification of AD (Habeck et al. (2008)).

In many studies, PCA is therefore employed to either reduce the high dimensionality of the data (Markiewicz et al. (2009; 2011a;b); Merhof et al. (2009; 2011)), to discriminate between dementia of Alzheimer type and asymptomatic controls (Fripp et al. (2008a); Habeck et al. (2008); Scarmeas et al. (2004)) or to assess the amount of variability of the data (Fripp et al. (2008a;b)).

The objective of this review is to present and discuss these applications of PCA, and also to give an insight into adequate preprocessing of the data and implementation of PCA:

Basically, any analysis of PET or SPECT data requires preprocessing of the data in a first step, comprising registration of each subject to a brain atlas (a.k.a. spatial normalization), smoothing of all voxel-values and normalization of intensities as briefly described in Section 2.2. This enables voxel-wise comparisons between images in univariate analysis (see Section 6.2.1) and the correlation (or interpretation of covariance) of all voxels within the whole-brain region in multivariate analysis.

After preprocessing, neuroimaging data is commonly reduced to a lower-dimensional subspace in the studies reviewed in this work. In most cases, this is achieved by PCA implemented as in Section 3.1, but also by the scaled subprofile model (SSM), which is a modification of PCA described in Section 3.2. Partial least squares correlation or regression (PLSC/ PLSR) is also related to PCA as it is based on the same decomposition procedure (Section 6.2.2).

The method to be used for dimensionality reduction and further analysis depends on the purpose of the study, and also on different criteria regarding stability of the dimensionality reduction. In Section 5, some criteria for the validation of PCA regarding stability and robustness are presented.

After PCA is accomplished on the neuroimaging data of AD patients and a CTR group, the resulting projections of all subjects can be used to train discrimination as described in Section 4.3. Employing MCI cases where AD is prognosed or was already confirmed, the disrimination can then be tested regarding its potential to detect AD in early stages.

A detailed outline of all methods presented in this review and a workflow for the analysis of PET and SPECT data is depicted in Figure 1.

2. Constitution of the data matrix

In all studies reviewed in this work, PET or SPECT images of asymptomatic controls and patients with AD are considered.

Both techniques generate three-dimensional images of the brain, depicting the aggregation of a radioactive tracer and therefore providing metabolic information (e.g., glucose metabolism, brain perfusion or plaque deposition) within distinct brain areas. Although PET produces images with higher resolution, SPECT is considered to be adequate to detect abnormalities of perfusion which are specific for AD (e.g., Caroli et al. (2007); Herholz et al. (2002b); Ishii et al. (1999); Matsuda (2007)). As SPECT is – in comparison to PET – more prevalent and economical, it is commonly the preferred imaging method according to Minati et al. (2009).

Overall, three tracers were used for the SPECT and PET data examined in this review:

SPECT-scans based on the tracer technetium-99m-ethyl cysteinate dimer (^{99m}Tc-ECD) show perfusion patterns of the brain. In Herholz et al. (2002b) it is observed, that superior results regarding the detection of AD and the assessment of affected brain regions can be achieved by ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG) PET-imaging, which measures the changes in glucose metabolism (Ishii et al. (1999)). The tracer ^{11}Carbon-Pittsburgh compound B (^{11}C PiB) is able to quantify β-amyloid deposition in the diseased brain as pointed out by Klunk et al. (2004).

2.1 Sample selection

If a groupwise comparison of subjects with AD and CTR is intended by statistical analysis of PET or SPECT images, not all datasets are apt to be included in the sample. Especially the CTR group should be gender- and age-matched to account for age-related atrophy within the brain. The effect of age-related changes of the brain on multivariate analysis such as PCA is discussed in Zuendorf et al. (2003), where at least two principal components, i.e., two independent directions of variance, could be correlated with age.

Subjects representing the AD group should not be affected by other neuro-degenerative diseases, and are also recommended to be in a stage of mild to moderate AD. Cases of late AD, where almost the whole brain is affected, would put too much emphasis on regions still unaffected by early-onset AD.

2.2 Preprocessing of the images

In each study, the PET or SPECT images selected for statistical analysis are registered to an atlas of the brain (a.k.a. spatially normalized), smoothed and intensity normalized. An optimized preprocessing method for SPECT images is presented in Merhof et al. (2011), where a dataset containing AD and CTR subjects is preprocessed by various methods, and subsequently tested by PCA and discrimination analysis. Best results regarding robustness and classification accuracy are achieved by affine registration (Bradley et al. (2002); Herholz et al. (2002a)), smoothing of voxel intensity values based on the standard isotropic Gaussian filter with full width half maximum (FWHM) of 12mm (Herholz et al. (2002a); Ishii et al.
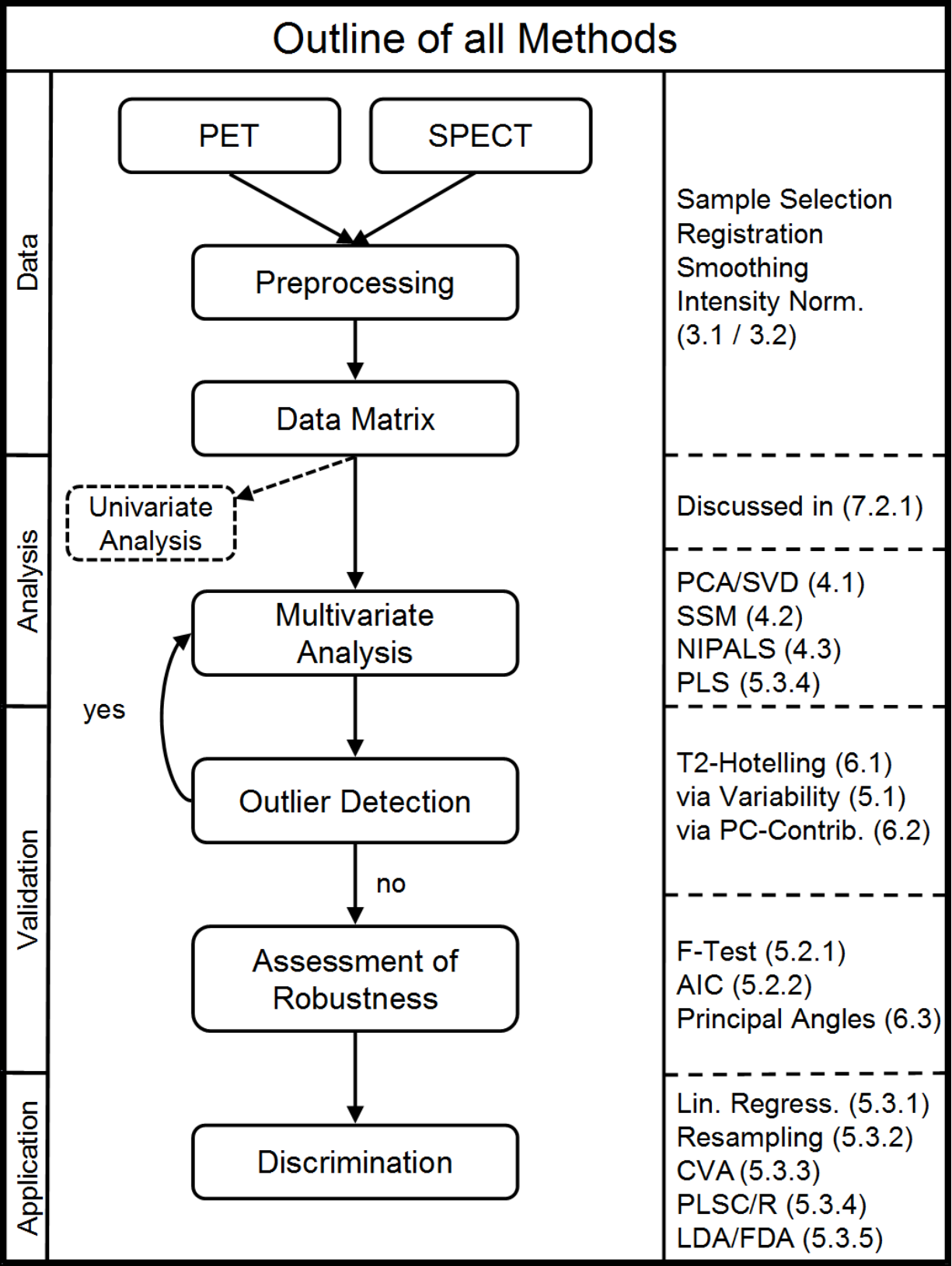
Outline of all Methods

Data

PET   SPECT

Preprocessing

Sample Selection
Registration
Smoothing
Intensity Norm.
(3.1 / 3.2)

Data Matrix

Analysis

Univariate
Analysis

Discussed in (7.2.1)

Multivariate
Analysis

PCA/SVD (4.1)
SSM (4.2)
NIPALS (4.3)
PLS (5.3.4)

yes

Outlier Detection

T2-Hotelling (6.1)
via Variability (5.1)
via PC-Contrib. (6.2)

Validation

no

Assessment of
Robustness

F-Test (5.2.1)
AIC (5.2.2)
Principal Angles (6.3)

Application

Discrimination

Lin. Regress. (5.3.1)
Resampling (5.3.2)
CVA (5.3.3)
PLSC/R (5.3.4)
LDA/FDA (5.3.5)

Fig. 1. Application flow and methods presented in this review

(2001); Matsuda et al. (2002)) and normalization according to the 25% brightest voxels within the whole-brain region.

To our knowledge, a detailed review of preprocessing methods and their impact on PCA applied to PET images (and subsequent analysis, with regard to discrimination of AD and CTR) has not yet been published. However, Herholz et al. (2004) present a detailed and effectual survey of the general handling of PET images in neuroscience.

After sample selection and preprocessing, the voxel-values of each scan are converted into a vector and all datasets are stored column-wise in a data matrix X as depicted in Figure 2. This enables univariate (i.e., voxel-wise) comparison, or multivariate analysis (e.g., PCA and in some cases subsequent discriminant analysis), as the observations for each voxel or brain region are now represented row-wise in X.

3. Principal component analysis

Two main implementations of PCA are considered in this review:

- The first and widely used approach is based on variance, where principal components (PCs) are determined by singular value decomposition (SVD) of the $m \times n$ data matrix X (e.g., Markiewicz et al. (2009); Merhof et al. (2011)). In this way, it is not necessary to compute the $m \times m$ covariance matrix XX^T which is time-consuming due to the very high dimensionality m of the data (in SPECT and PET images, the whole-brain region contains more than 10^5 voxels) and might even lead to a loss of precision.
- In a second implementation, PCA is modified to a scaled subprofile model (SSM) (e.g., in Habeck et al. (2008); Scarmeas et al. (2004)). SSM is also covariance-based, but also captures the regional patterns of brain function and thereby advances subsequent discriminant analysis. PCA is performed, and afterwards subject scaling factors are calculated to convey each subject's contribution to a fixed PC as described in Alexander & Moeller (1994) and Moeller et al. (1987).

Another framework is presented in Miranda et al. (2008) and Duda et al. (2001), where an approximation of PCA is achieved by minimizing the mean square error of approximation, also characterized as a total least squares regression problem (Van Huffel (1997)). However, to our knowledge this method has not been applied to SPECT or PET data of patients affected by AD and a CTR group so far and is therefore not considered further in this review.

As PCA is sensitive to outliers within the data, methods to perform a more robust PCA are also considered, e.g., in Serneels & Verdonck (2008). However, for analysis of SPECT or PET images the underlying data usually contains a manageable amount of subjects and can therefore be sorted manually or by applying tests as presented in Section 5. It is also advisable to visualize those PCs intended to remain in the subsequent analysis as explained in Section 3.4. Thereby, it can be assured that only those regions of the brain which explain the difference to CTR in mild to moderate AD are considered, and that there are no abnormally prominent regions identified by the PCA.

In this review, PCA via SVD and SSM are presented in Sections 3.1 and 3.2. During resampling, both of these methods may become unstable; therefore, an alternative implementation is indicated in Section 3.3.
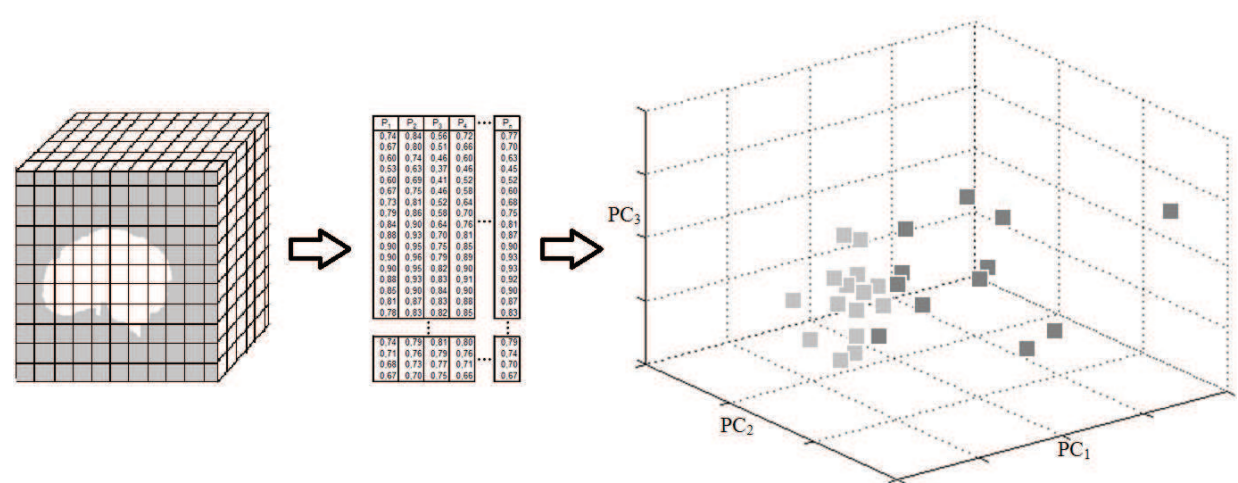
Fig. 2. Exemplary development of PCA on neuroimaging data. *Left*: Volume dataset. *Middle*: Data matrix X containing one volume dataset per column. *Right*: Projection into subspace spanned by three PCs.

A general outline of the PCA on neuroimaging data is depicted in Figure 2, where each image contained column-wise in the data matrix X is projected into a subspace spanned by the first three PCs.

3.1 PCA via singular value decomposition

Prior to multivariate analysis, the overall mean of the data matrix X is usually set to zero by subtracting the mean vector from each column. This is not compulsive but considerably simplifies further analysis (Habeck et al. (2010); Miranda et al. (2008)).

Singular value decompositon (SVD) of the data matrix is applied by $X = VSU^T$ (as in Markiewicz et al. (2009; 2011b)). As X is of size $m \times n$ with $m >> n$, it is sufficient to compute only the first n columns of V, i.e., the first n PCs. If the datasets contained in X were mean-centered beforehand, X is of rank $n - 1$ at most, so the number of PCs to be computed is furthermore reduced to $n - 1$ (this follows from the properties of the associated centering matrix, i.e., it is idempotent and therefore of rank $n - 1$).

The columns of V are sorted according to the magnitude of their associated singular values, i.e., the diagonal elements of S. PC scores for all subjects are computed by $V^T X$, i.e., each subject is projected into a PC-subspace as depicted in Figure 2. If all PCs were used, all variance of the data would be maintained, but a subset of only a few PCs is sufficient to represent more than 60% of the variance (see Section 4.2).

3.2 PCA modified to scaled subprofile model analysis

Scaled subprofile model (SSM) analysis enhances the discriminative powers of the PCA as it not only extracts the covariance structure within groups but also assesses the contribution of each subject to the covariance pattern. As explained in detail in Alexander & Moeller (1994), the data matrix X is natural log-transformed, and subsequently mean-centered over brain regions and subjects. Then PCA is performed on X as in Section 3.1 by $X = VSU^T$ and n PCs are contained in V. Furthermore, PCA via eigenvalue decomposition of the $n \times n$

(a) Principal Component 1   (b) Principal Component 2   (c) Principal Component 3
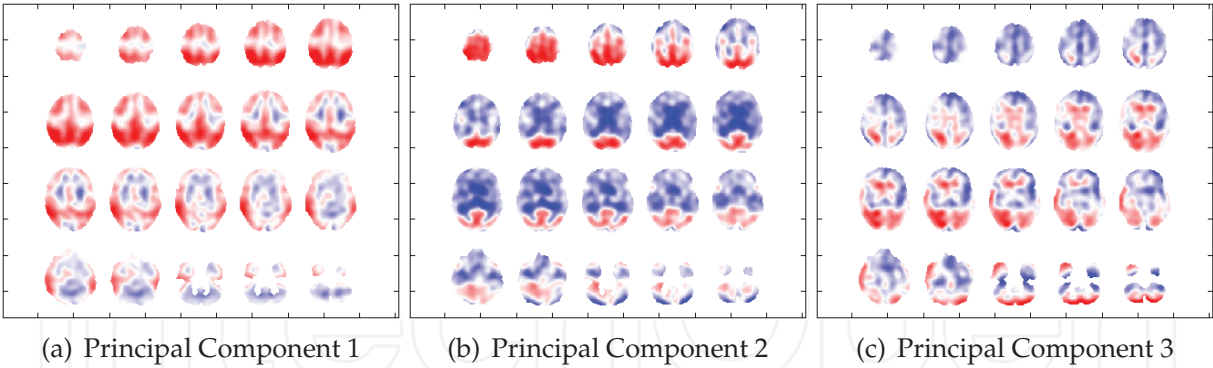
Fig. 3. Examples for the first three principal components of a dataset containing SPECT images of 23 asymptomatic controls and 23 patients with Alzheimer's disease.

covariance matrix $X^T X$ is applied, resulting in n eigenvectors which represent sets of subject scaling factors (SSFs). The associated eigenvalues to the PCs, SSFs respectively, of both decompositions are equal. Whereas the PCs describe the main directions of variance in the data, the SSFs describe the degree of subjects' expression of the fixed PC (Habeck et al. (2008)). The expression of the PC-scores $V^T X$ for each subject is quantified by the associated SSFs in accordance with the procedure described in Alexander & Moeller (1994) and Habeck et al. (2008). As above in Section 3.1, only a few PCs and associated SSFs are sufficient to reflect pathological differences within the data.

3.3 PCA via non-linear iterative partial least squares

During bootstrap resampling (e.g., to assess robustness of the PCA as described in Section 5.3), individual subjects may be repeatedly present within the resampled data matrix, thereby rendering the SVD unstable.

In this case, Markiewicz et al. (2009) propose the application of the non-linear iterative partial least squares (NIPALS) algorithm. The (resampled) data matrix X is decomposed by $X = v_1 t_1^T + R$, where v_1 denotes an estimate of the first PC of X, t_1 represents the appendent PC scores of each subject and R is the remaining residual. As an estimate for v_1, Wold et al. (1987) propose the (normalized) column of X with the largest variance, but the employment of other start vectors is possible as well (Miyashita et al. (1990)). The NIPALS algorithm is iterated with R acting as new start matrix until all PCs required for further analysis are computed. The NIPALS method is related to canonical correlation analysis (Höskuldsson (1988)), and thereby also to canonical variate analysis as presented in Section 4.3.3.

3.4 Visualization of PCs

Axial slices of PCs can be visualized as illustrated in Figure 3, where ^{99m}Tc-ECD SPECT images of 23 subjects with Alzheimer's disease and 23 asymptomatic controls were decomposed by PCA via SVD. As PCA seeks directions for representation (rather than discrimination), the displayed patterns are not to be mistaken with discriminant images.

The voxel-values of each PC are converted back into a three-dimensional image (reverse to the procedure in Section 2.1), and every third slice of the PC-image between slice 15 and 72 is depicted. The intensities of the voxel-values are mapped to a colormap ranging from

blue negative values to red positive values. Neutral voxel-values (= 0, as the data was mean-centered) correspond to white.

The main variance observed in the temporal lobes is captured in the first PC, whereas the second PC expresses changes in the area of the ventricles, which could be attributed to the expansion of ventricles in AD patients. A first intuitive conclusion might be to maintain only the first two PCs for further analysis, as those describe the regions usually affected by AD within the brain most distinctly. However, there are more reliable methods to decide which PCs to keep (see Section 4.2).

4. Applications

In the statistical evaluation of neuroimages, the main purpose of PCA is primarily an efficient reduction of the very high dimensionality and the removal of noise and redundant information within the data. The PCs produced during PCA represent the axes of the new subspace, into which the original datasets containing the voxel-values are transformed. The decision which PCs are suited to represent the data sufficiently has a great impact on all further analysis. Therefore, the contribution of each PC should be thoroughly evaluated. Different criteria for choosing a well-fitting subset of PCs are presented in Section 4.2. Also, the measurement of the amount of variability maintained within each PC is closely connected with the question of its significance (see also Section 4.1).

If the dataset at hand contains two (or more) groups of subjects, the PCs established to be relevant for further analysis are found to notably describe those regions within the brain, which differ significantly across groups. PCA can therefore be useful to train a discrimination or to provide the basis for subsequent discriminant analysis as presented in Section 4.3.

4.1 Explanation of the variability

Under the condition that the variables (voxels) of all subjects are on the same scale (this has to be ensured during preprocessing of the images), the variance of the ith PC equals its associated eigenvalue e_i (Massy (1965)). The percentage of the accumulated variance represented by any number n of all N PCs is then calculated by

$$var(n) = \frac{\sum_{i=1}^{n} e_i}{\sum_{i=1}^{N} e_i}. \tag{1}$$

In several studies it is observed that the first few PCs generally account for more than 60% of the variability (e.g., in Habeck et al. (2008); Markiewicz et al. (2009)). The percentage of the cumulative variance explained is used by Fripp et al. (2008a) to compare different methods for preprocessing of the data, e.g., spatial registration to different brain atlases.

4.2 Dimensionality reduction

In neuroimaging, the number of variables m (i.e., voxels of the whole-brain region) greatly outnumbers the number of observations n (i.e., subjects included in the study). For this reason, a dimensionality reduction of the data before further analysis, such as discrimination or correlation (as in Pagani et al. (2009)), is commonly applied. PCA is well suited for this purpose, as it reduces the variable space to a few dimensions only. It also helps to focus on

the main directions of variance within the data (i.e., the first few PCs) and treats unused PCs corresponding to lower eigenvalues as noise in the data.

In each of the reviewed studies, only the first few principal components (PCs) are used to represent the main variance of the data. In some cases, this is justified by execution of the partial F-test as presented in Section 4.2.1(Markiewicz et al. (2009)), by calculation of the cumulative variance explained by the PCs (e.g., Fripp et al. (2008a); Zuendorf et al. (2003), see also Section 4.1) or by application of Akaike's information criterion (Habeck et al. (2008); Scarmeas et al. (2004), see also Section 4.2.2).

4.2.1 Partial F-test

The partial F-test measures which PCs add significant variance to the data (Markiewicz et al. (2009)). In the beginning, the mean-centered data matrix $X = X_{start}$ is entered into a regression model, and its residual sum of squares $rss(1)$ is computed. In a first iteration, the first PC v_1 is added to the model and prediction values for the original data matrix are calculated by $v_1 v_1^T X_{start}$. Then the residual sum of squares of the deviation matrix $D = X_{start} - v_1 v_1^T X_{start}$ is calculated. In each of the following $N - 1$ iterations, D and the next PC are entered into the model.
F-values and p-values for each iteration are calculated by

$$F_n = \frac{(rss(n) - rss(n+1))(N - n)}{rss(n+1)} \qquad (2)$$

and

$$p_n = 1 - fcdf(F(n)), \qquad (3)$$

where $fcdf$ denotes the F cumulative distribution function with numerator and denominator degrees of freedom 1 and $N - n - 1$.
As the limiting factor for number of PCs, Markiewicz et al. (2009) propose p to be lower than 0.05, which is a standard level of significance.

4.2.2 Akaike's information criterion

Similar to partial F-test, Akaike's information criterion (AIC) determines the subset of PCs which represents the best fitting model (Akaike (1974)).

AIC-values are calculated by

$$A = -2log(L) + 2K, \qquad (4)$$

where L denotes the maximum value of the log-likelihood function of the model and K the number of estimable parameters (Burnham & Anderson (2002)). The model which scores the smallest AIC-value A is considered to be the best fitting one. As AIC may be biased if the ratio of sample size and number of parameters is too small (e.g., $\frac{n}{K} < 40$), Sugiura (1978) proposes a correction factor (AIC$_c$):

$$A_c = A + \frac{2K(K+1)}{n - K - 1}. \qquad (5)$$

Burnham & Anderson (2002) recommend the usage of AIC$_c$ in any case, as AIC and AIC$_c$ are similar for a sufficiently large ratio $\frac{n}{K}$.

In Habeck et al. (2008), AIC-values A are calculated only for models generated by the first six PCs (explaining more than 75% of all variance), and the best-fitting model with the lowest

AIC-value is chosen for subsequent analysis. However, it should be noted that the AIC does not recognize if none of the models is suited to represent the population, i.e., the PCs entered into the AIC need to be chosen carefully.

4.3 Discrimination methods

With regards to the early detection of AD, the discriminative power of PCA can be very valuable. Discrimination should be trained on subjects with mild to moderate AD and asymptomatic CTR, and afterwards be tested on MCI cases, thereby assessing the capability to detect early AD cases among the data collected for the study.

Due to the orthogonality of all eigenvectors, each PC is uncorrelated with all preceding PCs and therefore captures an independent feature of the dataset. As the main variance resides in the first PCs, they depict prominent features of the data (provided that there are no outliers). Hence, the PCs can be employed for the differentiation of groups within the dataset. Those PCs which best discriminate the subjects can either be determined in a linear regression model as presented in Section 4.3.1 (Habeck et al. (2008); Scarmeas et al. (2004)) or as in Section 4.3.2 by a leave-one-out approach (Fripp et al. (2008a)). If necessary, discrimination can be refined further, e.g., by Canonical variate analysis (Section 4.3.3) or Fisher's discriminant analysis (Section 4.3.5).

4.3.1 Linear regression

Linear regression is a subtype of general regression analysis and is widely used for the identification of those independent variables, which relate strongly to the dependent variable (e.g., group membership). After the successful completion of the regression, it can furthermore be applied to predict the group membership of a newly added value.

The achieved PC-scores $\tilde{X}$ of each subject are entered as independent variables into the linear regression model $y = \tilde{X}b + \epsilon$. The vector y of the subjects' group memberships, in this case AD and CTR, contains the dependent variables.

It is common to use only a subset of all PCs (determined by significance tests or the amount of variance they represent), but it should be noted that even a PC which captures little variance might still be related to a dependent variable (Jolliffe (1982)).

The regression results in a linear combination of those PCs which achieve the best differentiation of the two classes (e.g., Habeck et al. (2008); Scarmeas et al. (2004)).

If the dependent variables include more information than group membership (e.g., age or existence of genetic risk factors), partial least squares (PLS) regression can be applied (see also Section 4.3.4). This method generalizes PCA and multiple linear regression.

4.3.2 Leave-one-out resampling

In leave-one-out resampling, one subject is drawn from the underlying data sample per iteration and subsequent analysis is applied. This measures the individual contribution of each subject and can therefore be applied to sort out abnormal interference of particular subjects where necessary.

In Fripp et al. (2008a), $n - 1$ out of n images are decomposed by PCA in each iteration. Then, PC-scores of the subjects contained in the sample are plotted pairwise against each other.

Those PCs which generally provide the best cluster and separation of the groups within iterations are considered for further analysis.

4.3.3 Outline of canonical variate analysis

Canonical variate analysis (CVA) is another regression model considered to enhance the discriminative strength of PCA in neuroimaging. Similarly to linear regression, it identifies the best separation of groups depending on PC-scores. The first canonical variable (CV) is the best of all possible linear combinations of PC-scores for differentiation of the groups and – analogous to PCs – the following CVs are computed under the condition to be orthogonal to all precedent CVs.

PCA is applied for dimensionality reduction and removal of noise (i.e., discarded PCs). The within- and between-group sum-of-squares and crossproduct matrices W and B are computed for the PC-scores of all subjects. Then the CVs, i.e., the eigenvectors of $W^{-1}B$, are linear combinations of PC-scores and are sorted by their discriminative power (Borroni et al. (2006); Kerrouche et al. (2006)). CV-scores of all subjects are calculated analogous to PC-scores.

4.3.4 Outline of partial least squares correlation and regression

As in PCA, the main element of partial least squares (PLS) methods is the SVD, which is applied to the correlation matrix YX^T (rather than the data matrix X containing the mean-centered data, as in PCA). The independent variables are the mean-centered and normalized voxel-values of all brain images stored in X, and the n vectors of dependent variables for all subjects (also mean-centered and normalized) form the $k \times n$ matrix Y. SVD of YX^T produces VSU^T, where S is a diagonal matrix containing singular values and U and V column-wise contain the left respectively right singular vectors. Analogous to PCA, it is sufficient to compute only the first few columns of V. Then, the high-dimensional data of X is reduced by $T = X^T U$ (a.k.a. brain scores), and Y is reduced to $Y^T V$ (a.k.a. behavior scores).

It depends on the intention of the study, in which way these results are further analysed and applied. Krishnan et al. (2011) give an elaborate survey of the main PLS methods used in neuroimaging as well as of practical implementations. Generally, they present two basic approaches, i.e., PLS regression and PLS correlation. PLS regression is a generalization of multiple linear regression and PCA (Abdi (2010)), and is used to predict behavior on the basis of neuroimages, in this case PET or SPECT data. PLS correlation focuses on the analysis of the relation between two groups within the dataset and can be subdivided into more specific applications according to the design of the research.

4.3.5 Outline of linear and Fisher's discriminant analysis

Similar to CVA, linear discriminant analysis (LDA) seeks discriminative directions of the data rather than representative directions (as does PCA). It can be applied both to the original mean-centered voxel-values contained in the data matrix X or in a second step after performance of PCA to the PC-scores of all subjects. The latter approach is preferable when dealing with high-dimensional data, as either the inverse of an $m \times m$ scatter matrix has to be computed or a generalized eigenvalue decomposition of $m \times m$ matrices is required.

Fisher's discriminant analysis (FDA) is a special application of LDA, without the constraints of normal distributed groups and equal group covariance. It has lately been applied several

times to diffentiate between subjects with AD and normal controls, e.g., in Markiewicz et al. (2009; 2011a); Merhof et al. (2009; 2011).

The purpose of FDA is to maximize the ratio of the between- and the within-group scatter S_B and S_W, thereby projecting the data into a one-dimensional subspace. This is achieved by the projection vector w, i.e., the solution of the generalized eigenvalue problem $S_W^{-1} S_B w = \lambda w$ (Duda et al. (2001)). Subsequent classification can be computed with very limited effort by a threshold or nearest-neighbor approach.

5. Derivation of robustness of the PCA

So far, PCA and its applications in neuroimaging were introduced, but not yet validated and discussed. It is very important to assess the robustness of the PCA (and, where necessary, subsequent procedures) before interpretation of the results, as instability and overtraining may occur for various reasons. PCA is sensitive to conspicuous cases and it is therefore recommended to inspect the resulting PCs before further analysis. In order to ensure that no pathologically abnormal cases (outliers) remain in the training set, the T^2-Hotelling test is executed, e.g., by Pagani et al. (2009); Zuendorf et al. (2003) (see Section 5.1). Kerrouche et al. (2006) also propose further measurement of the individual contribution of one observation to each PC (see Section 5.2), to assess if the removal of one observation changes the outcome of PCA significantly. Habeck et al. (2010) also observe that if the first PC contains more than 90% of the variance to the data, it is very probable that the dataset X includes one or more outliers (see Section 4.1).

By bootstrap resampling of the dataset and subsequent PCA the instability caused by removal of a subset of subjects is measured (Markiewicz et al. (2009; 2011a); Merhof et al. (2011)) via principal angles between PC-subspaces.

5.1 Hotelling's T-square test

Hotelling's T-square test is an adaption of the Student's T-test to the multivariate case (Hotelling (1931)). As the F-distribution is more prevalent, the T^2-distribution is usually transformed to

$$T^2 \sim \frac{p(n-1)}{n-p} F_{p,n-p,} \tag{6}$$

where n denotes the number of subjects and p the number of PCs retained in the model. Let y_i denote the column vector of PC-scores of the ith subject, then its T^2-value is obtained by $T^2 = y_i^T y_i$. Zuendorf et al. (2003) propose a threshold of $p < 0.01$, and further assess the validity of the T^2-test by adding an abnormal case to a set of normal controls in 15 iterations. However, the T^2-test can also be applied to a dataset containing two or more groups (Kerrouche et al. (2006); Pagani et al. (2009)).

5.2 Contribution of subjects to PCs

The amount of the contribution $c_{i,j}$ of the ith subject to the jth PC is measured by

$$c_{i,j} = \frac{1}{n-1} \cdot \frac{y_i(j)}{e_j}, \tag{7}$$

where n denotes the number of all subjects, y_i the column vector of PC-scores of the ith subject and e_j the eigenvalue corresponding to the jth PC. An abnormally large value of $c_{i,j}$ indicates that the removal of the ith subject might significantly change the results of PCA (Kerrouche et al. (2006)).

5.3 Principal angles of PC-subspaces

In order to compare sets of PCs during resampling iterations, the use of principal angles between PC-subspaces of a fixed dimension is proposed by Markiewicz et al. (2009). If the largest principal angle between an original and resampled subspace is very small, the PCA can be considered to be sufficiently independent of the underlying training set. Otherwise, abnormal large principal angles can indicate that too many PCs (i.e., too much noise) are included in the analysis, or that the sample was not selected carefully enough with respect to outliers.

In bootstrap resampling, n subjects are drawn with replacement from the original training set (Efron & Tibshirani (1993)). For better replication of the original set, AD and CTR cases are stratified in the bootstrap sample (Markiewicz et al. (2009)).

For every sample, PCA is performed and the subspace spanned by the first i PCs is compared to the i-dimensional PC-subspace of the original set. This is achieved by calculating the largest principal angle between the two subspaces (Golub & van Van Loan (1996); Knyazev et al. (2002)). For any number of PCs, the mean angle across all iterations is computed. Increased angles indicate the destabilization of the PCA.

The same method can be applied with leave-one-out resampling, i.e., one subject of each group is dropped in every iteration (Markiewicz et al. (2009)).

The computation of prinipal angles should be treated with caution, as round-off errors might cause inaccurate estimates for small angles. A solution to this problem is proposed by Knyazev et al. (2002), where a combined sine and cosine based approach is presented and generalized.

6. Discussion

For analysis of SPECT and PET data, PCA is widely applied and commonly reported to deliver stable and efficient results when used correctly. However, some limitations of PCA outlined in Section 6.1 remain, which might interfere strongly with the outcome of the statistical analyses. In some cases it might even be advisable to apply alternative methods to obtain more reliable results. In Section 6.2, examples are given where the performance of PCA on neuroimaging data was investigated and compared to other approaches.

6.1 Limitations of the PCA in neuroimaging

As PCA is based solely on the decomposition of the covariance matrix, the underlying data must be dealt with carefully. The preprocessing of the images has a crucial impact on the outcome of the analysis as pointed out by Fripp et al. (2008a), where PCA on ^{11}C PiB PET data proved to be sensitive to inaccuracies originating from non-rigid registration and intensity normalization. On ^{99m}Tc-ECD SPECT, the classification accuracy of AD and CTR subjects via PCA and subsequent FDA depends significantly on the data preprocessing method (Merhof

et al. (2011)). Classification accuracy relies also on scanner type and reconstruction method of FDG-PET, if the data is aquired from a more heterogeneous dataset, e.g., from the ADNI database as in Markiewicz et al. (2011b).

Not only the preprocessing of the training set but also its composition is essential. This includes the stratification of groups, the sample sizes and the absence of outliers as described in Sections 6.1.1 and 6.1.2. Moreover, the number of PCs retained in the analysis is important and depends on the purpose of the study (see Section 6.1.3).

6.1.1 Sample size

The selection of subjects suited for training is constrained by many premises, such as age- and gender-matched CTR cases, stage of AD, the absence of other neurodegenerative disorders and the quality of the scan. It is also preferable for all images to be aquired by the same scanner, as this improves comparability of the data. Therefore, most studies only include less than 30 subjects of each group, except for Markiewicz et al. (2011b) where previous results on a smaller and more homogeneously selected training set were validated on more than 160 AD and CTR cases obtained from the ADNI database. ADNI provides generally accessible data of patients diagnosed with AD or MCI and of normal controls collected from various clinical sites (Mueller et al. (2005)).

An under-sized training set might cause the extraction of instable features (Markiewicz et al. (2009)) resulting in overly optimistic results of subsequent analysis (Markiewicz et al. (2011a)). This might be remedied by bootstrap resampling of the training set but must rely on the assumption that the sample is representative of the population (Markiewicz et al. (2009)).

6.1.2 Sensitivity to outliers

As the covariance matrix is calculated empirically, the estimates of eigenvectors (PCs) are heavily influenced by outliers, i.e., pathologically abnormal cases within the training dataset. The variance caused by only one outlier may be captured within the first PC, which will thereby not regard the variance within regular cases and dramatically change further results. Approaches which substitute the original covariance matrix by a more robust estimate (e.g., in Debruyne & Hubert (2006)) exist, but these methods are not practical for datasets of high dimensionality. For this reason, it is highly recommended to determine outliers by additional testing when applying PCA to neuroimaging data.

6.1.3 Number of PCs

Although several approaches are presented in Section 4.2, the determination of how many and which PCs are best suited to represent the original images remains subject to interpretation. So far, much relies on the purpose of PCA and the further analysis of the data. An elaborate overview over criteria for estimating the number of significant PCs and their application is presented in Peres-Neto et al. (2005), and some of these apply to covariance matrices as well as correlation matrices. The application of such methods can be ambivalent, as reviewed by Franklin et al. (1995). In most studies, PCs are chosen according to their potential to explain data and their impact on robustness. It is therefore advisable to determine the number of retained PCs not only based on one criterion but also on the best possible trade-off between the resulting accuracy and robustness.

6.2 Comparison to similar methods

In Section 4, different extensions to PCA such as linear regression or canonical variate analysis are presented, and also an outline of methods with similar properties or intentions as PCA is given. The decision which of these methods is best to employ always depends on the underlying research question, on the data available and the selected sample.

Sections 6.2.1 to 6.2.3 provide a review of the most important methods frequently applied to neuroimaging data and compare them to results obtained from PCA.

6.2.1 Univariate analysis of neuroimaging data

Univariate analysis measures the voxel-by-voxel correlation between groups (e.g., by a voxel-wise T-test in Habeck et al. (2008)) and thereby merely focuses on the identification of significant shifts between voxel-values. In neuroimaging, univariate methods are commonly used during image preprocessing, e.g., in Dukart et al. (2010) or Scarmeas et al. (2004) for intensity normalization. Voxel-wise analysis can also be used for differentiation between groups, but it is unanimously reported that multivariate approaches outperfom univariate analysis in this matter, especially in the detection of early-onset cases of dementia (Habeck et al. (2008); Scarmeas et al. (2004)). Another drawback of voxel-wise analysis is the sensitivity regarding the preprocessing of the data, and even under the assumption that an optimized normalization factor was applied, the interpretion of the results must be addressed in a multivariate fashion (McCrory & Ford (1991)). In another approach, Higdon et al. (2004) tried to apply a between-group T-test for dimensionality reduction, but this proved to be ineffective and even deteriorated accuracy results.

On the other hand, multivariate analysis is found to be more robust as it considers the entire covariance structure of the data (accounting for relations among regions) and withstands the deviation of individual voxel-values (Borroni et al. (2006); Habeck et al. (2008)). It thereby detects correlated alterations in a diseased brain, whereas univariate analysis might not be able to recognize these differences.

6.2.2 Partial least squares

When examining very high-dimensional data, and especially for the discrimination into groups within the dataset, PLS has been reported to perform better than PCA regarding the classification accuracy (Higdon et al. (2004); Kemsley (1996)). This is rather self-evident, as PCA does not take into account further behavioral data (e.g., neuropsychological data such as Mini-Mental State Examination (MMSE) scores, age, years of education). If PLS is applied for dimensionality reduction, Kemsley (1996) reports that fewer PLS dimensions than PCs were required for a successful subsequent differentiation of the groups. This implies that PLS will capture the most discriminative attributes of the subjects within the first dimensions, rather than the representative directions generated by PCA.

Nevertheless, there are certain drawbacks in the application of PLS methods. PLS tends to overfit the data, so the determination of the number of PLS dimensions kept in the analysis is of decisive importance (Abdi (2010)). In addition, PLS may detect differences which are not characteristic of the examined groups but were produced randomly by noise within the underlying dataset (Kemsley (1996)). Furthermore, PLS only works under the assumption that behavior relates linearly to neuroimaging data (McIntosh & Lobaugh (2004)).

Overall, if allowances are made for these effects and significant behavioral data is available, PLS can still be a favourable alternative to PCA.

6.2.3 Linear discriminant analysis

As explained above, performance of PCA (or any other dimensionality reduction method) prior to LDA is preferable in neuroimaging due to the high dimensionality of the data and the resulting expensive computation. To our knowlegde, LDA as described above in Section 4.3.5 has not yet been applied to discriminate AD from CTR using voxel-values of PET or SPECT images of the whole-brain region, although McEvoy et al. (2009) utilize a stepwise approach of LDA to identify brain regions significant for differentiation.

In other areas also dealing with high-dimensional data, such as object recognition in images, LDA is usually considered to perform superior to PCA. But this is not necessarily the case for small training sets, as pointed out by Martínez & Kak (2001). In the same study they also observe PCA to be less biased than LDA, i.e., less constrained to the training set.

The overall good results regarding accuracy and robustness of the PCA-LDA or PCA-FDA approach (e.g., as presented in Markiewicz et al. (2009)) also indicate, that a preceding PCA does not impair the discriminant analysis.

7. Conclusion

PCA applied to SPECT or PET data is well suited to reduce the high dimensionality of the original dataset containing voxel-values of the whole-brain region. It achieves best results when data is transformed into a subspace spanned by a well-chosen subset of PCs that represents the variability within all datasets and at the same time reduces noise and redundant information. PCA can also be used successfully to train discrimination between AD and a set of asymptomatic CTRs with the intention to enable an early detection of AD, or to provide a stable and effective basis for the subsequent application of discriminant analysis.

8. References

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression), *Wiley Interdisciplinary Reviews: Computational Statistics* 2(1): 97–106.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19(6): 716–723.

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.

Borroni, B., Anchisi, D., Paghera, B., Vicini, B., Kerrouche, N., Garibotto, V., Terzi, A., Vignolo, L., Luca, M. D., Giubbini, R., Padovani, A. & Perani, D. (2006). Combined 99mTc-ECD SPECT and neuropsychological studies in MCI for the assessment of conversion to AD, *Neurobiology of Aging* 27(1): 24 – 31.

Braak, H. & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes, *Acta Neuropathologica* 82: 239–259.

Bradley, K. M., O'Sullivan, V. T., Soper, N. D. W., Nagy, Z., King, E. M., Smith, A. D. & Shepstone, B. J. (2002). Cerebral perfusion SPET correlated with Braak pathological stage in Alzheimer's disease, *Brain* 125(8): 1772–1781.

Burnham, K. P. & Anderson, D. R. (2002 2nd edn). *Model Selection and Multi-Model Inference. A Practical Information-Theoretic Approach*, New York Springer-Verlag.

Caroli, A., Testa, C., Geroldi, C., Nobili, F., Barnden, L., Guerra, U., Bonetti, M. & Frisoni, G. (2007). Cerebral perfusion correlates of conversion to Alzheimer's disease in amnestic mild cognitive impairment, *Journal of Neurology* 254: 1698–1707.

Carr, D., Goate, A., Phil, D. & Morris, J. (1997). Current Concepts in the Pathogenesis of Alzheimer's Disease, *The American Journal of Medicine* 103(3, Supplement 1): 3S – 10S.

Debruyne, M. & Hubert, M. (2006). The Influence Function of Stahel-Donoho Type Methods for Robust Covariance Estimation and PCA.

Duda, R. O., Hart, P. E. & Stork, D. G. (2001). *Pattern Classification*, Wiley-Interscience.

Dukart, J., Mueller, K., Horstmann, A., Vogt, B., Frisch, S., Barthel, H., Becker, G., Möller, H. E., Villringer, A., Sabri, O. & Schroeter, M. L. (2010). Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies, *NeuroImage* 49(2): 1490 – 1495.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall/CRC.

Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T. & Fralish, J. S. (1995). Parallel analysis: a method for determining significant principal components, *Journal of Vegetation Science* 6(1): 99–106.

Fripp, J., Bourgeat, P., Acosta, O., Raniga, P., Modat, M., Pike, K. E., Jones, G., O'Keefe, G., Masters, C. L., Ames, D., Ellis, K. A., Maruff, P., Currie, J., Villemagne, V. L., Rowe, C. C., Salvado, O. & Ourselin, S. (2008a). Appearance modeling of 11C PiB PET images: Characterizing amyloid deposition in Alzheimer's disease, mild cognitive impairment and healthy aging, *NeuroImage* 43(3): 430 – 439.

Fripp, J., Bourgeat, P., Raniga, P., Acosta, O., Villemagne, V. L., Jones, G., O'Keefe, G., Rowe, C. C., Ourselin, S. & Salvado, O. (2008b). MR-less high dimensional spatial normalization of 11C PiB PET images on a population of elderly, mild cognitive impaired and Alzheimer disease patients., *Medical Image Computing and Computer-Assisted Intervention* 11(Pt 1): 442–449.

Golub, G. H. & van Van Loan, C. F. (1996). *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)*, 3rd edn, The Johns Hopkins University Press.

Habeck, C., Foster, N. L., Perneczky, R., Kurz, A., Alexopoulos, P., Koeppe, R. A., Drzezga, A. & Stern, Y. (2008). Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease, *NeuroImage* 40(4): 1503–1515.

Habeck, C., Stern, Y. & (2010). Multivariate Data Analysis for Neuroimaging Data: Overview and Application to Alzheimer's Disease, *Cell Biochemistry and Biophysics* 58: 53–67. 10.1007/s12013-010-9093-0.

Herholz, K., Herscovitch, P. & Heiss, W. (2004). *NeuroPET: PET in Neuroscience and Clinical Neurology*, Springer, Berlin.

Herholz, K., Salmon, E., Perani, D., Baron, J., Holthoff, V., Frölich, L., Schönknecht, P., Ito, K., Mielke, R., Kalbe, E., Zündorf, G., Delbeuck, X., Pelati, O., Anchisi, D., Fazio, F., Kerrouche, N., Desgranges, B., Eustache, F., Beuthien-Baumann, B., Menzel, C., Schröder, J., Kato, T., Arahata, Y., Henze, M. & Heiss, W. (2002a). Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET, *Neuroimage* 17(1): 302–316.

Herholz, K., Schopphoff, H., Schmidt, M., Mielke, R., Eschner, W., Scheidhauer, K., Schicha, H., Heiss, W. & Ebmeier, K. (2002b). Direct Comparison of Spatially Normalized PET and SPECT Scans in Alzheimer's Disease, *Journal of Nuclear Medicine* 43(1): 21–26.

Higdon, R., Foster, N. L., Koeppe, R. A., DeCarli, C. S., Jagust, W. J., Clark, C. M., Barbas, N. R., Arnold, S. E., Turner, R. S., Heidebrink, J. L. & Minoshima, S. (2004). A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging, *Statistics in Medicine* 23(2): 315–326.

Höskuldsson, A. (1988). PLS regression methods, *Journal of Chemometrics* 2(3): 211–228.

Hotelling, H. (1931). The generalization of student's ratio, *The Annals of Mathematical Statistics* 2(3): pp. 360–378.

Ishii, K., Sasaki, M., Sakamoto, S., Yamaji, S., Kitagaki, H. & Mori, E. (1999). Tc-99m Ethyl Cysteinate Dimer SPECT and 2-[F-18]fluoro-2-deoxy-D-glucose PET in Alzheimer's Disease: Comparison of Perfusion and Metabolic Patterns, *Clinical Nuclear Medicine* 24(8): 572–575.

Ishii, K., Willoch, F., Minoshima, S., Drzezga, A., Ficaro, E. P., Cross, D. J., Kuhl, D. E. & Schwaiger, M. (2001). Statistical Brain Mapping of 18F-FDG PET in Alzheimer's Disease: Validation of Anatomic Standardization for Atrophied Brains, *Journal of Nuclear Medicine* 42(4): 548–557.

Jolliffe, I. T. (1982). A Note on the Use of Principal Components in Regression, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31(3): pp. 300–303.

Kemsley, E. (1996). Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods, *Chemometrics and Intelligent Laboratory Systems* 33(1): 47 – 61.

Kerrouche, N., Herholz, K., Mielke, R., Holthoff, V. & Baron, J.-C. (2006). 18FDG PET in vascular dementia: differentiation from Alzheimer's disease using voxel-based multivariate analysis, *J Cereb Blood Flow Metab* 26: 1213–1221.

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., Bergström, M., Savitcheva, I., Huang, G.-F., Estrada, S., Ausén, B., Debnath, M. L., Barletta, J., Price, J. C., Sandell, J., Lopresti, B. J., Wall, A., Koivisto, P., Antoni, G., Mathis, C. A. & Långström, B. (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B, *Annals of Neurology* 55(3): 306–319.

Knyazev, A. V., Merico & Argentati, E. (2002). Principal angles between subspaces in an a-based scalar product: Algorithms and perturbation estimates, *SIAM J. Sci. Comput* 23: 2009–2041.

Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, *NeuroImage* 56(2): 455 – 475.

Markiewicz, P. J., Matthews, J. C., Declerck, J. & Herholz, K. (2009). Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease, *NeuroImage* 46(2): 472 – 485.

Markiewicz, P., Matthews, J., Declerck, J. & Herholz, K. (2011a). Robustness of correlations between PCA of FDG-PET scans and biological variables in healthy and demented subjects, *NeuroImage* 56(2): 782 – 787. Multivariate Decoding and Brain Reading.

Markiewicz, P., Matthews, J., Declerck, J. & Herholz, K. (2011b). Verification of predicted robustness and accuracy of multivariate analysis, *NeuroImage* 56(3): 1382 – 1385.

Martínez, A. M. & Kak, A. C. (2001). PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.* 23: 228–233.

Massy, W. F. (1965). Principal Components Regression in Exploratory Statistical Research, *Journal of the American Statistical Association* 60(309): pp. 234–256.

Matsuda, H. (2007). Role of Neuroimaging in Alzheimer's Disease, with Emphasis on Brain Perfusion SPECT, *Journal of Nuclear Medicine* 48(8): 1289–1300.

Matsuda, H., Kitayama, N., Ohnishi, T., Asada, T., Nakano, S., Sakamoto, S., Imabayashi, E. & Katoh, A. (2002). Longitudinal Evaluation of Both Morphologic and Functional Changes in the Same Individuals with Alzheimer's Disease, *Journal of Nuclear Medicine* 43(3): 304–311.

McCrory, S. J. & Ford, I. (1991). Multivariate analysis of spect images with illustrations in Alzheimer's disease, *Statistics in Medicine* 10(11): 1711–1718.

McEvoy, L. K., Fennema-Notestine, C., Roddey, J. C., Hagler, D. J., Holland, D., Karow, D. S., Pung, C. J., Brewer, J. B. & Dale, A. M. (2009). Alzheimer Disease: Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment1, *Radiology* 251(1): 195–205.

McIntosh, A. R. & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: applications and advances, *NeuroImage* 23, Supplement 1(0): S250 – S263.

Merhof, D., Markiewicz, P. J., Declerck, J., Platsch, G., Matthews, J. C. & Herholz, K. (2009). Classification Accuracy of Multivariate Analysis Applied to 99mTc-ECD SPECT Data in Alzheimer's Disease Patients and Asymptomatic Controls, *Nuclear Science Symposium Conference Record (NSS/MIC)* pp. 3721–3725.

Merhof, D., Markiewicz, P. J., Platsch, G., Declerck, J., Weih, M., Kornhuber, J., Kuwert, T., Matthews, J. C. & Herholz, K. (2011). Optimized data preprocessing for multivariate analysis applied to 99mTc-ECD SPECT data sets of Alzheimer's patients and asymptomatic controls, *J Cereb Blood Flow Metab* 31(1): 371–383.

Minati, L., Edginton, T., Grazia Bruzzone, M. & Giaccone, G. (2009). Reviews: Current Concepts in Alzheimer's Disease: A Multidisciplinary Review, *American Journal of Alzheimer's Disease and Other Dementias* 24(2): 95–121.

Miranda, A. A., Borgne, Y.-A. & Bontempi, G. (2008). New Routes from Minimal Approximation Error to Principal Components, *Neural Process. Lett.* 27: 197–207.

Miyashita, Y., Itozawa, T., Katsumi, H. & Sasaki, S.-I. (1990). Comments on the NIPALS algorithm, *Journal of Chemometrics* 4(1): 97–100.

Moeller, J. R., Strother, S. C., Sidtis, J. J. & Rottenberg, D. A. (1987). Scaled Subprofile Model: A Statistical Approach to the Analysis of Functional Patterns in Positron Emission Tomographic Data, *Jornal of Cerebral Blood Flow & Metabolism* 7(5): 649–658.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W. & Beckett, L. (2005). The Alzheimer's Disease Neuroimaging Initiative, *Neuroimaging Clinics of North America* 15(4): 869 – 877.

Pagani, M., Salmaso, D., Rodriguez, G., Nardo, D. & Nobili, F. (2009). Principal component analysis in mild and moderate Alzheimer's disease – A novel approach to clinical diagnosis, *Psychiatry Research: Neuroimaging* 173(1): 8 – 14.

Peres-Neto, P. R., Jackson, D. A. & Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited, *Computational Statistics & Data Analysis* 49(4): 974 – 997.

Petrella, J. R., Coleman, R. E. & Doraiswamy, P. M. (2003). Neuroimaging and Early Diagnosis of Alzheimer Disease: A Look to the Future, *Radiology* 226(2): 315–336.

Scarmeas, N., Habeck, C. G., Zarahn, E., Anderson, K. E., Park, A., Hilton, J., Pelton, G. H., Tabert, M. H., Honig, L. S., Moeller, J. R., Devanand, D. P. & Stern, Y. (2004).

Covariance PET patterns in early Alzheimer's disease and subjects with cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance, *NeuroImage* 23(1): 35 – 45.

Selkoe, D. J. (2001). Alzheimer's disease: Genes, proteins, and therapy, *Physiological Reviews* 81(2): 741–766.

Serneels, S. & Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements, *Computational Statistics & Data Analysis* 52(3): 1712 – 1727.

Sugiura, N. (1978). Further analysis of the data by akaike' s information criterion and the finite corrections, *Communications in Statistics - Theory and Methods* 7(1): 13–26.

Tabert, M. H., Manly, J. J., Liu, X., Pelton, G. H., Rosenblum, S., Jacobs, M., Zamora, D., Goodkind, M., Bell, K., Stern, Y. & Devanand, D. P. (2006). Neuropsychological Prediction of Conversion to Alzheimer Disease in Patients With Mild Cognitive Impairment, *Arch Gen Psychiatry* 63(8): 916–924.

Van Huffel, S. (1997). *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, SIAM, Philadelphia, PA.

Wold, S., Geladi, P., Esbensen, K. & Öhman, J. (1987). Multi-way principal components-and PLS-analysis, *Journal of Chemometrics* 1(1): 41–56.

Zuendorf, G., Kerrouche, N., Herholz, K. & Baron, J.-C. (2003). Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional imaging data sets as applied to FDG-PET and normal aging, *Human Brain Mapping* 18(1): 13–21.

Principal Component Analysis - Multidisciplinary Applications

Edited by Dr. Parinya Sanguansat

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as taxonomy, biology, pharmacy,finance, agriculture, ecology, health and architecture.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Elisabeth Stühler and Dorit Merhof (2012). Principal Component Analysis Applied to SPECT and PET Data of Dementia Patients – A Review, Principal Component Analysis - Multidisciplinary Applications, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0129-1, InTech, Available from:

http://www.intechopen.com/books/principal-component-analysis-multidisciplinary-applications/a-review-of-approaches-based-on-principal-component-analysis-and-multivariate-analysis-applied-to-sp

INTECH
open science | open minds