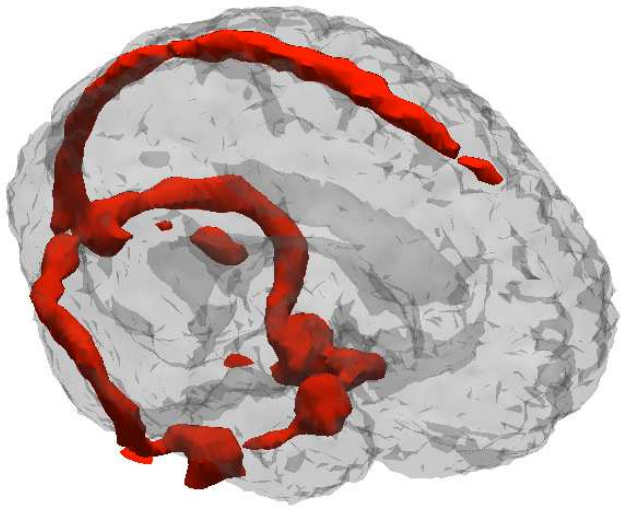
Analysis of Dynamic PET Data

Bjarni Bödvarsson & Martin Mørkebjerg



Kongens Lyngby 2006

IMM-2006-19

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Abstract

In dynamic positron emission tomography (PET) an artery sampling is needed for assessment and validation of parameters in kinetic models. The sampling can be uncomfortable and painful for the patient and technically demanding for the personnel performing the sampling. Noninvasive estimation of the artery time activity curve (TAC) is thus very useful, since the sampling then can be avoided.

Methods are tested on simulated data which is an approximation to PET data. The results from the simulated data are used to quantify how well the different methods perform and to illustrate their limitations. The methods are then used on real PET data, and the estimated TACs are compared to the sampled artery TAC.

Non-negative matrix factorization (NMF) and independent component analysis (ICA) show the best results with correlations around 0.95 with the artery sampled TAC. However, the scaling of the resulting components is lost in the deconvolution so a rescaling scheme is used to get the correct scale in the results. A factor is calculated to solve this scaling problem. The peaks found in the NMF and ICA components are higher than the ones found by the other methods.

The ICA and the NMF results are very similar when applied to real PET data. Therefore, the NMF is chosen as the most appropriate method as it is more stable and not as complicated as the ICA.

Resumé

I dynamisk positron emission tomografi (PET) er det nødvendigt med en arterie blodmåling for at estimere og validere parametre til kinetiske modeller. En blodmåling i arterien kan være ubehagelig og smertefuld for patienten. Samtidig er det teknisk besværligt for personalet der skal foretage målingen. Metoder til estimering af denne blodkurve er derfor meget brugbare, da en måling i arterien derved kan undlades.

Forskellige metoder bliver testet på simuleret data. Disse resultater bliver brugt til at kvantificere hvor godt metoderne klarer sig og hvilke begrænsninger de har. Derefter bliver metoderne kørt på rigtige PET data, og de estimerede blodkurver bliver sammenlignet med de målte blodkurver.

Metoderne Non-negative matrix factorization (NMF) og independent component analysis (ICA) viser de bedste resultater med korrelations koefficienter med de målte blodkurver på omkring 0.95. Metoderne estimerer kurver der har en væsentlig højere top end de andre metoder. Skalering af de estimerede løsninger er ikke kendt, da den forsvinder i estimeringen. Derfor er en måde hvorved de estimerede kurver kan reskaleres nødvendig. Dette bliver løst ved at udregne en faktor der ganges på de estimerede blodkurver.

ICA løsningerne er ikke væsenligt bedre end NMF løsningerne på rigtige PET data. NMF bliver valgt til den mest passende metode, fordi den er mere stabil og ikke ligeså kompleks som ICA metoden.

Preface

This masters thesis is made at the Informatics and Mathematical Modelling (IMM) institute at the Technical University of Denmark in collaboration with Neurobiology Research Unit (NRU) at Copenhagen University Hospital. The work was carried out from September 1 to February 28 2006.

Besides this thesis an abstract for *Human Brain Mapping 2006* has been produced from this work, it is included in Appendix F.

Acknowledgements

First of all we would like to thank our co-supervisor Claus Svarer at NRU and our supervisor Lars Kai Hansen at IMM, who have given us support and rewarding discussions on many issues. Secondly a thanks goes to NRU for providing us with data and also to all the people at NRU for always being helpful answering questions and participating in debate on especially medical issues. Last but not least we would like to thank Ole Winther at IMM for contributing with discussions and helping us to improve some of the methods used.

February 28 2006, Copenhagen.

Bjarni Bödvarsson

Martin Mørkebjerg

Contents

Abbreviations and nomenclature

Abbreviations

BBB	Blood brain barrier, page 6
BIC	Bayes information criterion, page 86
BP	Binding potential, page 6
CNS	Central nervous system, page 5
CSF	Cerebrospinal fluid, page 6
ICA	Independent component analysis, page 35
LOR	Line of response, page 2
MRI	Magnetic resonance imaging, page 8
NMF	Non-negative matrix factorization, page 29
PCA	Principal component analysis, page 25
PET	Positron emission tomography, page 2
ROI	Region of interest, page 6
TAC	Time activity curve, page 7

Nomenclature

Altanserin	A substance that binds to some neuroreceptors in the brain, page 3
Artery	Vessels where the blood flows from the heart, page 5
Grey matter	Tissue of the central nervous system, page 5
Randoms	Recorded events from unrelated coincidence, page 2
Reference region	Brain region where the tracer does not bind, page 6
Sinogram	Map of the projected lines of response, page 2
Vascular	The system of blood vessels, page 5
Vein	Vessels where the blood returns to the heart, page 5
White matter	Large nerve fibres in the brain, page 5

Chapter 1

Introduction

When studying some diseases or conditions the positron emission tomography (PET) imaging technique is a useful tool. The crucial advantage of PET compared with other imaging modalities is that it can detect metabolic changes before structural changes in tissues.

PET scans are commonly used in connection with the following diseases:

- Cancer
- Coronary Artery Disease
- Parkinson's Disease
- Epilepsy
- Huntington's Disease
- Alzheimer's Disease and other dementias

This thesis is limited to human brain PET scans. In order to make a scan of a human brain, the human subject must be injected with a radioactive tracer in a vein in the arm. From the vein the tracer is distributed via the blood to the rest of body including the brain. In order to use the information obtained scanning the brain optimally, the input to the brain via the artery has to be known. This is often done by continually taking blood samples during the scan, but this is associated with discomfort and is technically very demanding. In this thesis different methods for estimating the input to the brain to avoid taking continually blood samples are tested.

Because PET scans are very expensive and technically demanding the studies often suffer from the fact of having few subjects conversely the human anatomy and physiology vary a lot between subjects. This makes it hard to generalize the obtained results.

1.1 PET scanning

Positron emission tomography (PET) is a nuclear imaging technique and the basic principle is simple. As described by [23] the image is constructed by injecting a short-lived radioactive tracer isotope into a living subject. This tracer decays by emitting a positron which annihilates with an electron within a short distance and produces a pair of annihilation photons moving in opposite directions. These emission photons pairs are observed by a ring of scintillator crystals in the scanner by observing the produced light created when the photons energy is absorbed by the crystals. When two events are detected simultaneously by different crystals (within a few nanoseconds) they are recorded as coincidence events, hence the annihilation occurred somewhere on the line between the two crystals. These lines of responses (LOR) can then be used to map the density of the isotope in the subject. Some of the recorded events are coming from unrelated coincidences. This can happen if a photon is absorbed or missing the field of view. These erroneously events are called randoms and their rate can be measured. The procedure for PET scanning is illustrated in figure 1.1.

From the coincidence events information is obtained on how many annihilations occurred on any LOR between two crystals. This information is stored in raw data sinograms, that are constructed by projecting the number of counts (annihilations) for a given r and θ as illustrated in figure 1.2. Each row in the sinogram are all the parallel projections, Projection(r, θ), for a given θ. From these projections stored in the sinograms the underlying distribution can be calculated. Several algorithms have been designed to solve this problem. Here the so called *filtered back-projection* algorithm is applied because it is fast and

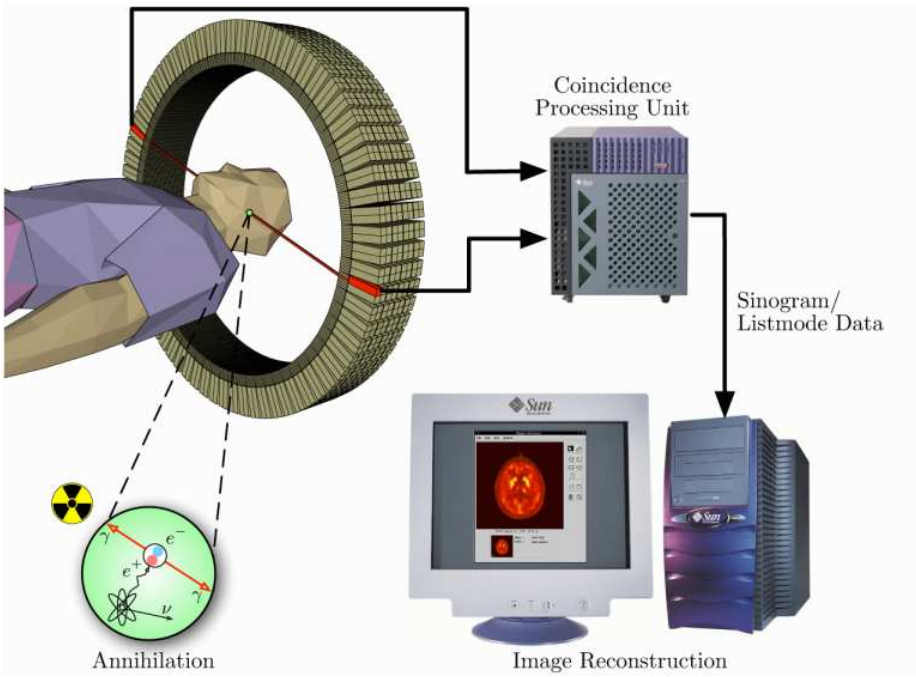


Figure 1.1: PET scan setup.[24]

robust. However, it suffers from artifacts around discontinuities of the tracer distribution. The result of the reconstructed PET data is an image with transaxial slices that represents the concentration for the investigated tracer.

The great benefit of PET is that simple biological molecules can be labeled with positron-emitting isotopes, and the distribution of these molecules can then be imaged. Labeling involves making molecules that have an unstable isotope of an element replacing the usual stable isotope of that same element, so the molecules behave in the same way chemically. The location of these labeled molecules in the body can be identified because a small fraction of them decay each second. The four commonly used isotopes are listed in table 1.1.

Fluorine-18, [^{18}F] is the most commonly used PET isotope even though it is not present naturally in the body at all. But fluorine can easily be incorporated into certain important biological molecules, such as glucose, without changing the way that the molecule behaves in the body. In this thesis the studied biological molecule is altanserin. Altanserin is a substance that binds to certain neuroreceptors in the brain.

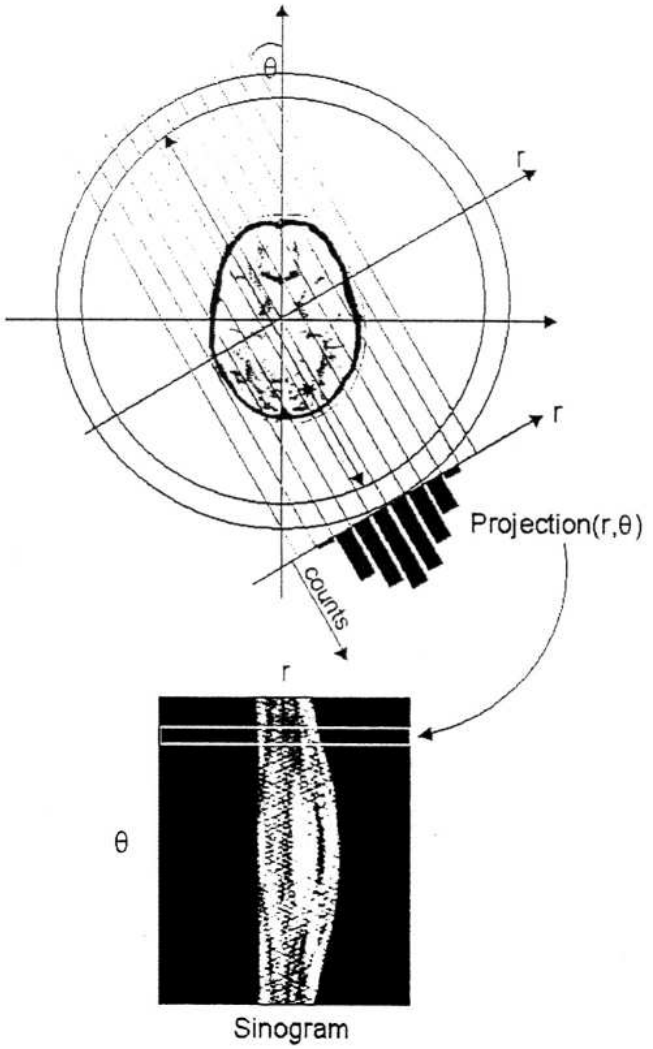


Figure 1.2: Geometry for obtaining PET raw data [23].

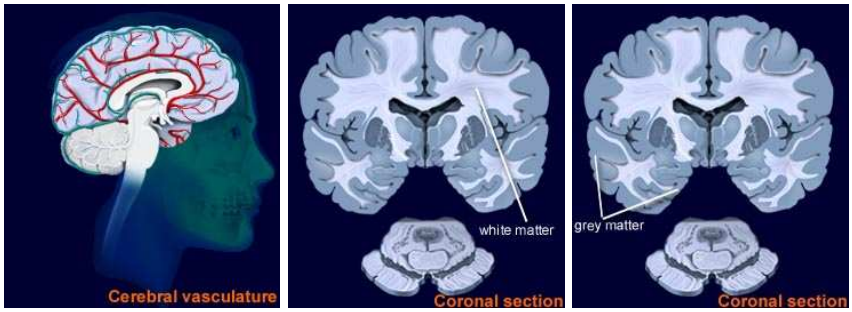
1.2 The brain anatomy

In this section a short introduction to brain anatomy is given to describe some of the concepts used in this thesis. As described in [10] the brain, also called

Name	Half life	Tracer example
Carbon-11 [^{11}C]	20 minutes	Methionine
Nitrogen-13 [^{13}N]	10 minutes	Ammonia
Oxygen-15 [^{15}O]	2 minutes	Water
Fluorine-18 [^{18}F]	110 minutes	Altanserin

Table 1.1: Commonly used isotopes.

cerebral, consists of many different materials and is often divided into regions which have different characteristics. The main focus in this thesis is the vascular part of the brain. The vascular part is the body's system of blood vessels that supplies the body, or a body part, with sufficient blood to allow normal body or organ function. The vascular system of the brain is seen on figure 1.3. It is divided into arteries and veins. The arteries are the vessels where the blood flows from the heart to the brain and the veins are the vessels where the blood returns. The vascular part of the brain covers approximately 5% of the total brain volume, and the main part of this is vein. The biggest vein, called sagittal sinus, is seen in figure 1.3, it starts at the forehead and runs all the way on the top of the head near the skull to the back of the head. Here it divides into two veins that run in each side of the brain. Most of the brains arteries are thin and the thickest are found in the center of the lower part of the brain.



(a) Vascular. Blue is vein, red is arterie (b) White matter (c) Grey matter

Figure 1.3: Some anatomy of the brain[4].

Apart from the vascular system the brain consists of some other materials where the two largest are white matter and grey matter. These are actual general terms referring to large nerve fibres in the brain and to tissue of the central nervous system (CNS). respectively. Grey matter is the areas where the actual "processing" is done whereas the white matter provides the communication

between different grey matter areas and between the grey matter and the rest of the body. White matter and grey matter are seen in figure 1.3. Approximately 60% of the total brain volume is white matter and approximately 40% is grey matter. Moreover, the brain consists of some cerebrospinal fluid (CSF). CSF is a fluid within the brain that cushions the brain inside the skull. CSF normally consists of glucose, salts, enzymes and a few white blood cells.

1.3 Kinetic modelling

Kinetic modelling of the data from the PET study is often involved when quantifying the brains uptake and binding of the tracer isotope. Tracer kinetic models are used to estimate the bindings or binding potentials (BP) which reflect the densities of neurotransporters or receptors in a brain region of interest (ROI). BP is influenced by a number of factors eg. the blood flow. Therefore, a compartmental system which includes the blood is used as model. As described by [8] two systems are set up, a one-tissue compartment configuration and a two-tissue compartment configuration. The one-tissue compartment model is used in the brain regions that do not have high-affinity receptors for the tracer isotope and models the blood brain barrier (BBB). These regions are called reference regions. The two-tissue compartment model is used in the brain regions that contain high-affinity receptors. The extra compartment models the binding of the substance by the neuroreceptors. These models can be seen in figure 1.4.

The tracer isotope is given intravenously in the arm and is then delivered by the artery by passing through the heart to the brain. From the artery the tracer passes through the blood-brain barrier (BBB) into the first compartment. Anatomically this compartment consists of several regions but it is approximated as a single compartment. The second compartment is the region of specific binding that contains high-affinity receptors. The one-tissue compartment system is as mentioned used to model the brain reference regions. The reference regions have no specific binding and therefore no second tissue compartment. There is also a nonspecific-binding compartment that exchanges with the first tissue compartment in both systems, but because of rapid equilibrium between these compartments they are treated as a single compartment.

In figure 1.4 the two compartment systems are shown. K_1, K_2, K_3 and K_4 are first order rate constants describing the potential directional exchanges between compartments. Physiologically, C_f represents [^{18}F] fluoride ion in an extravascular space unbound to tissue, while C_b refers to the ion bound to tissue.

The different models based on this compartment system for estimating the BP

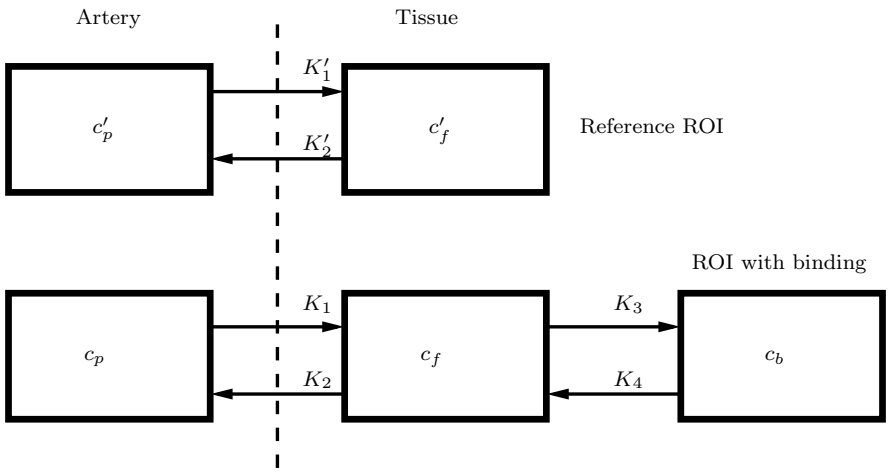


Figure 1.4: The one- and two-tissue compartment models, the broken line represents the blood brain barrier (BBB).

can be classified on the basis on whether arterial (invasive) or reference (noninvasive) information is needed. In general invasive models have fewer assumptions than noninvasive models but the disadvantages of invasive models are among other that arterial sampling is associated with discomfort, and is technically demanding.

With the noninvasive models such as reference tissue models there can be a possible loss in accuracy and increased bias. It depends on the assumption that nonspecific binding differences in the subjects are negligible.

Another possibility is to obtain the input function from the PET data itself by extracting a brain artery time activity curve (TAC). Apart from the obvious benefit of no artery cannulation, there are several other advantages, including the removal of issues regarding cross-calibration (between the PET data and that obtained from the blood samples), delay, dispersion and noise issues. To extract a relevant vascular TAC from the dynamic PET data, different methods are examined. We will look into methods for estimating vascular time activity curves from PET data. If possible it will reduce patient discomfort and improve accuracy.

1.4 Literature review

The main objective is to find the artery TAC using dynamic PET images instead of arterial sampling. The subject of non-invasive estimation has been tried before by using either MRI or PET images.

The PET data in this thesis has been used by [15] to estimate the artery TAC by the use of the K-means clustering algorithm. The results are used to compare different methods. In [11] and [9] the non-negative matrix factorization (NMF) method is used to extract the artery TAC from PET images. Here they use regions of interest (ROI) and only the initial 90 seconds of the scan. This method requires the user to define the region and the resulting TAC is only 90 seconds, hence leaving out information from remaining time of the scan.

An independent component analysis (ICA) method is implemented and used on PET images by [18] and [17]. The implemented method has no non-negative constraint. The method is tested on the PET data set used in this thesis giving estimates on artery TACs. Some of the curves have negative values making them unnatural and impossible to scale to the correct activity level. The results can be seen in Appendix E. Another use of ICA with no non-negative constraint on PET images is found in [2]. Here only a small part of the image and the initial 2 minutes are used. As with the NMF making it easier to estimate the peak but ignoring the end level. Choosing a small part of the image increases the uncertainty. The ICA method is also used on MRI images by [5].

The ICA methods documented have the weaknesses of no non-negative constraint. It is also important to note that none of the papers using deconvolution methods clarify the issues on the scaling of the estimated TACs.

1.5 Problem definition

The aim of this thesis is to estimate the artery TAC using clustering and deconvolution methods on PET data. This is needed so that arterial sampling during the scan can be omitted. Reasons for avoiding this sampling are among other the discomfort for the patient, risks for the patient, and the extra work involved for the personnel working with the sampling.

The analysis is carried out on a simulated data set and a data set consisting of 5 dynamic PET scans performed on healthy volunteers. Emphasis is put on extracting the vascular TACs with a minimum of human interaction, with the

aim of suggesting a method to estimate the vascular TAC for this kind of data set, using the [^{18}F] Altanserin tracer.

No region of interest (ROI) is defined on beforehand since human interaction is intended to be minimized, and all frames are used to estimated the entire TAC, and not just the first part. Since the TACs can not be negative, emphasis is put on non-negative constraints in the estimated TACs.

1.5.1 Approach

Five different methods are applied to the problem of extracting TACs, two clustering methods and three deconvolution methods. The five methods are K-means, Fussy C-means, principal component analysis (PCA), non-negative matrix factorization (NMF), and mean field independent component analysis (ICA). Each model is applied to simulated data and PET scans and the results are illustrated and analyzed. All results from PET scans are validated using the sampled artery TAC. The results are then compared and discussed, finally the most appropriate method is found.

1.5.2 Limitations

The optimization of each method is a study by itself, and strength is not put into optimization of implementations. Therefore, various toolboxes and algorithms, already implemented are used for each method. All implementations are executed in Matlab.

The effect on the parameter estimation in the metabolite analysis, by using the estimated vascular TAC is not tested. This would be an interesting subject to look into, but it is not done due to limitations in the scope of the thesis.

No ROI is used when applying methods to the PET data. This can be seen as both a strength and a weakness. The strength being that nothing is removed from a large data set thereby giving more significant results. The weakness is that it might be easier using only a part of the brain to perform the analysis. By choosing regions with known arterial or venous components, the result could be improved since a smaller number of insignificant components need to be estimated by the methods.

1.6 Thesis overview

The thesis consists of four main parts besides the introductory part:

- **Theory**, Chapters 3 to 6. In these Chapters the theory of the different methods used in the thesis are explained.
- **Results**, Chapters 7 to 11. All results are first tried on simulated data, and then on real PET data.
- **Discussion and conclusion**, Chapters 12 and 13. The results from the preceding chapters are discussed and then finally concluded upon.
- **Appendix**, In this part results can be seen for the PET data. Each chapter shows results from one method.

Chapter 2

Data

Two types of data are used in this thesis. A simple simulated PET data set and a real PET data set, provided by Neurobiology Research Unit (NRU), at Copenhagen University Hospital.

2.1 PET data

The main study was carried out on five subjects, two females and three males with an age ranging from 22 to 74 years. The data set is the same as used in [15]. The five subjects are named Pilot 2, Pilot 3, Pilot 4, Pilot 5 and Pilot 6. All subjects were healthy, meaning that they had a neurological examination and a cerebral MRI scan performed with normal results. They did not suffer from current physical illness or use medication with central nervous system (CNS) effects and had no history of alcohol or drug abuse. Furthermore their first and second-degree relatives had no history of psychiatric or neurological illness.

The scanner used was an 18-ring GE-Advance scanner, producing an image containing 35 slices with a distance of 4.25 mm in between. The approximate in-plane resolution has a lower bound of 5 mm and between-plane resolution of 6 mm.

800 MBq of [^{18}F]-altanserin was produced for each study. Four subjects received an intravenous bolus injection of [^{18}F] Altanserin diluted in ethanol over a 30 seconds period. The first subject (Pilot 2) was given a more rapid injection (less than 5 s). After the injection the dynamic PET data was recorded for 120 min (7200 sec). The scanning of the four subjects consists of 40 sequential scan frames in the following order: 6 x 5 s; 3 x 10 s; 3 x 20 s; 5 x 30 s; 5 x 1 min; 5 x 2 min; 4 x 5 min; 8 x 10 min. Pilot 2's scanning consists of 39 sequential scan frames in the order: 6 x 10 s; 3 x 20 s; 6 x 30 s; 5 x 1 min; 5 x 2 min; 8 x 5 min; 6 x 10 min.

While scanning a continuous measurement of the artery was carried out. The measuring was made by an automatic blood sampler, sampling from a radial artery. Blood samples were also taken in veins. Images were reconstructed as a sequence of 128 x 128 x 35 voxel matrices where each voxel measures 2.0 x 2.0 x 4.25 mm.

For each subject there exists 4 files: an .img-, .hdr-, .sif- and .tim-file. The .img-file consist of the dynamic PET image with the 128 x 128 x 35 voxel and 39 or 40 time frames for each voxel. The value in each voxel is simply the radioactivity in the time frame. The .hdr-file (header-file) is an information file which has the structure seen in table 2.1.

Variable	**Description**
name:	Name of image file
path:	Path of file
pre:	Precision for voxels in bit 1 - 1 single bit 8 - 8 bit voxels 16 - 16 bit integer 32 - 32 bit floats 32i - 32 bit complex numbers (64 bit pr. voxel) 64 - 64 bit floats
dim:	x,y,z,t, no. of pixels in each direction, [4x1 double]
siz:	Voxel size in mm, [3x1 double]
lim:	max and min limits for pixel values, [2x1 double]
scale:	Scaling of pixel values, [1 double]
offset:	Offset in pixel values, [1 double]
origin:	Origin for AC-PC plane (SPM notation), [3x1 double]
descr:	Description
endian:	Number format used 'ieee-be' or 'ieee-le'

Table 2.1: The information included in the header file.

The absolute value of each voxel can be calculated from the information contained in the header file. This is done the following way:

$$\text{abs\_pix\_val} = (\text{pix\_val} - \text{offset}) \cdot \text{scale} \tag{2.1}$$

All frames have been decay corrected to scanner start time. Slices from a typical PET scan can be seen in figure 2.1.

The .sif-file contains information about start and end time from the scanner start in seconds of each time frame. It also contains information on the total number of counts and the number of estimated randoms in each time frame. The .tim-file contains information calculated from the .sif-file, namely the frame mid-time again from the scanner start in seconds and the duration of each frame in seconds.

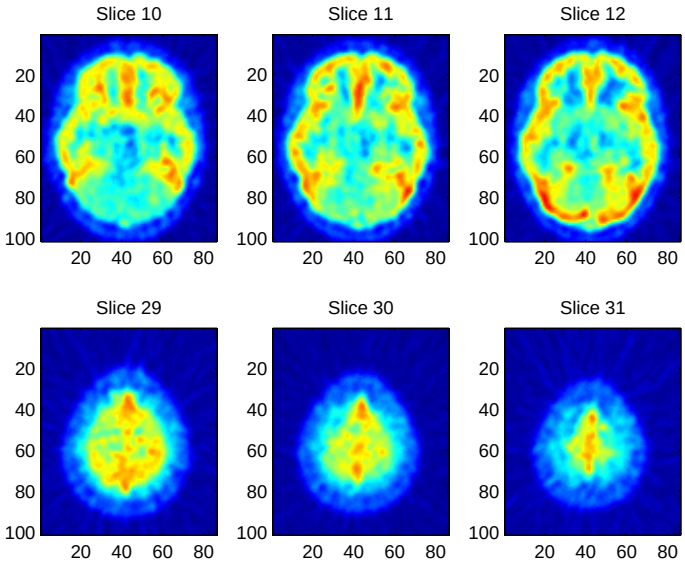


Figure 2.1: 6 mean slices from Pilot 6.

Before the transaxial PET slices are used by the different methods a threshold is made to remove the background from the images. After removing the background the images are reshaped into a matrix so each time depending voxel is a row as illustrated in the following matrix:

$$\mathbf{X} = \begin{pmatrix} x_1(1) & x_1(2) & \cdots & x_1(d) \\ x_2(1) & x_2(2) & \cdots & x_2(d) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(1) & x_n(2) & \cdots & x_n(d) \end{pmatrix}$$

where n is the number of voxels and d is the dimension or the number of frames. The typical size of the matrix is $90000 \times 40 = 3600000$. When using the non-negative constrained methods the negative values in the data are set to zero.

2.2 Sampled time activity curves

For each PET scan arterial and venous blood samples are taken. The vein samples are taken manually at different time points. The arterial samples are taken automatically and validated using manual samples. The manual arterial and venous samples are used for metabolite analysis. The manual arterial sample is also used for validation of the automatic sampler. An example of an arterial time activity curve (TAC) can be seen in figure 2.2. The autosampled curve has some periods where the activity is zero this is because the auto sampler is turned off when taking the manual samples.

The arterial TACs are manually resampled to the fit the 39 or 40 frame lengths for the 5 PET scans. This is done to be able to do comparisons between estimated TACs and the sampled TACs by using correlation. The resampled curve is also seen in figure 2.2.

2.3 Region of interest curves

The PET scans can be segmented into different regions which have different binding properties. For these regions the mean TACs are retrieved from the PET data. These TACs are used as references in this thesis when validating the clustering by the different methods. The comparison of the region of interest curves and the estimated curves from the methods, can give an indication of whether or not the method is able to segment the PET data into real regions. An example of ROI curves can be seen in figure 2.3. Since these curves are mean curves from the PET data and from the same subject, they can be compared directly to the estimated curves.

The ROI curves behave quite differently. There are regions that have high-

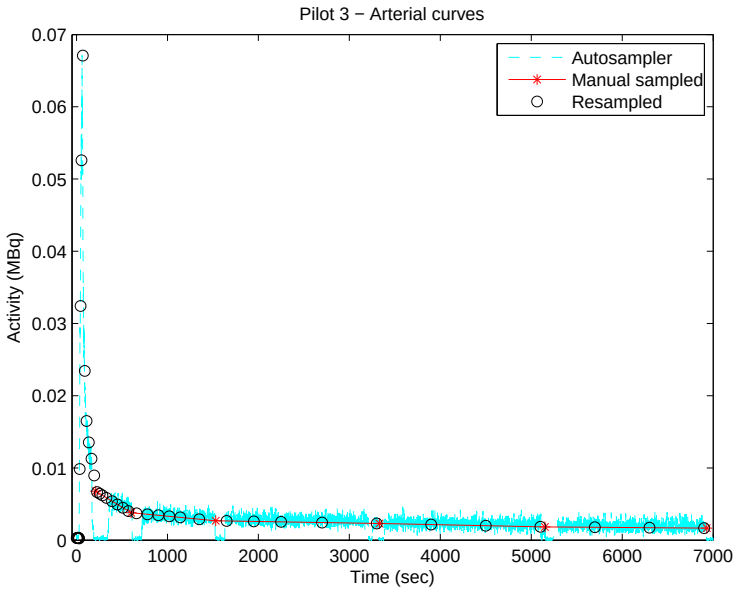


Figure 2.2: Automatically and manually sampled arterial TAC and the resampled curve, for Pilot 3.

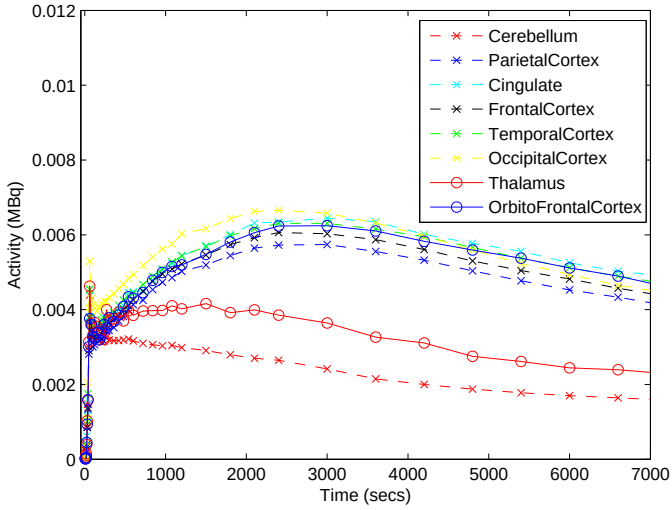


Figure 2.3: TACs from regions of interest for Pilot 6.

affinity receptors for the tracer isotope e.g. frontal cortex, and regions that do not have high-affinity receptors for the tracer isotope e.g. cerebellum.

2.4 Simulated data

A simulated data set is made to compare the methods and to evaluate the characteristics of the results from the methods. The data set has 4 signals shown in figure 2.4. From these signals a data set with 100000 observations is made, 10000 for each of the 10 time points. This is done by creating a mixing matrix $\mathbf{M}$, as seen in (2.2).

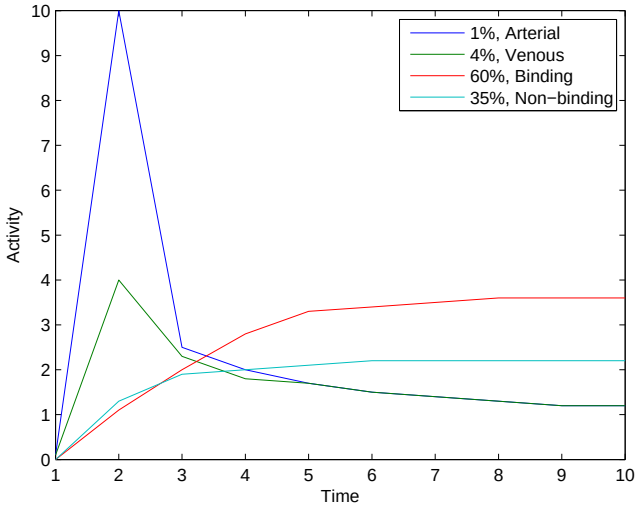


Figure 2.4: Simulated time activity curves.

$$\mathbf{X} = \mathbf{MS} + \epsilon \tag{2.2}$$

where $\mathbf{M}$ is the 10000×4 mixing matrix, $\mathbf{S}$ is the 4×10 time signal matrix. The resulting $\mathbf{X}, 10000 \times 10$, matrix is the simulated data set. $\mathbf{M}$ is made so that the parts of the different time signals are not equally represented in $\mathbf{X}$. How much of the different signals there is in $\mathbf{X}$ can be seen in figure 2.4. Each row in $\mathbf{M}$ sums to 1, so for each observation in $\mathbf{X}$ 4 coefficients in $\mathbf{M}$ linearly combine the sources of $\mathbf{S}$. The coefficients in $\mathbf{M}$ are independent. Noise is added to the

data by ϵ, which are independent and identically distributed random variables, with mean 0 and standard variation 0.1 times the standard variation of $\mathbf{MS}$.

The four signals are created so that they approximate the PET data. The simulated data has a signal with a high peak that is less represented in the data like the arteries, a signal that has a lower peak than the artery but with the same end-level as the veins, a signal that has high end level and is well represented in the data like the binding regions, and finally a signal that looks like the regions with non-binding. Because the simulated artery signal is only representing 1% of the data the highest percentage in a observation is only 33%. The highest percentage for the simulated vein signal is only 50%, and the last two signals have observations with 100% of the two signals.

The peak from the artery like signal is not fully there in the resulting data $\mathbf{X}$. This is because of the linear combination of the coefficients and signals, and the fact that the artery signal is only 1% of the data. The maximum value in $\mathbf{X}$ is 5.3 and not 10 as in the original signal. Moreover because of the fact the the vein and artery signals are not represented 100% in any observation the lowest end level in $\mathbf{X}$ is 1.6 and not 1.2 as in the two original signals.

The different methods used can then be applied to the simulated data to see if they are able extract the original sources in $\mathbf{S}$ even though the data does not contain the peak from the artery signal and the end level of the artery and vein signals.

Part I

Theory

Chapter 3

Clustering

In this thesis two different clustering algorithms are used. The two algorithms are the K-means and the Fuzzy C-means clustering. The main difference between them is that the K-means algorithm cluster the samples so each sample belongs to only one cluster, and the Fuzzy C-means algorithm cluster the samples so each sample belongs to all clusters but with different share in each cluster. The theory behind the two algorithms is described in this chapter.

3.1 K-means clustering

The data is represented by the n×d matrix $\mathbf{X}$.

$$\mathbf{X} = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \tag{3.1}$$

with n observation and d variables.

The K-means clustering [16] algorithm divides the data into k clusters. The aim is to make each cluster as homogenous as possible and at the same time making the differences between clusters as great as possible. Each observation is classified as one and only one type, the data point is therefore thought of as one type and not a mixture of different components. The algorithm is an iterative process and the cost function or the error measure is normally the sum of all the variances, or the within variances. This can be described as:

$$\text{Error}_{\text{K}-\text{means}} = \sum_{\text{i}=1}^{\text{n}} \sum_{\text{j}=1}^{\text{k}} \text{u}_{\text{ij}} \|\text{x}_{\text{i}}(\text{t}) - \text{c}_{\text{j}}\|^2 \tag{3.2}$$

where u_{ij} is 1 if observation i belongs to cluster j, else it is 0. c_j is the center of the j'th cluster, $x_i(t)$ is the i'th observation, n is the total number of observations in $\mathbf{X}$, and k is number of clusters. $\|x_i(t) - c_j\|$ is the distance between $x_i(t)$ and c_j. The error measure indicates the overall spread of data points about their centers. The within variance measure should be as small as possible to obtain a representative clustering.

The first step in the algorithm is selecting k points, these can be selected by supervision or simply by choosing k random observations from the data-set. The initial points serve as cluster centers in the first iteration of the algorithm. Each observation is set to the nearest cluster center by a chosen metric, e.g. Euclidian distance. Now the cluster centers are updated by setting the center to the mean of all its cluster members. The iteration now starts over and measures the distance from all observations to all cluster centers and so on. This process continues until the cluster centers converge.

If the distance measure is Euclidean, then K-means converges monotonically to a local minimum of within-class squared distortion. Hence, the algorithm is bound to converge to a solution but the solution might not be the optimal. Therefore a number of runs with different starting points can be performed to secure that the cost function does not end up in a local minima.

The algorithm can be described with the following steps:

- Initialize k clusters $i = 1..k$ with centres c_i. The cluster centre must be represented in the same space as the data.
- Assign each data point $x_{ij}(t)$ to the cluster with the closest centre c_i according to the distance metric used (i.e. Euclidian distance).

- Recalculate each cluster centre c_i so that it is now the average of all its members.
- Repeat steps 2 and 3 until the centroids stabilizes. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

As seen in the algorithm the number of clusters must be defined in advance. The correct number of clusters is not really defined, it depends on the data, the a priori information and the final application of the clustering result. Choosing the right number of clusters for the given data-set is important, because using the incorrect number of clusters can lead to meaningless results. One approach to validate the number of clusters is to plot the within variance. The within variance will of course decrease with increasing number of clusters. But hopefully there is a significant change in the decrease of the within variance when using more and more clusters. The point where this change occur is chosen as the optimal number of clusters.

The K-means algorithm is implemented as a standard function in Matlab (version 7.1), and this K-means version is used in this thesis.

3.2 Fussy C-means clustering

First developed by [6] and expanded by [3]. A model that can fit data points that have partial volume effects, could be the fussy C-means method. This method allows each data point to be described by several clusters, and not just one as in the K-means algorithm.

The fussy C-means algorithm clustering algorithm does not cluster the data as hard as the K-means. In fussy C-means each data point can belong to several cluster centers. The objective function to be minimized in this algorithm is:

$$\mathrm{Error}_{\mathrm{C-means}} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \sum_{\mathrm{j=1}}^{\mathrm{C}} \mathrm{u}_{\mathrm{ij}}^{\mathrm{m}} \|\mathrm{x_i(t)} - \mathrm{c_j}\|^2, \mathrm{m} = 1 < \mathrm{inf} \tag{3.3}$$

where m is the fuzziness parameter, u_{ij} describes how much the point i belongs to cluster j. The distance measure $\|x_i(t)-c_j\|^2$, could be the Euclidian distance.

The fuzziness parameter m describes how much one data point belongs to one cluster or more clusters. If m goes to 1 u_{ij} will be sparse for all data points i and the higher m is u_{ij} will become the same across j. This can be seen from the update of u_{ij} in (3.4).

The updated u_{ij} and c_j are calculated as:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i(t) - c_j\|}{\|x_i(t) - c_k\|} \right)^{\frac{2}{m-1}}} \tag{3.4}$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^m \cdot x_i(t)}{\sum_{i=1}^{n} u_{ij}^m} \tag{3.5}$$

Otherwise the algorithm is similar to the K-means. First the u_{ij} matrix is initialized and then the cluster centers c_j are calculated, and the u_{ij} matrix is updated. This is repeated a certain number of times or until the change in u_{ij} is small enough.

The fussy C-means method is included as a function in Matlab (version 7.1), and this implementation is used.

Chapter 4

Principal component analysis

Principal component analysis (PCA) described in [1], can be used to reduce the dimensionality of a large data set, such as a PET scan. The PCA can also be used as a tool to analyze the most important components in the data, found as the principal components that explain the most of the total variation. In [14] the components of facial images are extracted and the differences in components found using PCA and non-negative matrix factorization are shown.

A PET data set is very large and to extract the most important components a PCA is performed on the data. The PCA is commonly used for dimension reduction of large data set such as images, and the dimension can often be reduced greatly without loss of significant information if there is a high redundancy, which is the case with PET data. No specific assumptions are made on the probability density of the data set. Only first and second order statistics have to be calculated to perform the PCA.

The data is represented by the n×d matrix $\mathbf{X}$.

$$\mathbf{X} = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}, \hat{\mathbf{X}} = \mathbf{X} - E\{\mathbf{X}\} \tag{4.1}$$

with n observations and d variables. The data is centered in each frame and represented as $\hat{\mathbf{X}}$.

The principal component is the linear combination:

$$\mathbf{s}_1 = \sum_{i=1}^{d} w_1^T \hat{\mathbf{X}}_{.,d} = \mathbf{w}_1^T \hat{\mathbf{X}} \tag{4.2}$$

where $\mathbf{w}_1$ is an n-dimensional weight vector. $\mathbf{s}_1$ is the first principal component if the variance of the component is as large as possible. The first principal component is therefore found by using the $\mathbf{w}_1$ that maximizes the variation of $\mathbf{s}_1$. The criterion to maximize is:

$$E\{\mathbf{s}_1^2\} = E\{(\mathbf{w_1^T}\hat{\mathbf{X}})^2\} = \mathbf{w}_1^T E\{(\hat{\mathbf{X}}^T\hat{\mathbf{X}})^2\}\mathbf{w}_1 = \mathbf{w}_1^T \mathbf{C_X} \mathbf{w}_1 \tag{4.3}$$

under the constraint that the norm is equal to 1, $\|\mathbf{w}_1\| = 1$, where the norm is the Euclidian:

$$\|\mathbf{w}_1\| = (\mathbf{w}_1^T \mathbf{w}_1)^{1/2} = (\sum_{i=1}^{n} \mathbf{w}_{i1}^2)^{1/2} \tag{4.4}$$

The d×d covariance matrix $\mathbf{C_X}$ is estimated as $E\{\hat{\mathbf{X}}^T\hat{\mathbf{X}}\}$.

The solution to (4.3) is given by the unit-length eigenvectors of the covariance matrix $\mathbf{C_X}$.

$$\mathbf{w}_i = \mathbf{e}_i \tag{4.5}$$

where i=1,2...d. The eigenvectors, $\mathbf{e}_i$, are sorted in the order of decreasing eigenvalues, λ_i, which gives $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$

Therefore, by calculating the eigenvectors and eigenvalues of the estimated covariance matrix, the i'th principal components can be expressed as:

$$\mathbf{s}_i = \mathbf{e}_i^T \hat{\mathbf{X}} \tag{4.6}$$

A reduction of the data set can be achieved by using the j first principal components:

$$\hat{\mathbf{X}} \approx \sum_{i=1}^{j} \mathbf{e}_i \mathbf{s}_i \tag{4.7}$$

To decide how many principal components to include, the error of the principal component truncation can be evaluated. The error is defined as the sum of eigenvalues for the eigenvectors that are not included in the model.

$$\mathrm{error}_{\mathrm{pca}} = \sum_{\mathrm{m=j+1}}^{\mathrm{d}} \lambda_{\mathrm{m}} \tag{4.8}$$

Thereby if there is a lot of redundancy in the data set, the last λ's can be very small, and thereby explaining little variation in the data. These small λ's can therefore be discarded in the truncated PCA model.

Chapter 5

Non-negative matrix factorization

The non-negative matrix factorization method (NMF), proposed by [14], performs a deconvolution of some observed data matrix. The data is represented as a linear combination of two factors, **W** and **H**. The factors can be positively combined only, no subtraction is allowed. All elements in the data matrix and in the factors are positively constrained. This data structure is ideal for many observed signals in nature and in real life applications, since the data matrix can be a combination of different unknown signals that are mixed with each other in an unknown manner. The positive constraint is very useful when working with images, speech and many other applications where the original sources are sure to be added together in an observed signal. No statistical properties are assumed of the data.

The NMF performs a factorization of the data matrix, **V**, into two matrices with non-negative elements,**W** and **H**.

$$\mathbf{V} \approx \mathbf{WH} \tag{5.1}$$

where **V** has the dimension n×d, n samples with each of size d. **W** is n×K,

and $\mathbf{H}$ is K×d. If the problem is viewed as a blind separation problem, $\mathbf{V}$ is the observed signal, $\mathbf{W}$ is the mixing matrix and $\mathbf{H}$ contains the K original sources.

The number of sources to be estimated in the NMF method must be known in advance, since this parameter is given to the algorithm. The NMF does not give a qualitative measure of the factorization. The least squares measure will certainly decrease with the number of sources used. But the sources might not be "correct" in any sense, but just be noise or original sources split into non-existing signals.

Different cost functions can be used in the NMF algorithm. Using the squared Euclidian distance as shown in [13], the algorithm can be used as implemented in the Brede Toolbox, by [22].

$$\|\mathbf{V}, \mathbf{WH}\|^2 = \sum_{ij} (\mathbf{V}_{ij} - (\mathbf{WH})_{ij})^2 \tag{5.2}$$

The update rules given by [13] have been proved to make the cost function (5.2) converge to a local minimum. The advantage of this algorithm is that it is straightforward and easy to implement. The update rules are multiplicative factors applied to $\mathbf{W}$ and $\mathbf{H}$. The indices on the matrices indicate that the operations are element-wise.

$$\mathbf{H}_{sj} \leftarrow \mathbf{H}_{sj} \frac{(\mathbf{W}^T\mathbf{V})_{sj}}{(\mathbf{W}^T\mathbf{WH})_{sj}} \tag{5.3}$$

$$\mathbf{W}_{is} \leftarrow \mathbf{W}_{is} \frac{(\mathbf{VH}^T)_{is}}{(\mathbf{WHH}^T)_{is}} \tag{5.4}$$

The convergence criterions can either be the number of iterations or the change in Euclidian distance. The iterative improvement of the cost function from equation (5.2) can be evaluated as the algorithm is progressing, and the iterative process can be stopped if the slope of the cost function seems to flatten. Since this would be an indication of a minimum.

Two random matrices can be used to initialize the algorithm, or a prior can be put into $\mathbf{W}$ or $\mathbf{H}$.

The sources found by the NMF algorithm cannot be compared directly to the sampled blood curve, since the scale of the sources is unknown. This is because of the decomposition of the original matrix $\mathbf{V}$ into two matrices and it cannot be distinguished whether or not a factor is changed in $\mathbf{W}$ or $\mathbf{H}$. Additional information about the original sources has to be used if $\mathbf{W}$ and $\mathbf{H}$ need to be scaled correctly.

5.1 Weighted non-negative matrix factorization

The NMF method can be adjusted to take weights into account if a weighting of the data is known or estimated, when performing a deconvolution of the data.

To perform a weighted NMF the update rules have to be revised. The cost function is still based on the squared Euclidian distance:

$$\|\mathbf{V}, \mathbf{WH}\|^2 = \sum_{ij}((\mathbf{VQ})_{ij} - (\mathbf{WHQ})_{ij})^2 \tag{5.5}$$

where

$$\mathbf{Q} = \begin{pmatrix} w_1 & 0 & 0 & \vdots \\ 0 & w_2 & 0 & \vdots \\ 0 & 0 & \ddots & \vdots \\ \dots & \dots & \dots & w_d \end{pmatrix}$$

and w_j is the weight of the j'th column of the data matrix $\mathbf{V}$. This weighted cost function results in the following update rules:

$$\mathbf{H}_{sj} \leftarrow \mathbf{H}_{sj} \frac{(\mathbf{W}^T(\mathbf{VQ}))_{sj}}{(\mathbf{W}^T(\mathbf{WHQ}))_{sj}} \tag{5.6}$$

$$\mathbf{W}_{is} \leftarrow \mathbf{W}_{is} \frac{((\mathbf{VQ})\mathbf{H}^T)_{is}}{((\mathbf{WHQ})\mathbf{H}^T)_{is}} \tag{5.7}$$

The weighted update rules are implemented in Matlab as an expansion to the Brede Toolbox.

5.2 Scaling of W and H

As mentioned the sources are not scaled correctly, therefore a method is proposed to calculate a factor α that will rescale $\mathbf{W}$ and $\mathbf{H}$ to the original scale. The factor is elementwise multiplied with each row of $\mathbf{W}$, and the reciprocal transposed factor is elementwise multiplied with each column of $\mathbf{H}$. Thereby the cost function will not change with this rescaling.

$$\begin{aligned} \mathbf{H}'_{.i} &= \mathbf{H}_{.i} \cdot (\alpha^T)^{-1} && (5.8) \\ \mathbf{W}'_{j.} &= \mathbf{W}_{j.} \cdot \alpha && (5.9) \\ \mathbf{W}'\mathbf{H}' &= \mathbf{WH} && (5.10) \end{aligned}$$

where $\alpha = [\alpha_1 \ \ ... \ \ \alpha_K]^T$

Since $\mathbf{W}$ can be interpreted as the mixing matrix or as spatial images of the sources, a lemma can be defined for $\mathbf{W}$. In the mixing matrix each row must sum to 1. $\mathbf{W}$ is set to describe the mix of each source in each observation or pixel, and these mixing coefficients must sum to 1. This means that if $\mathbf{W}$ consists of images of the estimated sources, then the summed image must equal 1 in each pixel.

To return to the original scale of the signals in $\mathbf{H}$, the following lemma for $\mathbf{W}$ is defined.

$$\begin{aligned} \mathbf{W}_{j.}\alpha = 1 \Longleftrightarrow \sum_j (1 - \mathbf{W}_{j.}\alpha)^2 = 0 \Longleftrightarrow \\ \mathbf{W}_{j1} \cdot \alpha_1 + \mathbf{W}_{j2} \cdot \alpha_2 + ... + \mathbf{W}_{jK} \cdot \alpha_K = 1 \qquad (5.11) \end{aligned}$$

To acquire the correct scaling factor α, the following is derived from (5.11).

$$\frac{\delta}{\delta\alpha}\sum_j(1-\mathbf{W}_{j\cdot}\alpha)^2=0\Longleftrightarrow\sum_j((1-\mathbf{W}_{j\cdot}\alpha)\mathbf{W}_{j\cdot}^T)=0\Longleftrightarrow$$
$$\sum_j(\mathbf{W}_{j\cdot}^T-\mathbf{W}_{j\cdot}^T\mathbf{W}_{j\cdot}\alpha)=0\Longleftrightarrow\sum_j\mathbf{W}_{j\cdot}^T-\sum_j\mathbf{W}_{j\cdot}^T\mathbf{W}_{j\cdot}\alpha=0\Longleftrightarrow$$
$$\sum_j\mathbf{W}_{j\cdot}^T=\sum_j\mathbf{W}_{j\cdot}^T\mathbf{W}_{j\cdot}\alpha\Longleftrightarrow\sum_j\mathbf{W}_{j\cdot}^T=\alpha\sum_j\mathbf{W}_{j\cdot}^T\mathbf{W}_{j\cdot}\Longleftrightarrow$$
$$\alpha=(\sum_j\mathbf{W}_{j\cdot}^T\mathbf{W}_{j\cdot})^{-1}\sum_j\mathbf{W}_{j\cdot}^T\Longleftrightarrow$$
$$\alpha=(\mathbf{W}^T\mathbf{W})^{-1}\sum_j\mathbf{W}_{j\cdot}^T \tag{5.12}$$

5.3 Number of sources in NMF

The problem of deciding the number of sources to use in the deconvolution in the NMF is not straight forward, since no qualitative measures of the model is given. An analysis of the factorization can be done by inspection of the matrices $\mathbf{W}$ and $\mathbf{H}$. If values in the columns in the mixing matrix $\mathbf{W}$ are very small and/or the source matrix $\mathbf{H}$ has sources which do not seem natural or simply has very small values. This could be an indication of overfitting, yielding that too many sources are estimated in the NMF method.

Chapter 6

Independent Component Analysis

When looking for a better deconvolution model which hopefully describes the data better the choice naturally falls on independent component analysis (ICA). Compared with the NMF which simply factorizes the observed signals into two matrices the ICA also estimates the noise in the data. As described by [1] the definition of independent component analysis is

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_M(t) \end{pmatrix} + \mathbf{n} \tag{6.1}$$

where $x_i(t)$ is the time dependent observed signals, $\mathbf{A}$ is the unknown mixing matrix, $s_j(t)$ is the unknown time dependent sources, while $\mathbf{n}$ is a matrix of noise signals. N and M are the number of samples and the number of sources respectively.

ICA distinguishes itself from other methods by looking for components that are both statistically independent and nongaussian. In practice it is often impossi-

ble to find really independent components, but instead finding as independent components as possible. When applying the method to neuroimaging the data is analyzed for independent spatial patterns finding temporal signals that are as independent as possible.

6.1 Probabilistic ICA

There is a broad variety of ICA methods that have different approaches and different characteristics. For this application the non-negative constraint is very important. This constraint can be fulfilled with the probabilistic ICA namely the Mean Field ICA described and implemented by [25], [21], [20], [19] and [12]. Here the following formulation is used

$$\mathbf{X} = \mathbf{AS} + \mathbf{n} \tag{6.2}$$

where the noise $\mathbf{n}$ is assumed zero-mean gaussian with covariance $\mathbf{\Sigma}$, the observed signals are stacked into one matric $\mathbf{X}$ and the same goes for the unknown sources $\mathbf{S}$. The Likelihood $\mathcal{L}$ is then defined by $\mathcal{L} = p(\mathbf{X}|\mathbf{A}, \mathbf{S}, \mathbf{\Sigma})$. The source prior $p(\mathbf{S}|\phi)$ which in detail will be described in section 6.1.2 has some parameters which in shorthand are called ϕ. By defining the hyper-parameter ψ which is a shorthand of the models parameter $\psi = \{\mathbf{A}, \mathbf{\Sigma}, \phi\}$ the posterior probability is given by

$$p(\mathbf{S}|\mathbf{X}, \psi) = \frac{p(\mathbf{X}|\mathbf{A}, \mathbf{S}, \mathbf{\Sigma})p(\mathbf{S}|\phi)}{p(X|\psi)} \tag{6.3}$$

In this probabilistic approach version of ICA the unknown sources are integrated out, leaving the parameters ψ to be determined by maximizing the Likelihood $\mathcal{L}(\psi) = p(\mathbf{X}|\psi)$. Unfortunately, the Likelihood is too complicated in this approach and instead a lower bound $\mathcal{B}$ for the Likelihood is used as objective function. The lower bound is defined by

$$\mathcal{L}(\psi) = \ln p(\mathbf{X}|\psi) = ln \int q(\mathbf{S}|\phi) \frac{p(\mathbf{X}, \mathbf{S}|\psi)}{q(\mathbf{S}|\phi)} d\mathbf{S} \tag{6.4}$$

$$\geq \int q(\mathbf{S}|\phi) \ln \frac{p(\mathbf{X}, \mathbf{S}|\psi)}{q(\mathbf{S}|\phi)} d\mathbf{S} = \mathcal{B}(\psi|\phi) \tag{6.5}$$

The property of the bound holds for any choice of variational distribution $q(\mathbf{S}|\phi)$ and the misfit between $\mathcal{B}$ and $\mathcal{L}$ can be measured with Kullback-Leibler (KL) by measuring the divergence between them $\mathcal{L}(\psi) = \mathcal{B}(\psi|\phi) - KL(q,p)$, where $KL(q,p) \geq 0$ stand for divergence between the sources posterior and the variational distribution. Hence, if the bound is equal to the log likelihood KL equals zero.

For the optimization the derivatives of the lower bound are needed. This can easily be derived and the results are

$$\frac{\partial \mathcal{B}(\psi,\phi)}{\partial \mathbf{A}} = \mathbf{\Sigma}^{-1}(\mathbf{X}\langle\mathbf{S}\rangle_q^T - \mathbf{A}\langle\mathbf{S}\mathbf{S}^T\rangle_q) \tag{6.6}$$

$$\begin{aligned} & \frac{\partial \mathcal{B}(\psi,\phi)}{\partial \mathbf{\Sigma}} \\ = \; & \frac{N}{2}\mathbf{\Sigma}^{-1} - \frac{1}{2}\mathbf{\Sigma}^{-1}\langle(\mathbf{X}-\mathbf{A}\mathbf{S})(\mathbf{X}-\mathbf{A}\mathbf{S})^T\rangle_q\mathbf{\Sigma}^{-1} \\ = \; & \frac{N}{2}\mathbf{\Sigma}^{-1} - \frac{1}{2}\mathbf{\Sigma}^{-1}(\mathbf{X}\mathbf{X}^T - \mathbf{X}\langle\mathbf{S}\rangle_q^T\mathbf{A}^T - \mathbf{A}\langle\mathbf{S}\rangle_q\mathbf{X}^T + \mathbf{A}\langle\mathbf{S}\mathbf{S}^T\rangle_q\mathbf{A}^T)\mathbf{\Sigma}^{-1} \end{aligned} \tag{6.7}$$

$$\frac{\partial \mathcal{B}(\psi,\phi)}{\partial \phi_j} = \langle\frac{\partial \ln p(S|\phi)}{\partial \phi_j}\rangle_q = \frac{\partial \ln \delta_{[\mathbf{S}>0]}}{\partial \phi_j} + \frac{N}{\phi_j} - \sum_t \langle\mathbf{S}\rangle_{q_{jt}} \tag{6.8}$$

where $\langle\cdot\rangle$ denotes the source posterior average, that depends on $\mathbf{\Sigma}$. $\langle\mathbf{S}\rangle$ and $\langle\mathbf{S}\mathbf{S}^T\rangle$ can be seen as respectively the first and second moments. Equation (6.8) is calculated based on the specified source prior found in equation (6.12) in section 6.1.2.

6.1.1 Optimization of Parameters

The lower bound of the log Likelihood has to be optimized. This is done with the Expectation-Maximization (EM) Algorithm that in shorthand can be described by the following:

E-step: Maximize $\mathcal{B}(\psi|\phi)$ with respect to ϕ keeping ψ fixed.

M-step: Maximize $\mathcal{B}(\psi|\phi)$ with respect to ψ keeping ϕ fixed.

In the M-step the lower bound is maximized with respect to the model hyperparameters ψ by setting the derivatives (6.6), (6.7) and (6.8) equal to zero. The EM update obtained for the mixing matrix and the noise covariance is then

$$\mathbf{A} = \mathbf{X}\langle\mathbf{S}\rangle_q^T \langle\mathbf{S}\mathbf{S}\rangle_q^{-1} \tag{6.9}$$

$$\mathbf{\Sigma} = \frac{1}{N}(\mathbf{X}\mathbf{X}^T - \mathbf{X}\langle\mathbf{S}\rangle_q^T \mathbf{A}^T - \mathbf{A}\langle\mathbf{S}\rangle_q \mathbf{X}^T + \mathbf{A}\langle\mathbf{S}\mathbf{S}^T\rangle_q \mathbf{A}^T) \tag{6.10}$$

The E- and M-step is repeated back and forth updating the values of the parameters and continuously increasing the lower bound of the Likelihood. Unfortunately, this approach is very slow especially in the case where the system is very over complete hence the number of samples is much bigger than the number of sources which exactly is the case with images. Therefore an overrelaxed adaptive EM approach is used to speed up the algorithm. The modification is made to increase the step size in the M-step by taking larger and larger steps in the direction proposed by EM.

$$\psi^{n+1} = \psi^n + \gamma(\psi_{EM} - \psi^n) \tag{6.11}$$

where $\gamma \geq 1$. If $\gamma = 1$ the original EM algorithm is obtained. For each time a successful step is taken γ is increased, but if the Likelihood bound is decreased in a step γ is set to one again and the step is undone. This speeds up the process significantly, if multiple useful steps are found after each other.

6.1.2 Source prior and constrained mixing matrix

When choosing the source prior one must look for a distribution $p(\mathbf{S})$ which is sensible from an analytical point of view and with respect to the particular application at hand. For the application of images the non-negative constraint is naturally and the obvious choice falls on the exponential distribution. The density is

$$p(\mathbf{S}) \;=\; \left(\prod_i \phi_i^N\right) \exp\left[-\sum_{it} \phi_i S_{it}\right] \Theta(\mathbf{S}) \tag{6.12}$$

where $\Theta(\mathbf{S}) = 1$ for $S > 0$ and zero otherwise. The density is normalizable only if the parameters ϕ_i are all positive.

The mixing matrix is constrained to be positive which is natural for these images. This is done by a reparameterization of the mixing matrix $\mathbf{A}$. The positive constraint is obtained by using the exponential function $(\mathbf{A})_{ij} = \exp^{(\alpha)_{ij}}$ making the parameter $\mathbf{A}$ an underlying parameter α in the hyperparameter ψ. With this underlying parameter the derivative of lower bound is

$$\frac{d\mathcal{B}}{d[\alpha]_{ij}} = \frac{\partial \mathbf{A}}{\partial [\alpha]_{ij}} \frac{\partial \mathcal{B}}{\partial \mathbf{A}} = [\mathbf{A}]_{ij} \left[\frac{\partial \mathcal{B}}{\partial \mathbf{A}} \right]_{ij} \tag{6.13}$$

By setting this equal to zero the solution is given by a simple iterative formula

$$[\mathbf{A}]_{ij} := [\mathbf{A}]_{ij} \frac{[\mathbf{\Sigma}^{-1}\mathbf{X}\langle \mathbf{S} \rangle_q^T]_{ij}}{[\mathbf{\Sigma}^{-1}\mathbf{A}\langle \mathbf{S}\mathbf{S}^T \rangle_q]_{ij}} \tag{6.14}$$

6.2 Scaling of S and A

As with NMF $\mathbf{A}$ can be interpreted as the mixing matrix or as spatial images of the sources and a lemma can be defined for $\mathbf{A}$. In the mixing matrix each row must sum to 1. $\mathbf{A}$ is set to describe how the sources are mixed in each observation or voxel. This means that if $\mathbf{A}$ are several images of the original sources, then the summed image must equal 1 in each pixel. When finding the independent component of the $\mathbf{X}$ matrix the scale of $\mathbf{A}$ and $\mathbf{S}$ is unknown. To return to the original scale of the signals in $\mathbf{S}$ a similar lemma as defined in section 5.2 is defined for $\mathbf{A}$.

The correct scaling factor α is acquired the same way as in equation (5.12), which gives the following α formular.

$$\alpha = (\mathbf{A}^T\mathbf{A})^{-1} \sum_j \mathbf{A}_{j\cdot}^T \tag{6.15}$$

6.3 Information Criteria for model selection

Selecting the right number of sources hence the right model is crucial for all methods proposed. Therefore a good and robust measure is needed. With the ICA method the log likelihood ($\mathcal{L}$) is computed from the covariance matrix. This can be used in calculating different criteria for the selection of numbers of sources. Bayes Information Criterion BIC turns ($\mathcal{L}$) and the number of sources into a measure of the statistical optimal model.

$$\mathrm{BIC} = -2 \cdot \mathcal{L} + \mathrm{q} \cdot \log(\mathrm{N}) \tag{6.16}$$

where q is the numbers of parameters in the model and N is the number of samples. The criterion assumes that the N samples are statistically independent. Unfortunately, this is not the case in PET imaging, because of the point spread function neighboring voxels correlate. Hence, the criterion has to be adjusted to cope with the true number of independent samples. This is done by multiplying the total $\mathcal{L}$ with the fraction of true number of samples $\hat{N}$ over the number of samples N and substituting the number of samples N with the true number of samples $\hat{N}$. This gives the following modified Bayes information criteria

$$\widehat{\mathrm{BIC}} = -2 \cdot \mathcal{L} \cdot \frac{\hat{N}}{N} + q \cdot \log(\hat{N}) \tag{6.17}$$

which then can be calculated for different number of sources to find the number to use. $\widehat{\mathrm{BIC}}$ is minimized to find the optimal model.

Part II

Results

Chapter 7

Clustering results

In this and the following Chapter the results obtained using the different methods will be presented. In this Chapter the results from the K-means and the fussy C-means algorithm are shown. The methods are first tested on simulated data and then on PET data.

7.1 K-means results

In this section results from performing the K-means clustering method to extract vascular TAC from PET data are presented. The results are partly reproduced from [15]. However, the weighting is not performed here, since the weighting is high in the last frames and the arterial information is mainly in the first ones. These results are going to be compared with more advanced deconvolution methods later in the thesis.

7.1.1 K-means on simulated data

The characteristic of the K-means method is first tested by using the method on simulated data. The data is described in section 2.4. The K-means algorithm

requires the number of cluster K to be defined in advanced. Therefore, a comparison across K is needed. This is not straightforward, as the cost function will decrease towards zero with increased numbers of clusters. It can not be used directly to compare across different values of K. However, a simple method is to look for large changes in the cost function as the value of K increases. This indicates a possible significant change in the partitioning, hence a good choice of K. For the simulated data the within variance (cost function) is plotted for different K values in figure 7.1. Here it is clearly seen that the optimal choice for K is 4.

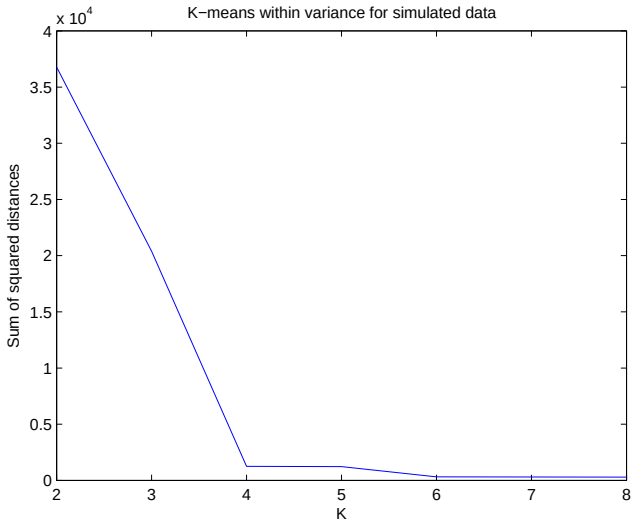


Figure 7.1: Within variance for different values of K.

As mentioned in the theory in section 3.1 a number of runs can be performed to secure that the cost function does not end up in a local minimum. The run with the lowest cost is chosen as the optimal solution across the different random initializations. In figure 7.2 the result of K-means clustering with 3 clusters and 2 different initial starting points is seen. It shows a big difference between the clusters depending on the initial cluster centers, hence the cost function ends up in different local minima. In this case a good choice of clusters is 4 which the robust results in figure 7.3 shows. Here both initial starting points end up in nearly the same result and with the same percentage of samples in each cluster. That was not the case with 3 clusters.

Because the process converges in a finite number of steps and because of the relatively fast convergence the K-means is an excellent initial choice for clustering. As indicated the K-means clustering has some disadvantages. But with

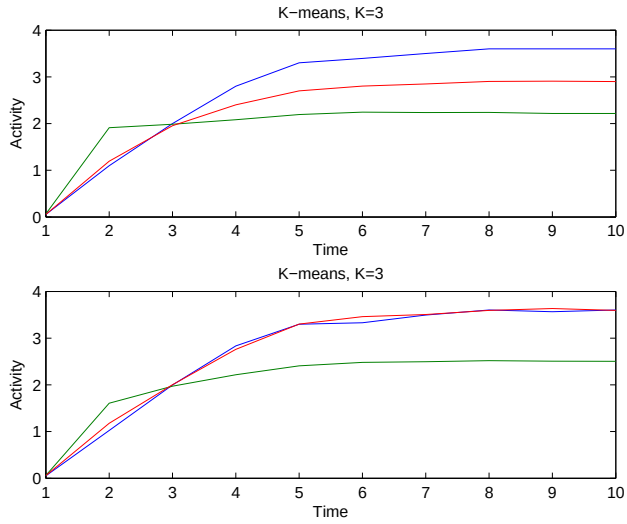


Figure 7.2: K-means with 3 and different initial starting points.

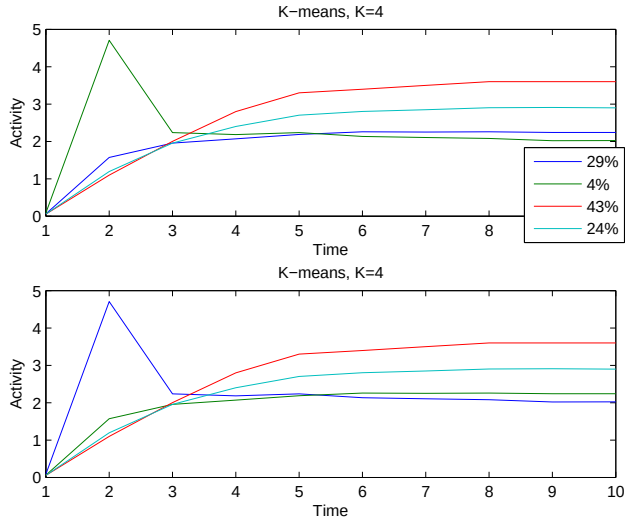


Figure 7.3: K-means with 4 and different initial starting points.

some safeguards, such as repeating the clustering several times for each choice of K, choosing the initializations in each case as a random subset of the data points, and varying the value of K over a range of relevance, the method shows very stable and fast performance.

Two of the main problems when using K-means on the simulated data are that each sample is clustered as one type, and that each cluster center is calculated as a mean of the samples belonging to the cluster. The first problem is that the method assumes that each sample consist of only one type, but in this case each sample has some fractions of the different types. Therefore, the method assumption is too crude. The second problem is that the method can not have clusters which have higher or lower values than the values found in data. This is seen in figure 7.4, where the vascular like cluster is plotted together with the underlying arterial like source. Because the highest value in the simulated data is around 5 the cluster centers can never reach the peak of the underlying arterial like source.

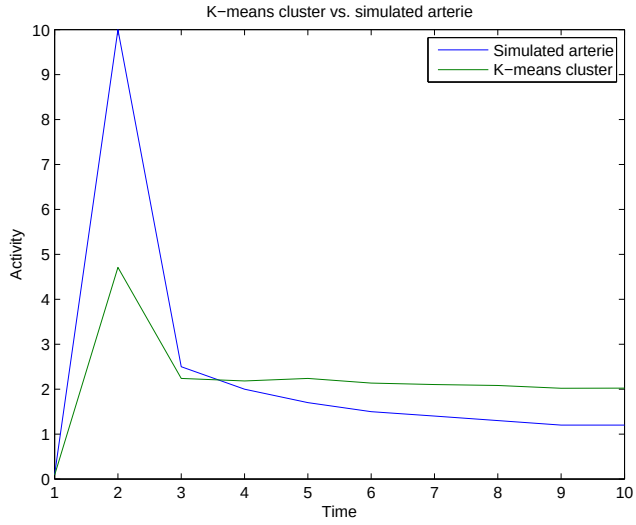


Figure 7.4: K-means cluster vs simulated artery.

The results from simulated data shows, that the K-means method has some limitations when working with data that has partial volume effects.

7.1.2 K-means on PET data

The K-means method is used to cluster the PET images into K clusters. In an optimal solution each cluster should contain only one component e.g. venous blood, arterial blood, white matter, grey matter or cerebrospinal fluid (CSF). This would be a straightforward and elegant way to segment the PET image and thereby identifying the vascular regions. In real life however this is not quite feasible with PET images. The resolution is not high enough to show the smaller details of the brain such as the arteries. This being the case the signals that are to be estimated can be expected to be noisy and a mixture of several components, as not many voxel will consist of blood alone. Not meaning that the voxels containing blood can not be found in the PET image, because if a voxel is 50% blood and 50% other components the TAC of that particular voxel is most likely to be different from those voxel containing 0% or almost no blood.

The PET image and the arterial sampled TAC are compared before the K-means algorithm is used to cluster the PET image. The maximum value in the sampled blood curve is 0.067 MBq for Pilot 6, and the maximum value in the PET image is 0.047 MBq. Therefore, it can be concluded in advance that the K-means method will not be able to extract the peak in the blood curve. However, the method is still going to be applied to the PET data, to see how the method performs and to compare the results to more advanced methods.

Choosing the number of clusters, K

An important issue when working with the K-means algorithm is to choose the right number of clusters, K. The correct number being the K where the features of the TACs become clear, and not necessarily the anatomical number of components in the brain.

To get an indication of the number of clusters to use, the cost function, the sum of squared distances between cluster centers and its members is shown for different K values in figure 7.5. It can be seen that for K larger than 3 or 4 the distance does not change very much. This reveals that K=3 or K=4, might be a good number of clusters to use. To use the same number of clusters for all Pilots, K=4 is chosen.

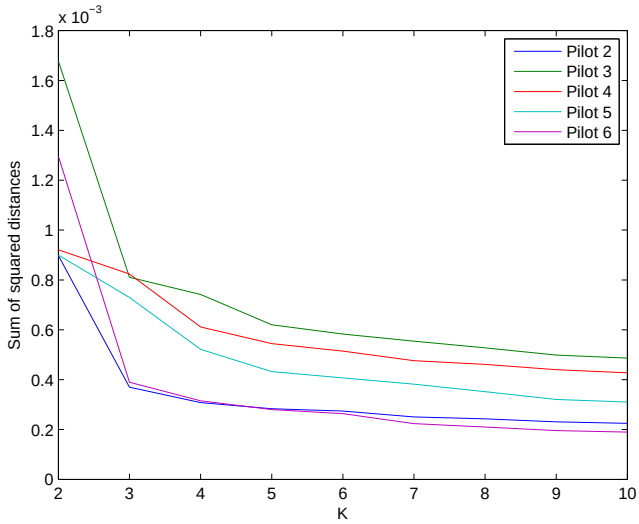


Figure 7.5: The summed distances for all 5 Pilots.

Results

The cluster centers for the clustering with K=4 are shown for Pilot 6 in figure 7.6. The blue cluster center resembles a vascular TAC and 4.51% of the voxels belong to that cluster for Pilot 6, which is not far from the true 5% vascular volume of the brain.

To examine where the clusters are located in the brain, a spatial image can be made where each voxel is assigned to one cluster. Figure 7.7 shows 6 transversal slices of the brain where each cluster is illustrated by a color, the same colors are used in figure 7.6. The blue cluster is the vascular component and is located where the veins are in the brain. In slice 10-12 the vein goes from the sides of the brain and to the back where they join and continue to the top of the brain. In slice 29-31 the cluster is in the middle, where there is a large vein in the brain. Thus it can be be concluded that the vascular looking TAC found, is indeed from the venous regions of the brain.

The vascular cluster found has some percentage of other tissues because of partial volume effect. Therefore, a second clustering is performed on the vascular cluster from the first clustering. This is to get the purest vascular source. Again the within variance is plotted to find the optimal number of clusters. When using K=4 in the first clustering, the sum of squared distances for the second

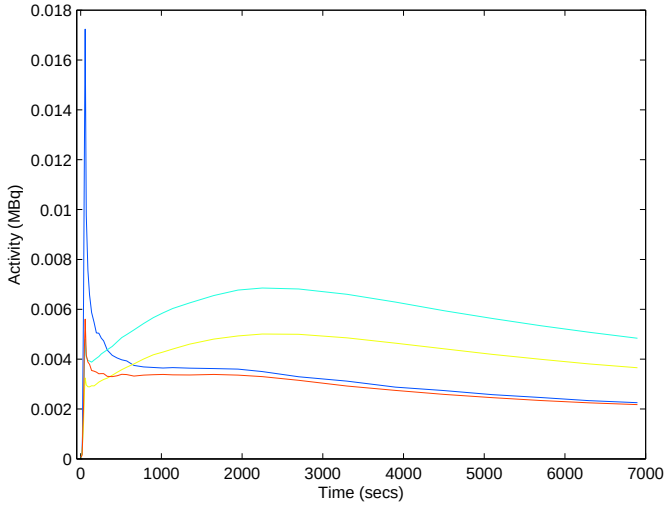


Figure 7.6: TACs for K=4, Pilot 6

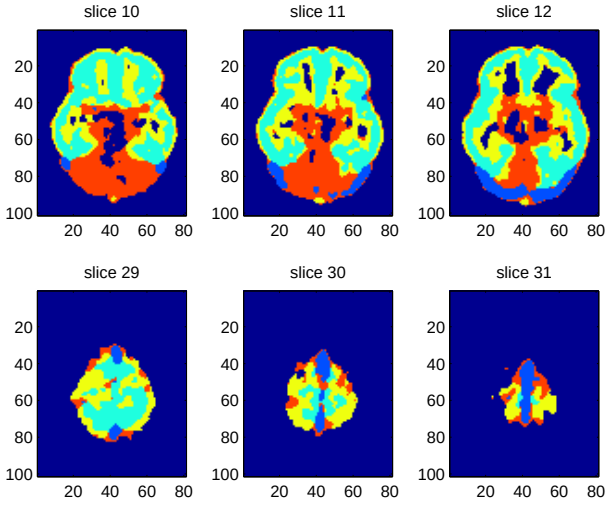


Figure 7.7: Spatial image for K=4, Pilot 6

clustering are calculated and shown in figure 7.8. Here it can be seen that the slope of the curves does not change much after K=4. Therefore, K is set to 4 in the second clustering for all Pilots.

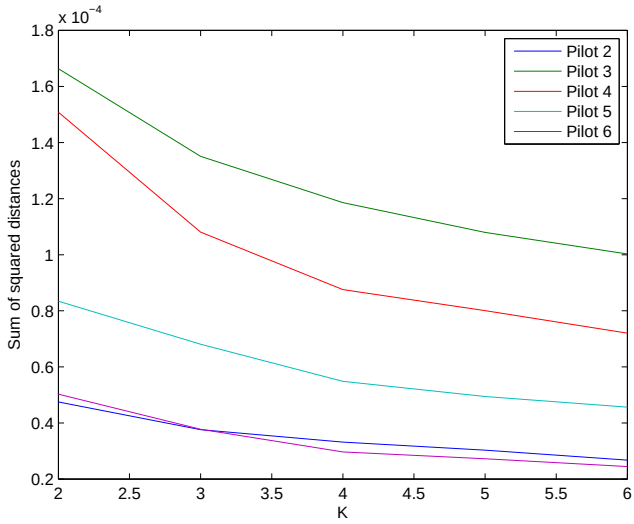


Figure 7.8: The summed distances in 2^{nd} clustering, for all 5 Pilots.

For the 2^{nd} clustering the vascular voxels comprise 0.36% of the total number of voxels, this means that the vascular cluster is trimmed to only include the most vascular ones, and therefore the TAC from this cluster should be the one that is closest to the sampled arterial TAC.

Validation

To validate different methods it is necessary to have a PET image which is segmented into vascular regions or to use the TAC from an arterial sampling, either way comparing the estimated result to the true solution.

The sampled arterial TAC is compared to the vascular cluster found by the K-means method. As seen in figure 7.9 the peak is very low in the clustered TAC compared to the sampled TAC and the level of the cluster after the peak is a little high. This is as expected since the maximum value in the PET image is lower than the maximum of the arterial sampled TAC. This is also consistent with the results found using simulated data.

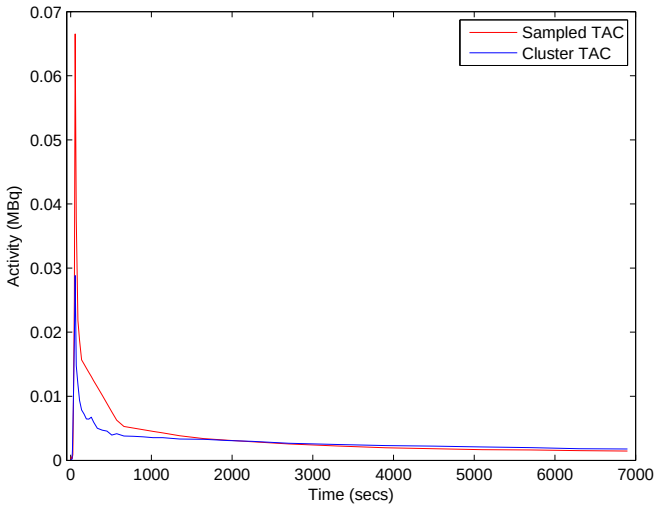


Figure 7.9: Sampled and hierarchical clustered vascular TACs for Pilot 6.

All the results shown are from Pilot 6. The other Pilots have similar results, which can be seen in Appendix A.

Weighting of frames

The weighting scheme used in [15] emphasizes the latter part of the TACs, therefore the end level in the resulting TACs is very close to the level in the sampled arterial TAC. Since no statistical reason for this weighting is given, this weighting scheme is not used when clustering with the K-means method.

7.2 Fussy C-means results

The K-means has the limitation that each data point can only belong to one cluster and because of partial volume effect this is not a good model for PET data. This problem might be solved with the Fuzzy C-means where each data point has a membership degree to each cluster. To test this property the C-means algorithm is used to cluster the simulated data set.

7.2.1 C-means on simulated data

To find the number of clusters to use on the simulated data set the within variance is calculated for different number of clusters. This is done several times with different fuzziness, m, all given the similar result as seen in figure 7.10. The curve indicates that 4 clusters should be used.

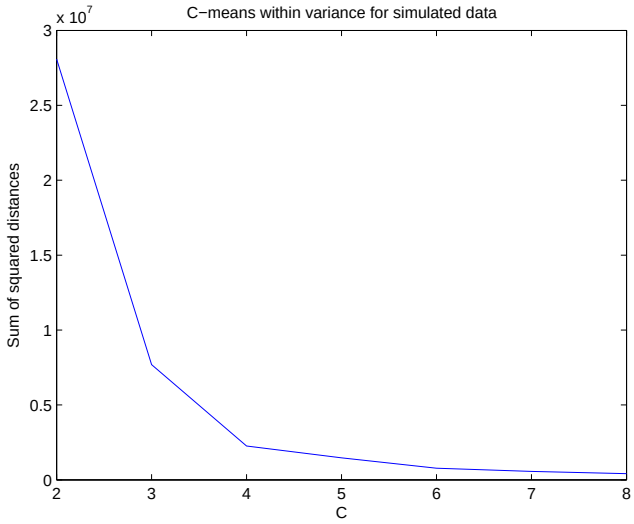


Figure 7.10: Fussy C-means cluster for different no. of clusters, $m = 2$.

The fussy C-means algorithm is then used to cluster the simulated data into 4 clusters with different fuzziness. In figure 7.11 and 7.12 the 4 centers of respectively $m = 2$ and $m = 3$ is seen. When using low fuzziness the centers are very close to the centers from the K-means method as expected. This is because each sample is given close to only one cluster membership. Using very high fuzziness the cluster membership is close to equal shared between the clusters. Here the method has difficulties estimating the small parts in the data as seen in figure 7.12 where the peak is not found as clearly.

The method has difficulties separating the small distributions because the different clusters have tendency to be equal sized as m increases. As seen in figure 7.12 the percentage of the vascular like curve is 10% that is double the real percentage.

The fussy C-means method is not used to cluster the PET data, since the method

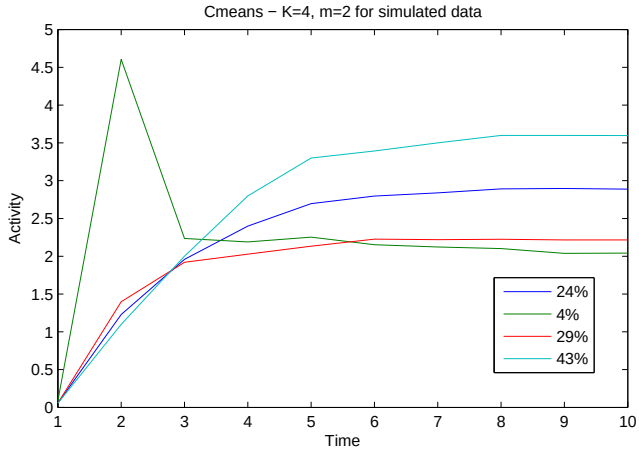


Figure 7.11: Fussy C-means cluster, $C = 4$ and $m = 2$.

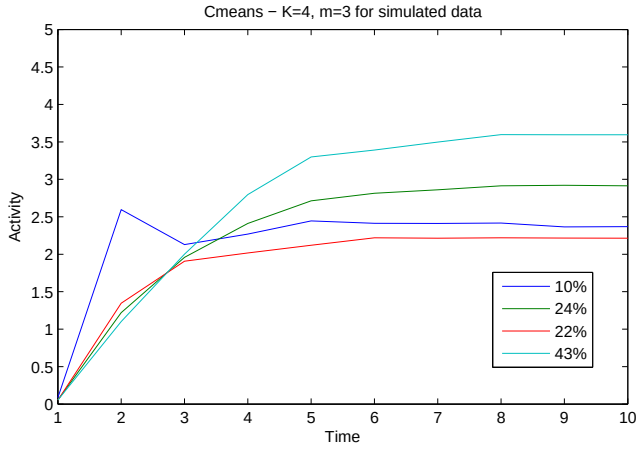


Figure 7.12: Fussy C-means cluster, $C = 4$ and $m = 3$.

can not give clusters that are very small compared with the other clusters. Furthermore, the method does not solve the problem, that cluster centers can not have values higher than the ones in the data. Hence, the C-means does not solve the problems from the K-means method.

7.3 Concluding on clustering

The biggest disadvantage when using the clustering methods is that the clusters can not have values lower or higher than the ones in the given data. The K-means method performs well taking into consideration the low resolution of the PET image, and since the maximum in the PET image is much lower than the peak in the sampled blood curve.

Segmentations of the brain is possible, and the venous regions are found. The arterial regions can not be distinguished from the venous, and this is probably because of the small diameter of the arteries. Hence, the maximum resolution in the PET image is to small. Since most of the voxels in the vascular cluster are from veins, and not arteries, the peak in the TAC is much lower in the cluster TAC than in the sampled TAC. If arterial information is to be extracted the K-means algorithm is probably inefficient since it does only cluster a voxel to one cluster. Therefore, if the arterial part of a voxel is never large, then no arterial cluster is extracted. The partial volume effects in the PET data makes the extraction of arterial TAC hard, since these regions are small.

For clustering the K-means method is the most appropriate, as the C-means does not perform any better on the PET data. C-means also has an extra parameter that needs to be tuned. However, using the K-means is not able to extract arterial clusters. This is shown for both simulated and real PET data.

Chapter 8

Scaling deconvolution results

The limitations found using clustering can be solved by different deconvolution methods. The deconvolution methods have the consequence that the scale of the results is unknown because, a factor can be applied to the coefficients and the sources without a change in the cost function. This scaling problem is looked into in this Chapter.

In the α factor scaling method explained in section 5.2 on page 32, the model at hand must explain data well enough on average, to assume that the spatial images voxel-wise sum to one. This method can only be used if there is a non-negative constraint on the results. Therefore, the α factor method can not be used on the PCA results. Since the purpose of this project is to avoid arterial sampling while doing the PET scan, the information from the arterial sampling can not be used. The vein sampling done during scanning is necessary since analysis of blood for metabolites is required. Therefore the vein information can be used to scale the results.

The vein blood curve has the same activity level as the arterial blood curve in steady state. Therefore the last arterial and vein samples can be assumed to be identical. This can be used to scale the vascular dynamic source back to the original activity(MBq) scale.

To evaluate this statement on the [^{18}F]Altanserin PET scans, the artery and

vein TACs are compared. In figure 8.1 the artery and vein sampled TACs are shown, it can be seen that the vein and artery curves follow each other nicely in the last frames.

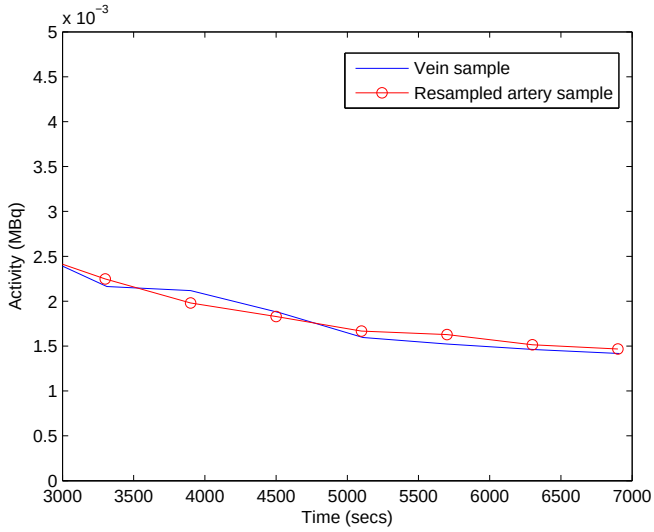


Figure 8.1: Sampled blood curves for one subject.

The end level of the vein curve can then be used to scale the estimated artery TAC into the correct scale. However, this will only hold if the estimated TAC is estimated correctly as an artery or vein TAC, and that the end level of the estimated is TAC is stable.

Chapter 9

PCA results

A principal components analysis (PCA) is performed on the simulated data and then on the PET data. This is a method that runs very fast so if relevant information can be extracted using first and second order statistics, the PCA is an appropriate tool to use.

The resulting principal components are analyzed and compared to the arterial TAC. For the PET data the solution can also be validated using the spatial information, given by the coefficients for the principal components.

9.1 PCA on simulated data

The PCA is performed on the simulated data set and the 1^{st} and the 2^{nd} principal component can be seen in figure 9.1. These two components describe 94% of the total variation in the data set. The 2^{nd} principal component looks like the negative vascular time signal. As mentioned one of the challenges with the PCA is that there is no non-negative constraint on the components and on the combination of these components, and this complicates the analysis of the solution. Furthermore, the scaling of the components is a problem since the end level of component can be negative.

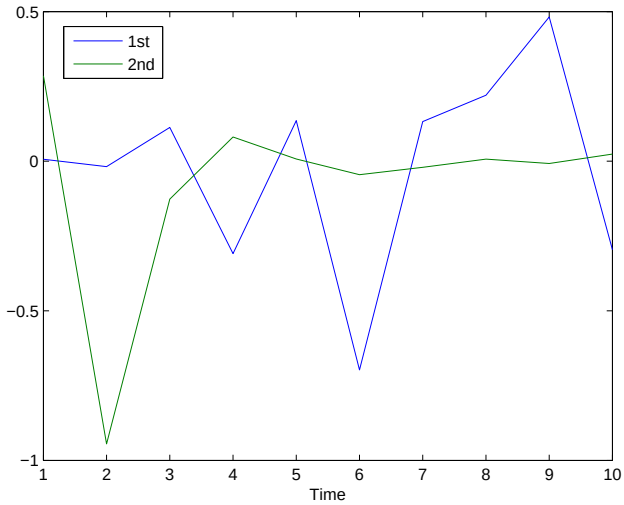


Figure 9.1: 1^{st} and 2^{nd} principal component.

The last 8 components are hard to analyze, and like the 1^{st} component they do not have any resemblance to a specific signal in the simulated data.

The PCA is very good at reducing the dimension of the data set, as 94% of the variation is explained by only 2 components. Since the aim is to extract natural time signals from the data, this method does not seem to be a good choice. A method that extracts more natural properties and easily interpreted components is therefore desired.

9.2 PCA on PET data

PCA is used to extract features from the PET data. Both spatial and dynamic results are given from the PCA. It is analyzed what spatial regions are explained by the different components and how the dynamic TACs for these components behave, to see if the principal components can be interpreted as true sources of some kind from the original data. All results in this chapter have been generated with Pilot 6 as data, results from the other data sets are illustrated in Appendix B.

Figure 9.2 shows the explained part of the total variation in the original data, as a function of the number of principal components. It is clearly seen that by excluding some of the variation in the PET data the size of the data set can be reduces a great deal. E.g. 5 principal components explain 70.7% of the total variation, this reduces the size of the data set to 12.5% of the original size.

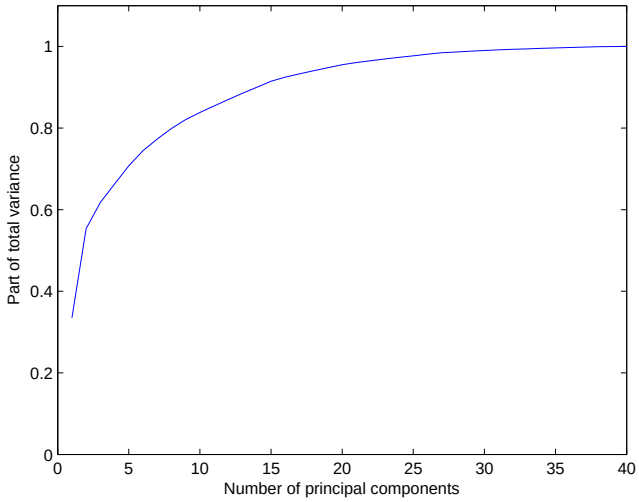


Figure 9.2: Part of total variation explained in PCA.

The two first principal components are shown in figure 9.3. The first component seems to describe a binding of some kind, since it increases with time at first and then decreases slowly. The second component resembles a vascular signal because of the peak in the beginning and then a very low activity level.

To see where the 2^{nd} principal component is located, the spatial image of that component is illustrated in figure 9.4. It is clearly seen that the location of the 2^{nd} principal component, matches the vascular region of the brain. The latter 38 components, do not have a well defined spatial location. The last 38 principal components are hard to identify with any natural signal of the brain. This is probably because of the cancelation effect of the components thereby, making analysis of the components very difficult.

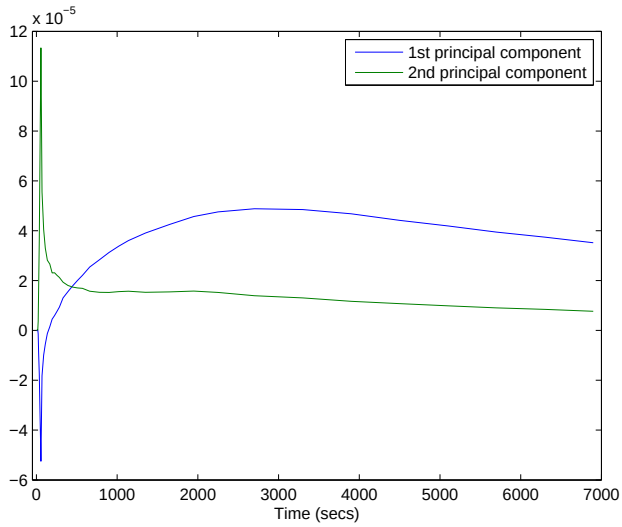


Figure 9.3: Two first principal components.

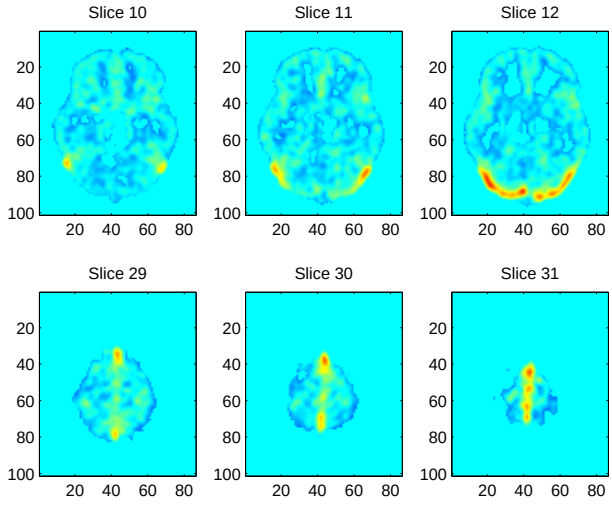


Figure 9.4: Spatial location of 2^{nd} principal component.

9.2.1 Validation

The 2^{nd} principal component is scaled to match the end level of the arterial and venous sampled TACs, all three curves can be seen in figure 9.5. Comparing the vascular principal component to the arterial sampled TAC, there is a great difference between the peaks of the two curves, but the levels fit very well to each other. However, the component is not far from the vein TAC. This is a clear indication of that the 2^{nd} principal component is venous, and that no components extract pure arterial information.

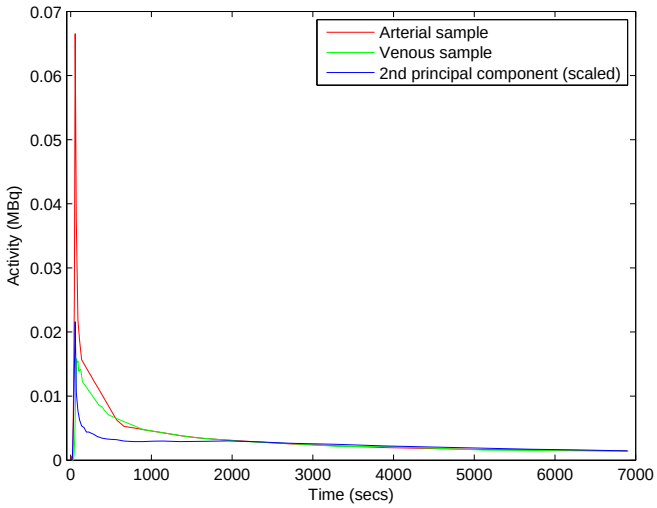


Figure 9.5: Comparison of arterial- and venous sampled TACs and scaled 2^{nd} principal component.

9.3 Concluding on PCA results

The PCA is able to estimate the venous regions of the brain by the 2^{nd} principal component, and no arterial information is extracted from the data. Therefore, the PCA can not be used to estimate the arterial TACs.

When comparing the 2^{nd} principal component with the sample arterial curve, it shows that the peak of the estimated TAC is much too low. This is probably because of the fact that so little of the brain is arterial, and therefore it is not

extracted as a clear component in PCA. And since the components can cancel each other out and be linearly combined in any way, the interpretation of the principal components becomes very hard. The factors that explain the most of the variation, binding and venous regions, are extracted as the first two principal components, but all the 38 other principal components are hard to analyze.

The reason the PCA is unable to estimate the arterial TAC could be because of the way the PCA represent the components. The components found are not individual and sparse, but more like a mixture of different artifacts, and to combine them to one factor, say arteries, then maybe all principal components need to be linearly combined. Also, as shown in [14], the PCA does not extract natural components from the data. To achieve this a non-negative constraint might be a good idea. The first and second order statistics of the PCA method is not enough to extract arterial TACs. This is shown for simulated and real PET data.

Chapter 10

NMF Results

In [11] the NMF is used to extract vascular curves, where they use a ROI and only perform the NMF on the initial 90 seconds. In this chapter the NMF is used on the whole PET data set, so no region needs to be defined and all frames are used to get the entire time activity curve (TAC) and not just the peak.

In this chapter the NMF method is used to deconvolute the simulated and the [^{18}F]Altanserin PET data set. The results are validated using the sampled artery TACs, to find the most appropriate method to extract TACs from the PET images.

Deciding the number of sources to use in the NMF method is a non-trivial task, and no exact method of doing this is available. No statistical information, like the log likelihood function ($\mathcal{L}$), is given when using this model. However, the error of the model can be investigated for different number of sources.

There are two ways to evaluate the results from the PET data, since the method gives both spatial and dynamic information. The dynamics of the solution are the TACs that should resemble the TACs of different regions of the brain. One or more of these being the vascular part of the brain, namely the venous and arterial regions. The sampled and the estimated TACs can be compared with each other, by visual inspection and by correlation. The spatial information of the results of the NMF method are images of the origin of the different sources

found. Therefore, these images should, if the method gives a good solution, segment the brain into different regions, each with different bindings towards the tracer. This is not possible with the simulated data, since the data is randomly distributed, thereby making the spatial image uninteresting.

10.1 NMF on simulated data

A deconvolution of the simulated data is performed using the NMF. The cost function is shown in figure 10.1, this is to show how well the NMF describes the data for different number of sources. This clearly indicates that 2 sources is a good choice, since the slope does not change much after this point.

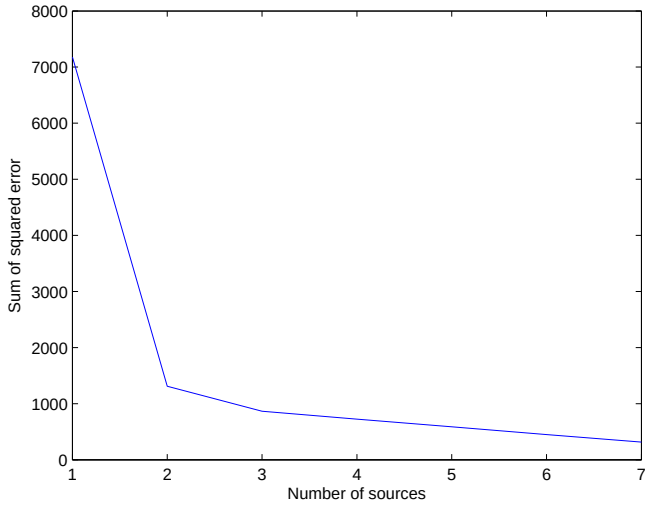


Figure 10.1: Cost function for different number of sources.

Using 2 sources in NMF, the resulting TACs is a vascular curve and a binding curve, they can be seen in figure 10.2. The estimated TACs are rescaled by the α factor, see equation (5.12) on page 33. The two TACs have the correct level except for the latter part, where both curves are too high. The scaling can naturally never be correct, when estimating the wrong number of sources using the NMF. This explains the incorrect scaling using the α factor. An interesting observation is that the estimated binding TAC is zero at Time=2, where the vascular signals peak.

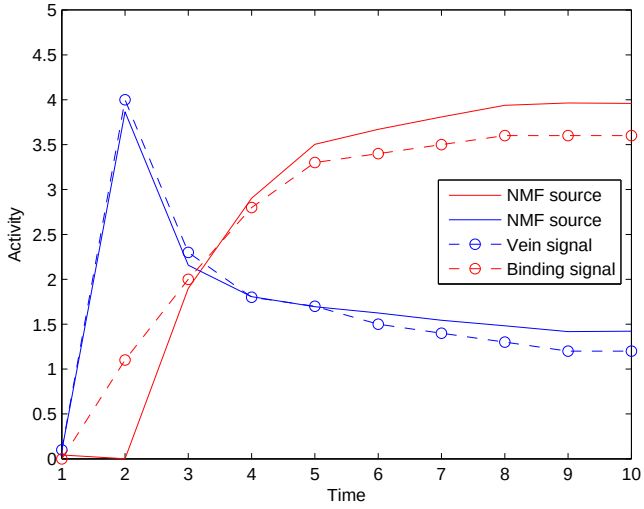


Figure 10.2: NMF estimated α scaled sources, 2 sources.

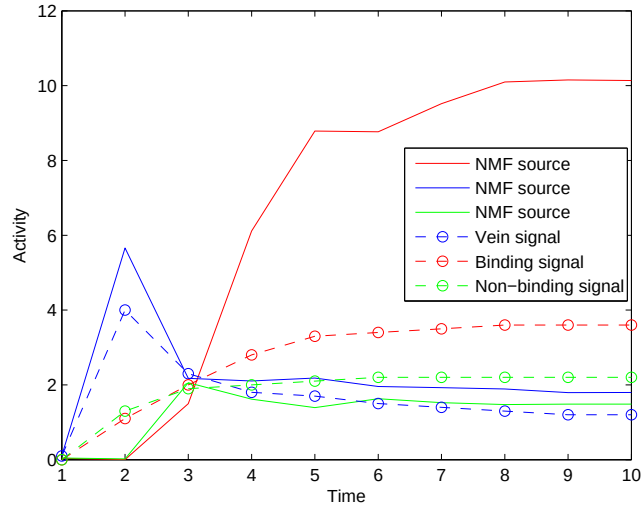


Figure 10.3: NMF estimated α scaled sources, 3 sources.

To see if the NMF method is able to extract more than 2 sources, the NMF is performed using 3 sources, the results is shown in figure 10.3. Again the sources are rescaled using the α factor. The shape of the 3 estimated curves looks fairly good, and the peak in the vascular curve is higher than before. But there is a problem at Time=2, where both the binding and the non-binding estimated TACs are zero. This indicates that there is a tendency that other curves are underestimated where a curve has a peak. The peak in the estimated vascular curve somehow explains the data well enough in point Time=2, so that the other curves are set to zero.

The maximum value in the vascular NMF source is 5.7 and the maximum value in the simulated data set i 5.3. This shows that the NMF has potential to extract sources that are not available directly in the data, but only partially in some observations.

The red NMF estimated source is not scaled correctly. It is too high with a maximum level of 10. This failure in the α scaling method also has the consequence that the coefficients will be erroneous. To compensate for the high level of the estimated binding signal, the coefficients for this source become very small. Thereby distorting the parts from the other sources. Since the binding source is much larger than the other sources, it can influence the other parts in the rescaled solution.

The simulated data set has 4 sources, but since the two vascular signals are similar and very small (1% and 4%) these signals might be described as one by the NMF, this seems to be the case with 2 and 3 sources.

It can be concluded that the NMF works fairly well when estimating two sources, which is also what the cost function indicates. Here the sources found are rescaled correctly and the vein and binding TACs are found. There is a problem with the end level of the estimated curves, this is probably because of the fact that the cost function is not sensitive enough to the curve level, and also because there are two signals that are not estimated, which will give some error.

The NMF shows promising results on the simulated data, but the number of sources is underestimated and the α scaling is only correct when using 2 sources.

10.2 NMF on PET data

Now the NMF is performed on PET data and the number of sources to use is investigated. In figure 10.4 the sum of squared error (SSE) is shown for 1 to

6 sources for all 5 Pilots. The SSE is naturally decreasing with the number of sources used, but any large change in the slope of the curves might be an indication of how many sources to use. For all Pilots the largest change in the SSE is between 1 and 2 sources, and from 2 to 5 sources the slope seems to be stable. This is an indication that 2 sources could be a reasonable number of sources to use in the NMF model, this however is investigated further in this section.

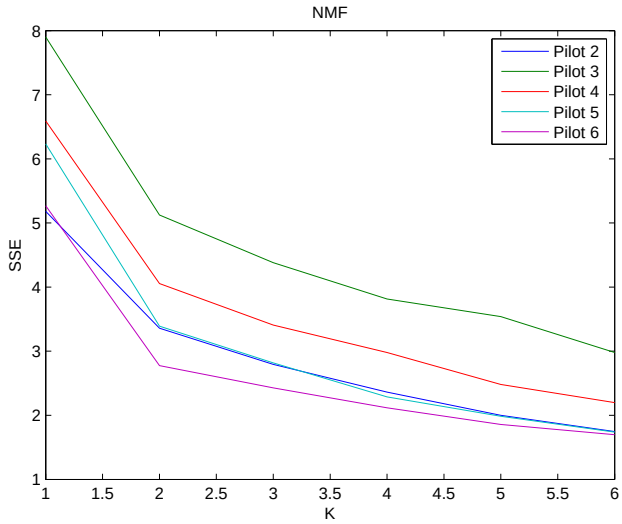


Figure 10.4: Sum of squared error, as a function of number of sources.

Using the NMF algorithm to deconvolute the observed PET image into two sources gives the TACs shown in figure 10.5. One of the TACs has a clear vascular shape and the other TAC looks like a curve describing binding. The vascular curve has a peak and then a low level in the end of the curve. As the SSE plot indicates, 2 sources seems to be a good choice, but to see if this is the optimal number of sources, TACs using more sources are generated.

Figure 10.6 shows the TACs based on 3 sources. There is one vascular like curve and two other curves, which look like binding and non-binding TACs. This result looks very promising, since using 3 instead of 2 sources actually yield good results, as the TACs look natural and not like noise or overfitted sources. One of the differences in the vascular curve when using 3 instead of two sources is that the end level of the curve is lower for 3 sources.

To validate the results of the NMF the spatial images of the sources are illus-

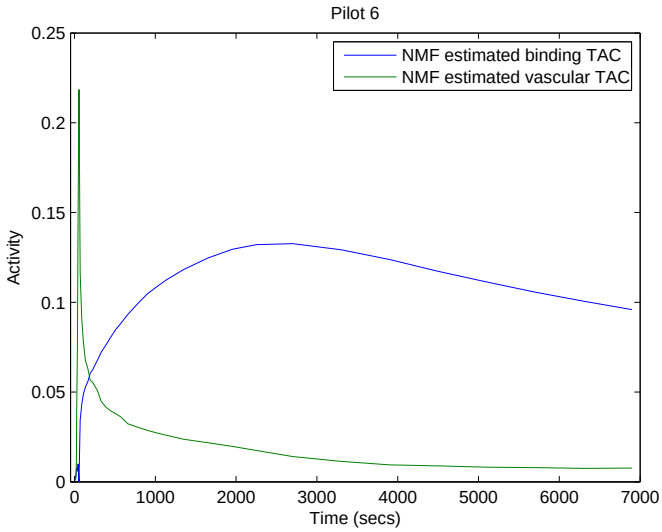


Figure 10.5: TAC found by NMF, for K=2 in Pilot 6.

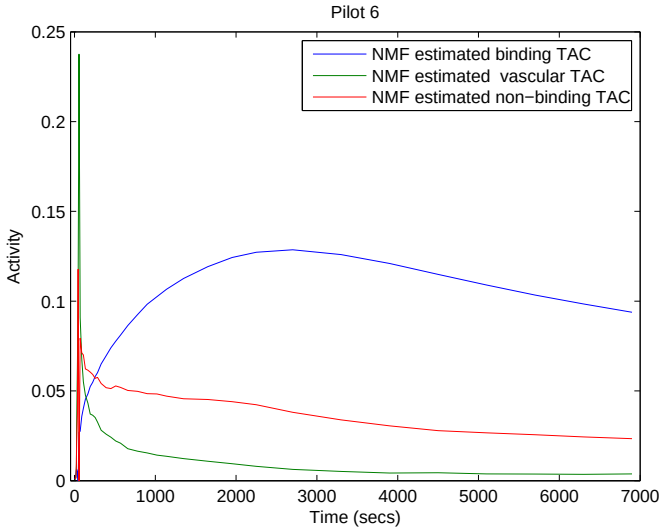


Figure 10.6: TAC found by NMF, for K=3 in Pilot 6.

trated. In figure 10.16, 10.17, and 10.18 the spatial location of the 3 sources are shown on page 80. The three spatial images consist with the dynamic curves shown in figure 10.6, as the sources seem to describe binding, non-binding and vascular regions.

Using 4 sources in the NMF gives inconsistent results and unnatural TACs, which is a indication of overfitting. The optimal number of sources to use is therefore chosen to be 3 sources, since the sources found are natural looking and there are no signs of overfitting.

The results shown in this section are for Pilot 6, results for all Pilots can be seen in Appendix C.

Correlations

The correlations between the arterial TAC and the vascular TACs are shown in table 10.1 to estimate the quality of the found TACs. For K=2 and K=3 the correlations are high, all above 0.94. For K=4 the correlations are lower for all Pilots. K=3 can therefore not be rejected as a good number of sources to use in the NMF. However, the correlation is higher for 2 sources in Pilot 2, 3, and 4. This might be because of the low end level of the curves, with 3 sources. Intuitively using 2 sources in the NMF algorithm might seem to few, since the method should extract vascular information in one source and then all the rest in the last source. The tradeoff between 2 and 3 sources is the peak height, and end level height in the vascular curve. Using 2 sources, the peak is not very high, and with 3 the peak is high and the end level is very low.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
K = 2	0.9872	0.9784	0.9854	0.9490	0.9559
K = 3	0.9679	0.9684	0.9697	0.9667	0.9688
K = 4	0.9091	0.7453	0.8297	0.7997	0.9229

Table 10.1: Correlations between estimated and arterial sampled blood curves.

To evaluate the two non-vascular components found by using 3 sources in the NMF, they can be correlated to the ROI curves shown in section 2.3. There are ROI with and without binding. These are compared with the estimated TACs in table 10.2 and 10.3. This is only done for K=3, since in that case there seems to be an estimated binding and a non-binding TAC. The binding ROI is the frontal cortex region, and the non-binding ROI is the cerebellum region. When evaluating the correlation coefficients for binding and non-binding regions of the brain, shown in table 10.2 and 10.3, one must look at the estimated curves

represented in the **H** matrix to see if any artifacts are present. Sometimes one of the curves goes to zero as the other curves have peaks. If one of the values in the TAC is suddenly zero as seen with the simulated data. This will greatly effect the correlation coefficient, although the curves might be very alike for all but one time point. This can explain some of the lower values in table 10.2 and 10.3. Otherwise the binding TACs have high correlations, while the non-binding TACs are generally lower. This also consists with the fact that the non-binding TAC is estimated as the 3^{rd} component, therefore being the least significant of the three sources.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
K = 3	0.7808	0.9520	0.8309	0.8680	0.9575

Table 10.2: Correlations between binding TAC and frontal cortex ROI curve.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
K = 3	0.8083	0.6569	0.7935	0.7720	0.7171

Table 10.3: Correlations between non-binding TAC and cerebellum ROI curve.

Scaling

The α scaling method is used to re-scale the TACs found using the NMF. All three TACs can be evaluated using this re-scaling scheme. Figure 10.7 shows the ROI TACs and the NMF TACs. It can be seen that the NMF estimated non-binding TAC fits the ROI TAC well. The binding TAC does not fit as well, it is higher than the ROI TAC.

The α factor seems to re-scale the NMF solution into the approximately correct activity scale (MBq). Therefore, this method is used to scale the NMF TACs back to the correct scale.

Although the correlations between the estimated TACs and the artery TACs are very high, the vein scaling methods mentioned in Chapter 8 can not be used. This is because of the low end level in the estimated artery TACs, this level seems unstable and often too low.

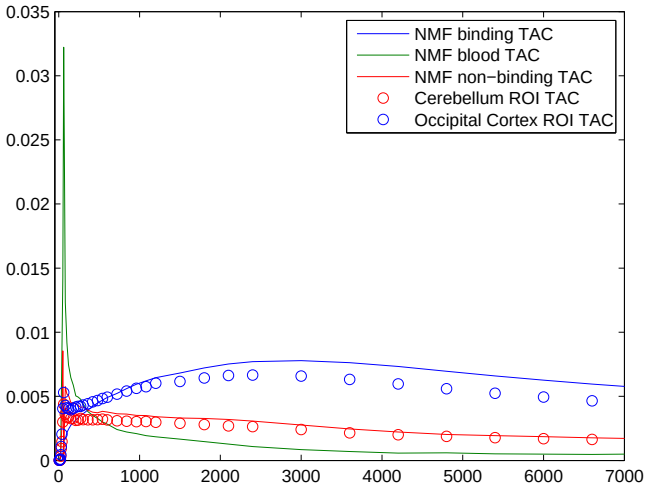


Figure 10.7: Re-scaled TACs, ROI TACs shown as circles.

Parts of total brain volume

After rescaling the results, the parts of binding, non-binding and vascular regions of the total brain volume can be calculated. The parts for the NMF results are shown in table 10.4. It can be seen that for the binding component the part is between 0.57 and 0.66, the non-binding part is between 0.24 and 0.30, and the vascular part of the brain is between 0.09 and 0.14. The vascular like component part is clearly too large since only approximately 5% of the brain is vascular. Otherwise the other parts match the white- and gray-matter parts of the brain a little better. They are approximately 65% and 35% respectively. This also indicates that the scaling is correct, as the binding is mostly in the white matter, and white matter is approximately 60% of the total brain volume.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
Binding	0.6597	0.5729	0.6139	0.6641	0.5995
Non-binding	0.2492	0.2852	0.2966	0.2352	0.2691
Vascular	0.0911	0.1419	0.0895	0.1007	0.1314

Table 10.4: The parts of the brain volume for each Pilot, with K=3.

10.2.1 NMF error

The residuals of the NMF solution are easily extracted.

$$\text{residuals} = \mathbf{V} - \mathbf{WH} \tag{10.1}$$

The residual term can be display as an image, since it is the error for each voxel in PET scan. The residual images are inspected to see if the model describes the PET fairly, in that case the residuals should be normal and uniformly distributed around zero. And the residuals should be randomly distributed in the residual image, and not mainly being present in a certain region of the brain.

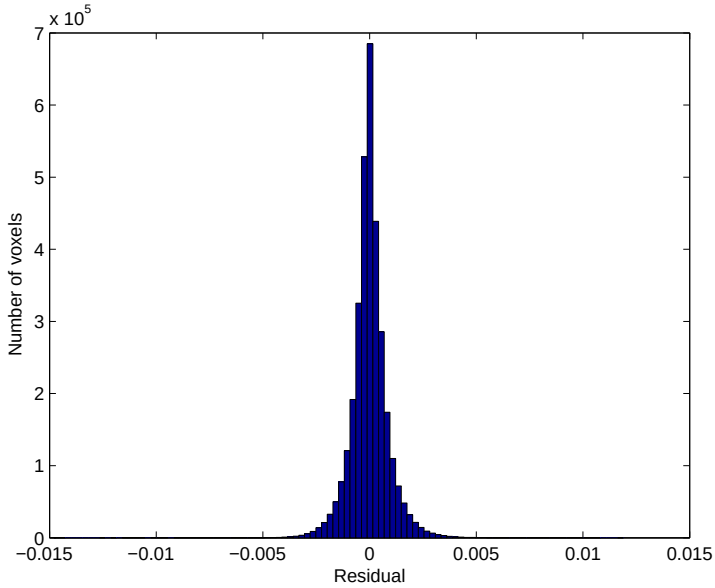


Figure 10.8: Histogram of the residuals for Pilot 6 using 3 sources.

The histogram in figure 10.8 shows the residuals for the NMF method for one subject. It can be seen that the residuals are centered around zero and look normally distributed. This error evaluation is still valid when α scaling the NMF solution, since the term **WH** will be the same after the α factor scaling process. A spatial image of the residuals can be seen in figure 10.9. From this image the residuals look randomly distributed in the brain, and no region with high residuals is seen.

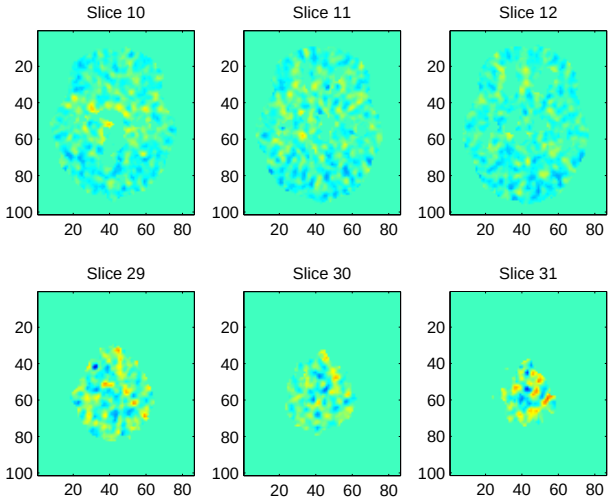


Figure 10.9: Mean over time residuals for Pilot 6, using 3 components.

10.2.2 Weighted NMF

Statistically weighted NMF

The level of the end of the vascular TACs found with the NMF method, seems to be very low. The uncertainty in the last time points therefore seems to be high. This contradicts with the fact that the last frames are long and have many counts compared to the first frames. In fact the first frames do not have the required 500000 counts to perform the reconstruction. Since the peak of the arterial TAC is in the first few frames, these frames can not be discarded. One way of solving this problem is to weight the frames as shown in [7]. The weighting used is:

$$\text{weight}_f = \frac{\text{duration}_f}{\sqrt{\text{counts}_f}} \tag{10.2}$$

where f is the frame number. This weighting assumes the number of counts to be poisson distributed, and therefore it can be rewritten to:

$$\text{weight}_f = \text{duration}_f \frac{1}{\sigma_f} \tag{10.3}$$

where σ_f is the standard deviation of the frame f. This results in the weighting function that can be seen in figure 10.10, for this PET data set. To choose the number of sources to use in WNMF the sum of squared errors are evaluated. In figure 10.11 the SSE is shown, and it can be seen that for most of the Pilots the curve flattens after 3 sources. This is a strong indication that 3 is a reasonable number of sources to use, which is the same as in the un-weighted NMF.

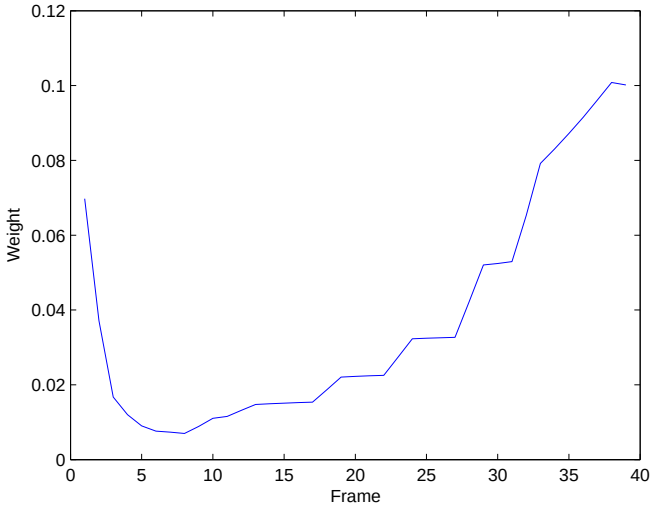


Figure 10.10: Weighting applied to frames.

This weighting scheme is applied to the [^{18}F]Altanserin PET data. A comparison of the estimated α scaled vascular TAC and the arterial TAC can be seen in figure 10.12.

Weighting the data using equation (10.2), the last frames are weighted the most, and very little emphasis is put on the frames 5-15 where the peak in the TAC appears. This is consistent with the results seen in figure 10.12 where the end of the vascular TAC is close the level of the sampled artery TAC, but the peak of the vascular like TAC found by the WNMF is much lower than the artery TAC. The peak in the resulting blood TAC is also lower than the peak estimated by

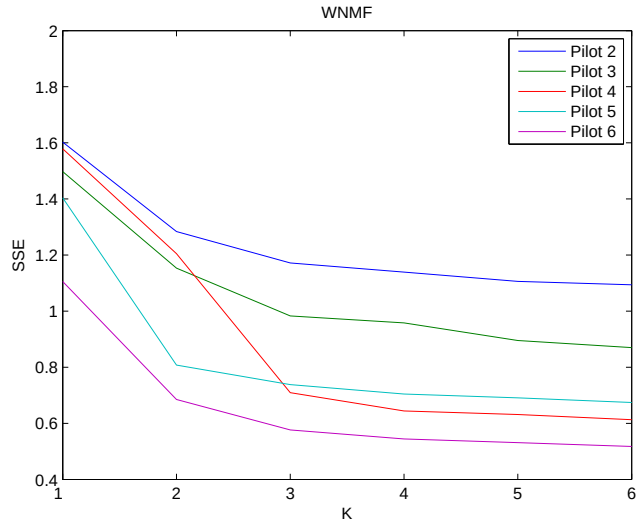


Figure 10.11: Sum of squared error, as a function of number of sources.

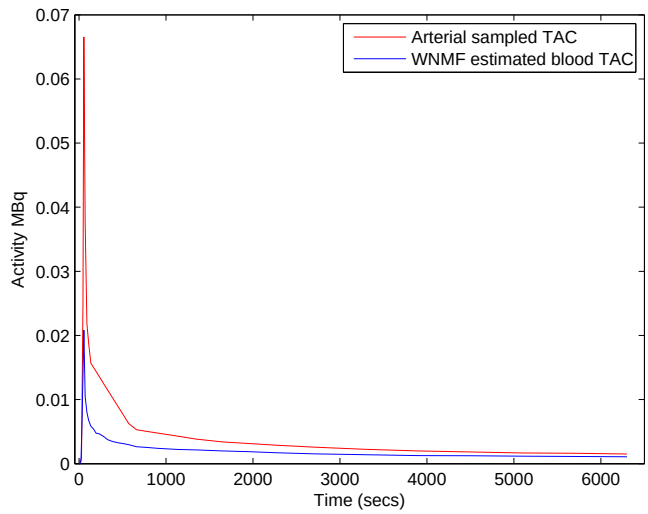


Figure 10.12: Comparison of sampled and estimated arterial TAC, for Pilot 6.

the K-means method. The weighting of the frames is therefore not useful when estimating artery TACs from PET data.

Alternatively weighted NMF

The peak is underestimated using the statistically weighted NMF. An alternative weighting scheme is considered. Although the first frames in the PET scans are the most noisy ones, these frames contain the information where arterial blood TACs are significantly different from the tissue TACs.

A weighting is proposed where the frames containing the peak are weighted the most, and the other frames are less important. The last few frames are weighted high to get a good segmentation of the PET scan. This weighting can be seen in figure 10.13. The peak in the arterial blood curve is between the 5^{th} and the 15^{th} frame.

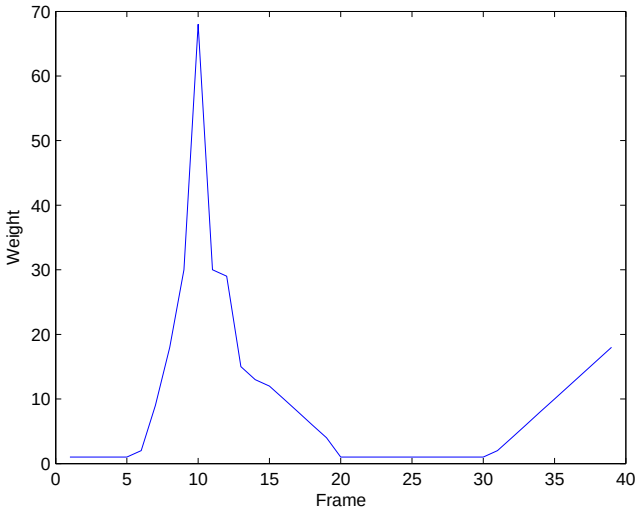


Figure 10.13: Alternative weights of frames.

The dangers of this alternative weighting is that the weight function might be a way of overfitting the data at hand. The error in the estimated TACs is now after the peak, where the data points are weighted the least. The TACs found are more vascular, since the emphasis on the peak focuses on vascular tendencies in the dynamic part of the data set. When evaluating the spatial images of this

alternative weighting, several vascular components occur and only one other component. Therefore this solution does not describe the data as a whole but it mainly extracts vascular information. Furthermore, the alternative weighting function can not be statistically supported.

10.2.3 NMF stability

To test the stability of the NMF algorithm used 15 runs are made on the same data, and the results are compared to each other. In figure 10.14 the two TACs are shown with ± three standard deviations.

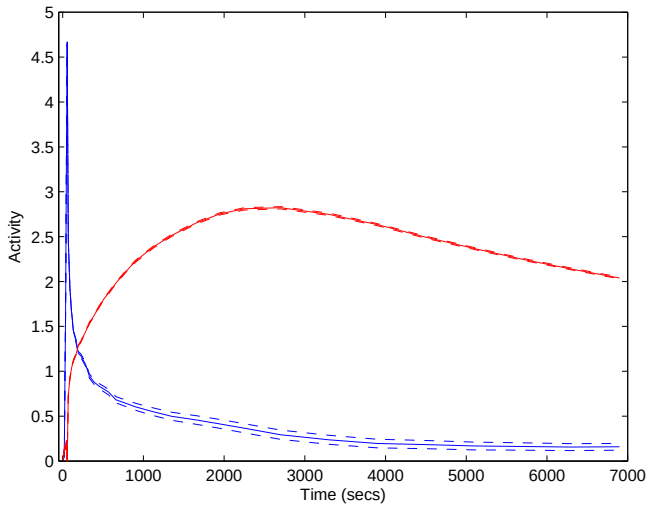


Figure 10.14: TACs shown with ± 3 standard deviation.

The results of the NMF algorithm seem to be very stable, so there is no need to do several runs to find the optimal solution.

10.2.4 Comparison with K-means

To illustrate the differences in the estimated vascular TACs from the K-means and the NMF method, the two TACs and the sampled arterial TAC are shown in figure 10.15.

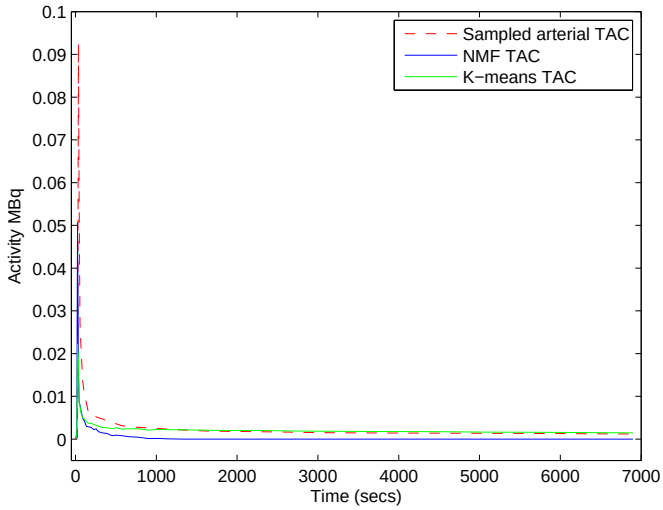


Figure 10.15: Comparison of K-means, NMF and arterial sampled TAC, for Pilot 2.

Is is clearly seen that the peak in the NMF TAC is higher than the K-means peak. There is also a great difference in the end level of the two curves. The K-means follows the arterial sampled TAC well, but the NMF TAC is very close to 0, and well below the K-means TAC.

10.3 Concluding on NMF

The results from the simulated data shows, that the NMF method can be used to extract two sources, the vein- and the binding-signal. These signals are scaled correctly using the α factor. However, the NMF is unable to extract more than two sources. For larger number of sources the method scales erroneously.

The NMF methods generates stable results, and there does not seem to be a need for several runs to find the optimum solution of the iterative algorithm. Using the α re-scaling method the mixing and TACs from the NMF can be used directly in modeling and can be compared to sampled or ROI TACs.

The vascular TACs estimated by the NMF are better than the ones found by

the K-means clustering and by PCA. The weighted NMF does not perform very well, since the arterial information is in the first frames where the noise is the highest.

An interesting thing about the NMF results is the spatial images of the sources. This clearly indicates that the method is able to perform a deconvolution of the PET scan into meaningful components. It is also obvious that estimating the true underlying sources using only 3 components, gives a very crude solution, as the approximately 90000×40 voxels are described by a positive linear combination of 3 sources, a total of 3×40 points.

With both simulated and PET data the NMF method is able to extract reasonable vascular TACs that have higher peaks, than available in data. The method is not able to distinguish between arterial and venous signals, this is probably because the arteries are to small compared to the resolution of the data. To get really good arterial TACs the method will most likely need to have a larger part of arteries in some samples.

In both simulated and PET data, values of zero appear where the vascular signal peaks, and there is a problem with the end level of the curves, which seems hard to estimate using the NMF. This indicates that the NMF has the tendency to underestimate the number of sources, or viewed from another angle, the sources need to be more dissimilar than those in the data in this thesis, to be able to deconvolute into more sources.

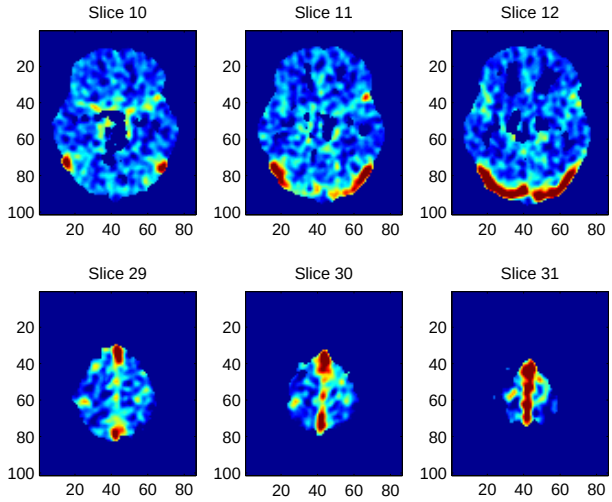


Figure 10.16: Spatial part of vascular component, K=3 for Pilot 6.

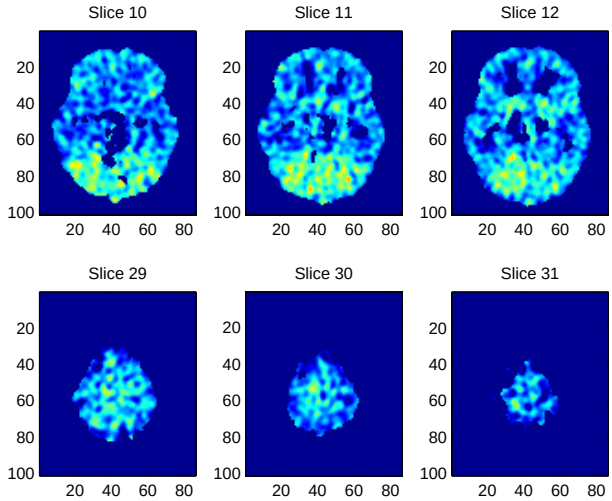


Figure 10.17: Spatial part of non-binding component, K=3 for Pilot 6.

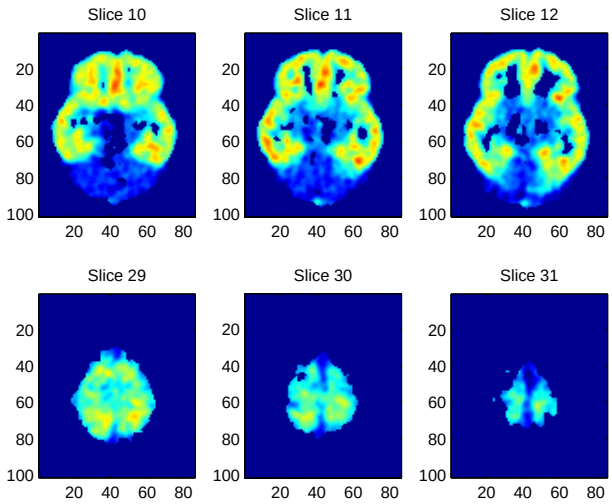


Figure 10.18: Spatial part of binding component, K=3 for Pilot 6.

Chapter 11

Results from independent component analysis

The mean field independent component analysis (ICA) method is used to deconvolute the data into spatial and dynamic sources, as with the NMF. When using this method the negative values are truncated to zero and each time dependent sample is divided by the global maximum to transform data in to the domain from 0 to 1. Again the number of sources has to be known in advance, but as mentioned in chapter 6 the $\mathcal{L}$ is calculated. With this at hand it is possible to estimate the statistically optimal number of sources.

11.1 ICA on simulated data

The ICA is performed on the simulated data described in section 2.4 to find the characteristics of the method. All results are rescaled with the α factor, described in equation (5.12). Initially the log likelihood, $\mathcal{L}$, is calculated for different number of sources. Doing this the first characteristics of the method are found. Running the mean field ICA with positive constraints is slow compared to the NMF, moreover sometimes the method ends up in a local maximum when using more than 4 sources and therefore the model has to be recalculated, and after a few runs for different number of sources the maximum log likelihood is

found. With this at hand Bayes information criterion (BIC) can be calculated according to equation (6.16) in section 6.3. The result is seen in figure 11.1. The statistical optimal number of sources is 2, but for 3 and 4 sources the BIC term is very close to the minimum.

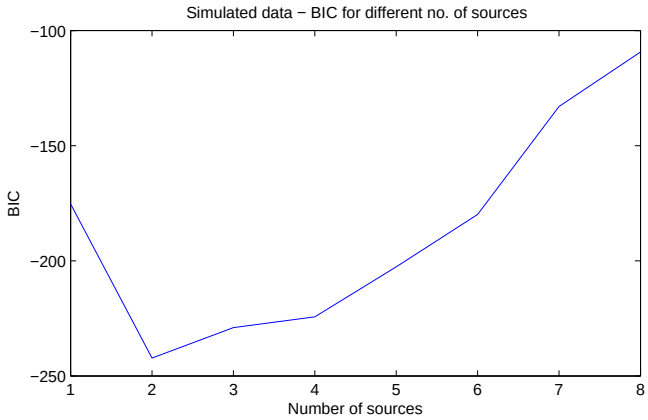


Figure 11.1: BIC for different no. of sources.

When using 2 sources on the data the resulting sources are seen in figure 11.2. The sources are multiplied with the global maximum and scaled according to the equation (6.15) in section 6.2. The TACs are plotted together with the underlying signals from the simulated data. The 2 sources are very similar to the simulated binding curve and to the simulated vein curve. With only 2 ICA sources it is possible to extract a vein like curve even though it only represents 4% of the data. The source resembling a vein is too high in the last time points, and the binding like source is a little low in the last 6 frames. In the binding source the activity is not zero in at Time=2, as in the NMF results, although the value is a little too low.

With 3 ICA sources the same plot is made in figure 11.3. Now the first source looks like the simulated binding but with an end level too high, the second source matches the simulated non-binding completely, and the third source is vascular like where the end level is too high, as with 2 sources. It is clearly seen that the levels in the latter parts of the sources are more erroneous than when using 2 sources. The peak in the vascular source is higher than before, and the values in the other sources at Time=2 are still not zero, as in the NMF results.

When using 4 sources to describe the data, the 4^{th} source describes a signal that is not present in the original signal. The method overfits the data. It can therefore be concluded that the ICA can not extract all 4 sources of the

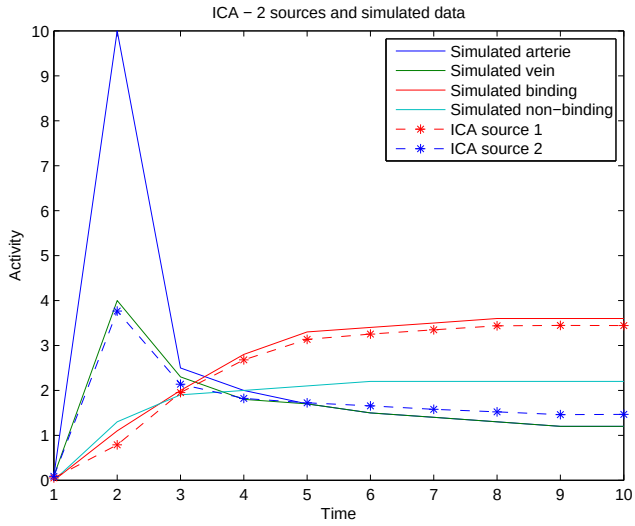


Figure 11.2: ICA with 2 sources and the simulated curves.

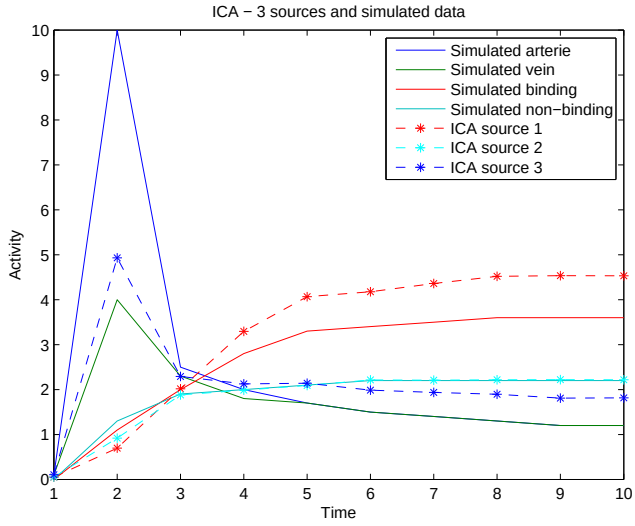


Figure 11.3: ICA with 3 sources and the simulated curves.

original data. Only the vein signal of the two vascular sources can be described by an ICA source. The best results are found using 2 sources, here the vein and the binding source are extracted and fit the original signal very well, after being rescaled with the α factor. This indicates that the ICA underestimated the number of original sources. Moreover, when extracting the vascular signal the best fit is found when using 2 sources, although the peak is lower when using 2 instead of 3 sources. It can be concluded that only the vein signal can be extracted as a source. The arterial signal has to be represented in more observations or the parts has to be larger in the observations, to be extracted.

11.2 ICA on PET data

Although the extraction of the arterial signal in the simulated data was not possible, the ICA method is used to deconvolute the PET data in order to see how it performs, and to compare the results to the NMF solution.

11.2.1 Loglikelihood function $\mathcal{L}$ and no. of sources

The number of components in the ICA is estimated by evaluating the log likelihood as a function $\mathcal{L}(K)$ of number of sources, K.

The $\mathcal{L}$ shown in figure 11.4 shows as expected that the log likelihood increases with increasing number of sources. It also shows that the slope decreases with number of sources. A great increase is seen from 1 to 2, after that the slope becomes lower. Therefore, it is hard to argue directly from $\mathcal{L}(K)$ that a specified number of sources should be used, except for K=2 maybe.

Bayes information criterion

Bayes information criterion (BIC) can be used to choose the right number of sources in a statistical model where $\mathcal{L}(K)$ can be estimated. However, when estimating BIC for all Pilots the minimum that should indicate the statistical right number of sources does not appear. In this case though the voxels are not independent and therefore the number of independent samples must be estimated, as described in section 6.3 on page 40. The estimated number of independent samples are found by evaluating the reconstruction performed when making the PET scan. The point spread function is 6mm×6mm×8.5mm wide,

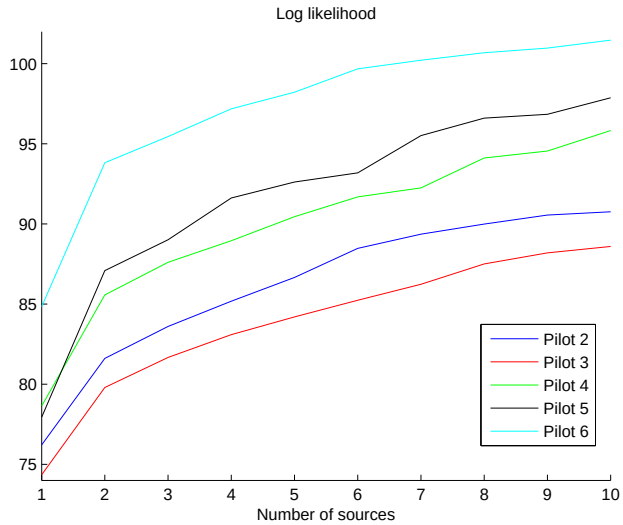


Figure 11.4: Log likelihood function for all Pilots for different no. of sources.

and since the resolution in the PET scan is 2mm×2mm×4.25mm, this gives that the number of samples must be divided by $3 \times 3 \times 2 = 18$.

The BIC is then calculated for a different number of sources with this correction, to see if this criterion can be used to find a reasonable number of sources to use in the ICA model. It is seen in figure 11.5 that although the number of independent samples are corrected in the BIC calculation the criterion is not able to give information about how many sources to use. An explanation for this could be that the model is trying to describe data that has approximately 80000 samples that are very noisy and therefore it statistically needs more than 10 sources. But when using e.g. 8 sources they already look very noisy and unnatural as seen in figure 11.6 and when looking at the spatial information of these sources in the mixing matrix they are randomly distributed. This indicates that the number of sources should be smaller. With all these indications at hand the number of sources must be found by visual inspection of the dynamic sources and their respective spatial distribution.

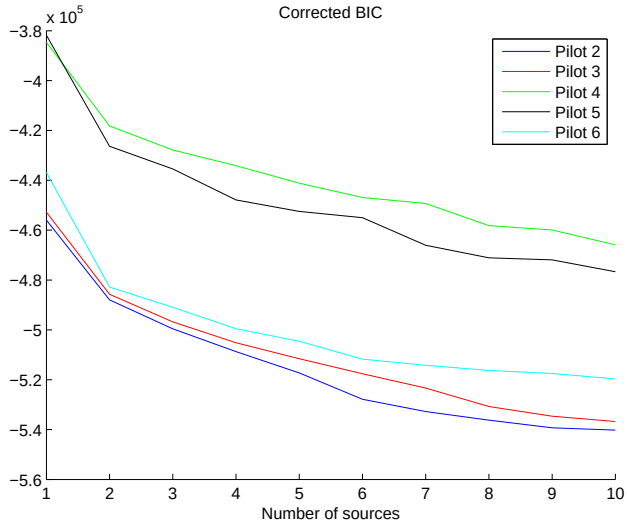


Figure 11.5: Corrected BIC for all Pilots

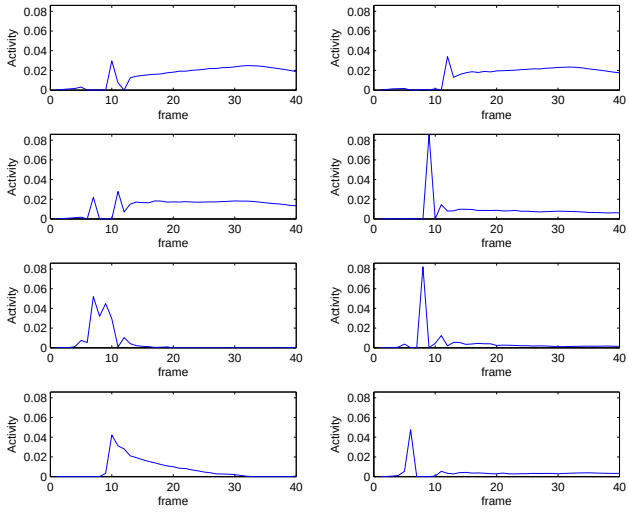


Figure 11.6: ICA with 8 sources for Pilot 3.

11.2.2 Visual inspection of no. of sources

When doing a visual inspection to determine the optimal number of sources one must look for dynamic sources that are as smooth and natural as possible. The spatial distribution found in the mixing matrix must be different for different sources, since the sources should describe different regions. By using this technique the optimal number of sources for data can be found. In figure 11.7 the dynamic sources using 2 sources are shown. The 2 sources represent the binding TAC and the vascular TAC. Using the same method with 3 sources, the TACs are as shown in figure 11.8. The 3 sources seem to be binding, non-binding and vascular TACs.

A vascular TAC is extracted with 2 and 3 sources. The difference between them is mainly the height of the peak and the level of the curve after the peak. The peak is higher using 3 sources, however the level in the latter part of the curve is very close to zero.

Using 4 sources in the ICA, the vascular part is divided into 2 sources. The spatial image of the two vascular sources is generated where the sources are located. The spatial images in figure 11.9 confirm that the 2 vascular sources are 1 true source that has been subdivided, and not a artery or vein source. This shows that the ICA is able to extract 3 independent sources from the PET data.

This is the case for all Pilots except for Pilot 6. When using 4 sources on Pilot 6 the method does not subdivide the vascular source in 2, instead it finds two 2 sources with different binding levels as seen in figure 11.10. When looking at the spatial information the 2 sources are distributed differently in the brain indicating that the sources come from different regions with different binding potentials.

From this analysis the optimal number of sources is found. It is 3 sources for Pilots 2, 3, 4, and 5. For Pilot 6 the optimal number of sources is 4. All the ICA estimated TACs can be found in Appendix D.

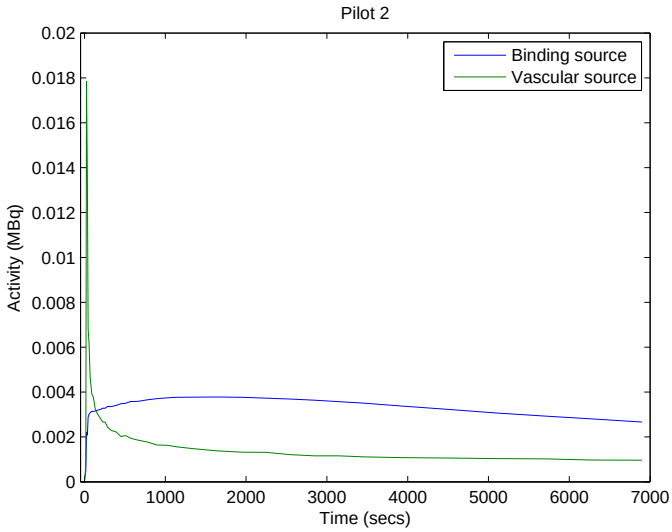


Figure 11.7: TACs from Pilot 2, ICA with 2 sources.

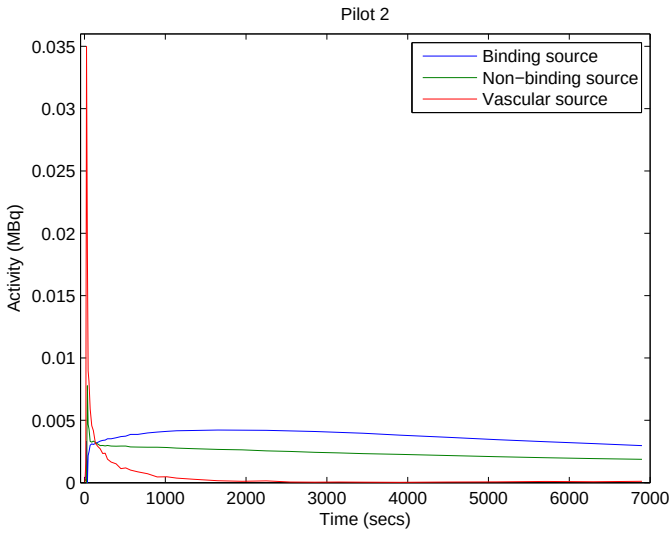
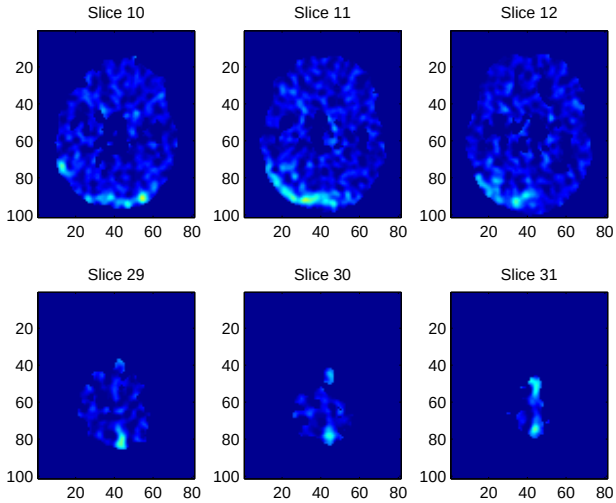
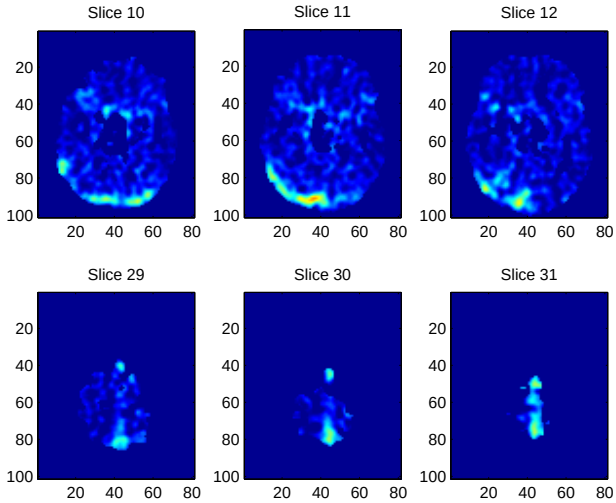


Figure 11.8: TACs from Pilot 2 with zoom, ICA with 3 sources.



(a) Source 3 - Vascular



(b) Source 4 - Vascular

Figure 11.9: The 2 sources subdivided from the true vascular.

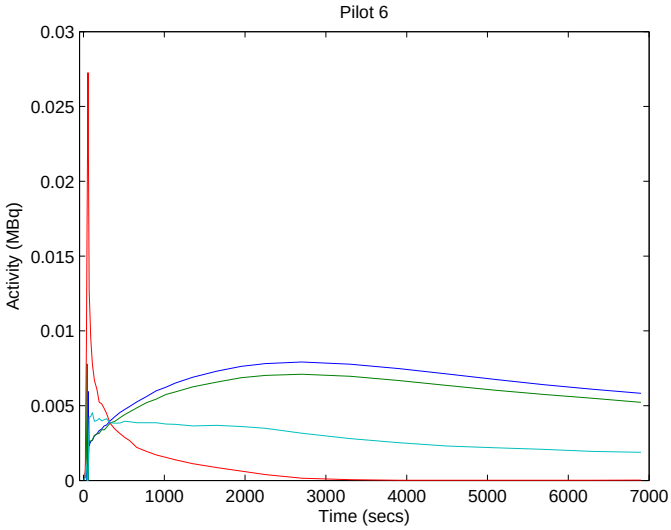


Figure 11.10: TACs from Pilot 6 with, ICA with 4 sources.

11.2.3 Correlation of the TAC curves

To measure how well the sources describe the true region TACs and the sampled arterial TACs the correlations are calculated. In table 11.1 the correlations between the sampled arterial TACs and the ICA vascular TACs are shown. All values are very high, the lowest being 0.9299. The different Pilots have the highest correlation using either 2 or 3 sources. This is probably a tradeoff between the peak height and the end level of the vascular source.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
K = 2	0.9858	0.9754	0.9855	0.9469	0.9563
K = 3	0.9941	0.9592	0.9620	0.9517	0.9554
K = 4	0.9519	0.9559	0.9585	0.9299	0.9519

Table 11.1: Correlations between estimated and arterial sampled blood curves.

However, the non-vascular TACs need validation too. This is done by correlating the binding and non-binding sources with different ROI curves, seen in figure 2.3 on page 14. The correlations between binding ROI and the binding TAC from the ICA solution are displayed in table 11.2, and correlations between the non-binding ROI and the non-binding TACs are displayed in figure 11.3. The

correlations for 2 sources is not calculated for binding and non-binding since there are no specified sources for these, using only 2 sources.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
K = 3	0.7832	0.9001	0.8130	0.8697	0.9677
K = 4	0.8214	0.8891	0.8321	0.8146	0.9519

Table 11.2: Correlations between binding TAC and frontal cortex ROI curve.

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
K = 3	0.8719	0.8301	0.8166	0.7595	0.6320
K = 4	0.8883	0.8067	0.7831	0.8542	0.9461

Table 11.3: Correlations between non-binding TAC and cerebellum ROI curve.

It seems to be a good compromise to use 3 sources in the ICA method, thereby getting a fair segmentation of components that look like true sources of vascular, binding and non-binding components. However, as the visual inspection indicated Pilot 6 is different, this is also seen in table 11.3 where the absolut lowest correlation is seen namely for Pilot 6 with 3 sources. Here the correlation is 0.6320 again indicating that 4 sources gives better results. However, the vascular component is the most critical, since it is this TAC that should be estimated for use in parameter estimations in analysis of binding potentials.

The vascular TAC extracted from the PET scans have high correlations with the sampled arterial TACs, but there is one crucial issue; the level of the TAC at the end of the curves. In the estimated TACs the end level is very close to zero, which makes it hard to do re-scaling of the TACs into the original activity scale. This is necessary because the scale of the PET scans are lost when performing a deconvolution of the data, as mentioned in Chapter 8. Since the end levels are so close to zero, a re-scaling using the information from the vein sample becomes impossible.

11.2.4 Parts of total brain volume

To analyze the part of the components of the total volume of the brain the parts in table 11.4 are shown.

It can be seen that the binding is close to 60% for some Pilots, which is the part of the brain that is white matter. The Non-binding is between 27% and 39% and that is not too far from the 35% of grey matter in the brain. The vascular

Subject	Pilot 2	Pilot 3	Pilot 4	Pilot 5	Pilot 6
Binding	0.5588	0.4887	0.6278	0.4761	0.6063
Non-binding	0.3382	0.3530	0.2763	0.3997	0.2779
Vascular	0.1030	0.1583	0.0959	0.1242	0.1158

Table 11.4: The parts of the brain volume for each Pilot, with K=3, except for Pilot 6, K=4.

part however is much too high, between 10% and 16%, and only 5% of the brain is vascular.

11.2.5 Validation of scaling

To see if the α factor scales the TACs correctly, the ROI TAC are compared with the sources from the ICA. In figure 11.11 the ICA binding and non-binding TACs are shown with the ROI TACs which they should resemble. It is clearly seen that the non-binding TAC is not scaled according to the non-binding ROI. The binding TAC however is close to the level of the binding ROI TAC.

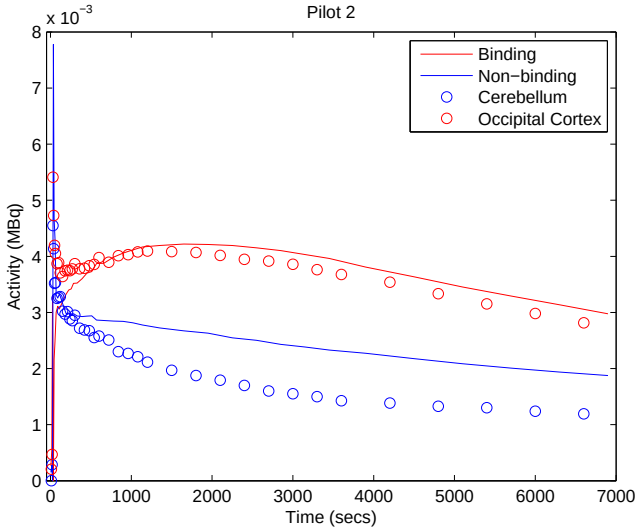


Figure 11.11: ROI curves compared with ICA alpha scaled TACs.

In figure 11.12 the vascular TAC for Pilot 2 from the ICA is shown. The peak

of the curve is not quite high enough and the level after the peak is much too low.

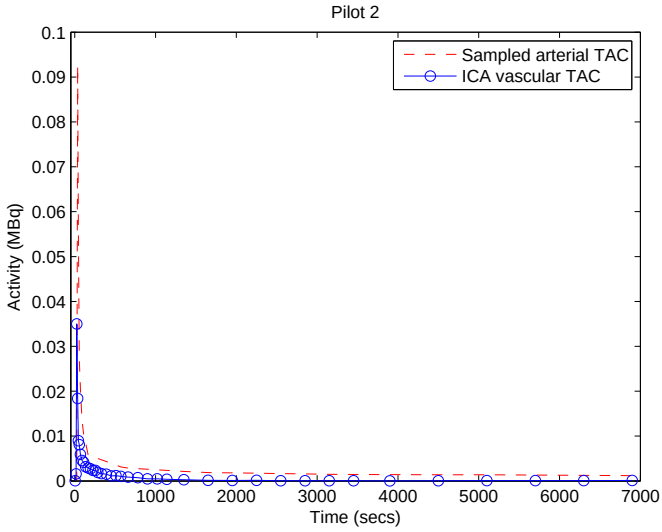


Figure 11.12: Arterial samples compared with ICA alpha scaled TAC.

In general there is a problem with the exact level of the components, as the vascular TAC is too low and the non-binding TAC is too low. The α scaling can therefore be working correctly, but since the ICA solution is not exact and the levels are a bit off the true levels, the re-scaled TACs have some errors due to the uncertainty of the ICA solution.

11.3 Concluding on ICA

The log likelihood is calculated when using the independent component analysis. With the log likelihood it is possible to do a statistical comparison on different models, which should give a measure on how many sources to use. However, when using the statistical measure on the PET data the statistical optimal number of sources is not found. This is due to the high number of samples that are correlated. By using visual inspection the optimal number of sources was found to be 3 for Pilot 2-5, and 4 for Pilot 6. The vascular sources have very high correlations with the arterial TAC, the lowest being 0.95 when using the optimal number of sources. The sources are scaled to be used as arterial TAC

estimations by using the α scaling method. Because of the low end level of the vascular sources a scaling using the vein TAC is not possible.

The end level of the estimated vascular TAC is too low for all Pilots, and the peak is not high enough. The results are very similar to those found using NMF, however the peak in the vascular TACs in the NMF is higher than for ICA.

The ICA performs better on the simulated data than the NMF, but on real PET data the two result are very similar. This is probably because the samples in the simulated data are independent of each other, which the ICA assumes. This is not the case with the real PET data, here the voxels are dependent on each other.

Part III

Discussion and conclusion

Chapter 12

Discussion

In this chapter the results from the various methods applied to simulated and real PET data sets are compared and discussed.

12.1 Comparison of results

Simulated data

The results from the simulated data show that the clustering methods have some limitations because of the partial volume effect, and the PCA is not able to give better results. The most appropriate methods are NMF and ICA. The NMF extracts two of the four original sources and the ICA extracts three sources in the simulated data. This is evaluated by scaling with the α factor and comparing the results with the original TACs. In the NMF the scaling was wrong if more than two sources were estimated. The ICA shows better results, as the values for the non-vascular components are not zero during the peak of the arterial signal, as in the NMF results. However, the ICA is unstable compared to the NMF method, and needs several repetitions to find the optimal solution. For both ICA and NMF the α scaling becomes incorrect when overfitting using too many sources.

PET data

Four methods have been applied to the PET data set to extract arterial information. The most simple one, the K-means method, was able to extract venous information. But since the maximum intensity in the PET data is much lower than the peak of the arterial curve, the K-means method will never be successful at extracting the peak of the artery TAC. Although this clustering method is not able to find arterial regions, because of the partial volume effect, the method can be used to extract features for larger regions in the brain. E.g the venous region is extracted very nicely, and the spatial image of the cluster looks good, since the regions of the clusters are well defined, and match the true underlying characteristics of the regions of the brain. The fussy C-means method is not able to solve the limitations in the K-means method, since the values in the clusters can never be higher than values in the data. The fussy C-means has the disadvantage that the cluster sizes becomes equally sized the more fussy the algorithm clusters, and was therefore not applied to the PET data.

The PCA is able to extract venous information as a principal component, but not arterial information. Binding and venous components are extracted as the most important components. The evaluation of the principal components is difficult, since coefficients and components can be negative. The scaling of the components is also a problem, as the α factor can not be applied to this method. Instead the vein end-level can be used to scale the vascular principal component. The result from this is not better than the K-means result. The PCA is therefore not a good choice for extracting arterial information from PET data. It can be concluded that the 1^{st} and 2^{nd} order statistics are not enough to extract arterial TACs.

The ICA and NMF solutions from the PET data are not very different from each other, but the main differences of the methods are the complexity and the stability of algorithms. The optimal solution using ICA is not found by using the same number of sources for all Pilots, as with the NMF. A characteristic both models have is that there is no constraint on the shape of the sources, to make the TACs smoother and more natural. Especially in the beginning of the TACs during the peak in the arterial blood curve, the non-vascular components behave strangely by having either peaks, or by going to zero. This is an artifact because the ICA and the NMF do not make the exact factorization or simply because of the high amount of noise in the initial frames. Both methods have difficulties extracting the correct level of the TACs, this is probably because the slope of all the curves is close to zero in the latter part. Or because of overfitting, simply estimating too many sources. Overfitting will make the scaling incorrect, which is shown in the analysis of the simulated data.

The NMF and ICA methods, are able to extract sources despite of no observations where the whole source is present. Therefore, they could be able to estimate the arterial regions although they are very small compared to the resolution of the PET data. The correlation between the arterial sampled TAC and the estimated TAC is very high for the NMF and ICA results, approximately 0.95. Since the correlations are very close to one, the extracted TAC have the same characteristics as the sampled artery TAC. When evaluating these correlations the NMF and ICA perform equally well.

12.2 Scaling challenges

When performing a deconvolution another challenge appears since the scale of the spatial and dynamic factors is lost. Although the correlations are very high, which indicates that the estimated TACs are good, the estimated arterial TAC must be scaled to fit the scale of the sampled TAC.

The α factor is used to solve this problem, thereby the dynamic components can be compared directly to the arterial sampled TAC. The α factor scaling will only make sense if the model describes the data well. And if there is not any region in the brain that is very badly explained by the model, since this would influence the scaling parameter for the sources. A way to avoid the influence from the other sources could be to only scale the estimated vascular TAC. This can be done by using the end level of the sampled vein TAC. Since the level of the vein and the artery are very close to each other in the last frames. However, there is a problem with the results since the estimated TACs have very low end levels. Therefore, the scaling becomes very unstable, since a value close to zero has to be multiplied by a high number to reach the correct level. This will influence the whole curve and can make the peak in the beginning much too high. This makes the vein scaling very unreliable to use on these results.

12.3 Choice of method

The vascular TACs found by the K-means and the PCA have the correct levels in the latter part of the frames, but these methods are not able to extract the peak. The NMF and the ICA extract the peak better, but the end level of the TACs are close too zero.

The deconvolution methods, are clearly better at extracting peaky TACs because of the partial volume effect. Making these methods the best choice when

estimating artery TACs since these regions are small. The NMF is chosen over the ICA method, because of the complexity of the ICA and since the results in the PET data from the ICA are not better than those from the NMF. Also the implementation of the NMF is much simpler, more stable, and requires no parameter tuning.

The results from the NMF are very promising, this is stated by the correlations between the estimated and the sampled TACs.

Chapter 13

Conclusion

The problem of estimating the artery TAC from dynamic PET data has been investigated using several different methods.

For estimating the end level of the arterial TAC, which is approximately the same as the level of the venous TAC, the K-means can be used. The PCA also finds the level very nicely, but there are issues with the evaluation and scaling of the principal components which makes this method more complicated to work with. However, these methods are not able to estimate the peak of the artery TAC.

The scaling issue with the deconvolution methods is solved by using the α factor to scale the estimated TACs, and the methods can be used on the entire PET data without defining regions of interest or removing some of the last frames.

The most appropriate method to extract the arterial TAC is the NMF. The NMF is chosen because of its simplicity compared to the ICA. The results from this method show that the peak is extracted for both simulated and real data. The correlation coefficients between the estimated artery TACs and the sampled TACs are around 0.95 for all Pilots. This is a very high correlation considering that they are obtained with medical data. These coefficients confirm that the results are valid.

Part IV

Appendix

Appendix A

K-means results

A.1 K-means clusters

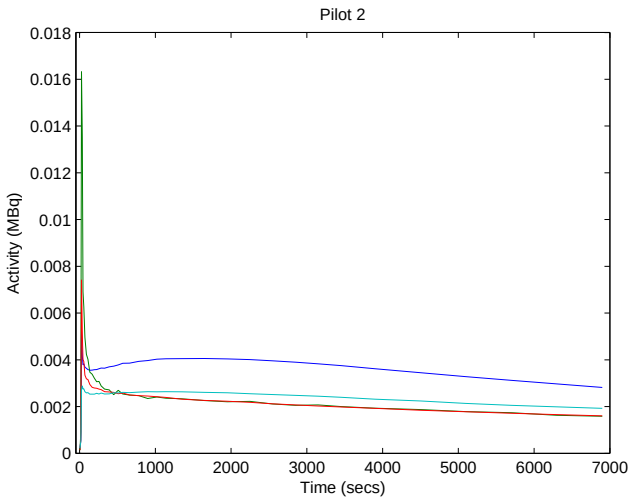


Figure A.1: Pilot 2, K=4.

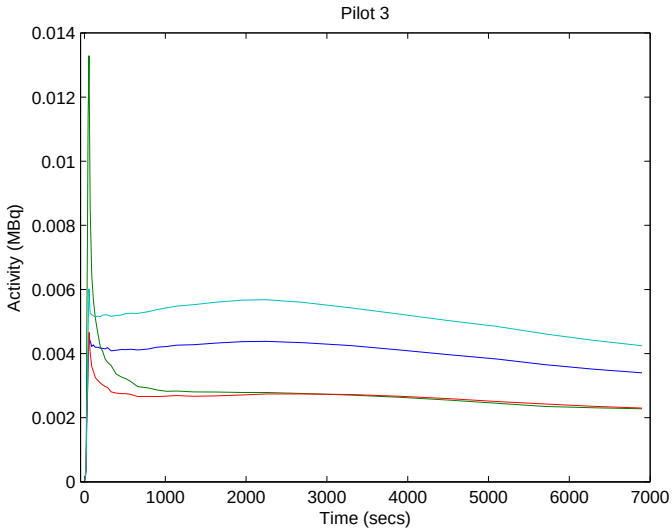


Figure A.2: Pilot 3, K=4.

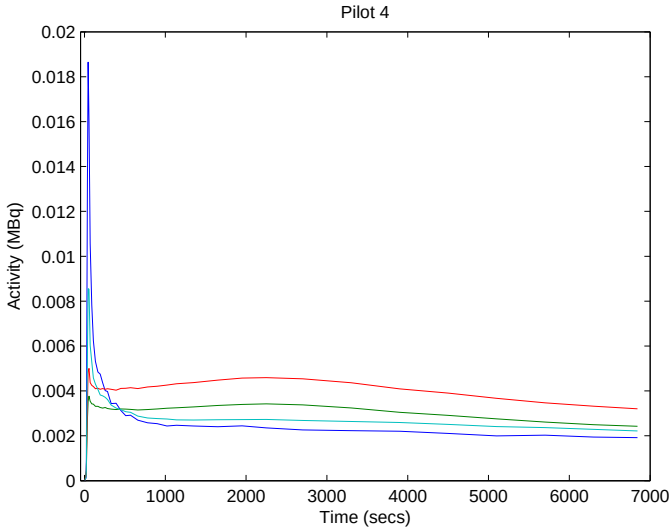


Figure A.3: Pilot 4, K=4.

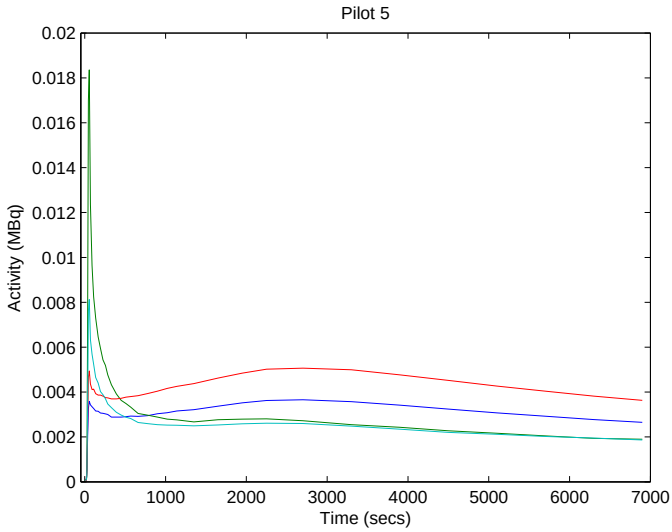


Figure A.4: Pilot 5, K=4.

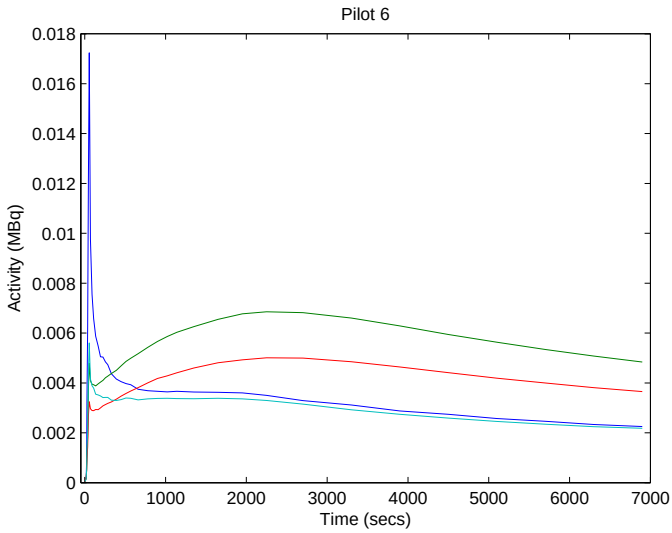


Figure A.5: Pilot 6, K=4.

A.2 Comparison of TACs

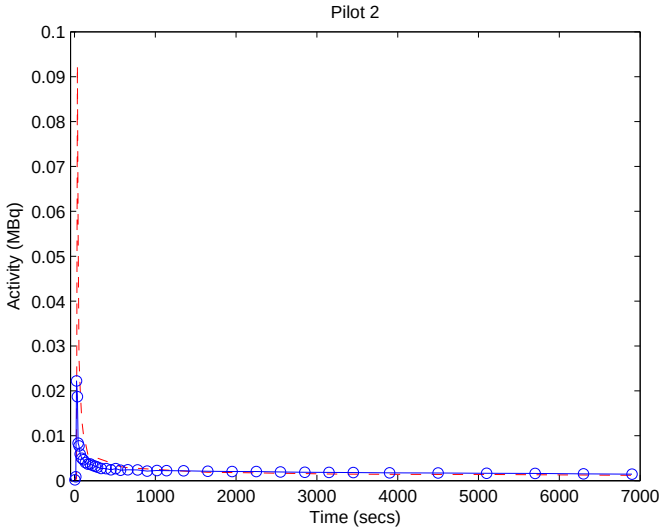


Figure A.6: Pilot 2, K=4 and K2=4.

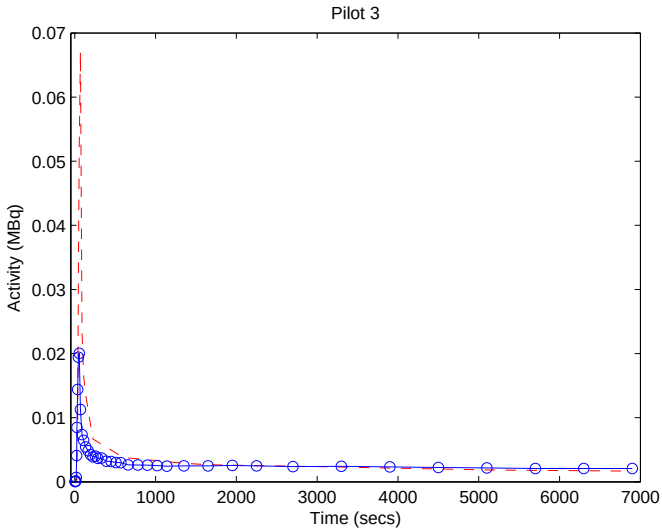


Figure A.7: Pilot 3, K=4 and K2=4.

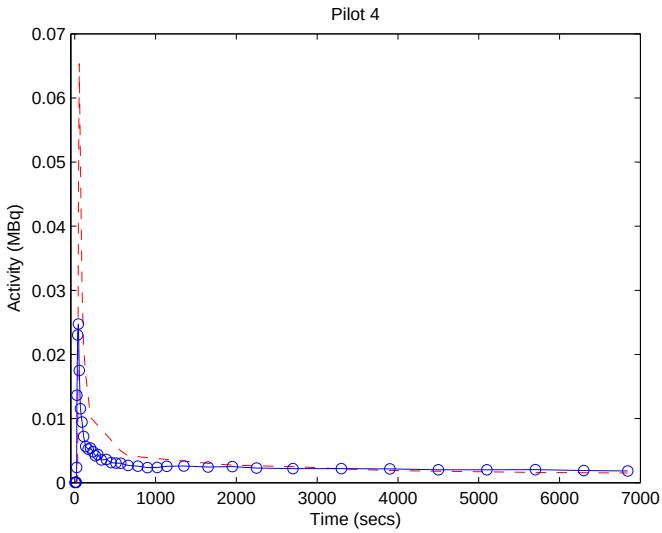


Figure A.8: Pilot 4, K=4 and K2=4.

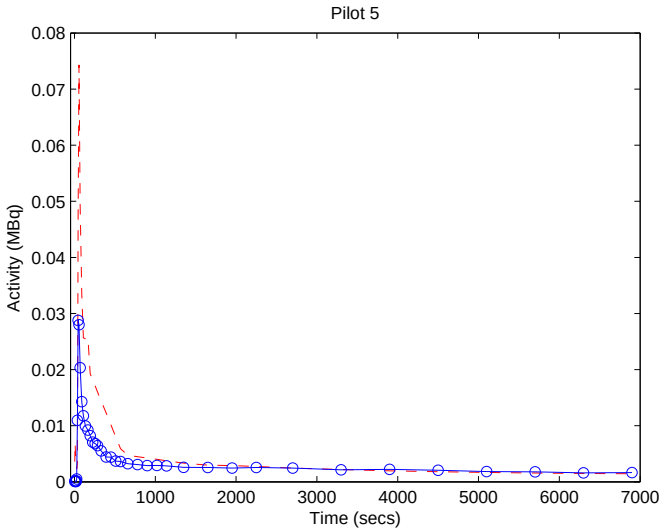


Figure A.9: Pilot 5, K=4 and K2=4.

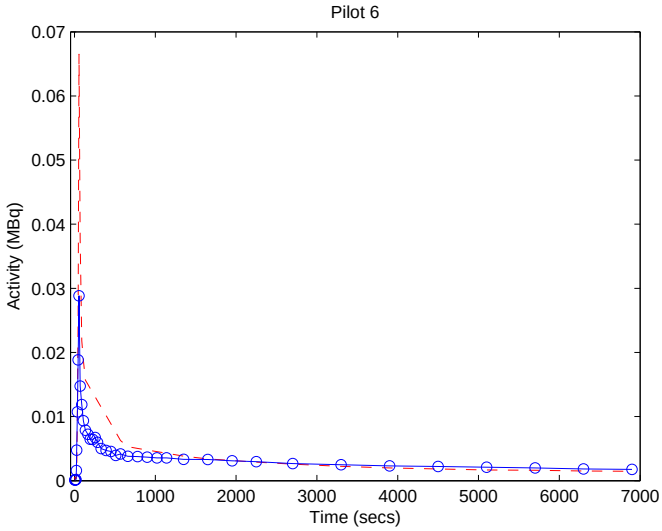


Figure A.10: Pilot 6, K=4 and K2=4.

Appendix B

PCA results

B.1 Principal components

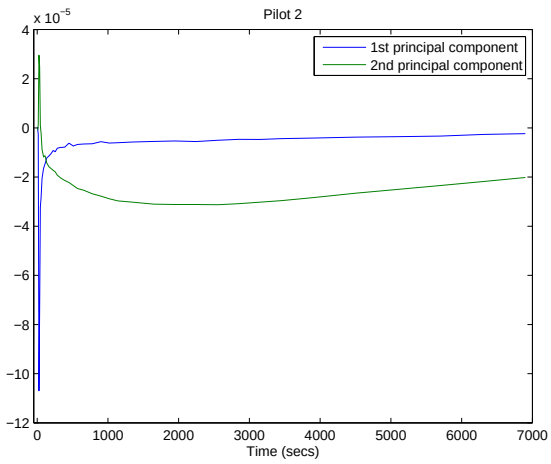


Figure B.1: Pilot 2, first two principal components.

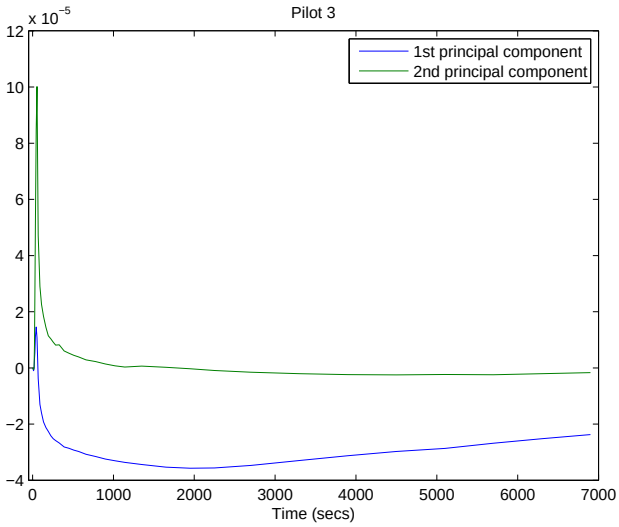


Figure B.2: Pilot 3, first two principal components.

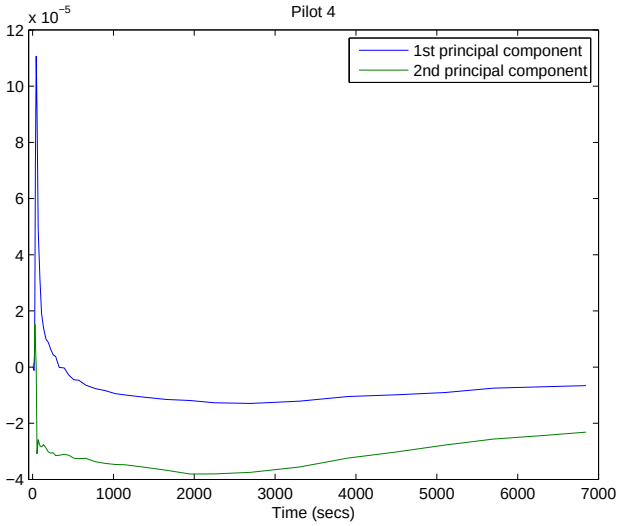


Figure B.3: Pilot 4, first two principal components.

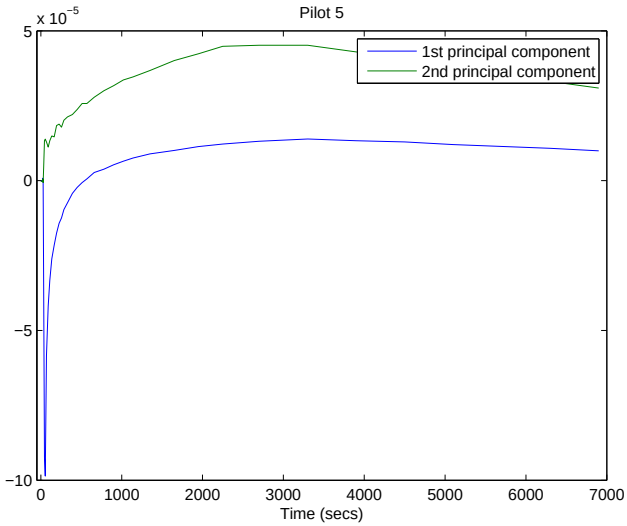


Figure B.4: Pilot 5, first two principal components.

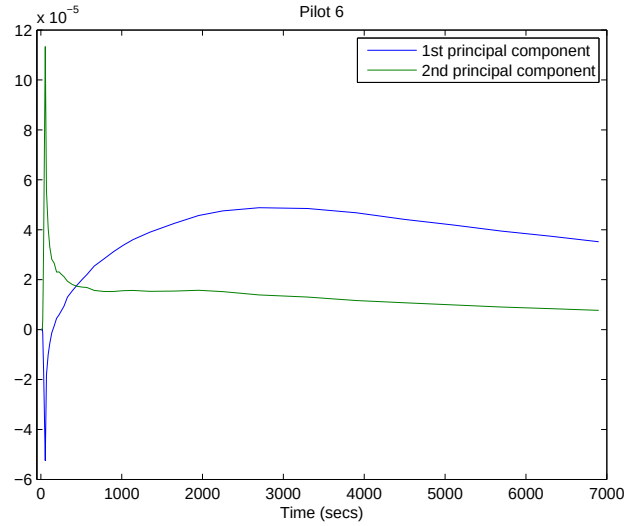


Figure B.5: Pilot 6, first two principal components.

B.2 Comparison of TACs

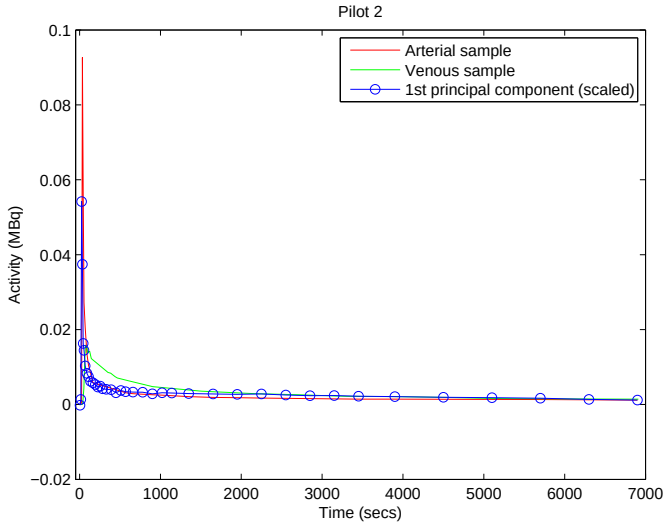


Figure B.6: Pilot 2, comparison of scaled principal component and artery TAC.

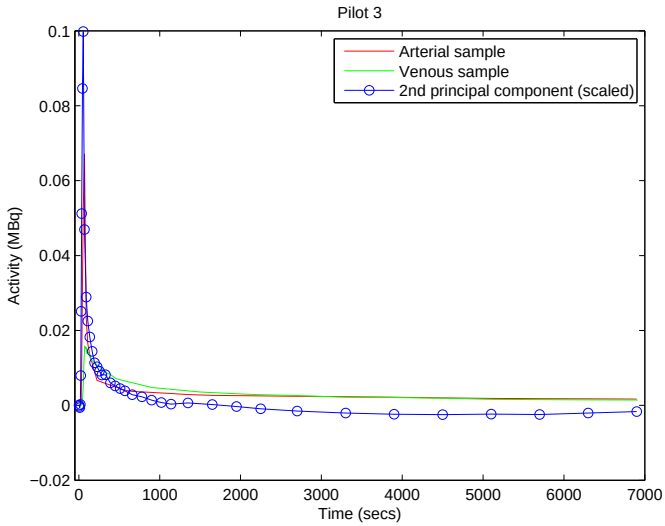


Figure B.7: Pilot 3, comparison of scaled principal component and artery TAC.

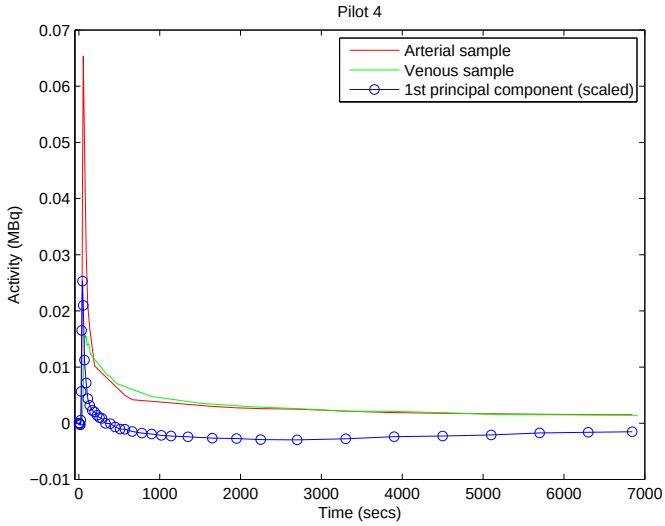


Figure B.8: Pilot 4, comparison of scaled principal component and artery TAC.

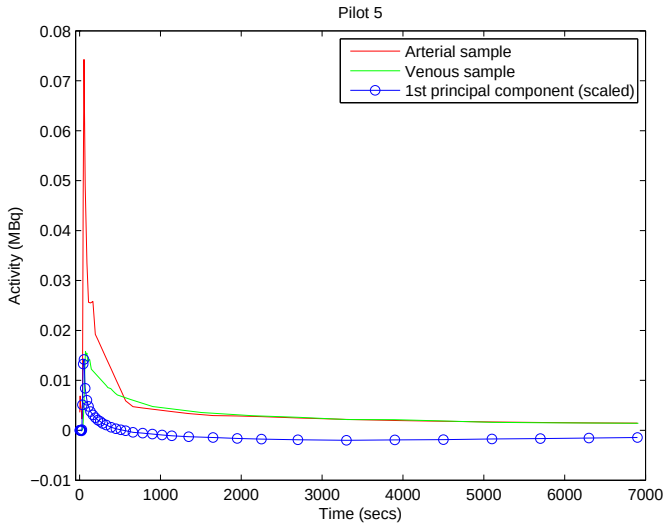


Figure B.9: Pilot 5, comparison of scaled principal component and artery TAC.

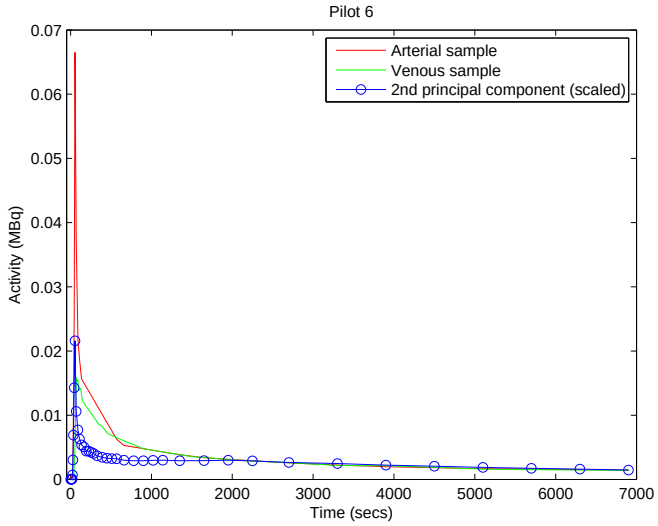


Figure B.10: Pilot 6, comparison of scaled principal component and artery TAC.

Appendix C

NMF results

C.1 NMF sources

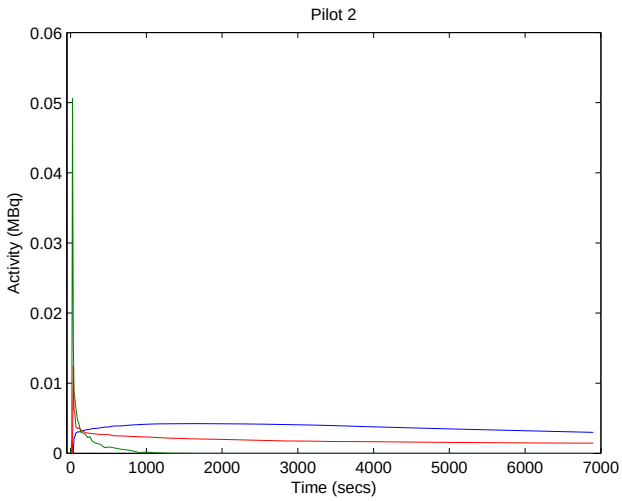


Figure C.1: Pilot 2, K=3.

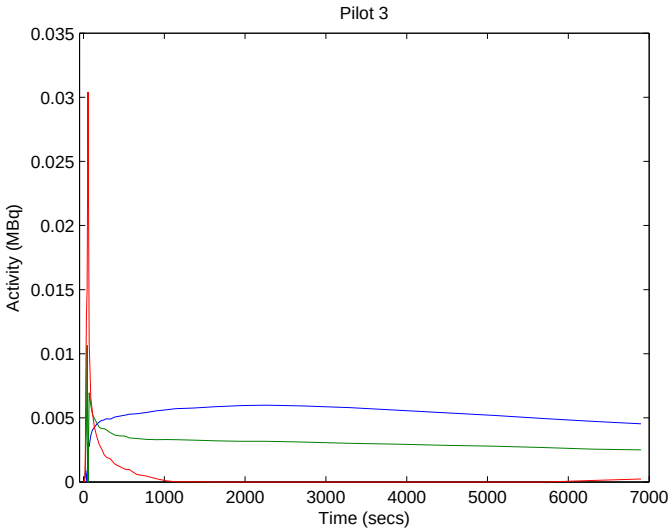


Figure C.2: Pilot 3, K=3.

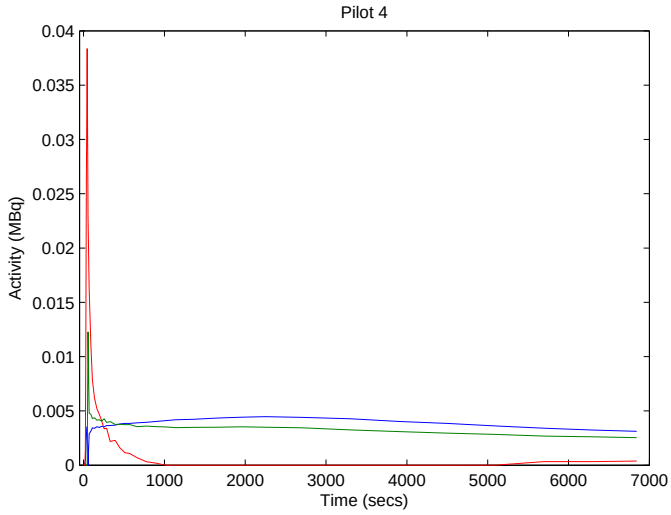


Figure C.3: Pilot 4, K=3.

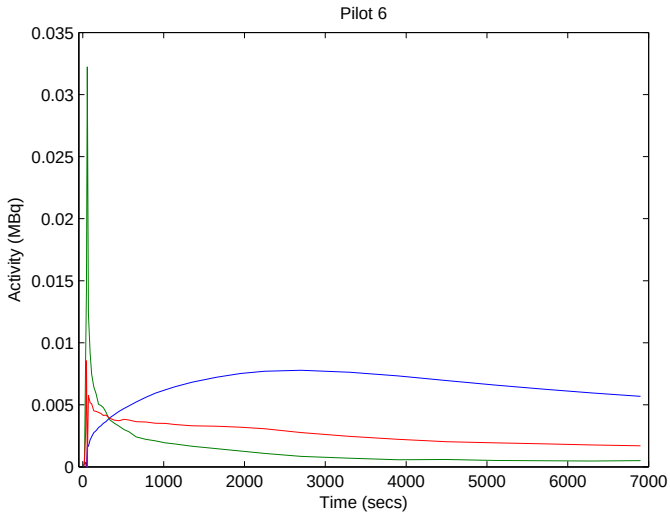


Figure C.4: Pilot 5, K=3.

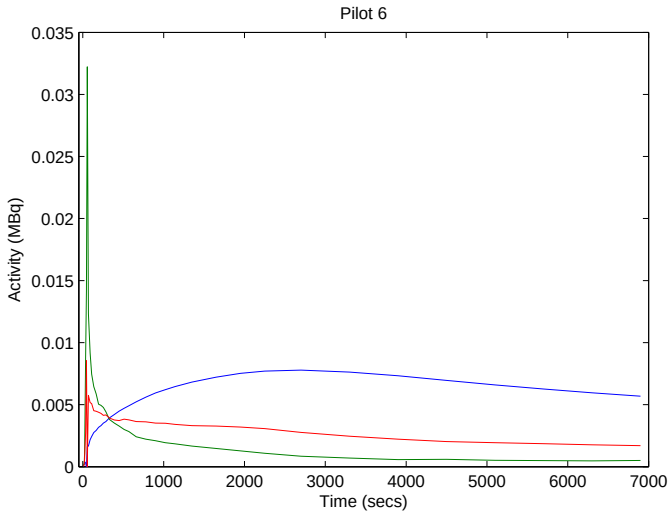


Figure C.5: Pilot 6, K=3.

C.2 Comparison of TACs

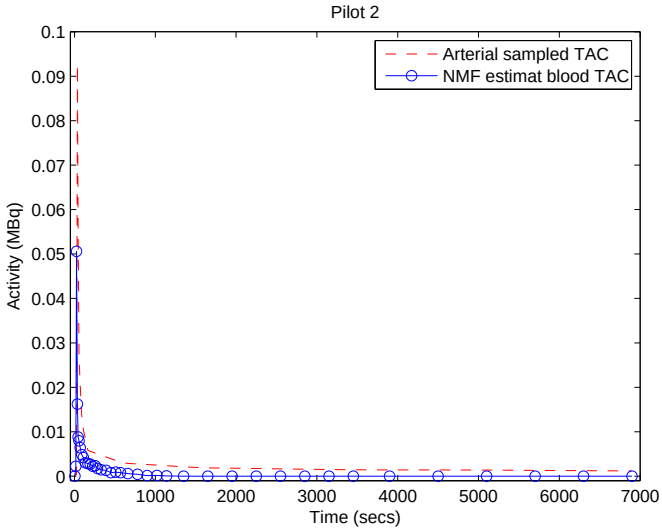


Figure C.6: Comparison of TACs, Pilot 2.

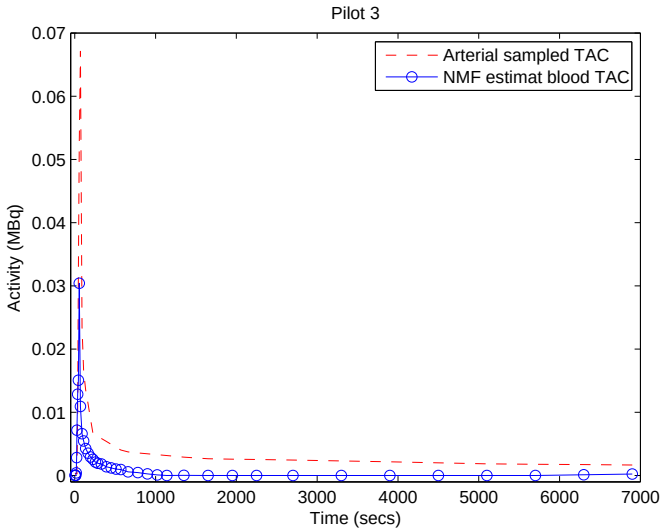


Figure C.7: Comparison of TACs, Pilot 3.

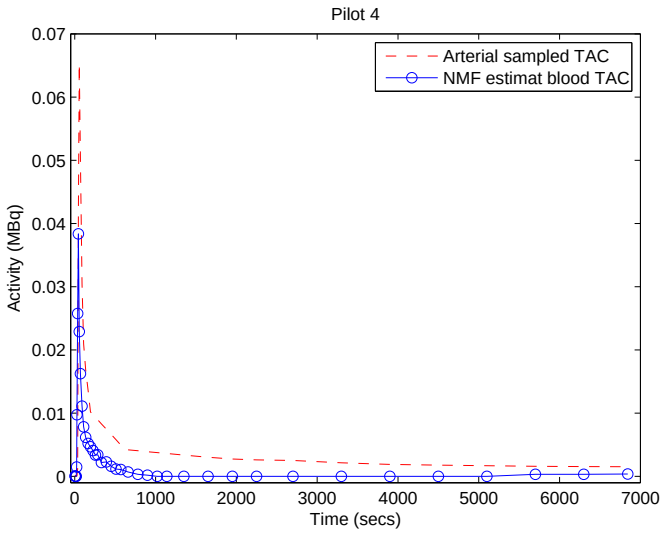


Figure C.8: Comparison of TACs, Pilot 4.

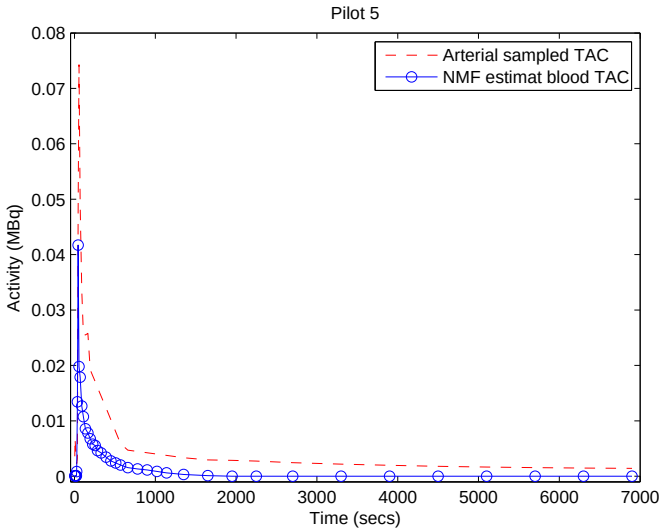


Figure C.9: Comparison of TACs, Pilot 5.

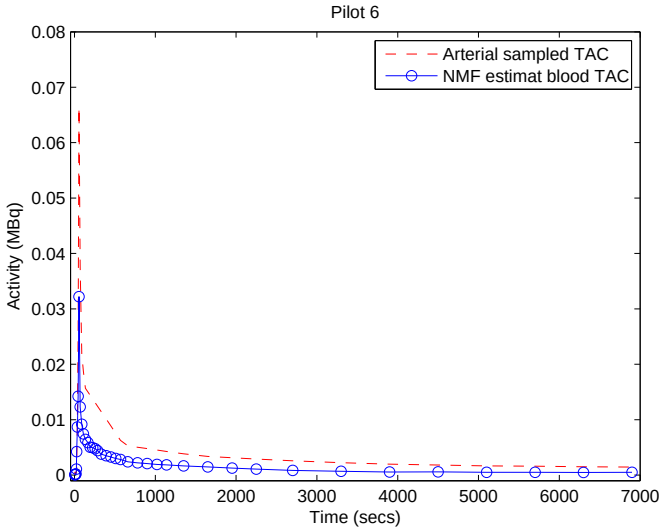


Figure C.10: Comparison of TACs, Pilot 6.

Appendix D

ICA results

D.1 ICA sources

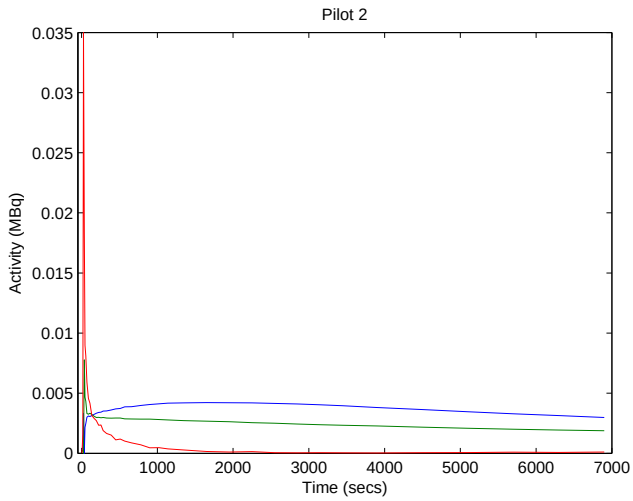


Figure D.1: Pilot 2, K=3.

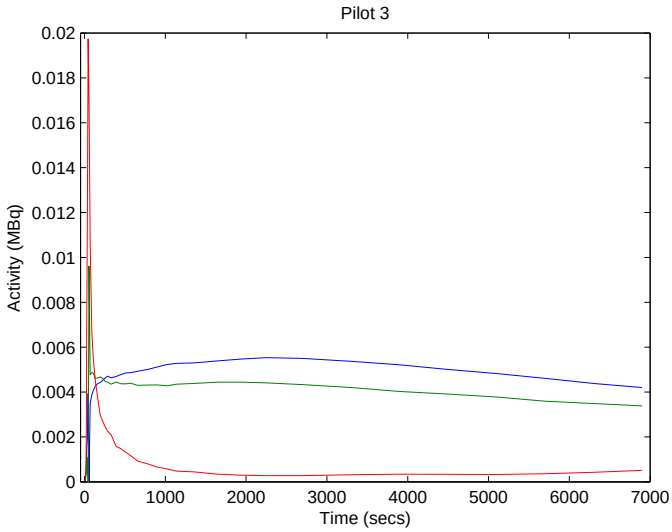


Figure D.2: Pilot 3, K=3.

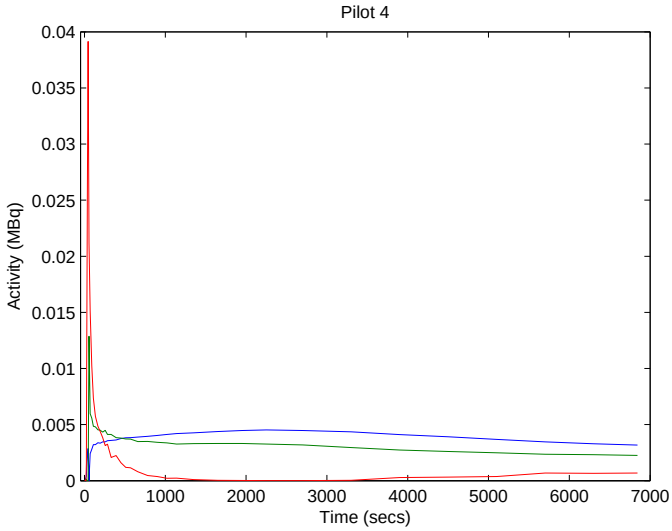


Figure D.3: Pilot 4, K=3.

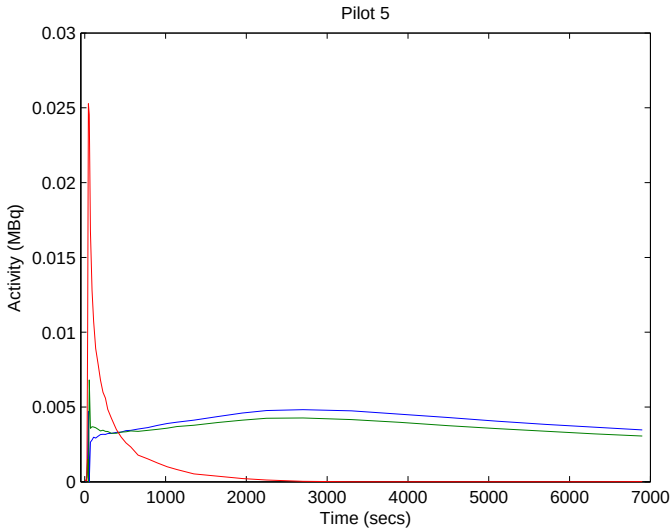


Figure D.4: Pilot 5, K=3.

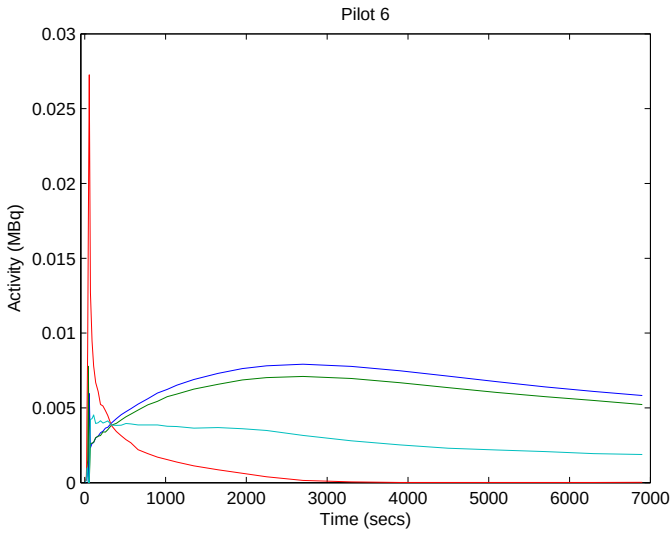


Figure D.5: Pilot 6, K=4.

D.2 Comparison of TACs

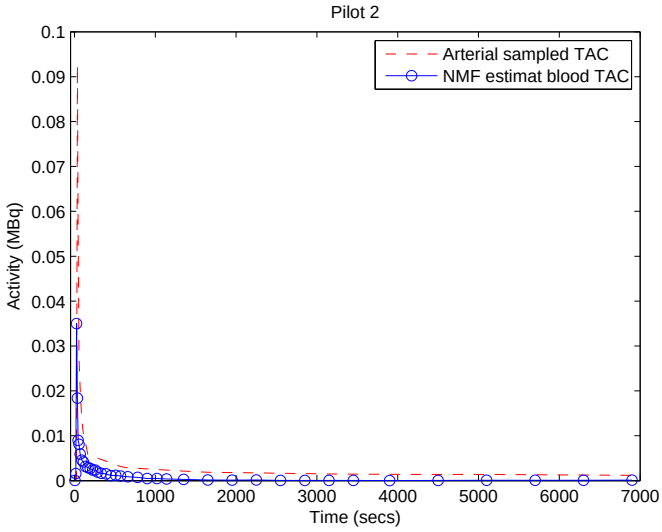


Figure D.6: Comparison of TACs Pilot 2.

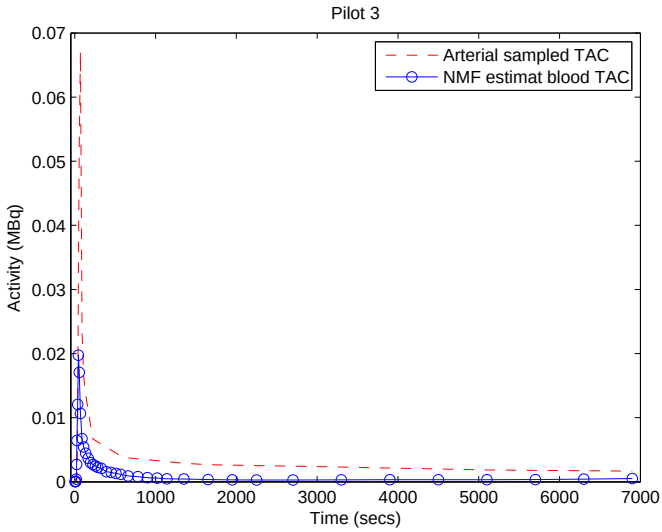


Figure D.7: Comparison of TACs Pilot 3.

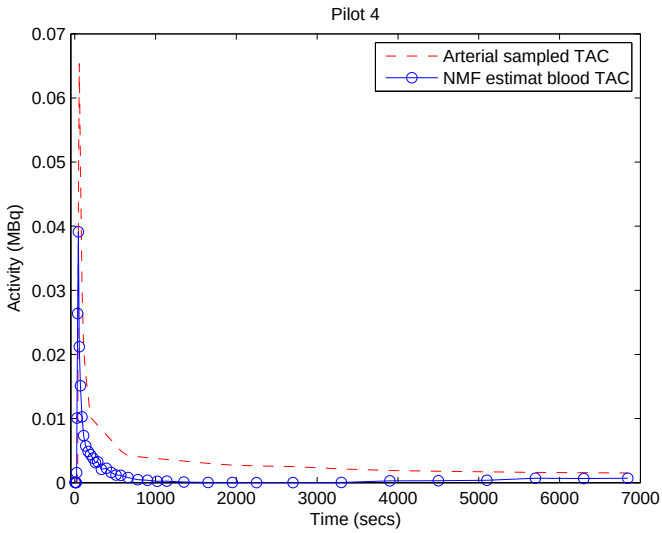


Figure D.8: Comparison of TACs Pilot 4.

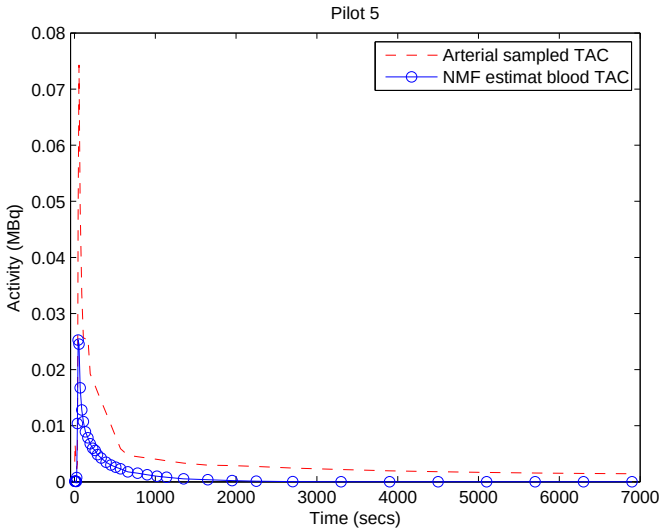


Figure D.9: Comparison of TACs Pilot 5.

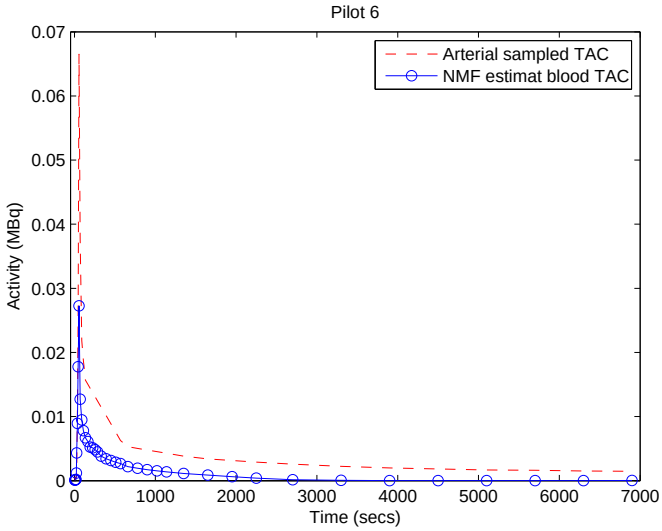


Figure D.10: Comparison of TACs Pilot 6.

Appendix E

EPICA results

E.1 Comparison of TACs

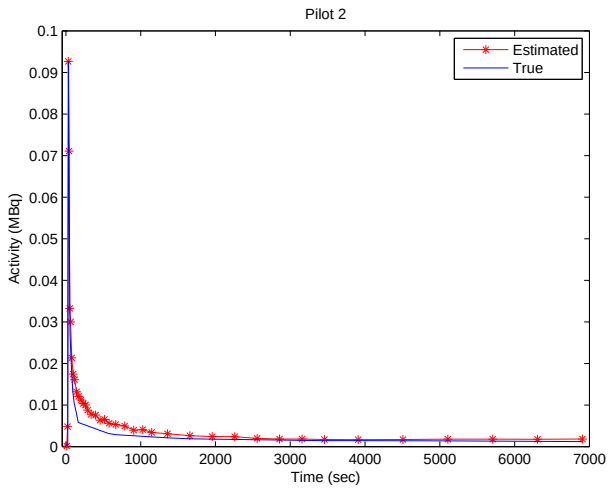


Figure E.1: EPICA - Scaled to peak - Pilot 2, K=2.

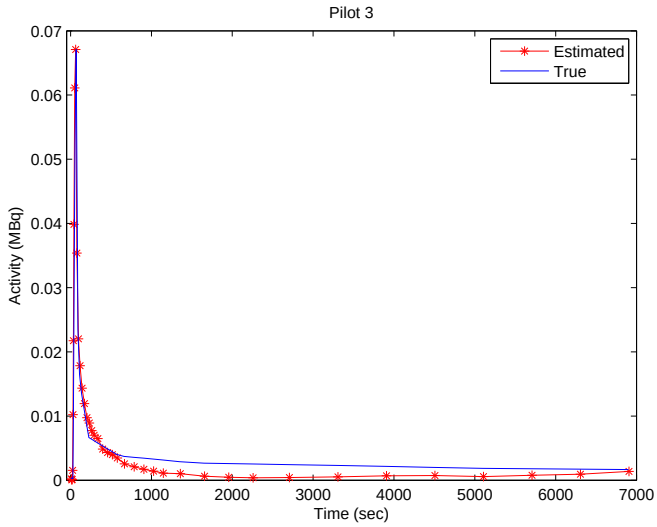


Figure E.2: EPICA - Scaled to peak - Pilot 3, K=2.

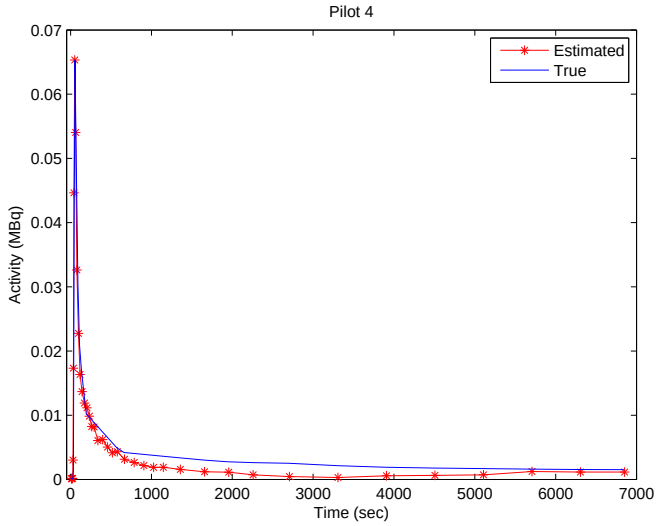


Figure E.3: EPICA - Scaled to peak - Pilot 4, K=2.

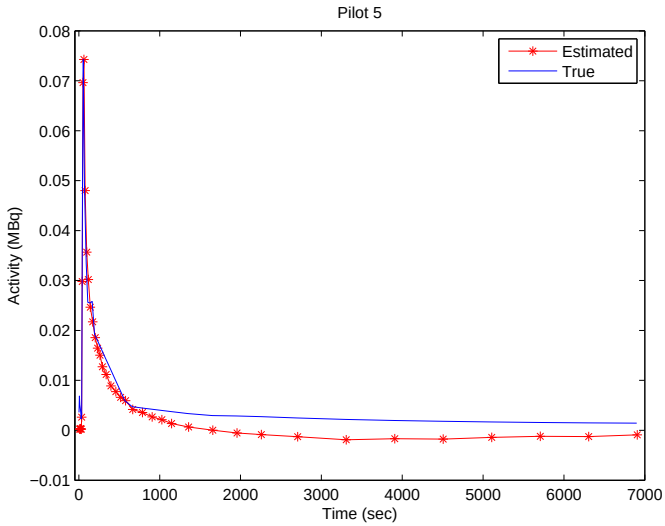


Figure E.4: EPICA - Scaled to peak - Pilot 5, K=2.

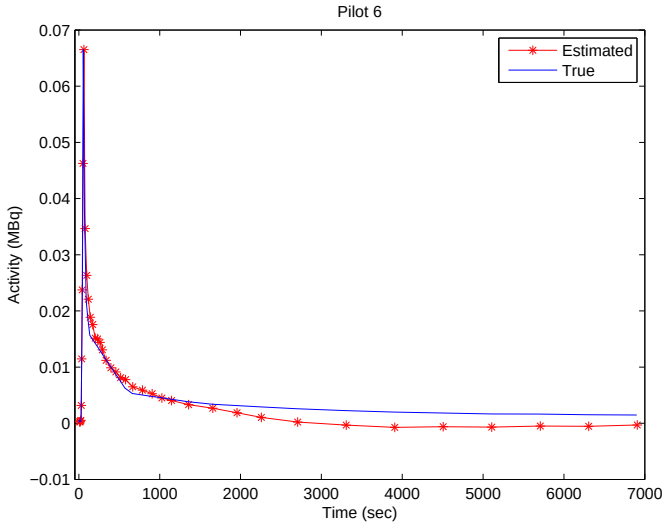


Figure E.5: EPICA - Scaled to peak - Pilot 6, K=2.

Appendix F

Abstract for Human Brain Mapping 2006

Extraction of time activity curves from positron emission tomography: K-means clustering or Non-negative matrix factorization.

Bjarni Bödvarsson[1], Martin Mørkebjerg[1], Lars Kai Hansen[1], Gitte Moos Knudsen[2], Claus Svarer[2]

[1] *Intelligent Signal Processing, IMM, Technical University of Denmark, Kgs. Lyngby, Denmark.* [2] *Neurobiology Research Unit, Copenhagen University Hospital, Copenhagen Denmark.*

Objective

Measurement of the cerebral input function in terms of the arterial plasma input time activity curve (TAC) is required for assessment and validation of the quantification of receptor binding parameters on the basis of dynamic positron

emission tomography (PET) images and kinetic models [1]. Alternative non-invasive methods for estimation of the input function have been described based on cluster analysis [1], based on linear decomposition methods such as independent component analysis (ICA) [2, 3], and by non-negative matrix factorization (NMF) of the signals in a region of interest [4]. The linear decomposition methods are preferred because they do not assume that voxels are dominated by the vascular signal, where the NMF method is attractive because of the relative straightforward estimation procedure. Furthermore, the NMF method is attractive because it - in contrast to the ICA methods - does not assume a specific form of the spatial distribution.

In this study the NMF method is used to extract the input function without the use of a set of predefined regions of interest (ROI) and the results are subsequently compared to the results from K-means clustering used in [1].

Methods

Five healthy subjects were investigated with dynamic PET-scanning after rapid intravenous 18F-altanserin bolus injection. Arterial and venous blood samples were withdrawn automatically and manually and counted in an external coincidence counter.

The non-negative matrix factorization describes the matrix $\mathbf{V}$, as a linear combination of the factors $\mathbf{W}$ and $\mathbf{H}$. All elements in $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{H}$ are non-negative.

$$\mathbf{V}=\mathbf{WH}$$

This can be applied to the problem at hand of extracting TACs; $\mathbf{V}$ being the dynamic PET image, $\mathbf{W}$ is the mixing matrix, and $\mathbf{H}$ contains the basis TACs.

Results

Figure F.1 shows the outcome of the NMF approach in comparison to the measured arterial blood samples. It is clearly seen that the peak in the blood curve as identified by NMF is higher than the one found by clustering. In addition, it matches well with the sampled blood TAC.

The reason for the reduced amplitude of the clustering TAC is that it represents the average of voxel TACs some of which are only partly vascular, while the enhanced signal in the NMF TAC is due to the deconvolution of the signal from

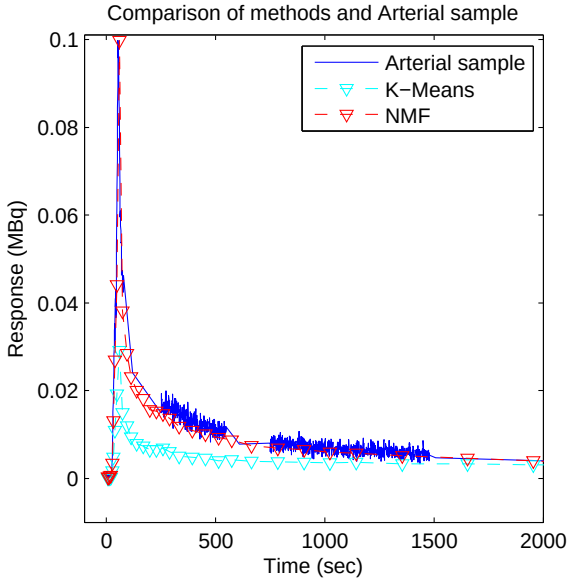
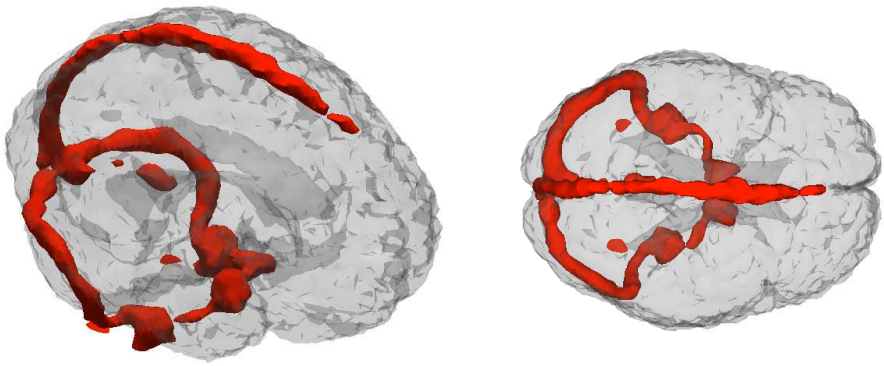


Figure F.1:



(a) (b)

Figure F.2:

each voxel into different components.

To visualize and validate the results the spatial information from the vascular source is used to generate a 3D image (figure F.2) by thresholding the column of the vascular component of the **H** matrix. Figure F.2 demonstrates that the vascular TAC found by the NMF is spatially located corresponding to the larger cerebral veins.

Conclusion

For these data, NMF is excellent for extraction of TACs in PET data without using a predefined ROI for the analysis. For this particular set of data the NMF method is superior to the K-means clustering approach.

References

1.Liptrot, M. et al., 2004, NeuroImage, 21:483-493.
2.Naganawa, M. et al., 2005, NeuroImage, 26:885-890.
3.Ahn, Y. J. et al., 2000, proc. ICA 2000, 493-497.
4.Kim, Kyeong Min et al., 2001, IEEE, 3:1282-1285.

Bibliography

[1] J. Karhunen A. Hyvärinen and E. Oja. *Independent Component Analysis*. Wiley Interscience, 2001.

[2] Ji Young Ahn, Jae Sung Lee, Myoung Jin Jang, and Dong Soo Lee. Non-invasive extraction of input function from carotid artery in $h_2^{15}o$ dynamic brain positron emission tomography using independent component analysis. *Second International Workshop on Independent Component Analysis and Blind Signal Separation. Proceedings*, pages 493–7, 2000.

[3] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms (Advanced Applications in Pattern Recognition)*. Springer, July 1981.

[4] BrainExplorer. http://www.brainexplorer.org/brain_atlas/Brainatlas_index.shtml, [Online; accessed 20-February-2006].

[5] F. Calamante, M. Mørup, and L. K. Hansen. Defining a local arterial input function for perfusion MRI using independent component analysis. *Magnetic Resonance in Medicine*, 52(4):789–797, 2004.

[6] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernatics*, 3:32–57, 1974.

[7] Roger N. Gunn, Adriaan A. Lammertsma, Susan P. Hume, and Vincent J. Cunningham. Parametric imaging of ligand-receptor binding in pet using a simplified reference region model. *NeuroImage*, 6(1):279–287, 1997.

[8] M. Ichise, J.H. Meyer, and Y. Yonekura. An introduction to pet and spect neuroreceptor quantification models. *Journal of Nuclear Medicine*, 42(5):755–63, 2001.

[9] S. Choi D. S. Lee J. S. Lee, D. D. Lee. Application of non-negative matrix factorization to dynamic positron emmision tomography. *Proc. ICA*, pages 629–632, December.

[10] Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell. *Principles of neural science*. McGraw-Hill Companies, 2000. Fourth edition.

[11] Kyeong Min Kim, H Watabe, M Shidahara, Ji Young Ahn, Seungjin Choi, N Kudomi, K Hayashida, Y Miyake, and H Iida. Noninvasive estimation of cerebral blood flow using image-derived carotid input function in $h_2^{15}o$ dynamic pet. volume 3, pages 1282–1285, 2001.

[12] M. Opper L. Csato and O. Winther. Tap gibbs free energy, belief propagation and sparsity. *Advances in Neural Information Processing Systems*, 14, 2002.

[13] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000.

[14] Daniel D. Lee and Sebastian H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.

[15] Matthew Liptrot, Karen H. Adams, Lars Martiny, Lars H. Pinborg, Markus N. Lonsdale, Niels V. Olsen, Soren Holm, Claus Svarer, and Gitte M. Knudsen. Cluster analysis in kinetic modelling of the brain: a noninvasive alternative to arterial sampling. *NeuroImage*, 21(2):483–493, 2004.

[16] J. MacQueen. Some methods for classification and analysis of multivariate observations. volume 1, pages 281–297, 1967.

[17] Mika Naganawa, Y. Kimura, K. Ishii, K. Oda, K. Ishiwata, and A. Matani. Extraction of a plasma time-activity curve from dynamic brain pet images based on independent component analysis. *Biomedical Engineering, IEEE Transactions on*, 52(2):201–210, 2005.

[18] Mika Naganawa, Yuichi Kimura, Tadashi Nariai, Kenji Ishii, Keiichi Oda, Yoshitsugu Manabe, Kunihiro Chihara, and Kiichi Ishiwata. Omission of serial arterial blood sampling in neuroreceptor imaging with independent component analysis. *NeuroImage*, 26(3):885–890, 2005.

[19] M. Opper and O. Winther. Adaptive and self-averaging thouless-anderson-palmer mean field theory for probabilistic modeling. *Phys. Rev. Lett*, 64(056131).

[20] M. Opper and O. Winther. Tractable approximations for probabilistic models: The adaptive thouless-anderson-palmer mean field approach. *Phys. Rev. Lett*, 86:3695–3699, 2001.

[21] K. B. Petersen. *Mean Field ICA*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2005. Supervised by Ole Winther, IMM.

[22] Finn Årup Nielsen and Lars Kai Hansen. Experiences with matlab and vrml in functional neuroimaging visualizations, in visualization development environments (vde2000), princeton plasma physics laboratory, princeton, new jersey, usa, 2000 april 27-28. 2000.

[23] Gustav K. von Schulthess. *Clinical positron emission tomography (PET): correlation with morphological cross-sectional imaging*. Lippincott Williams and Wilkins, 1999.

[24] Wikipedia. http://en.wikipedia.org/w/index.php?title=Positron_emission_tomography\&oldid=38315150, [Online; accessed 6-February-2006].

[25] O. Winther and K. B. Petersen. Flexible and efficient implementations of bayesian independent component analysis. *Preprint submitted to Neurocomputing*, 2005. December.