Cluster analysis of the signal curves in perfusion DCE-MRI datasets

Mojgan Mohajer

Dissertation
an der Fakultät für Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Mojgan Mohajer
aus Teheran

München, den 02.07.2012

Aus dem Institut für Biologische und Medizinische Bildgebung
(Direktor: Prof. Dr. V. Ntziachristos)
Helmholtz Zentrum München

Erstgutachter: Prof. Dr. Volker J. Schmid

Zweitgutachter: Prof. Dr. Karl-Hans Englmeier

Tag der mündlichen Prüfung: 18. Oktober 2012

Contents

List of Abbreviations

AIC	Akaike's information criterion
AIF	Arterial input function
ASL	Arterial spin labeling
CT	Computed tomography
DCE-MRI	Dynamic contrast-enhanced MRI
DSC-MRI	Dynamic susceptibility contrast MRI
EES	Extracellular extravascular space
EM	Expectation maximization
Gd-DTPA	Gadolinium-diethylene triamine pentaacetic acid
HCC	Hepatocellular carcinoma
ICA	Independent component analysis
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
NSF	Nephrogenic systemic fibrosis
PCA	Principal component analysis
Pdf	Probability density function
PET	Positron emission tomography

PM	Parallelism measure
SOM	Self organizing maps
TE	Time to echo
TR	Repetition time

Abstract

Pathological studies show that tumors consist of different sub-regions with more homogeneous vascular properties during their growth. In addition, destroying tumor's blood supply is the target of most cancer therapies. Finding the sub-regions in the tissue of interest with similar perfusion patterns provides us with valuable information about tissue structure and angiogenesis. This information on cancer therapy, for example, can be used in monitoring the response of the cancer treatment to the drug. Cluster analysis of perfusion curves assays to find sub-regions with a similar perfusion pattern. The present work focuses on the cluster analysis of perfusion curves, measured by dynamic contrast enhanced magnetic resonance imaging (DCE-MRI). The study, besides searching for the proper clustering method, follows two other major topics, the choice of an appropriate similarity measure, and determining the number of clusters. These three subjects are connected to each other in such a way that success in one direction will help solving the other problems. This work introduces a new similarity measure, parallelism measure (PM), for comparing the parallelism in the washout phase of the signal curves. Most of the previous works used the Euclidean distance as the measure of dissimilarity. However, the Euclidean distance does not take the patterns of the signal curves into account and therefore for comparing the signal curves is not sufficient. To combine the advantages of both measures a two-steps clustering is developed. The two-steps clustering uses two different similarity measures, the introduced PM measure and Euclidean distance in two consecutive steps. The results of two-steps clustering are compared with the results of other clustering methods. The two-steps clustering besides good performance has some other advantages. The granularity and the number of clusters are controlled by thresholds defined by considering the noise in signal curves. The method is easy to implement and is robust against noise. The focus of the work is mainly the cluster analysis of breast tumors in DCE-MRI datasets. The possibility to adopt the method for liver datasets is studied as well.

Zusammenfassung

Pathologische Untersuchungen haben gezeigt, dass Tumore während ihres Wachstums aus verschiedenen Teilregionen mit homogenen Eigenschaften bestehen. Außerdem ist die Zerstörung der Blutversorgung des Tumors Ziel vieler Therapiemöglichkeiten. Die Analyse dieser Teilregionen des Tumorgewebes mit ähnlichem Perfusionsmuster, gibt uns wertvolle Information über das Tumorgewebe und seiner Angiogenese. Damit wird die Charakterisierung bzw. die Klassifikation von Tumoren und die Bewertung von Therapieerfolgen möglich. Mit Hilfe der Cluster-Analyse der Perfusionskurven können Teilregionen mit ähnlichen Perfusionsmustern gefunden werden. Ziel dieser Arbeit ist die Cluster-Analyse von DCE-MRI-Signalverläufen und die Bestimmung von Sub-Regionen im Tumor. In dieser Studie folgen neben der Suche nach dem richtigen Clustering-Verfahren zwei andere wichtige Themen, nämlich die Wahl eines geeigneten Ähnlichkeitsmaßes und die Bestimmung der Anzahl der Cluster. Diese drei Themen sind in der Weise miteinander verbunden, dass der Erfolg in einer Richtung zur Lösung der anderen Probleme führt. Es wird gezeigt, dass der euklidische Abstand als Ähnlichkeitsmaß für den Vergleich der Signalverläufe nicht ausreichend ist. Diese Arbeit stellt ein neues Ähnlichkeitsmaß, nämlich das Parallelitätsmaß (PM), für den Vergleich der Parallelität in der Auswaschungsphase der Signalverläufe vor. Um die Vorteile des euklidischen Abstandes mit denen vom Parallelitätsmaß zu kombinieren, wurde ein Zwei-Stufen-Clustering entwickelt. Das Zwei-Stufen-Clustering verwendet zwei verschiedene Ähnlichkeitsmaße in zwei aufeinanderfolgenden Schritten. Die Ergebnisse von Zwei-Stufen-Clustering werden mit den Ergebnissen von anderen Clustering-Verfahren verglichen. Das Zwei-Stufen-Clustering hat neben einer guten Leistung einige weitere Vorteile gegenüber anderen Verfahren. Die Granularität und die Anzahl der Cluster werden von Schwellwerten gesteuert, die unter Berücksichtigung des Rauschens in Signalverläufen bestimmt werden. Die Methode ist einfach zu implementieren und ist gegen Signalrauschen robust. Der Schwerpunkt der Arbeit ist die Cluster-Analyse

von Brusttumoren in DCE-MRI Datensätze. Die Möglichkeit einer Erweiterung des Verfahrens für Leber Datensätze ist ebenfalls untersucht worden und wird dargestellt.

Chapter 1

Introduction

1.1 Background

Many factors have effect on microvascular perfusion in a living tissue, such as medical procedures, drugs, and diseases [Hayat, 2007]. Therefore, the assessment and monitoring of microvascular perfusion in tissues are of the great interest. The perfusion can be monitored by an imaging modality that measures the signal changes in the tissue of interest after a contrast agent injection. The patterns of these changes are indirectly connected to the underlying perfusion. In addition, areas of tissue with a similar perfusion pattern have related perfusion properties. Pathological studies show that tumors consist of different sub-regions with more homogeneous vascular properties during their growth [Choyke et al., 2003, Gianfelice et al., 2003]. On the other hand, destroying tumor's blood supply is the target of most cancer therapies [Jain, 2005]. Finding the sub-regions in the tissue of interest with similar perfusion patterns provides us with valuable information about tissue structure and angiogenesis. This information in cancer therapy, for example, can be used in monitoring the response of the cancer treatment to the drug. Cluster analysis of perfusion curves assays to find sub-regions with a similar perfusion pattern. Present work focuses on the cluster analysis of perfusion curves, measured by dynamic contrast enhanced magnetic resonance imaging (DCE-MRI). Perfusion studies by means of DCE-MRI have some advantages compared to other imaging modalities.

Magnetic resonance imaging (MRI) has established itself as an important and useful medical imaging technique. In contrast to computed tomography (CT) and X-ray technologies, MRI does not use ionizing radiation, and thus is not associated with the same

health hazards. Furthermore, in its relative short time of use, since the early 1980s, no long-term health issues related to its strong static magnet field are known. Therefore, in contrast to CT and X-ray, a higher number of MRI scans can be performed on the same individual. That is the main reason why MRI is used in perfusion studies.

To analyze the perfusion in biological tissue *in vivo*, MRI can be used in two ways. The first method is based on using contrast agents that change the magnetic susceptibility of blood [Jackson et al., 2005]. The contrast-enhanced measurement can be performed as T_1 weighted or as T_2^* weighted imaging, in each of which different aspects of the tissue are visualized. The T_1 weighted perfusion MRI is normally known as dynamic contrast-enhanced MRI (DCE-MRI), while the T_2^* weighted perfusion imaging is noted as dynamic susceptibility contrast MR imaging (DSC-MRI). As an alternative, in the arterial spin labeling (ASL), blood water is used as a freely-diffusible tracer [Jackson et al., 2005]. That means the patients' own blood takes the role of the contrast agent. ASL imaging entails many artifacts; thus, an accurate quantization of blood flow is still difficult [Crawley et al., 2003]. For this reason, most of the perfusion analyses in MRI are focused on DCE-MRI.

DCE-MRI in combination with pharmacokinetic modeling is widely applied in oncology to calculate parameters related to tumor vascular physiology [Brasch and Turetschek, 2000, Griebel et al., 1997, Neeman, 2000]. Commonly, the small molecular contrast agent gadolinium-DTPA (Gd-DTPA), is applied as a bolus injection during the serial acquisition of MR images. A set of MRI volumes is acquired after injection. As a result, at each voxel in the volume, we obtain dynamic information about functional and vascular properties of the tissue. If tumors are growing beyond several millimeters in diameter, they induce surrounding vessels to create neo-vessels, i.e. angiogenesis [Hanahan and Folkman, 1996]. The screening of angiogenesis by means of DCE-MRI is extensively studied by various authors [Brasch and Turetschek, 2000, Griebel et al., 1997, Neeman, 2000]. In addition, many cancer treatments are based on either stopping the process of angiogenesis or destroying existing tumor vasculatures. Therefore, a non-invasive method, to monitor the tumor vascular changes, is of great interest. To determine tissue changes during cancer treatment, one can analyze the temporal changes of the signal intensities, the so-called signal curves, in DCE-MRI of the tissue of interest. This can either be done by means of perfusion analysis, which studies the physiological parameters of the tissue of interest, or based on machine learning methods.

Clustering methods are one of the most important approaches of machine learning

methods. The goal of clustering methods is to divide the region of interest into sub-regions with more similar characteristics. This kind of analysis is also of interest for perfusion analysis and in tumor classification, tumor characterization and therapy monitoring. For example, in the case of therapy monitoring, changes in tumor patterns before and after therapy can be compared. Furthermore, the representative curve of each cluster is the mean curve, averaged over all signal curves inside the cluster, and therefore has enhanced signal to noise ratio. Due to this noise reduction, representative curves depict proper candidates for pharmacokinetic analysis for a more accurate characterization of tumor microcirculation.

Various clustering methods have been used to cluster signal curves. A schema of classification using pre- and post-patterns has been widely adopted following its initial clinical evaluation by Kuhl et al. [1999]. For example, Kubassova et al. [2007] and Lavini et al. [2007] classified tumors into benign or malignant by pre-defined characteristics of signal curves, which are basically derived from the classification of initial enhancement and the late pattern. In both methods, the classification of extracted features was performed using fixed thresholds. As a similarity measure, Euclidean distance was used in the feature space. Both methods successfully segment and roughly classify the tumors, but they are too rough to distinguish between different areas inside a tumor.

Other methods are vector quantization [Leinsinger et al., 2006, Schlossbauer et al., 2008, Wismüller et al., 2006] and minimization of energy function [Zheng et al., 2007]. The first one has the disadvantage of using too many parameters and becomes more complex for high dimensional vector spaces, while the latter suffers from general optimization problems and is again too steady and complex for high dimensional vector spaces.

Most methods are based on fixed parameters like the number of clusters, that is the number of groups of different patterns in signal curves. Some examples are k-means clustering [Baumgartner et al., 2005, Lee et al., 2006], fuzzy c-means [Kannan et al., 2011], neuronal gas [Meyer-Baese et al., 2008, 2009], neuronal network [Lucht et al., 2001], support vector machine [Daducci et al., 2009], and independent component analysis [Koh et al., 2008]. Neuronal network was implemented for classification of carcinoma tumors, which was based on training data with three defined classes [Lucht et al., 2001]. In another study [Daducci et al., 2009], DCE-MRI datasets were trained by means of support vector machine. As it is usual for training-based clustering methods, large numbers of training data and well-selected expert's knowledge are required. Another group of cluster-

ing methods performed on DCE-MRI datasets are mode-seeking algorithms, such as mean shift [Castellani et al., 2006, Stoutjesdijk et al., 2007]. The mean shift algorithm suffers from difficulties to choose proper initial parameters.

Selecting a proper distance measure is another important factor in most clustering methods. In almost all previous works on clustering of DCE-MRI datasets of breast tumors, Euclidean distance is the measure of dissimilarity. These works can be divided into two general categories. In the first category, DCE-MRI signal curves are considered as vectors in an m-dimensional vector space. These vectors are used directly in clustering methods [Castellani et al., 2006, Meyer-Baese et al., 2009, Twellmann et al., 2008, Varini et al., 2006, Wismüller et al., 2006]. In the second category, m' features are extracted for each signal curve and signal curves are clustered based on the m'-dimensional feature vectors [Daducci et al., 2009, Lee et al., 2010, Leinsinger et al., 2006, Nattkemper et al., 2005, Stoutjesdijk et al., 2007].

In this work, the cluster analysis has been studied in three directions, selecting the appropriate clustering method, providing the proper similarity measure, and the choice of the underlying number of clusters. These three subjects are thoroughly connected, and success in one direction will help solving the other problems.

1.2 Outline

Cluster analysis of DCE-MRI signal curves is the topic of this thesis. Major problems of the clustering in general and in the case of DCE-MRI datasets have been discussed. A new similarity measure has been introduced, and a two-steps clustering has been developed. Furthermore, in several experiments, different clustering methods have been compared. In addition, the possibility to predict the proper number of clusters by means of Gap Statistic has been studied.

A summary of previous works on cluster analysis of DCE-MRI signal curves has been discussed in this chapter. The further structure of the chapters can be summarized as follows. The concept of perfusion DCE-MRI has been discussed in chapter two. For this, the MRI technique has been discussed briefly and a brief introduction to perfusion analysis and some well-used perfusion models are given. In the third chapter, a selected number of clustering methods have been introduced and topics of similarity measure, dimension reduction and number of clusters have been discussed. In chapter four, the clustering of

DCE-MRI of the breast tumors has been discussed. Different clustering scenarios have been performed on real datasets, and the results have been compared. Chapter five demonstrates the advantages of clustering in the case of liver datasets. A new clustering method has been introduced and performed on nine DCE-MRI datasets. The heterogeneity of liver tissue in the neighborhood window of each voxel has been compared to the heterogeneity of the clusters. In chapter six, the introduced Gap statistics from chapter two have been performed on the simulated and real datasets. The thesis is closed with discussion of the results and the conclusion in chapter seven.

Chapter 2

Introduction to perfusion DCE-MRI

In this chapter, a brief introduction of the basic principles of DCE-MRI is given. Started by MR-imaging technique, it continues on introducing contrast agents, describing different DCE-MRI techniques. The chapter ends up with a brief introduction to perfusion analysis and some well-used perfusion models.

2.1 MRI

Magnetic resonance imaging is a medical imaging technique. It is based on the principles of nuclear magnetic resonance (NMR) [Hornak, 2000]. Although details of MRI are beyond the scope of this dissertation, a brief description of this technique is given here for a better understanding of the major factors and properties important for image acquisition. The MRI signal arises from interaction of core spin of atoms with an external magnetic field. Usually, ^1H nucleus or ^{13}C nucleus is used in MRI due to the strong present in biological tissue.

A nucleus, when placed in an external magnetic field B_0, is not aligned with its magnetic

Nucleus	$2\pi/(\gamma)MHzT^{-1}$
^1H	42.576
^{13}C	10.705
^{31}P	17.235

Table 2.1: Approximate values of γ for some common nuclei.

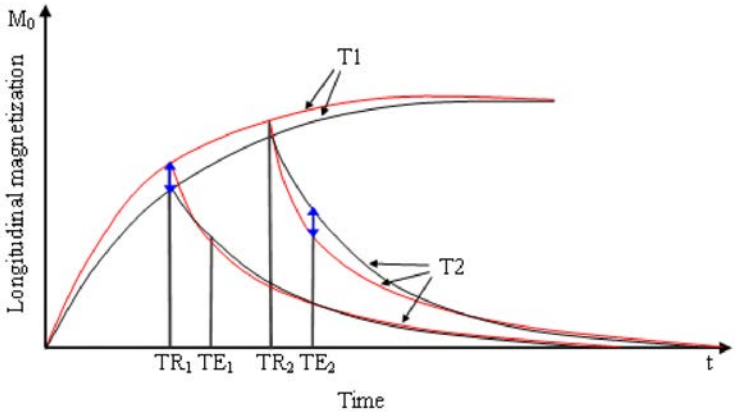


Figure 2.1: The red and black curves are recovery and decay curves for two different tissues. Different values of TR and TE result in different weighting of T_1 or T_2 properties of the tissues.

moment, and therefore, will precess at the Larmor frequency

$$w_L = \gamma B_0.$$

The ratio γ is called gyromagnetic ratio and is determined by terms of proton mass, charge, and spin angular momentum. Approximate values of γ are given in Table 2.1 for some common nuclei.

Different orientations of spin angular momentum, in a static external magnetic field, describe different magnetic energy levels. According to quantum physics, only certain orientations of spin are allowed. Applying a radio frequency magnetic field, transition between different spin orientations can be induced. The most efficient transfer occurs when the applied radio frequency is in resonance with Larmor frequency.

The spatial information in MRI measurement can be achieved by applying an inhomogeneous static magnetic field B_0 to the object of interest. The origin of MRI signal can be evaluated from the mentioned resonance condition. The measured signal S at each position of the volume of interest is given by [Hashemi and Bradley, 2004]:

$$S \propto N \cdot (e^{-TE/T_2}) \cdot (1 - e^{-TR/T_1}). \tag{2.1}$$

The signal is related to density of protons N, that is, the number of protons per unit volume, spin-lattice relaxation time T_1, and spin-spin relaxation time T_2. The parameter T_1 is related to the way the protons absorb energy from lattice. The parameter T_2, in contrast,

TR	TE	Result
long	long	T_2 weighting
short	short	T_1 weighting
long	short	N dominates, proton density weighting
short	long	both T_1 and T_2 weighting

Table 2.2: Summary of different choices for TR and TE.

is related to the way the protons release the absorbed energy. Spin-lattice relaxation generally takes longer than spin-spin relaxation. Both T_1 and T_2 are properties of the tissue. The parameters TE and TR are constants selected for the actual measurement. The time interval between applications of radio frequency pulse is called repetition time TR. The parameter TE stands for echo delay time or time to echo.

Different tissues have different values of N, T_1, and T_2. These lead to a variation of MR signal over the entire image. To enhance the image contrast, this signal variation needs to be emphasized by selecting proper TR and TE values. Figure 2.1 depicts how different choices of TR and TE values result in different weighting of T_1 or T_2 properties of the tissues. The red and black curves are recovery and decay curves of two different tissues. In general, long TR and TE result in T_2 weighted and short TR and short TE result in T_1 weighted images. Table 2.2 summarizes the possible choices for TR, TE, and the resulted images [Jackson et al., 2005].

In reality, instead of the relaxation time T_2 in Eqn. 2.1 an effective value T_2^* is used. The relation between T_2^* and T_2 can be described by the following formula [Hashemi and Bradley, 2004]:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B. \tag{2.2}$$

The parameter ΔB measures the inhomogeneities in the surrounding magnetic field and γ is the gyromagnetic ratio.

2.2 Contrast agents

Contrast agents are chemical substances, which are introduced to the body to increase the contrast among tissues [Hornak, 2000]. Gadopentetic acid in the form of gadopentetate

dimeglumine or Gd-DTPA was one of the first contrast agents for MR imaging. It is paramagnetic and reduces the relaxation time T_1, and to some extent also the relation time T_2 depending on the used concentration [Jackson et al., 2005]. As a side effect, it disturbs local magnetic field homogeneity and consequently, has T_2^* shortening effect. The T_1 and T_2 shortenings are related to the interaction with the out shell electrons of the agent. Therefore, T_1 and T_2 shortenings are phenomena caused by short-range effect in nm scale. The T_2^* shortening is related to field disruption in mm scale. The field disruption is associated with the paramagnetic, super paramagnetic or ferromagnetic nature of the agent. In contrast to X-ray and CT, MRI does not measure directly the concentration of contrast agent but rather the effect of agent on local water protons [Jackson et al., 2005]. Most of the clinically used agents are non-selective, also called extracellular. Non-selective extracellular agents do not bind to protein, and thus are not tissue specific. Extracellular agents are excreted by the kidney. Gd-DTPA is also an extracellular agent. It is injected intravenously, and in general, does not enter tissue cells. In contrast to contrast agents used in X-ray and CT, Gd-DTPA has no radiation effects, but it may cause a toxic reaction in patients with severe kidney problems called nephrogenic systemic fibrosis (NSF) [Thomsen et al., 2006].

2.3 DCE-MRI

Dynamic contrast-enhanced MRI is an expansion of MRI technique, which adds an extra temporal dimension to the existing spatial dimensions. As a result of the dynamic acquisition, for each voxel in DCE-MRI, instead of one single measurement, a sequence of measurements over the time is performed. Each sequence of measurements is called a signal curve. Figure 2.2 schematically demonstrates DCE-MRI acquisition. A signal curve corresponding to a voxel inside the tumor area is also depicted in Figure 2.2.

DCE-MRI is based on the fact that the concentration of extracellular agent, e.g. Gd-DTPA, in the extracellular extravascular area varies over the time. The variation in shape of signal curves is strongly related to blood flow and physiological properties of tissue. Figure 2.3 demonstrates, as an example, a schematic illustration of three different types of signal curves. The signal curves from left to right can be categorized as, rapid initial and sustained late enhancement, rapid initial and stable late enhancement, and rapid initial and decreasing late enhancement [Daniel et al., 1998].

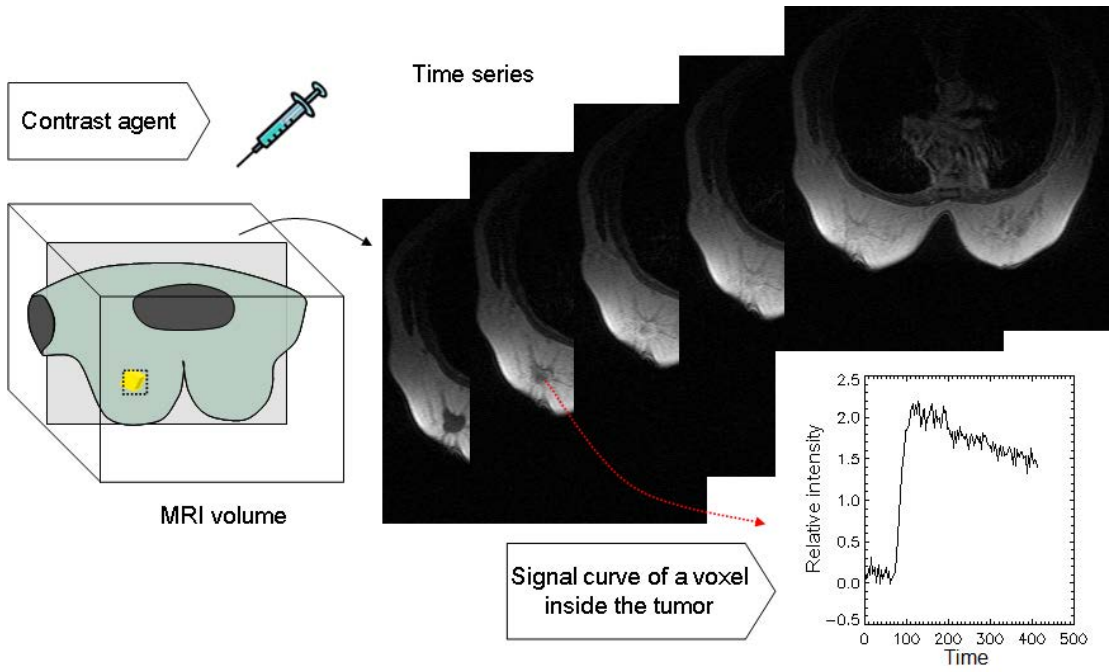


Figure 2.2: Schematic illustration of DCE-MRI acquisition. A signal curve corresponding to a voxel inside the tumor area is depicted.

MR imaging can be done T_1 weighted or T_2^* weighted. Characteristics of these two approaches can be summarized as follows [Jackson et al., 2005]:

- T_1 weighted imaging:
  - Acquisition time is circa 5 to 7 minutes.
  - Spatial resolution should be reduced due to the long acquisition time if high temporal resolution is desired.
  - The method results in increased local signal intensities.
- T_2^* weighted imaging:
  - The acquisition time is about 1 to 2 minutes.
  - Higher spatial resolution is possible due to the short acquisition time.
  - The method results in decreased local signal intensities.
  - The method suffers from spatial geometric distortions and signal abnormalities.

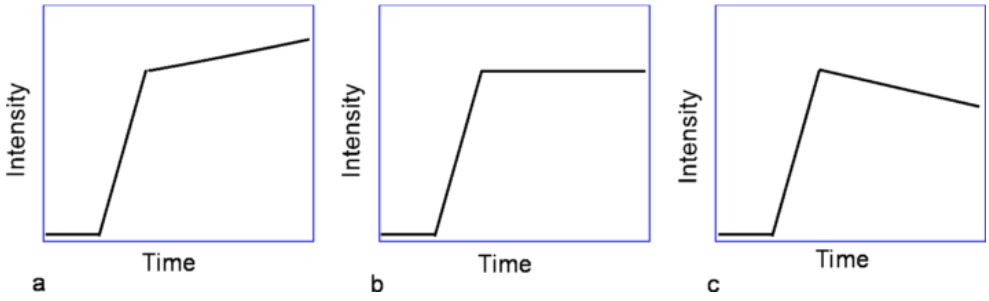


Figure 2.3: Schematic illustration of three different types of signal curves.

Signal intensities in DCE-MRI depend in a complex manner on parameters such as tissue perfusion, microvascular permeability, blood volume, and extravascular volume [Tofts, 1997]. In general, there are two different contemplations of DCE-MRI. One approach is to include dependencies between signal curves and tissue parameters by incorporating the blood flow and relevant physical laws. This approach is known as perfusion analysis. Another approach tries to extract proper features from signal curves and the association of these features with different clinical aspects.

For perfusion analysis, the accurate characterization of arterial input function (AIF) is important (see section 2.4). For an accurate measurement of the AIF, a temporal resolution in the order of 1 second is required [Henderson et al., 1998]. At the same time, a good spatial resolution is required to find the exact position of the artery in relation to the tissue of interest. Due to geometric distortions and signal abnormalities in T_2^* weighted images, T_1 weighted protocols are preferred for perfusion analysis. On the other hand, the longer acquisition times required in T_1 weighted protocols do not allow to achieve high temporal and high spatial resolution at the same time. In clinical practice, a high spatial resolution at the expense of temporal resolution is routine. Thus, in the mostly used DCE-MRI protocols for clinical tumor detection in breast cancer, for example, 5 to 6 high spatial resolution volumes are acquired after bolus injection, while scanning time of a whole volume is about 100 seconds. In contrast, in case of perfusion DCE-MRI, to reduce the scanning time of the volume to few seconds, normally a complete volume acquisition of the whole breast is waived.

The study of signal curves is the aim of perfusion analysis. Since the signal curve of a single voxel is not very reliable, it is common to use the average signal curve over a group of voxels in the tissue of interest. However, the selection of an appropriate group of voxels

for analysis is difficult. For example, in a cancerous tumor it is quite common that the tissue is very heterogeneous so that the average over the whole tumor results in a signal curve, which has nothing to do with the real signal curve of the tissue. The main subject of the present work is to find more homogeneous sub-regions in a heterogeneous tumor.

2.3.1 DCE-MRI in breast cancer

Breast cancer is the most common cancer among women [Jackson et al., 2005]. Accordingly, breast cancer is the most common cause of death among all other cancers in women [Husmann et al., 2010]. Early detection and advances in therapy monitoring are very important in diagnosis and therapy of breast cancer. Mammography and ultrasound are commonly used in diagnosis of breast cancer. Nevertheless, both have their limitations [Jackson et al., 2005].

From the diagnosis point of view, DCE-MRI offers, besides the morphological information, functional characteristic of the tissue. It is known that extra vascular support is needed in solid tumors growing beyond millimeter size [Hanahan and Folkman, 1996]. Pathological studies show that breast tumors consist of different sub-regions with more homogeneous properties during their growth [Choyke et al., 2003, Gianfelice et al., 2003]. In breast cancer, this information can be used to improve tumor diagnosis. DCE-MRI can also be used to monitor tumor response to treatments. A decrease in the rate and magnitude of signal enhancement indicates a successful treatment. In contrast, a poor response results in persistent abnormal enhancement [Hayes et al., 2002, Knopp et al., 1994]. An exact comparison of the signal curves before and after therapy delivers information about the response of the tumor to the therapy.

Data acquisition

The T_1 weighted MR breast imaging is normally performed with gradient-echo techniques. The contrast agent is dosed about 0.1 $mmol$ per kg body weight. The choice of high temporal or spatial resolution strongly depends on the goal of imaging. For tumor detection, normally high spatial resolution 3D technique with a voxel size less than 1 mm is used. The acquisition time is between 2 and 4 minutes. In this kind of imaging usually only three or four volumes are acquired. The dynamic information of datasets is restricted to pre-enhanced, post- and late post-enhanced volumes. However, in the most clinical DCE-

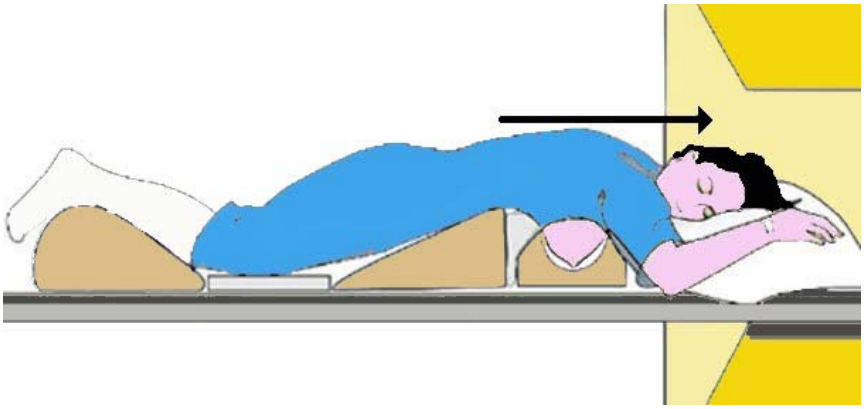


Figure 2.4: Prone, head first positioning during DCE-MR imaging of the breast (image from [Protocol, 1999]).

MRI in breast, the spatial resolution is chosen so that every 60 to 90 s an acquisition is possible. In this case, during the first 6 minutes after agent injection, 4 to 5 acquisitions can be performed. Thus, more characteristics from signal curves can be derived for tumor differentiation [Jackson et al., 2005].

However, temporal intervals of 60 s or more are not sufficient for an exact perfusion analysis and pharmacokinetic modeling of the tissue. The first 4 to 6 minutes after contrast agent injection, is the most informative time for the pharmacokinetic analysis of the tissue. The kinetic information is very important for specificity and tumor differentiation [Jackson et al., 2005]. It has been shown that for pharmacokinetical analysis, especially for a correct estimation of AIF a temporal resolution better than 1 s is required [Jackson et al., 2005]. In general, perfusion analysis of DCE-MRI requires a temporal resolution better than 30 s during the first 90-150 s after bolus injection [Henderson et al., 2000, Jackson et al., 2007]. This kind of DCE-MRI is also known as perfusion MRI. To achieve the required temporal resolution, usually the 3D volume is reduced to a few slices through the region of interest, e.g., tumor.

Patient movement during the acquisition of DCE-MRI is a major problem. To minimize the motion, a prone head first positioning of the patient, as illustrated in Figure 2.4, is usual. The breast mostly consists of soft tissue, and therefore, is easily deformed. To reduce the deformation of the breast during scanning, the breasts are positioned in an empty area under the chest of the patient. Thus, the breast obtains a constant form during the scanning.

2.3.2 DCE-MRI in liver cancer

In comparison to its use for the breast, the central nervous system, and the musculoskeletal system, where the use of MRI is massively growing, its usage for the liver was limited due to strong movement of the organ and its complex vascular system. The improvement in the MRI technique and the shortening of acquisition time increase the interest in the usage of MRI scan in liver [Bartolozzi et al., 1999, Rummeny and Marchal, 1997, Van Beers et al., 1997]. Totman et al. [2005] and Tsushima et al. [2001] have also shown the potential of using DCE-MRI to detect and characterize the tumors in liver. Model-based perfusion techniques may allow quantification of microvascular parameters. This may contribute to a better understanding and a better differential diagnosis of focal liver lesions [Thng et al., 2010]. The liver has a dual blood supply, derived from the hepatic artery and the portal vein. It has only one venous output vessel. These make liver one of the most complicated and challenging organs in perfusion modeling and analysis [Pandharipande et al., 2005, Thng et al., 2010].

In the case of hepatocellular carcinoma (HCC), DCE-MRI can help in tumor identification due to the predominately arterial blood supply of the tumor tissue compared to the portal venous blood supply of the liver tissue. Furthermore, DCE-MRI might be developed into a bio-marker for treatment monitoring [Thng et al., 2010, Totman et al., 2005, Tsushima et al., 2001].

The liver is an organ which is exposed to different sources of motion such as peristalsis of the stomach, breathing motion and heart beats. There are several studies trying to reduce motion artifacts with help of registration techniques [Caldeira et al., 2008, Melbourne et al., 2007, 2008, Rohlfing et al., 2004]. Also, voluntary breath-hold techniques have shown success to reduce respiratory motions [Jackson et al., 2002, Kimura et al., 2004]. Unfortunately, voluntary breath-holding results in lower temporal resolution. The gap between two successive acquisitions can be up to 7 s. The method is limited to those patients who are able to hold their breath voluntarily. Another possibility is the use of navigator echo techniques [Vasanawala et al., 2010].

Due to the biological complexity of the liver, the tendency toward more complex models with more compartments is growing [Mescam et al., 2010, Thng et al., 2010]. However, a more complex, multi-compartment model is more sensitive to noise and motion artifacts. For perfusion analysis, an average curve of a group of neighboring voxels is typically se-

lected. It reduces the noise and increases the reliability of the signal curve in comparison with the single voxel signal curve. On the contrary, the liver is a very heterogeneous tissue [Jackson et al., 2007]. This heterogeneity is related to different biological properties of the local tissue and is reflected in dynamic signal curves. Averaging signal curves of a selected window inside the tumor or of any other parts of the liver means that signal curves of a heterogeneous area with different enhancement patterns are averaged, which is of disadvantage. As a solution, a careful grouping of the voxels with similar enhancement patterns can be suggested. The corresponding signal curves can be averaged in order to reduce motion artifacts and noise. This method also segments the liver into more homogeneous sub-regions with similar enhancement patterns. In this way, a better understanding of the liver and its tumors can be achieved. The method also leads to more accurate and suitable design of perfusion models. Although this kind of clustering has been successfully used in perfusion analysis of other organs [Baumgartner et al., 2005, Leinsinger et al., 2006, Wismuller et al., 2006], there are no similar studies on DCE-MRI of the liver to our knowledge.

2.4 Perfusion analysis

In physiology, perfusion is known as the blood flow to a capillary bed in the biological tissue. More precisely, it is defined as the volume of blood that flows into one gram tissue [Westbrook et al., 2005]. Signal changes in DCE-MRI are related to the change in concentration of contrast agent in the tissue. At the same time, concentration change of agent depends on perfusion in the tissue. Thus, the analysis of signal curves provide information about tissue perfusion. The signal curves have been studied in two directions, by extracting characteristics features such as the area under the curve or peak of the signal curves [Kubassova et al., 2007, Lavini et al., 2007] or by pharmacokinetic modeling.

Pharmacokinetic modeling tries to describe the process of blood perfusion. This kind of analysis of the tissue becomes extremely attractive in characterizing the biological properties of the tumors. The goal of kinetic modeling is quantitative estimation of microvascular characteristics, particularly the fractional blood volume and the microvascular permeability of the tumor vessels. It has been demonstrated that tumor microvessels compared to normal non-tumor vessels are more permeable for the transendothelial diffusion of large molecular solutes like Gd-DTPA [Jackson et al., 2005].

During the last decades, several kinetic models have been proposed. Renkin [1959] and Crone [1965] described a useful and widely accepted model of this process. According to Renkin and Crone:

$$FLR = \frac{F}{PV}\left(1 - e^{\frac{-PS}{F}}\right), \tag{2.3}$$

where FLR represents fractional leak rate, F flow, PV plasma volume, and PS the so called permeability surface area product. For $PS << F$, Eqn. 2.3 is in first order equal to:

$$FLR = \frac{PS}{PV}. \tag{2.4}$$

2.4.1 Contrast agent concentration in T_1 weighted DCE-MRI

The relationship between relaxation rates $1/T_1$, $1/T_2$, and contrast agent concentration can be predicted by the Solomon-Bloembergen equations [Gowland et al., 1992]:

$$\frac{1}{T_1} = \frac{1}{T_{10}} + r_1 C(t), \tag{2.5}$$

$$\frac{1}{T_2} = \frac{1}{T_{20}} + r_2 C(t), \tag{2.6}$$

where r_1 and r_2 are spin-lattice and spin-spin relaxation constants of the contrast agent, T_{10} and T_{20} are spin-lattice and spin-spin relaxation times in the absence of contrast material, and C(t) is the average contrast agent concentration in tissue at the time point t. To monitor the kinetic behavior of a contrast agent *in vivo* it is necessary to link changes in contrast agent concentration to changes observed in MR images.

In Eqn. 2.1, it is assumed that a spin flip of 90° is induced by magnetic field. This equation can be extended for an arbitrary flip angle α to the following formula:

$$S \propto N \cdot e^{-TE/T_2^*} \cdot \frac{\sin(\alpha)(1 - e^{-TR/T_1})}{1 - \cos(\alpha)e^{-TR/T_1}}. \tag{2.7}$$

It is assumed that Gd-DTPA has no effect on the proton density N. Then, the changes of signal intensities after administration of the contrast agent are related to the shortening effect of T_1 and T_2 relaxation times, as described in Eqn. 2.5 and Eqn. 2.6. During the

imaging process and between subsequent acquisitions, the signal intensity is altered due to the various number of parameters such as the signal loading of the coil, receiver settings at MR console, and image reconstruction parameters. Therefore, the measured signal intensity is related to an internal standard. Due to this nature of MR signals, the absolute values of measured signals are not comparable. A common way to relief this issue is to relate the post-contrast signal intensities to pre-contrast signal intensities. This is normally done by calculating the so called relative intensity:

$$S_r = \frac{S_{Gd} - S_0}{S_0}, \tag{2.8}$$

where S_0 and S_{Gd} are pre- and post-contrast signals, respectively. However, for very small values of S_0 this formula leads to strong enhancement of the signal. This problem appears specially for noisy background signals.

Combining Eqn. 2.5, Eqn. 2.8, and Eqn. 2.10, the relationship between the relative signal intensity and average contrast agent concentration for a voxel at time point t can be obtained by [Brix et al., 2004]:

$$C(t) = \frac{-1}{TR \cdot r_1} \cdot \ln \left[1 + S_r(t) \cdot \left(1 - e^{\frac{TR}{T_{10}}}\right)\right] \tag{2.9}$$

with

$$S \propto N \cdot (1 - e^{-TR/T_1}). \tag{2.10}$$

Eqn. 2.10 is the simplified version of Eqn. 2.7. For simplicity, a flip angle of 90° and a strongly T_1 weighted imaging are assumed, so that e^{-TE/T_2^*} is ignorable due to small TE.

2.4.2 General kinetic model

The goal of kinetic modeling is to find physiological relevant properties of the tissue from variation of contrast agent over time in the tissue of interest. For this, the tissue structures and the functional processes that affect the distribution of the tracer should be defined. Normally, the tissue is presented as two or more compartments, each of which is a bulk tissue characteristic. These compartments are, for example, the vascular plasma space, the extracellular extravascular space (EES), and the intracellular space [Tofts et al., 1999]. A schematic illustration of these compartments is depicted in Figure 2.5. Transfer rate of

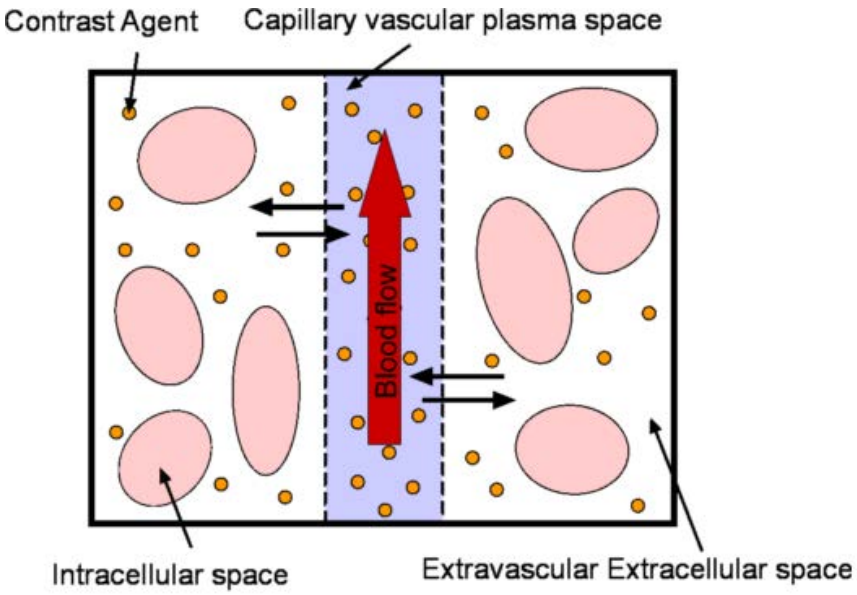


Figure 2.5: A schematic illustration of the vascular plasma space, the extracellular extravascular space (EES), and the intracellular space is depicted [Jackson et al., 2005].

contrast agent between the vascular plasma space and the extravascular extracellular space is associated with the following parameters: the blood flow to the tissue, the permeability of the blood vessel walls, and the surface area of the perfusing vessels. Contrast agents like Gd-DTPA do not enter the cell and therefore, EES is the volume of distribution of contrast agent. The rate of concentration change of contrast agent in the EES, can be estimated by the modified general rate equation [Kety, 1951]:

$$v_e \frac{dC_e(t)}{dt} = K_{trans}(C_p(t) - C_e(t)), \tag{2.11}$$

where C_e is the concentration of agent in extracellular space v_e, C_p is the concentration of agent in plasma space v_p, and K_{trans} is the volume transfer constant between v_p and v_e [Tofts et al., 1999]. The Eqn. 2.11 can be extended to include the concentration of contrast agent in blood plasma:

$$C(t) = v_p C_p(t) + v_e C_e(t). \tag{2.12}$$

Solving Eqn. 2.11 and Eqn. 2.12, we get:

$$C(t) = v_p C_p(t) + K^{trans} \int_0^t C_p(t') e^{\frac{-K^{trans}(t-t')}{v_e}} dt', \tag{2.13}$$

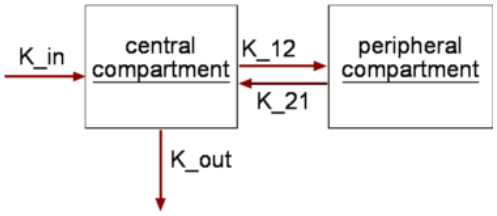


Figure 2.6: Schematic illustration of early Brix model.

which is equivalent to:

$$C(t) = v_p C_p(t) + C_p(t) \otimes H(t), \tag{2.14}$$

where $\otimes$ is the convolution operation and $H(t)$ is the impulse response or residue function.

There are a range of compartmental modeling approaches, which mainly differ based on applied assumptions. One of the main sources of differences is the determination of an arterial input function (AIF) which describes $C_p(t)$ in Eqn. 2.13. In the following sections, some of the modeling schemata are introduced, which are mostly used in literature and also in this work.

2.4.3 Early Brix model

In this model, the extracellular space is considered a single peripheral compartment and the plasma space is the central compartment (see Figure 2.6). The AIF is defined as an exponentially decaying function on a patient-by-patient basis [Brix et al., 1991]. The Brix model describes the relationship between these two compartments with two mass differential equations:

$$\frac{dM_1}{dt} = K_{in} - (K_{12} + K_{out})M_1 + K_{21}M_2, \tag{2.15}$$

$$\frac{dM_2}{dt} = K_{12}M_1 - K_{21}M_2, \tag{2.16}$$

where M_1 and M_2 are the amounts of contrast agent in central compartment and peripheral compartment, respectively. The parameters K_{12} and K_{21} are first-order rate constants, which describe contrast agent transfer between two compartments. They are related by the equation $K_{12}v_p = K_{21}v_e$ with v_p being the volume of plasma space and v_e the volume of extracellular space. The parameter K_{out} is the first-order rate constant of the elimination

of contrast agent. The parameter K_{in} the zero-order rate constant and is equal to the contrast agent infusion rate.

The Eqn. 2.15 and Eqn. 2.16 are simplified by assumption that v_e is small, and therefore, both $K_{12}M_1$ and $K_{21}M_2$ can be neglected in Eqn. 2.15. Using $C_p = M_1/v_p$ and $C_e = M_2/v_e$ we have:

$$\frac{dC_p}{dt} = \frac{K_{in}}{v_p} - K_{out}C_p \tag{2.17}$$

and

$$\frac{dC_e}{dt} = K_{12}\frac{v_p}{v_e}C_p - K_{21}C_e. \tag{2.18}$$

The solution of the above differential equations system for initial conditions $C_p(0) = 0$ and $C_e(0) = 0$ is:

$$C_p(t) = \frac{K_{in}}{v_p K_{out}}(e^{K_{out}t'} - 1)e^{-K_{out}t}, \tag{2.19}$$

and

$$C_e(t) = \frac{K_{in}K_{12}}{v_e}\left[v(e^{K_{out}t'} - 1)e^{-K_{out}t} - u(e^{K_{21}t'} - 1)e^{-K_{21}t}\right], \tag{2.20}$$

with $u = \frac{1}{K_{21}(K_{21}-K_{out})}$ and $v = \frac{1}{K_{out}(K_{21}-K_{out})}$. During contrast agent infusion $0 \leq t \leq \tau$, we have $t' = t$. After infusion, we have $t' = \tau$, where τ is the infusion duration.

For $TR \cdot r_1 \cdot C(t) << 1$ Eqn. 2.9 can be simplified to:

$$S_r(t) = FC(t), \tag{2.21}$$

where F is a factor depending on TR, T_{10} and r_1. Combining the equations Eqn. 2.20 and Eqn. 2.21, we obtain:

$$S_r(t) = A\left[v(e^{K_{out}t'} - 1)e^{-K_{out}t} - u(e^{K_{21}t'} - 1)e^{-K_{21}t}\right], \tag{2.22}$$

where A is a constant and depends on TR, T_{10}, r_1, K_{in}, K_{12}, and v_e.

2.4.4 Tofts model

Tofts model consists of three compartments, plasma volume, extracellular space, and the lesion. Figure 2.7 depicts schematically these three compartments. It is assumed that after

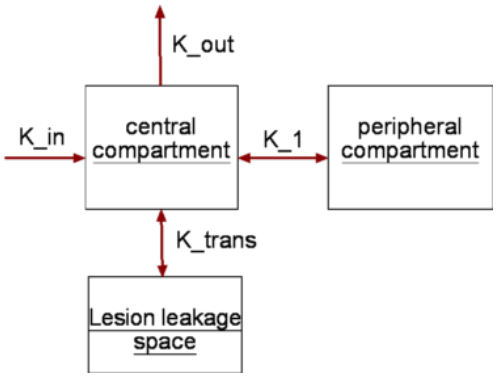


Figure 2.7: Schematic illustration of Tofts model.

a bolus injection the tracer is fast mixed within the plasma and is well-mixed. The flow of contrast agent from plasma volume to extracellular space and kidney is described as [Tofts and Kermode, 1991, Tofts, 1997]:

$$-v_p \frac{dC_p}{dt} = K_1(C_p - C_e) + K_{out}C_p, \tag{2.23}$$

where K_1 is the flow constant into extracellular space and K_{out} is the constant elimination rate into the kidney. The flow into the extracellular space is described as [Tofts and Kermode, 1991, Tofts, 1997]:

$$v_e \frac{dC_e}{dt} = K_1(C_p - C_e). \tag{2.24}$$

A solution of the linear differential system of Eqn. 2.23 and Eqn. 2.24 is a biexponential function [Tofts and Kermode, 1991]:

$$C_p(t) = D(a_1 e^{-m_1 t} + a_2 e^{-m_2 t}), \tag{2.25}$$

where a_1 and a_2 are the amplitudes of the components and m_1 and m_2 are the rate constants. This biexponential function was fitted to an average AIF obtained from several experiments for a dose of $D = 0.1$ $mmol$ per kg body weight resulting in parameters, $a_1 = 3.99$ $kg/liter$, $a_2 = 4.78$ $kg/liter$, $m_1 = 0.144$ min^{-1}, and $m_2 = 0.0111$ min^{-1} [Whitcher and Schmid, 2011]. The flow of tracer from plasma into the third compartment in Tofts model, the lesion leakage space, is described similar to Eqn. 2.24:

$$v_{le} \frac{dC_{le}(t)}{dt} = K_{trans}(C_p(t) - C_{le}(t)), \tag{2.26}$$

where v_{le} is the volume of leakage space and k_{trans} is the transfer coefficient between plasma space and leakage space. Together with Eqn. 2.25, a solution of Eqn. 2.26 can be written as:

$$C_{le}(t) = C_p(1 - e^{-K_{trans}t/v_{le}}). \tag{2.27}$$

In fatty tissues, the signal is weak. Thus, the average concentration of tracer in tissue can be considered the concentration of the tracer in the leakage space, that is $C_t = v_{le}C_{le}$. Combining Eqn. 2.25 and Eqn. 2.27, the mean concentration in the lesion tissue C_t can be described by a triple exponential function:

$$C_t(t) = D(\sum_{i=1}^{3} b_i e^{-m_i t}) \tag{2.28}$$

with $b_1 = K_{trans}a_1/(m_3 - m_1)$, $b_2 = k_{trans}a_2/(m_3 - m_2)$, $b_3 = -(b_1 + b_2)$, and $m_3 = K_{trans}/v_{le}$. The parameters a_1, a_2, m_1 and m_2 are derived experimentally as explained before.

In the Tofts model, it is assumed that K_{trans} is the same in both directions, from the central compartment to leakage space and contrariwise. Later, the model was extended to a version, where similar to the Brix model, the transport is bidirectional, with K_{trans} being the flow constant into the leakage space and K_{ep} the other direction [Tofts, 1997]. Consequently, the Eqn. 2.26 can be written as:

$$v_{le}\frac{dC_{le}(t)}{dt} = K_{trans}C_p(t) - K_{ep}C_{le}(t) \tag{2.29}$$

and the mean concentration in the lesion tissue C_t can be described as:

$$C_t(t) = DK_{trans}\sum_{i=1}^{2} a_i \frac{e^{-(K_{ep}/v_{le})t} - e^{-m_i t}}{m_i - (K_{ep}/v_{le})} + v_p D\sum_{i=1}^{2} a_i e^{-m_i t} \tag{2.30}$$

The mean concentration in lesion tissue, given either by Eqn. 2.28 or Eqn. 2.30 is related to relative signal intensity S_r by Eqn. 2.21.

2.4.5 Comparison between Brix and Tofts models

Table 2.3 summarizes the most important differences between Brix and Tofts models [Kiessling et al., 2007]. One of the main differences between these two models is that in Brix model,

	Tofts model	Brix model
C_p	biexponential	single rate constant
Gd-DTPA injection	Bolus	Infusion
Estimated parameters	K_{trans}, K_{ep}/v_e	A, K_{21}, K_{out}
Is T_{10} needed?	Yes	No

Table 2.3: Summary of the most important differences between Brix and Tofts models.

the AIF is not predefined, but rather is defined as an exponentially decaying function on a patient-by-patient basis and is included in model parameters. Tofts model, on the other hand, uses a predefined AIF, which is obtained experimentally.

In Brix model, T_{10} is implicit in the model parameters. But, Tofts model estimates T_{10} from pre-enhanced data. One of the common ways to estimate T_{10} for different tissues is the multiple flip-angle acquisitions technique [Wang et al., 1987, Whitcher and Schmid, 2011]. In this method, multiple acquisitions are performed with different flip angles. Measured signal intensities from different acquisitions are used to estimate the equilibrium signal intensity m_0 and the pre-injection longitudinal relation rate R_{10} with $T_{10} = 1/R_{10}$, using the equation:

$$S(\theta) = \frac{m_0 \sin(\theta)(1 - e^{-TR \cdot R_{10}})}{1 - \cos(\theta) e^{-TR \cdot R_{10}}}, \tag{2.31}$$

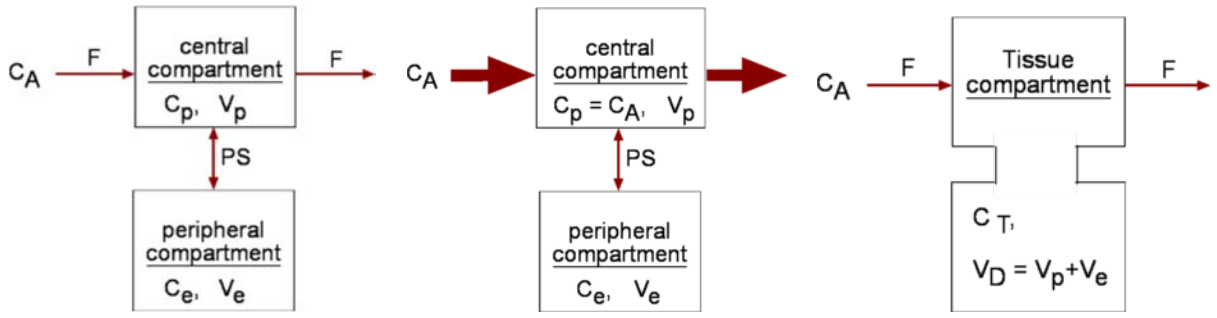
where θ is the flip angle and $S(\theta)$ is the corresponding observed signal intensity.

2.4.6 Brix open two-compartment model and nested compartment models

In contrast to the early Brix model [Brix et al., 1991], in the Brix open two-compartment model [Brix et al., 1999, 2009, 2010] the transfer constants K_{12} and K_{21} in Figure 2.6 are equal in both directions. Thus, we have $K_{12} = K_{21}$ and $K_{in} = K_{out} = \tilde{F}$. The Eqn. 2.17 and Eqn. 2.18 are changed to:

$$V_p \frac{dC_p}{dt} = \tilde{F}(C_A - C_p) - PS(C_p - C_e), \tag{2.32}$$

$$V_e \frac{dC_e}{dt} = PS(C_p - C_e), \tag{2.33}$$



(a) (b) (c)

Figure 2.8: A schematic illustration of the open two-compartment model and the two reduced models. (a) open two-compartment model, (b) permeability-limited compartment model, and (c) flow-limited compartment model. [Brix et al., 2009]

where the concentration of contrast agent in an artery C_A is obtained directly from dataset. The parameter $\tilde{F}$ is the capillary plasma flow and PS is the permeability-surface area product. Using Eqn. 2.32 and Eqn. 2.33 together with Eqn. 2.12, the tissue parameters $\tilde{F}/V_T$, PS/V_T, $v_p = V_p/V_T$, and $v_e = V_e/V_T$ can be estimated for a given C_A and C_t. The volumes V_p, V_e, and V_T are plasma compartment volume, central compartment volume, and examined tissue volume, respectively.

The open two-compartment model is later extended to a nested compartment model, which consists of several compartment models. A full two-compartment model, and two or more reduced models are introduced to describe different tissue scenarios [Brix et al., 2009, 2010]. Brix et al. [2009] included, for example, two reduced models, the permeability-limited compartment model Figure 2.8(b) and the flow-limited compartment model Figure 2.8(c). The permeability-limited compartment model describes a model in which the plasma flow is so high that the concentration-time curve in the intravascular plasma compartment cannot be distinguished from the arterial input function $C_p \cong C_A$. Thus, the estimated parameters are reduced to three parameters PS/V_T, v_p, and v_e. The flow-limited compartment is for the complementary scenario in which the transfer of the administered contrast agent between the intravascular and central compartment is fast when compared to the capillary plasma flow. Therefore, the distribution space of the contrast agent in the tissue of interest can be approximated by a single tissue compartment with the relative distribution volume $v_D = v_P + v_e$. The number of parameters is reduced to two parameters, $\tilde{F}/V_T$, v_D. A schematic illustration of the open two-compartment model and the two reduced models is

depicted in Figure 2.8.

For the model ranking from the set of nested models, the Akaike's information criterion (AIC) [Akaike, 1973] is used. The AIC penalizes for the number of model parameters in order to balance bias against variance.

Due to the increased number of parameters in open two-compartment model and the complexity of the nested compartment models, the parameters' estimation can be unstable and strongly dependent on the initial parameters of the fitting algorithm. The noise of signal curves influences this instability. Thus, for nested compartment models, an enhancement in the signal to noise ratio is of a great interest. As mentioned before, the representative of a group of signal curves has an improved signal to noise ratio. Therefore, in a complex pharmacokinetic model, it makes sense to cluster the signal curves into the groups of 'similar' signal curves instead of voxel-wise analysis of the signal curves.

2.5 Summary

In this chapter, the concept of perfusion DCE-MRI has been discussed. As mentioned earlier, contrast agent changes the signal in MRI by shortening the T_1 and T_2 tissue parameters. The shortening effect is related to the concentration of contrast agent in tissue. Thus, in DCE-MRI, signal changes related to the agent concentration changes are available in signal curves. On the other hand, the concentration change of the agent is related to micro circulation and angiogenesis of the tissue. Tumors develop a peculiar vascular structure during their growth. The relationship between signal changes and microcirculation in the tissue seems to provide a non-invasive possibility to characterize the tumors. The signal curves have been studied in two directions. Either arbitrary features are selected from signal curves and studied if they are characteristics of different types of tumors or pharmacokinetic models are used to extract tissue parameters from signal curves. In both kinds of studies, it is of great interest to be able to group the voxels with a similar course of signal curves together. With clustering of similar signal curves, a better signal to noise ratio can be achieved. The pattern of the sub groups and the number of clusters in the tissue of interest could be themselves characteristics of different types of tumors. The number of clusters and the representative signal curves of the clusters provide valuable information for understanding the complexity of the tissue and designing more accurate pharmacokinetic models. The goal of present work is to provide and discuss methods for

cluster analysis of DCE-MRI signal curves.

Chapter 3

General cluster analysis for DCE-MRI

Each voxel in the DCE-MRI image is related to a signal curve. The signal curve demonstrates the enhancing pattern measured over time for the corresponding voxel. Thus, the DCE-MRI dataset consists of a set of signal curves. Hence, a signal curve is measured in m subsequent time points, each signal curve can be considered a data point in an m-dimensional data space. In this sense, the dataset can be described with a matrix, where the columns of the matrix are the vectors of signal curves. An example for a voxel in DCE-MRI and its corresponding signal curve are depicted in Figure 3.1.

DCE-MRI signal curves have been studied by various authors, mainly in two directions. A part of these investigations has been concerned with extracting physiologically meaningful parameters by fitting the DCE-MRI curves to pharmacokinetic models. In the other direction, signal curves have been studied based on their shapes and extracted features

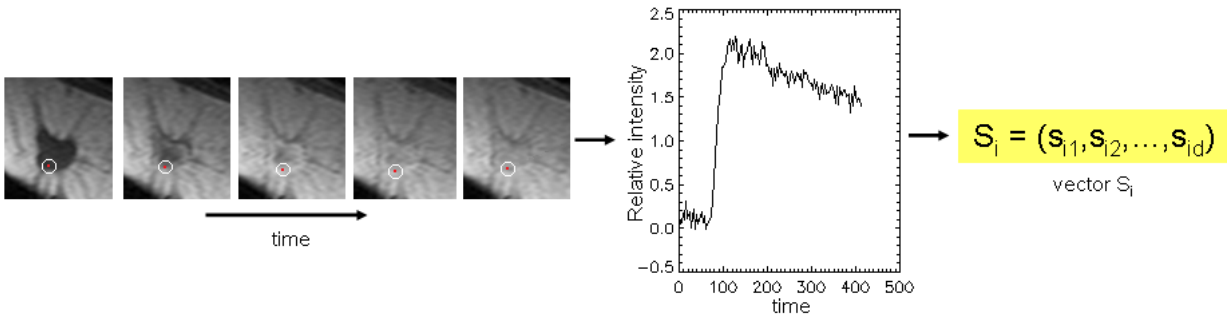


Figure 3.1: A voxel in DCE-MRI and its corresponding signal curve

without using any pharmacokinetic models.

Examples of pharmacokinetic models are discussed in section 2.4. The pharmacokinetic approach, besides difficulties to define the proper model, suffers from general fitting problems such as proper initialization of fitting algorithm, numerical uncertainty, and finding a proper arterial input function.

Non-pharmacokinetic analysis of DCE-MRI signal curves can be divided into two main subgroups of supervised and unsupervised clustering methods. These methods can be used directly on DCE-MRI signal curves either by considering signal curves as vectors in m-dimensional vector space or by extracting m' features ($m' < m$) and considering m'-dimensional subspace of features. Dimension reduction can be achieved by extracting features from signal curves such as area under the curve or slope of rapid enhancement phase, by principal component analysis, or vector quantization.

In cluster analysis of DCE-MRI signal curves, three major questions are of interest. The first question is related to the proper choice of the clustering method. During the last decades, a huge number of clustering methods have been developed. Each of these clustering methods has its own advantages and drawbacks. The second major question is related to the number of clusters in a dataset. Actually, it is *a priori* unknown how many clusters should be distinguished in a tumor area or other surrounding tissues. The third major question is how to measure the similarity or dissimilarity between two signal curves. The conventional clustering methods, especially those based on Euclidean distance as a dissimilarity measure, are not satisfying. A comprehensive study using Euclidean distance as similarity measure and comparison with other measures can be found in Chapter 4.

This chapter gives an overview of some of the clustering methods, and the dimensionality reduction methods mostly used in literature in the case of DCE-MRI clustering. The theoretical aspects of these methods are discussed and compared. The two other important aspects of clustering would be the choice of the number of clusters and the choice of proper similarity measures, these are also discussed. Three popular similarity measures, Euclidean distance, cosine measure, and correlation coefficient, are described as well. As alternative to these three measures, a new similarity measure, named as the parallelism measure (PM), is introduced.

3.1 Dimension reduction

Since computation is less expensive in a reduced vector space, it is common to reduce the dimension of a dataset before the clustering algorithm is performed. Furthermore, it is often the case that many variables of a sample are strongly correlated and thus redundant. There are several methods to reduce the dimensionality of a dataset. In general, the goal is to eliminate the unnecessary dimensions and minimize the loss of significant information at the same time. For this, the principal component analysis is a common choice. The principal component analysis transforms the dataset linearly to a lower dimension. Another possibility is to extract arbitrary features from original dataset.

The following example demonstrates how feature extraction can be used to reduce the dimension of a dataset. For each voxel in a DCE-MRI dataset, the corresponding signal curve $v = \{v_1, ..., v_m\}$ can be considered as a function $f : \mathbb{R} \to \mathbb{R}$, which maps the indices $\{1, ..., m\}$ to the vector components $\{v_1, ..., v_m\}$. Now, one can fit a polynomial to this function. A cubic polynomial, for example, is described by four parameters. Thus, fitting each signal curve with a cubic polynomial reduces the m-dimensional vector space of dataset to a four-dimensional vector space.

In the following two sections, the principal component analysis is described briefly and the possibility of feature extraction of DCE-MRI dataset is discussed.

3.1.1 Principal component analysis

The principal component analysis (PCA) transforms a dataset with a linear orthogonal transformation, so that the transformed dataset has the largest variance along the first coordinate in the new coordinate system [Shlens, 2005]. With other words, for a dataset X orthogonal matrix P needs to be calculated, so that:

$$Y = PX. \tag{3.1}$$

The resulting matrix Y is the transformed dataset in the new coordinate system. The rows of P are called the principal components of X. The samples of the transformed dataset, i.e. columns of Y, have the maximal variance along the first principal component. Moreover, the components of the samples in Y are less correlated. In general, the correlation coefficients of the components of the vectors in a dataset are presented in out of diagonal

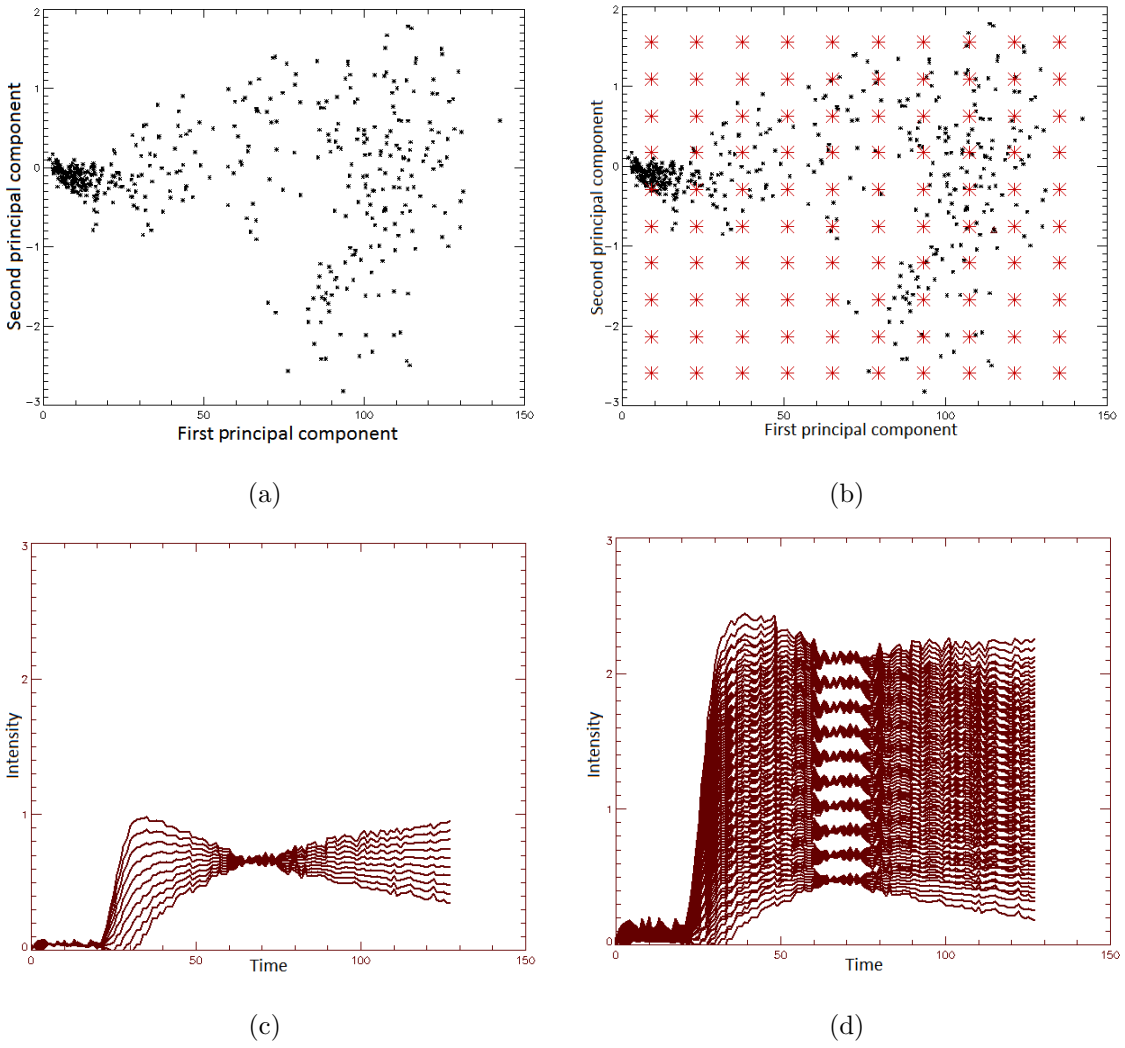


Figure 3.2: (a) data points of the tumor region of a breast DCE-MRI dataset are projected on the plane of the first and second principal components. (b) Red stars are points on a grid on the plane of first and second principal components. (c) The first column of grid points, from left, is transformed into original data coordinates and the corresponding signal curves are depicted. (d) Corresponding signal curves of all grid points.

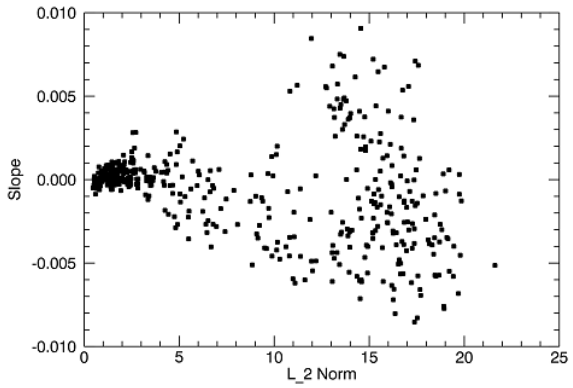
positions of the correlation matrix C of that dataset. For transformed dataset Y, the correlation matrix is given by the equation $C_Y = YY^T$. To achieve the minimal correlation between the vector components of Y, C_Y needs to be a diagonal matrix. Substituting $Y = PX$ in $C_Y = YY^T$, we have:

$$C_Y = PC_XP^T, \tag{3.2}$$

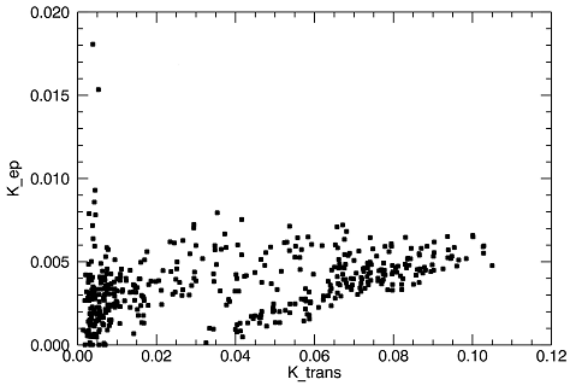
where $C_X = XX^T$ is the correlation matrix of X. To diagonalize the matrix C_Y, it is sufficient to set P equal to the transposed matrix of eigenvalues of C_X. The matrix C_X is a symmetric matrix, and therefore, diagonalizable by its eigenvectors E and a diagonal matrix D. As a result, we have $C_X = EDE^T$ and thus, $C_Y = P(EDE^T)P^T$. By selecting $P = E^T$ and due to the fact that P is orthogonal, we have $P^T = P^{-1}$ and thus, $C_Y = D$.

Sorting the rows of P properly, so that the corresponding eigenvalues decrease with the row number, the first principal component corresponds to the largest eigenvalue and is the best-fit line of the dataset. It is supposed that dataset X is zero centered. If not, the center of the dataset should be translated to zero. Dataset X is an $m \times n$ matrix, where n is the number of samples or data points and m is the number of variables. Accordingly, Y is an $m \times n$ matrix. Considering only the first $m' < m$ rows of Y instead of the full data, we obtain the reduced dataset $Y_{m'}$ as a projection on the subspace of first m' principal components. Analog, we get reduced dataset $X_{m'} = P_{m'}^T Y_{m'}$, with $P_{m'}^T$ being the transpose of the reduced matrix $P_{m'}$ of the first m' principal components.

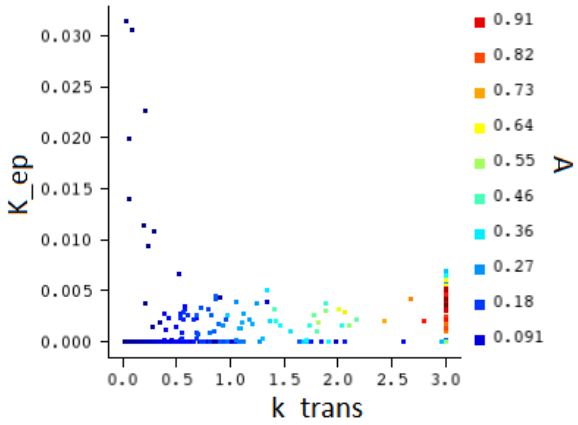
Principal component analysis and its nonlinear extension, kernel PCA, were performed on DCE-MRI of breast datasets by Twellmann et al. [2004] for a better visualization of the tumor region in a reduced dataset. Eyal et al. [2009] used PCA as preprocessing and dimension reduction for further analysis. In principal component analysis of the DCE-MRI datasets of breast, listed in Table C.1, can be observed that the first and second principal components, corresponding to the two largest eigenvalues, are strongly related to the peak variation of signal curves as well as to the shape variation during washout phases. Figure 3.2 demonstrates this observation in a DCE-MRI dataset. Figure 3.2(a) shows the projected data points on the plane of first and second principal components. The red stars in Figure 3.2(b) represent a 10×10 grid of points on the same plane over the data domain. The grid points of the first column at the left side in Figure 3.2(b) are transformed by P^T in the original data coordinates, and the corresponding signal curves are depicted in Figure 3.2(c). Signal curves of all grid points are depicted in Figure 3.2(d). While the



(a) 2D feature space of L_2-norm and slope of the washout phase for a dataset.



(b) 2D feature space of Tofts model for a dataset.



(c) 3D feature space of the Brix parameters for a dataset.

Figure 3.3: Three feature spaces for signal curves of a tumor region in DCE-MRI of breast.

peaks of the signal curves vary along the first principal component, moving along the second principal component results in different washouts. As demonstrated in this example, PCA can be a promising technique to perform dimension reduction in DCE-MRI datasets.

3.1.2 Features extraction

Feature extraction is an alternative method besides PCA to reduce the dimension of a dataset. Since DCE-MR imaging exists, many investigations have been performed to extract features from signal curves which are characteristic of different kinds of tumors [Da-

ducci et al., 2009, Eyal et al., 2009, Guo and Reddick, 2009, Meyer-Baese et al., 2008, Schlossbauer et al., 2008]. Area under the curve, peak of the curve, and slope of the rapid enhancement phase are common choices.

In a recent paper [Mohajer et al., 2010], a feature space consisting of two features has been discussed. One of these features is the so-called l_2-norm of the signal curves, which is related to the area under the curve. The other one is an extracted feature called reverse value of the signal curve. The reverse value tries to take the diversity of the shapes into account. For clustering, the 2D feature space of the signal curves is divided into sub-regions by means of a grid. Another simple feature, which tries to distinguish between different forms of the washout phases, is the slope of the linear regression line of the washout phase.

Other possibilities are fitting of parameterized functions to the signal curves. The space of parameters can be considered as the feature space. Figure 3.3 demonstrates three different feature spaces of the same set of signal curves. Figure 3.3(a) shows the feature space with l_2-norm used as one parameter and the slope of the washout phase as the second parameter. The results are quite similar to those obtained from projection of the same signal curves on the plane of first and second principal components depicted in Figure 3.2(a). In Figures 3.3(b) and 3.3(c), for the same signal curves the fitting parameters calculated from Tofts model and early Brix model are presented, respectively. The Brix model has three parameters. Values of the third parameter A are shown in Figure 3.3(c) with different colors.

3.2 Clustering methods

During the last decade, different authors either tried different cluster analysis on DCE-MRI datasets or performed different machine learning algorithms to classify tumors. In the case of breast imaging, experiments have been performed only on low temporal resolution datasets consisting of only 5 to 6 slices [Daducci et al., 2009, Eyal et al., 2009, Guo and Reddick, 2009, Meyer-Baese et al., 2008, Schlossbauer et al., 2008]. In the majority of these works, clustering has been applied on features extracted from signal curves, such as area under the curves or the slope of the washout phase. In all these investigations, Euclidean distance is the measure of dissimilarity.

On the other hand, datasets from perfusion imaging have high temporal resolution, with volume acquisition in time intervals of less than 4 seconds. High temporal resolution

images introduce new aspects in clustering of DCE-MRI, which were not present and therefore, not discussed in previous studies. In summary, these aspects introduce a higher complexity of the clustering problem due to the higher temporal and spatial resolution. Higher temporal resolution leads to a higher information content of the signal curves. This implies the necessity of defining new features or searching for new distance measures.

The clustering methods used in literature for cluster analysis of DCE-MRI datasets can be divided into two major groups of supervised and unsupervised clustering methods. In supervised clustering methods, the dataset is trained according to *a priori* knowledge gained by experts. Therefore, a large number of training datasets and an objectively gained experts' knowledge are required. Examples of supervised clustering methods [Bishop, 2006, Hastie et al., 2005] are support vector machine and neuronal networks. For example, neural network was implemented for classification of carcinoma tumors based on training data with three defined classes [Lucht et al., 2001].

Providing the experts' knowledge is a difficult and time-consuming task. Supervised methods have problems to handle irregularity and new phenomena in the case of new untrained datasets. For these reasons, supervised clustering is not used in present work.

In contrast to supervised clustering methods, unsupervised clustering methods do not need any experts' knowledge. Unsupervised clustering methods are an important part of unsupervised learning. The idea behind these methods is to find underlying structures in unlabeled data. Every unsupervised clustering is based on some assumptions or the choice of one or more parameters. One of the main parameters in unsupervised clustering is the number of clusters. Accordingly, unsupervised clustering methods can be divided into two subcategories. The category of those methods, where the number of clusters is given directly as a parameter. The other category are clustering methods in which the number of clusters is not predefined. In a series of well-used clustering methods, such as EM, k-means, and ICA, the number of clusters should be given as a parameter. In contrast, there are clustering methods in which the number of clusters depends on the choice of a threshold. Examples of threshold based clustering methods are hierarchical clustering methods and mean shift clustering.

3.3 Unsupervised clustering methods with number of clusters as a parameter

In this section, four different clustering methods are described. Section 3.3.1 introduces the expectation maximization algorithm as a basis for mixture based clustering algorithms. The popular and well-used k-means algorithm is discussed in section 3.3.2. Section 3.3.3 continues with fuzzy c-means clustering as an improvement of k-means algorithm. Independent component analysis is introduced as an alternative to mixtures based clustering algorithms in section 3.3.4.

3.3.1 Expectation maximization algorithm

Expectation maximization algorithm (EM) is a fundamental and general method of clustering, which is the basis of many other clustering methods. In this work, the EM algorithm is not used as a clustering method. It is discussed here for a better understanding of other algorithms such as k-means and mean shift clustering methods. The expectation maximization algorithm assumes that the dataset is described by a statistical model consisting of a set of parameters Θ and probability functions P_Θ of the form,

$$P_\Theta(x_1, ..., x_N, z_1, ..., z_N) = \prod_{t=1}^{N} P_\Theta(z_t) P_\Theta(x_t|z_t). \tag{3.3}$$

In this equation, $x_1, ..., x_N$ are the observations from a dataset with underlying mixture distributions and $z_1, ..., z_N$ are the set of unobserved latent data. The missing latent data determine from which distribution the observation originates. For each observation x_i, we have the conditional probability $P_\Theta(x_i|z_i = j)$, where $j = 1, .., k$ and $P_\Theta(z_i = j) = \tau_j$. The goal is to determine parameters Θ and τ, so that for each observation, the corresponding distribution can be determined.

As a simple case, let consider a probability model that is a mixture of Gaussian distributions: $\Theta = \langle \mu_1, ..., \mu_k, \Sigma_1, ..., \Sigma_k \rangle$, with μ_i center of mixture i and Σ_i covariance matrix of mixture i. The conditional probability $P_\Theta(x_i|z_i = j)$ can be rewritten as $\mathcal{N}(x_i|\mu_j, \Sigma_j)$. The marginal distribution function of observation x_i is then the sum of the joint distributions over all possible values of z_i:

$$P_\Theta(x_i) = \sum_{j=1}^{k} \tau_j \mathcal{N}(x_i|\mu_j, \Sigma_j). \tag{3.4}$$

It follows that the joint probability for x_i and z_i, which is the same as likelihood function, can be written as:

$$P_\Theta(x_1, ..., x_N, z_1, ..., z_N) = \prod_{i=1}^{N} P_\Theta(x_i)$$
$$= \prod_{i=1}^{N} \sum_{j=1}^{k} \tau_j \mathcal{N}(x_i|\mu_j, \Sigma_j). \tag{3.5}$$

The parameter set Θ can be estimated by maximizing the likelihood function expressed in Eqn. 3.5 for a given state of z_i. For simplicity, normally the logarithm of the likelihood (log likelihood) is used instead of likelihood function itself:

$$lnP_\theta(x|\mu, \Sigma, \tau) = \sum_{i=1}^{N} ln \left\{ \sum_{j=1}^{k} \tau_j \mathcal{N}(x_i|\mu_j, \Sigma_j) \right\}. \tag{3.6}$$

The maximization is achieved by setting the derivatives of Eqn. 3.6 to zero. The derivative of Eqn. 3.6 with respect to the means μ_j can be written as [Bishop, 2006]:

$$-\sum_{i=1}^{n} \underbrace{\frac{\tau_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}{\sum_{k=1}^{K} \tau_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}}_{\gamma(\tau_j)} \Sigma_j (x_i - \mu_j) = 0. \tag{3.7}$$

The expression $\gamma(\tau_j)$ is called responsibility, which is the prior probability of $z_i = j$ for the given observation x_i. Multiplying both sides in Eqn. 3.7 by the inverse of the covariance matrix Σ_j, we obtain:

$$\mu_j = \frac{1}{N_j} \sum_{i=1}^{n} \gamma(\tau_j) x_i, \tag{3.8}$$

where:

$$N_j = \sum_{i=1}^{n} \gamma(\tau_j). \tag{3.9}$$

Equivalently, applying the derivation with respect to covariance Σ_j in Eqn. 3.7, we obtain [Bishop, 2006]:

$$\Sigma_j = \frac{1}{N_j} \sum_{i=1}^{n} \gamma(\tau_j)(x_i - \mu_j)(x_i - \mu_j)^T. \tag{3.10}$$

In the same manner, by applying the derivation with respect to mixing coefficients τ_j in Eqn. 3.7, we obtain [Bishop, 2006]:

$$\tau_j = \frac{N_j}{n}. \tag{3.11}$$

For EM algorithm, as an unsupervised clustering method, two main assumptions are necessary. The first assumption is that the dataset can be described by the chosen probability model. The second assumption is about the number of underlying mixture distributions. The number of mixtures can be understood as the number of clusters. In the case of Gaussian mixture, the algorithm should be initialized with some guess values for μ_j, Σ_j, and τ_j, where μ_j and Σ_j are mean value and covariance matrix of the distribution j, and τ_j is the mixing coefficient. The algorithm consists of the following steps:

1. Initialization: Choose some starting values for the means μ_j, covariances Σ_j and mixing coefficients τ_j and evaluate the initial value of the log likelihood.

2. E step: Evaluate the responsibilities $\gamma(\tau_j)$ from Eqn. 3.7 by using current parameters Σ_j, and μ_j.

3. M step: Re-estimate the parameters Σ_j, μ_j, and τ_j using the new calculated responsibilities from step 2 and Eqn. 3.8, Eqn. 3.10 and Eqn. 3.11.

4. Evaluate the log likelihood expressed in Eqn. 3.6 and check for convergence of either the parameters or the log likelihood. Convergence could be verified by testing whether the changes in the log likelihood or the parameters are smaller than a selected threshold. If the convergence criterion is not satisfied, return to step 2.

Another way to look at the EM algorithm is to consider X, the set of all observations together with Z, the set of all unknown coefficients z_i as the actual dataset. As before, for each observation x_i, we have the conditioned probability $P_\Theta(x_i|z_i = j)$, where $j = 1, .., k$ and $P_\Theta(z_i = j) = \tau_j$ so that the joint density function for Z the set of all z_i is:

$$P_\Theta(Z) = \prod_{j=1}^{k} \tau_j, \tag{3.12}$$

and the conditional density function of X the set of all observations for a given Z can be written as:

$$P_\Theta(X|Z) = \prod_{j=1}^{k} P_\Theta(X|\mu_j, \Sigma_j), \tag{3.13}$$

and therefore, the joint density function of all observations X takes the form:

$$P_\Theta(X) = \prod_{j=1}^{k} \tau_j P_\Theta(X|\mu_j, \Sigma_j). \tag{3.14}$$

The likelihood function of the sets X and Z for a given set Θ of Gaussian mixture parameters can be written as:

$$P_\Theta(X, Z|\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{k} \tau_j \mathcal{N}(x_i|\mu_j, \Sigma_j). \tag{3.15}$$

Thus, the log likelihood is relieved from the logarithm of the inner summation as was given in Eqn. 3.6:

$$\begin{aligned} \ln P_\Theta(X, Z|\Theta) &= \sum_{i=1}^{N} \sum_{j=1}^{k} \ln \{\tau_j \mathcal{N}(x_i|\mu_j, \Sigma_j)\} \\ &= \ln(\tau(j)) - d/2 \ln(2\pi) - \frac{1}{2} \ln(|\Sigma_j|) - \frac{1}{2}(x_i - \mu_j)^T (\Sigma_j)^{-1}(x_i - \mu_j). \end{aligned} \tag{3.16}$$

The computation complexity of each step of EM algorithm can be up to $O(nkp^2)$, where n is the number of observations, k is the number of clusters and p is the dimension of data space. This depends on required matrix multiplication and on the type of matrix (diagonal or not). Thus, for r iterations, the computation complexity of the algorithm can be in the order $O(rnkp^2)$. For large n and p, EM algorithm can be quite time consuming.

In the case of DCE-MRI datasets, there are two reasons besides the computational complexity which make EM algorithm inappropriate. In fact, the number of distributions and the underlying mixture model for the dataset should be known.

3.3.2 K-means

K-means algorithm is a well-used and easy to implement algorithm. It is a partitioning clustering method, where the number of clusters k is a given fixed parameter. The k-means algorithm is based on the following consideration. Let D be a dataset with n observations, and C_1, C_2, ..., C_k, k disjoint clusters in D. The error function E, also called objective function, is defined as:

$$E = \sum_{j=1}^{k} \sum_{x \in C_j} d(x, \mu(C_j)), \tag{3.17}$$

where $d(x, \mu(C_j))$ is the distance between observation x and the centroid $\mu(C_j)$ of the cluster C_j. In Eqn. 3.17, the Euclidean distance is a common choice for the distance measure. The goal of k-means algorithm is to minimize the objective function. Thus, k-means algorithm can be considered as an optimization problem, which tries to find the centroids of the clusters so that Eqn. 3.17 is minimized.

The algorithm can be described in following steps:

1. Initialization: The cluster centroids $\mu(C_j)$ are initialized to starting values. For example, they could be initialized with uniformly random values from the domain of the data points.

2. Assignment: Each data point is assigned to the closest centroid $\mu(C_j)$ by minimizing $\|x - \mu(C_j)\|$. Finally, the new clusters C_j are built as:

$$C_j = \{\forall\ x_i \mid \|x_i - \mu(C_j)\| \leq \|x_i - \mu(C_l)\|\,,\ l = 1, ..., k\}\,. \tag{3.18}$$

3. Update: The centroids $\mu(C_j)$ are updated for the newly built clusters C_j from the previous step,

$$\mu(C_j) = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i. \tag{3.19}$$

The k-means algorithm is a special case of the expectation maximization algorithm. In k-means clustering, each z_i from Eqn. 3.3 is a class label for the data point x_i, and therefore, $z_i \in \{1, ..., K\}$. The k-means probability model is a $\mathcal{N}(\mu_{z_i}, I)$. That is, each cluster is a Gaussian distribution with mean value μ_{z_i} and identity matrix I as covariance matrix [Bishop, 2006]. Simplifying Eqn. 3.16 for this special case, we get:

$$\ln P_\Theta(X, Z|\Theta) = \sum_{i=1}^{N}\sum_{j=1}^{k} \ln(\tau_j) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\underbrace{\sum_{i=1}^{N}\sum_{j=1}^{k} \|x_i - \mu_j\|}_{A}\,. \tag{3.20}$$

Maximizing the log likelihood function, is equivalent to minimize the term A in Eqn. 3.20. The term A acts as the error function in the k-means algorithm.

There are some considerations about k-means algorithm. First of all, k-means clustering with the objective function represented in Eqn. 3.17 and Euclidean distance as the similarity measure is looking for convex spherical clusters. The clusters are expected to be of equal size. As a heuristic algorithm, the result depends on the initial clusters. Thus, there is no guarantee that the algorithm converges to the global optimum.

The example shown in Figure 3.4 demonstrates problem of k-means algorithm related to initialization step. The dataset consists of three not well-separated clusters. The first and second clusters are of equal size, and each of them contains $n = 50$ data points. The data points are driven from 2D normal distributions $\mathcal{N}(\mu_1, \delta I)$ and $\mathcal{N}(\mu_2, \delta I)$ with

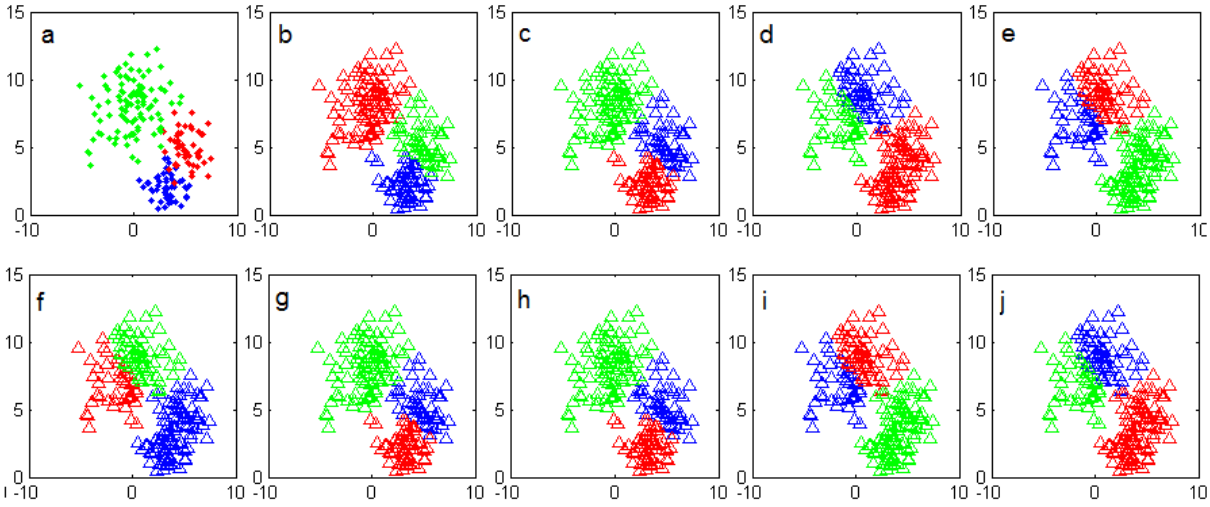


Figure 3.4: Image (a) depicts the original clusters of the simulated dataset. The next 9 images (b-j) depict the results of 9 executions of k-means clustering for $k = 3$.

$\mu_1 = (3, 4)$, $\mu_2 = (5, 5)$ and $\delta = 1$. The third cluster consists of $2n$ data points driven from a normal distribution $\mathcal{N}(\mu_3, 2\delta I)$ with $\mu_3 = (0, 7)$. In Figure 3.4, the first image (a) depicts these clusters. The other nine images (b-j) in Figure 3.4 show the results obtained from different runs of k-means algorithm with $k = 3$ and random starting points on the simulated dataset. It can be observed that in five cases out of nine, the algorithm was not able to find the correct clusters. Performing the k-means algorithm 100 times on same dataset, in 50% of the cases a correct labeling could be achieved.

The computation complexity of the k-means algorithm for n observations, k clusters, and r iterations can sum up to $O(rnk)$, which is almost linear for small r and k. The algorithm is easy to implement and is relative fast comparing to other algorithms. However, in complex datasets with unknown mixtures, the k-means clustering can easily lead to unrealistic results, which do not represent the correct underlying patterns in the dataset.

3.3.3 Fuzzy c-means

One of the main problems of the k-means algorithm is that it tends to convert to a local optimum. This behavior depends strongly on the choice of starting points. The main reason, why k-means algorithm converts into local optima, is that only data points inside a cluster are considered, after cluster centers are calculated and labeled. The fuzzy c-means

clustering can be considered as an extended version of k-means clustering, which tries to overcome this shortcoming. The c-means clustering takes account of all data points to determinate cluster centers.

Comparing the error function of fuzzy c-means clustering, for k clusters and N data points,

$$E = \sum_{j=1}^{k} \sum_{i=1}^{N} w_{ij} \left\| x_i - \mu_j \right\| \tag{3.21}$$

to that of k-means clustering presented in Eqn. 3.17, it can be recognized that weighting parameters w_{ij} are introduced. In addition, the interior sum of the objective function is not restricted to the data points inside the cluster j but rather is extended to all data points in the dataset. The weights w_{ij} for the observation x_i in relation to the cluster C_j are defined as:

$$w_{ij} = \frac{1}{\sum_{m=1}^{k} \left(\frac{\|x_i - \mu_j\|}{\|x_i - \mu_m\|} \right)^{2/(d-1)}}, \tag{3.22}$$

where the parameter d controls the influence of observations in the whole clustering process.

Considering the case $d = 1$. For $d \to 1$ the exponent $2/(d-1)$ in Eqn. 3.22 goes to infinity. Let x_i be a data point and μ_j the closest cluster center to x_i. For $m = j$, we have $\left(\frac{\|x_i - \mu_j\|}{\|x_i - \mu_m\|} \right) = 1$ and for $m \neq j$ we have $\left(\frac{\|x_i - \mu_j\|}{\|x_i - \mu_m\|} \right) < 1$. Thus, for $d = 1$, w_{ij} is equal to 1. If μ_j is not the closest center, there is at least one m with $\left(\frac{\|x_i - \mu_j\|}{\|x_i - \mu_m\|} \right) > 1$, so that w_{ij} is equal to zero. This explanation shows that for $d = 1$ the fuzzy c-means algorithm converts to k-means clustering. If $d \to \infty$, $w_{ij} \to 1/k$ and all w_{ij} convert to the same value. Obviously, the parameter d controls the influence of data points in the resulting cluster centers. In this sense, parameter d describes the fuzziness of the algorithm.

The algorithm has similar steps as by k-means:

- Initialization: Cluster centers μ_j are initialized to starting values. For example, they are initialized with uniformly random values within the domain of data points.
- Assignment: Weights w_{ij} are calculated for each data point x_i according to Eqn. 3.22.
- Update: Centroids μ_j are recalculated using w_{ij} from previous step,

$$\mu_j = \frac{\sum_{i=1}^{N} w_{ij} x_i}{\sum_{i=1}^{N} w_{ij}}. \tag{3.23}$$

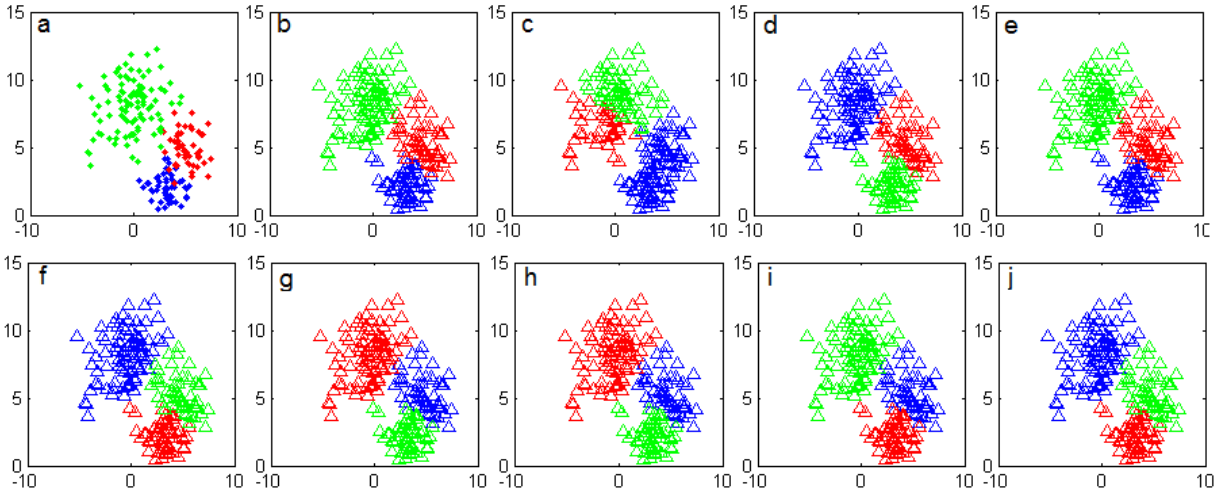


Figure 3.5: Image (a) depicts the original clusters of the simulated dataset. The next 9 images (b-j) depict the results of 9 executions of fuzzy c-means clustering with $k = 3$.

Figure 3.5 shows the results achieved by performing fuzzy c-means clustering with $d = 2$ on the simulated dataset from section 3.3.2. The first image (a) depicts the clusters. The other nine images (b-j) represent the results obtained from different runs of c-means algorithm. In contrast to k-means clustering, performing the fuzzy c-means algorithm, 100 times on the same dataset, in 85% of the cases a correct labeling could be achieved. Best results have been achieved for $d = 5$ with 100% performance.

This example clearly demonstrates the improved performance of c-means clustering compared with k-means clustering. In addition, since all data points are included in updating the cluster centers, the computation time is increased. The complexity of the algorithm for n observations and k clusters performed in r iterations can sum up to $O(rnk^2)$ while r is usually bigger than in k-means clustering due to the slower convergence of the algorithm. Although fuzzy c-means introduces a more general form of k-means clustering, it still suffers from some of the problems of k-means algorithm. Fuzzy c-means also searches for spherical normal distributions of the equal size. The weights w_{ij} are closely related to the probabilities τ_j in EM. Thus, fuzzy c-means clustering can also be considered as an efficient implementation of EM algorithm for mixture of Gaussians with covariance $\Sigma = \delta I$.

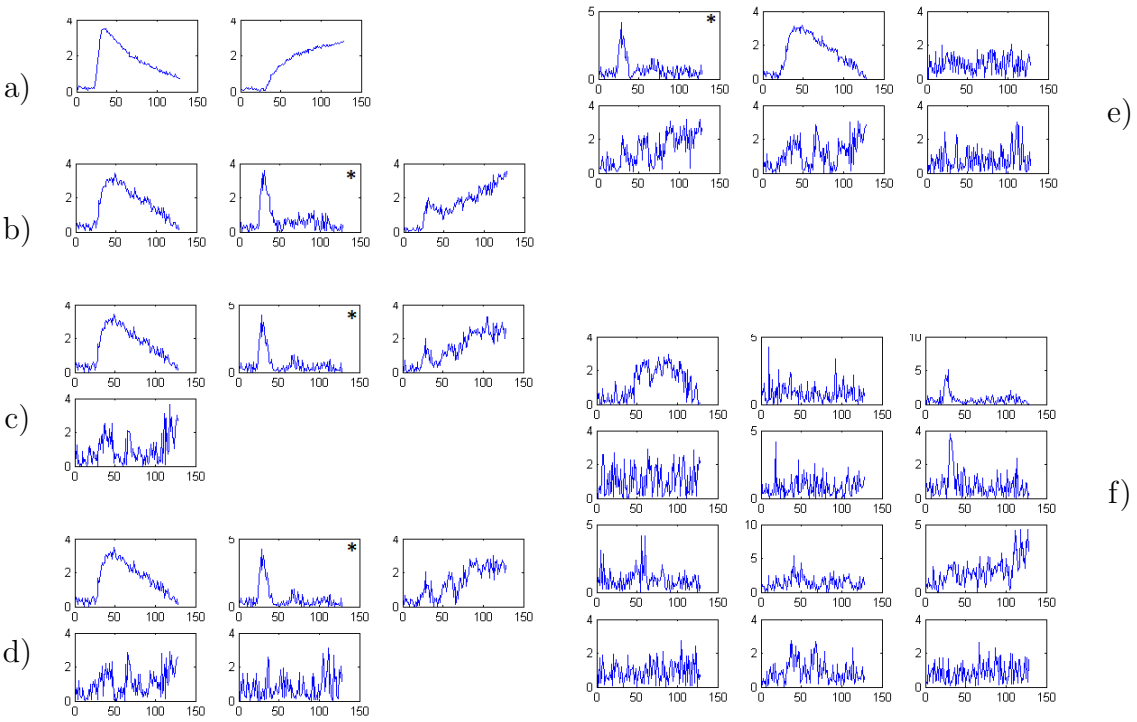
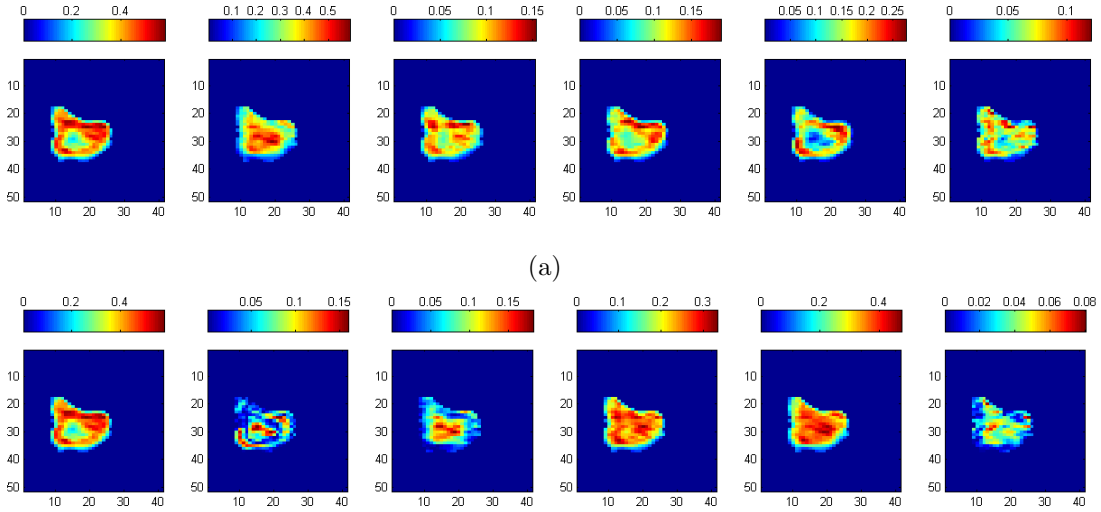


Figure 3.6: Temporal ICA components with different numbers of components are depicted for a tumor region of a DCE-MRI dataset. In images (a) to (f) numbers of ICA components are $2, 3, 4, 5, 6$, and 12 respectively.

3.3.4 Independent component analysis

Independent component analysis (ICA) is a special case of so called blind source separation. Under the assumption of mutual statistical independence of the non-Gaussian source signals, independent component analysis tries to find the linear components of a multivariate signal [Hyvärinen et al., 2001b, Stone, 2004]. It is quite common to use whitening, usually with the eigenvalue decomposition, and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem. The Whitening converts the covariance matrix C of the dataset into the identity matrix I. Whitening and dimension reduction can be achieved with principal component analysis or singular value decomposition.

Linear noiseless ICA can be defined as follows. Suppose observed data consists of n vectors $x_1, ..., x_n$ and these vectors are linear combinations of n components $s_1, ..., s_n$. The goal of ICA is to transform the observed data into the components s_i by a linear



(a)

(b)

Figure 3.7: Two executions of FastICA algorithm in temporal domain are depicted in Figures (a) and (b). The depicted color at each voxel position in the image reflects the coefficient value of the corresponding ICA component for that voxel.

transformation W:

$$s = Wx, \tag{3.24}$$

while $x_i = a_{i,1}s_i + ... + a_{i,k}s_k + ... + a_{i,n}s_n$. The matrix W and the unknown components s_i should be estimated.

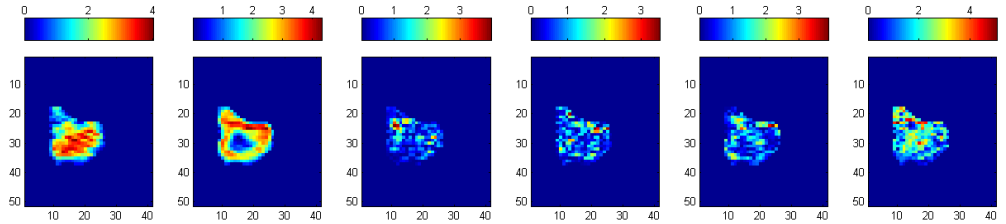
To solve this problem, one answer among others, is to find a linear transformation W, so that the random variables s_i are as independent as possible. It has been shown [Hyvärinen et al., 2001b, Stone, 2004] that the problem can be solved if and only if the components s_i are non-Gaussian. Let $p(s_1, s_2, ..., s_n)$ be the joint probability density function (pdf) and $p_i(s_i)$ the marginal pdf of s_i. Then, s_i are independent if and only if the joint pdf is factorable:

$$p(s_1, s_2, ..., s_n) = p_1(s_1)p_2(s_2)...p_n(s_n). \tag{3.25}$$

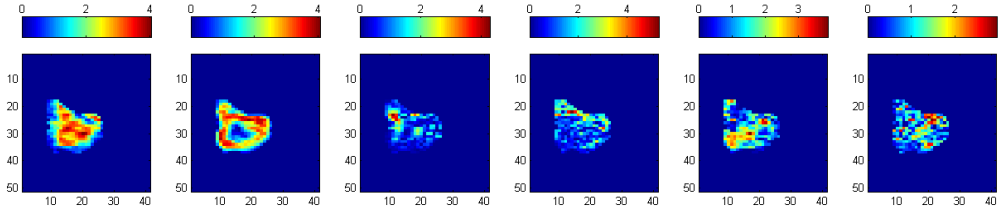
There are several ways to maximize non-Gaussianity, such as kurtosis of a random variable or negentropy as a measure of non-Gaussianity [Hyvärinen et al., 2001b]. Negentropy, e.g., is defined as:

$$J(p_x) = S(\phi_x) - S(p_x), \tag{3.26}$$

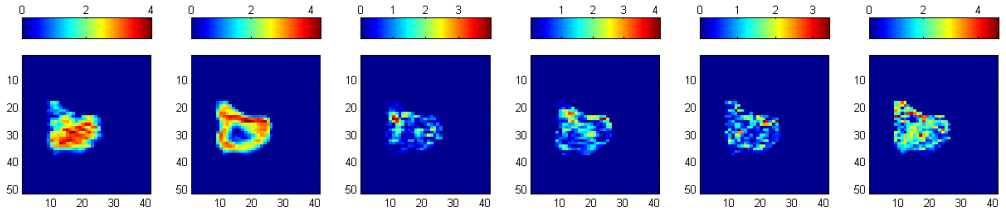
where $S(\phi_x)$ is the differential entropy of the Gaussian density with the same mean and



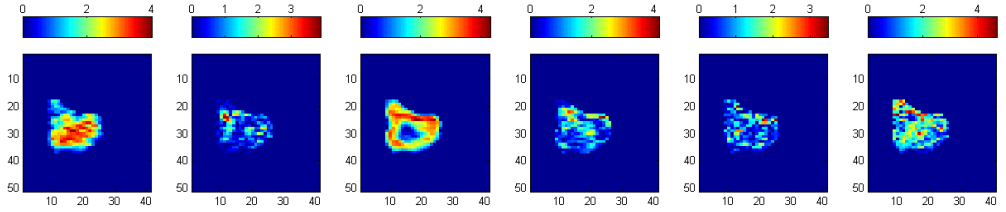
(a) spatial ICA with kernel 'pow'



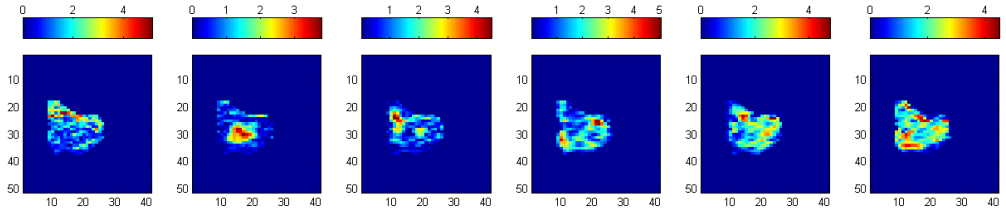
(b) spatial ICA with kernel 'pow'



(c) spatial ICA with kernel 'tahn'



(d) spatial ICA with kernel 'gauss'



(e) spatial ICA with kernel 'skew'

Figure 3.8: Five executions of FastICA algorithm in spatial direction are depicted in Figures (a)-(e). Each image displays an independent component.

the same variance as p_x. $S(p_x)$ is the differential entropy of $p_x = -\int p_x(u) log p_x(u) du$.

The FastICA algorithm is one of the most popular implementations for finding one maximally non-Gaussian direction. It consists of the following steps:

1. Center the mean of the data to zero.
2. Whiten the data to z.
3. Choose an initial vector w of unit norm.
4. Let $w \leftarrow E\left\{zg(w^T z)\right\} - E\left\{g'(w^T z)\right\} w$, where g is, e.g., kurtosis or negentropy.
5. Let $w \leftarrow w/\left\|w\right\|$.
6. If not converged, go back to step 4.

In general, the orders of ICA components are not distinguishable. Furthermore, the number of ICA components is unknown. In addition, each algorithm finds only one of the various but quite different possible answers. Even a single algorithm can result in different components if performed several times with different initializing values. Due to these problems, if expected components and the number of source components are unknown, the application of ICA is limited. However, the main problem of ICA remains the strong assumption of independency among the components.

To overcome these limitations, there are extensions of ICA toward more general forms such as topographic ICA [Hyvärinen et al., 2001a] and tree-dependent ICA [Bach and Jordan, 2002]. Topographic ICA allows for dependencies in the form of known topology among the components while tree-dependent ICA considers dependencies in the form of a tree relationship.

In the case of DCE-MRI datasets, ICA can be applied in two considerable directions: temporal and spatial. Suppose a dataset S consists of n signal curves s_i, where each s_i is a vector in $\mathbb{R}^m$. We can think of each s_i as a linear combination of non-Gaussian independent components c_j such as $s_i = Ac_j$. This is a temporal analysis of the dataset. Alternatively, we can consider dataset S consisting of m vectors $v_{i'}$ in $\mathbb{R}^n$, where each $v_{i'}$ is a linear combination of non-Gaussian independent components $c'_{j'}$ such as $v_{i'} = Bc'_{j'}$. This is the spatial analysis of the dataset.

In temporal analysis, it is supposed that there are different sources of signals, so that the measured signal at each voxel is a combination of these signals. This method can be

used to separate perfusion signal curves from background noise. In spatial analysis, it is supposed that voxels from regions with similar properties can be considered as separate signal sources. Thus, spatial ICA divides each vector $v_{i'}$, a single slice in a dataset S, into several components.

Only a few reports exist about usage of ICA on DCE-MRI datasets. Yoo et al. [2002] used ICA to detect the lesions and to distinguish between benign and malignant. Later, Koh et al. [2008] used ICA for the segmentation of tumor area from the surrounding tissue. Yoo et al. [2002] and Koh et al. [2008] both used ICA in the temporal case. Saalbach et al. [2009] compared the results obtained by ICA, with topographic ICA, and tree-dependent ICA. According to their results, tree-dependent ICA and FastICA with a stability analysis lead to better performances than topographic ICA due to the fact that it is difficult to design a proper topography. In addition, they observed no significant differences between spatial and temporal application of ICA. Even though, the application remains limited to detecting tumors. Additional reports [Calamante et al., 2004, Mehrabian et al., 2011] are published about the possibility to detect local atrial input function (AIF) by means of ICA separation. The evidence of an AIF similar component among temporal ICA components of DCE-MRI have been observed also in the present work. However, such a component is detected only when sufficient numbers of components are selected.

Figures 3.6(a) to 3.6(f) demonstrate six executions of FastICA on a dataset for $2, 3, 4, 5, 6,$ and 12 number of components in the temporal domain. In each of the cases with 3, 4, 5 and 6 components, an AIF similar signal component can be recognized (marked with * in the Figures 3.6b- 3.6e). In ICA with 6 components, the orders of first and second components are interchanged. In ICA with 12 components, no similar components to the components in Figures 3.6(b) to 3.6(e) can be recognized. The ICA with only 2 components shows a strong suppressing of noise. The AIF similar signal is not recognizable in this case.

To study the reproducibility of the results, several executions of ICA in the temporal and spatial domain have been performed on a DCE-MRI dataset. Figure 3.7 demonstrates two consecutive executions in temporal domain using FastICA package [FastICA, 2005]. The Nonlinearity function $g(u)$ was set to 'pow3', which means $g(u) = u^3$. The 'stabilization' was set to 'off'. The dimensionality of dataset was reduced to six by preprocessing the dataset to the space of the six first principal components. In each execution, ICA algorithm has determined six components for the given dataset in temporal domain.

In each image of 3.7(a) and 3.7(b), the coefficient value w_{ij} of each voxel for the corre-

sponding ICA component is depicted as a color. In other words, the color red for a voxel means that the corresponding ICA component for that image is more strongly represented in that voxel. These examples demonstrate that the calculated ICA components and the corresponding separation matrix W can vary from one execution to the other.

Similar experiments in the spatial domain have been performed on the same DCE-MRI dataset. Figure 3.8 compares five executions of FastICA algorithm. In contrast to the temporal domain, the depicted images are the calculated ICA components themselves. Figures 3.8(a) and 3.8(b) demonstrate results obtained from two consecutive executions with the same configuration used in the case of experiments in the temporal domain. In Figures 3.8(c) and 3.8(d), the ICA executions have been performed with different nonlinearity functions, but all other configurations have been the same. In Figure 3.8(c), the nonlinearity function is $g(u) = \tanh(a_1 u)$, called 'tahn' in the FastICA package. In Figure 3.8(d), the nonlinearity function is $g(u) = u \exp(a_2 u^2/2)$, called 'gauss'. In the last experiment, depicted in Figure 3.8(e), $g(u) = u^2$. Here, the 'stabilization' was turned on, otherwise, the algorithm was not able to convert.

Except for the last execution shown in Figure 3.8(e), it can be observed in other cases that the calculated ICA components in spatial domain are very similar. Thus, the results indicate a higher stability of the spatial domain comparing to the temporal domain. As it was demonstrated also in the example shown in Figure 3.6, the challenge is to find the proper number of components.

3.4 Unsupervised clustering methods without predefined number of clusters

Unsupervised clustering methods without predefined number of cluster are rather based on similarity matrix than the mixture modeling. This kind of unsupervised clustering method does not require the number of clusters as an input parameter. The number of clusters is normally specified by other parameters. The most famous example of such a parameter is a threshold, which defines tolerable dissimilarity between the data points inside a cluster. Hierarchical clustering methods are famous examples of such similarity based clustering methods.

In the sections 3.4.1 and 3.4.2, hierarchical clustering and mean shift algorithm are

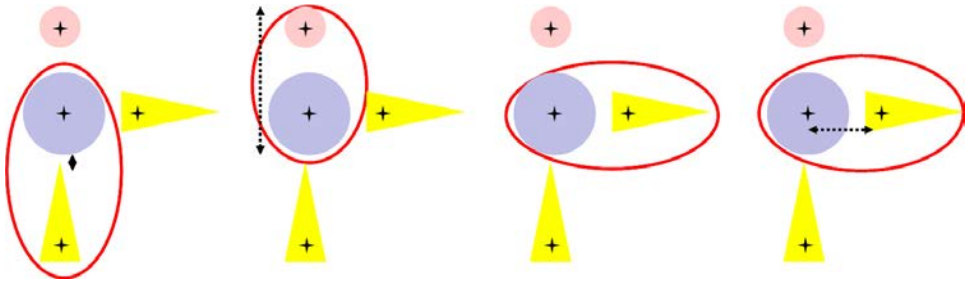


Figure 3.9: A schematic illustration of different linkage methods from left to right: Single linkage, complete linkage, average linkage and centroid linkage.

discussed. Both of these algorithms do not require a predefined number of clusters as an input parameter. In section 3.4.3 the algorithm self-organizing maps is introduced, which can also be considered as an unsupervised clustering method. However, self-organizing maps neither belongs to clustering methods with a predefined number of clusters nor to the opposite category.

3.4.1 Hierarchical clustering

Hierarchical clustering is a kind of a so called agglomerative clustering. It uses similarity measure $d(.,.)$ to compare two vectors. At the beginning, each vector is considered as a single cluster. The clustering method uses a strategy to merge pairs of clusters. In each step of clustering, two clusters are merged together until a threshold is achieved.

The most famous strategies are single linkage, complete linkage, average linkage, and centroid linkage. A schematic illustration of these four strategies is demonstrated in Figure 3.9. A brief description of these methods are presented below:

1. Single linkage: Two clusters C_i and C_j are merged, if they contain the pair of points showing the strongest similarity. Two clusters C_i and C_j are merged by the single linkage if they minimize the SL value given as:

$$SL = \min \{d(u, v) : u \in C_i, v \in C_j\} \tag{3.27}$$

for all i and j with $i \neq j$, $i \in \{1, ..., k\}$, $j \in \{1, ..., k\}$ and k number of clusters.

If the clusters are not well-separated, the so called chaining phenomenon can occur. This means two clusters are merged just because of a single pair of data points have

been close to each other. On the other hand, if the clusters are well-separated, this method is able to find clusters of arbitrary form and distribution. Even non-convex clusters, such as rings, can be found.

2. Complete linkage: For each pair of clusters, the most dissimilar pair of points is determined. Those two clusters C_i and C_j are merged that have the most similar of all dissimilar pairs. The clusters C_i and C_j are merged by the complete linkage if they minimize the value CL described as:

$$CL = \max\{d(u,v) : u \in C_i, v \in C_j\} \tag{3.28}$$

for all i and j with $i \neq j$, $i \in \{1,...,k\}$, $j \in \{1,...,k\}$ and k number of clusters.

Complete linkage tends to find compact clusters of approximately equal diameters. Thus, two small well-separated clusters might be merged together just because the diameter of the merged cluster is still less than another larger cluster.

3. Average linkage: Those two clusters C_i and C_j are merged that exhibit the minimum of average pairwise distances. Two clusters C_i and C_j are merged, if they minimize the average linkage:

$$AL = \frac{1}{|C_i|\,|C_j|} \sum_{u \in Ci} \sum_{v \in C_j} d(u,v) \tag{3.29}$$

for all i and j with $i \neq j$, $i \in \{1,...,k\}$, $j \in \{1,...,k\}$ and k number of clusters.

Average linkage is a compromise between single linkage and complete linkage methods. Therefore, in the case of real data, it may work better.

4. Centroid linkage: In the first step, centroids $\mu_i = \frac{1}{|C_i|} \sum_{u \in Ci} u$ of clusters C_i are calculated. Two clusters are merged, if they have the smallest distance $d(\mu_i, \mu_j)$ between two centroids.

In hierarchical clustering, the whole dataset can be described by a dendrogram (see Figure 3.10). A dendrogram is a binary tree. It shows level of similarity at which two clusters are merged together. To calculate the dissimilarity matrix or array, at least $n(n-1)/2$ calculations are required, where n is the number of data points. Therefore, the computation complexity as well as the memory usage of the hierarchical algorithms is approximated by $O(n^2)$. Due to the different linkage strategies, these algorithms are not based on a specific mixture model.

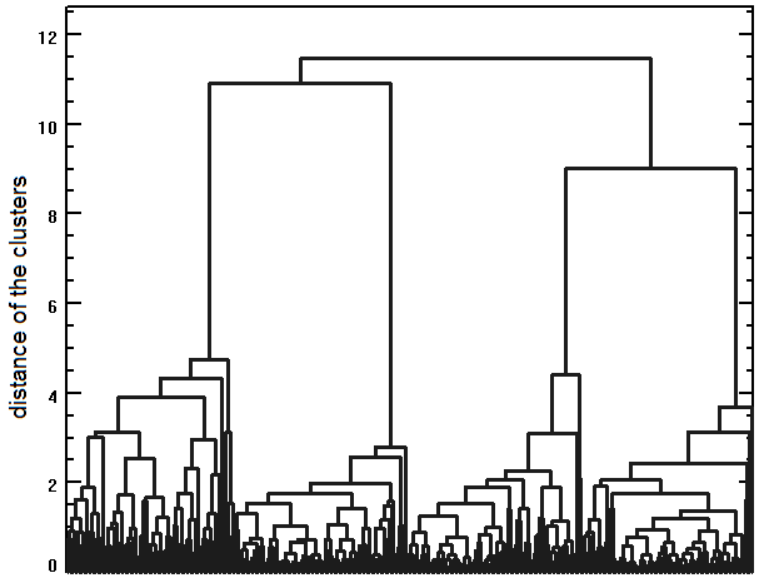


Figure 3.10: Denderogram of a simulated dataset with four clusters is depicted. The vertical axis gives at which dissimilarity index two clusters are merged.

In the case of average linkage method, the group average dissimilarity $d(G, H)$ between two clusters G and H is defined as:

$$d(G, H) = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{i' \in H} d_{ii'}, \tag{3.30}$$

where N_G and N_H are the numbers of samples in each cluster. For $N \to \infty$, Eqn. 3.30 becomes the following form:

$$\int \int d(x, x') p_G(x) p_H(x') dx dx'. \tag{3.31}$$

Eqn. 3.31 is a characteristic of the relationship between the densities $p_G(x)$ and $p_H(x')$ of samples in clusters G and H [Hastie et al., 2005].

The average linkage method attempts to produce both relatively compact clusters and distant clusters [Kaufman and Rousseeuw, 1990]. Hierarchical clustering has some advantages over the widely used k-means clustering. Hierarchical clustering produces a hierarchical representation in which clusters at each level of the hierarchy are created by merging clusters at the prior lower level. In k-means clustering, the choice of different numbers of clusters might lead to different assignments of data points to clusters. However, in hierarchical clustering, sets of clusters are nested into each other [Hastie et al., 2005].

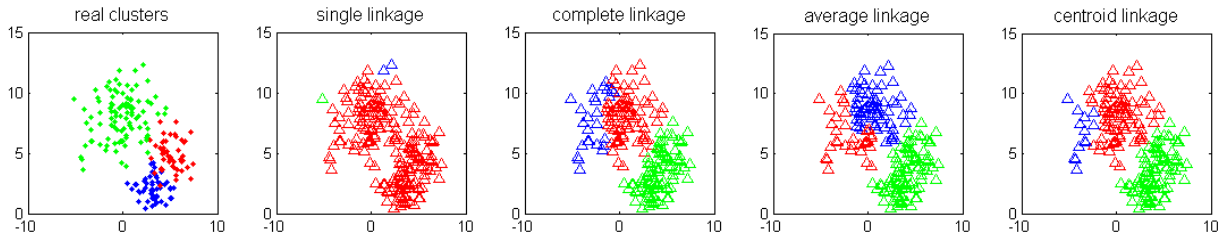


Figure 3.11: The hierarchical clustering with four introduced linkage methods is performed on the same simulated dataset from section 3.3.2.

Different hierarchical linkage methods have been performed to the same simulated dataset used in section 3.3.2. The results are presented in Figure 3.11. None of the four linkage methods was able to recognize the original three clusters of the dataset. Nevertheless, all linkage methods can be quite useful in recognizing a specific kind of pattern in a dataset.

Hierarchical clustering methods are more flexible when looking for clusters in datasets sampled from non-Gaussian mixtures. The relationships among data points in a dataset are described through the given similarity measure. By selecting a linkage method, we decide what kind of ordering of the similarity matrix is desired. The number of clusters is not necessarily given as a parameter. The granularity of clustering can be controlled by the maximal dissimilarity allowed inside each cluster.

3.4.2 Mean shift

Mean shift algorithm is a mode seeking algorithm. Mean shift algorithm clusters a dataset by assigning each data point to its corresponding mode. This is done by slightly shifting the data point toward the mode. For this, a neighborhood is considered around the data point. The mean point within this neighborhood is determined. In the next step, the neighborhood area is centered to the newly calculated mean point. In this way, the centers of high density areas, so called modes, are found. There is no need for *a priori* definition of the number of clusters. The number of clusters is controlled by the radius of the neighborhood area, so called kernel size.

The idea of the algorithm is strongly connected and inspired by Parzen windows [Comaniciu and Meer, 2002, Fukunaga and Hostetler, 1975]. The Parzen windows technique

is a well-known and popular density estimation method [Parzen, 1962]. In this technique, a neighborhood is defined around each data point. This neighborhood, in 2D space for example, can be a rectangular window. Data points with a large number of neighbors in the surrounding window are selected as modes or centers of densities. In the original implementation of Parzen windows, the space of the dataset was divided into a grid of sub areas. Each grid cell was considered as a Parzen window. While this kind of implementation is very easy and practical in one- or two-dimensional space, it is quite difficult to implement this method in higher dimensions. The result of Parzen windows varies strongly with the window size and with the decision, what the 'large' number of points in a neighborhood is. The modes found by Parzen windows can be considered the centers of clusters.

In mean shift algorithm, centers of gravity are found simultaneously with the path of each point to the corresponding mode. Tracking of data points to modes, that is to the center of gravity, is done by shifting each data point to the actual center of its neighborhood. The shifting continues until it converges. It has been shown by Comaniciu and Meer [2002] that if the neighborhood function, the so called kernel, has a convex and monotonically decreasing profile, the mean shift algorithm converts.

A proper kernel for mean shift algorithm can be described as follows. For a dataset consisting of n data points x_i, $i = 1, ..., n$ in the m-dimensional vector space $\mathbb{R}^m$ a multivariate kernel density estimate is defined by:

$$\widetilde{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - x_i) \tag{3.32}$$

with

$$K_H(x) = |H|^{-0.5} K(H^{-0.5}x). \tag{3.33}$$

The function $K(x)$ is the m-variate kernel function and satisfies the regularity constraints:

$$\begin{aligned} &\int_{R^d} K(x)dx = 1, \\ &\lim_{\|x\| \to \infty} \|x\|^d K(x) = 0, \\ &\int_{R^d} xK(x)dx = 0, \\ &\int_{R^d} xx^T K(x)dx = c_K I, \end{aligned} \tag{3.34}$$

where c_K is a constant and H is a symmetric positive define $m \times m$ band width matrix. $K(x)$

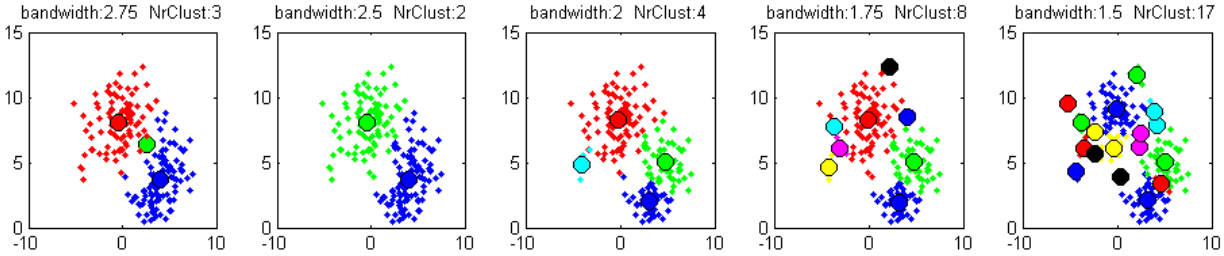


Figure 3.12: Mean shift clustering with different bandwidths is performed on the same simulated dataset from section 3.3.2.

can be, for example, a Gaussian kernel $c_t e^{-x/2}$ or an Epanechnikov kernel $c_t max(1-x, 0)$ with the normalizing constant c_t.

Mean shift algorithm is performed in following steps:

1. Select a data point x from the dataset.

2. Find all data points x_i in the neighborhood of x:

$$N(x) = \{\forall x_i | \, \|x_i - x\| < \lambda\}, \tag{3.35}$$

   where λ is the bandwidth of the algorithm.

3. Shift each data point to the center $m(x)$ of neighborhood N(x):

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)}, \tag{3.36}$$

   where $K(u)$ is the kernel from Eqn. 3.34.

4. Repeat from step two until the algorithm converts.

The result of the steps 1 to 4 is the mode which controls data point x. These steps can be applied to each data point in the dataset. At the end, points are merged to a cluster, if their modes lie at distances smaller than a threshold.

Figure 3.12 demonstrates results obtained from different runs of the mean shift algorithm with different bandwidths performed on the simulated dataset from section 3.3.2. Centers of the three main density distributions are discovered successfully in all executions. The result of the mean shift algorithm strongly depends on the size of the chosen

Figure 3.13: A schematic illustration of the effect that by increasing the bandwidth new clusters can appear.

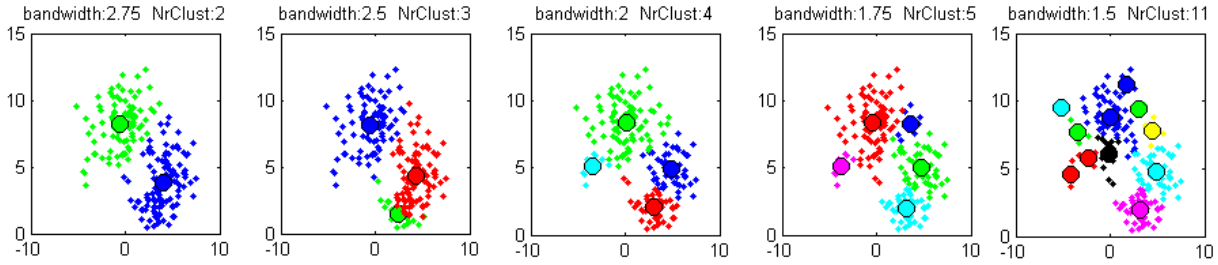


Figure 3.14: The randomized version of mean shift clustering with different bandwidths is performed on the same simulated dataset from section 3.3.2.

bandwidth. In general, decreasing the bandwidth in mean shift algorithm the number of clusters increases. In some situations it might happen that the number of clusters decreases by decreasing the bandwidth. An example is shown in Figure 3.13. In the first image from left, the bandwidth is smaller and the center point is shifted toward the upper six points. In the image on the right, the bandwidth is larger, and therefore the three points on the bottom left fall in the neighborhood. As a result, the center point of the neighborhood converts and builds a new cluster.

The computation complexity of the mean shift algorithm, in the given implementation, is $O(rn^2)$, with r being the number of iterations and n the number of data points. To accelerate the algorithm, normally an alternative implementation is used. In the new implementation, the steps 1 to 4 are performed on a randomly selected data point x. During the shifting phase, all data points in the neighborhood of x are marked as visited, and they belong to the same mode that x belongs. The next random point will be selected among the unvisited points. The computation complexity is reduced to $O(rkn)$, where k is the number of modes. The drawback of the latter implementation is that depending on the selected random points, for the same bandwidth the algorithm can result in different

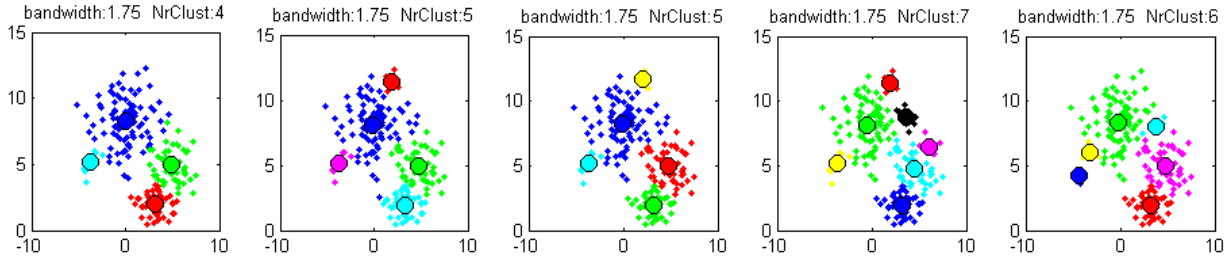


Figure 3.15: The randomized version of mean shift clustering with a fixed bandwidth is several times executed on the same simulated dataset from section 3.3.2.

clusters.

Figure 3.14 shows the results obtained with the new implementation performed on simulated data for different values of bandwidth. Again, the three centers of original densities are found successfully. Figure 3.15 depicts five executions of the random implementation for fixed bandwidth 1.75. The result of clustering varies in each execution, but the three centers of the original densities can be found correctly.

Mean shift algorithm is a powerful technique in finding the clusters of an arbitrary shape. The original implementation is computationally very expensive. In the randomized version, the correct assignment of all points to the corresponding mode is not guaranteed. However, the main drawback of the algorithm, in both implementations, remains the difficulty of selecting the proper bandwidth.

3.4.3 Self organizing maps

Self organizing maps (SOM), also called Kohonen maps [Somervuo and Kohonen, 1999], is an algorithm that maps a dataset into a set of prototypes. The prototypes are arranged in relation to the neurons in a topological order [Haykin, 2008]. The main idea of the algorithm is that the neurons, which are points in usually 1 or 2-dimensional space, are related to the high dimensional data space through some weights or prototypes. A schematic illustration of the SOM algorithm is depicted in Figure 3.16. The neurons are arranged in a topological order, e.g. in the form of quadratic or hexagonal grids. The corresponding weights are vectors in the data space, and can be chosen randomly or distributed uniformly on the main principal component plane of the dataset. At each iteration of the algorithm, one data point is selected at random. In the next step, the closest prototype to the selected data

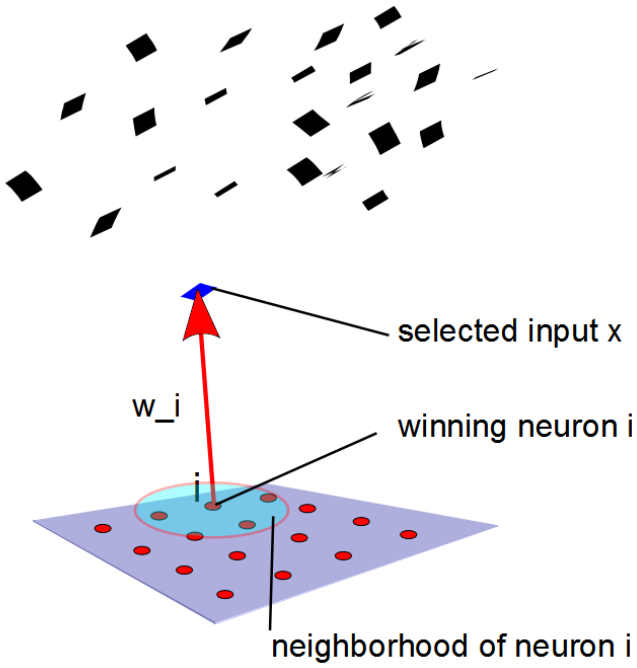


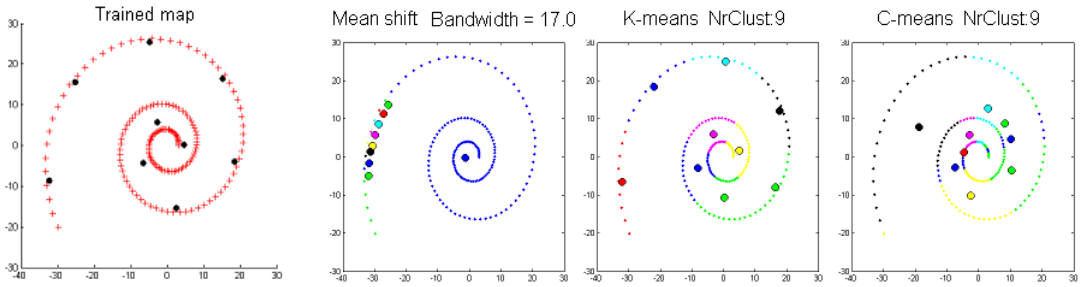
Figure 3.16: A schematic illustration of the SOM algorithm.

point is found, and the corresponding neuron is activated. The selected neuron activates the other neurons in its scope. The weights of all activated neurons are moved toward the selected data point. Due to the random selection of the data points, the data points from areas with a higher density are selected more often. Thus, after enough iterations, the weights of the neurons will have a similar distribution as dataset.

Mathematically, SOM can be described as follows. Suppose dataset S consists of n data points s_i, with $s_i \in \mathbb{R}^d$ and $i = 1, ..., n$. The n' neurons c_j are arranged e.g. on a quadratic grid. Each neuron has a corresponding weight w_j or prototype in the data space, with $w_j \in \mathbb{R}^d$ and $j = 1, ..., n'$. The algorithm consists of three main steps:

- Initializing: The synaptic weights are set to initial values. The initial values can be random values from data space, or be uniformly distributed values over the plane of the first two principal components of the dataset. SOM algorithm, initialized with latter values, converts faster.

- Competition: In this step, for a randomly selected input s_i, the neuron c_j with its weight being closer to the selected input s_i, is found. The index j of the neuron with



(a) SOM with a 3×3 grid (b) Results of 3 clustering methods with number of clusters = 9.

Figure 3.17: The position of prototypes after training the SOM algorithm is compared with three clustering methods on the same dataset.

the closest weight to the s_i is:

$$j(s_i) = \{k \mid min(\|s_i - w_k\|),\ k = 1, ..., n'\}\,. \tag{3.37}$$

- Cooperation: The weights of all neurons in the neighborhood of the winning neuron from previous steps are updated so that they become closer to selected data point s_i.

$$^{t+1}w_k =^t w_k +^t \alpha\eta(c_j, k, t)(s_i -^t w_k). \tag{3.38}$$

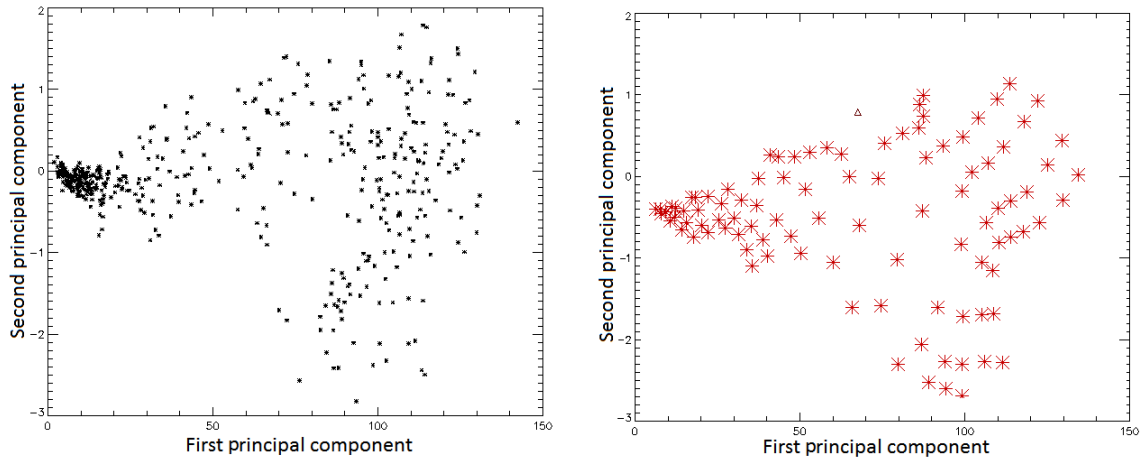
The value t indicates the current iteration. The parameter α represents the learning rate and decreases over time. The neighborhood function or kernel of the algorithm is $\eta(c_j, k, t)$. It can be, for example, of a Gaussian form:

$$\eta(c_j, k, t) = \exp\left(\frac{\|c_j - c_k\|}{^t\delta^2}\right). \tag{3.39}$$

The parameter $^t\delta$ controls the scope of the kernel function and decreases over time.

The "Competition" and "Cooperation" steps are executed iteratively until the weights do not change significantly. The convergence of the algorithm is strongly supported by decreasing the values α and δ. The choice of α and δ, and the speed of altering these parameters affect the results of the SOM algorithm.

In Figure 3.17 four clustering methods, SOM, mean shift, k-means, and c-means clustering are compared on a 2D dataset of a spiral form. It can be observed that prototypes in SOM methods have a quite similar arrangement as cluster centers of k-means clustering. The k-means clustering result has a well arrangement of cluster centers along the spiral.



(a) (b)

Figure 3.18: (a) Data points of the tumor region of a breast DCE-MRI dataset are projected on the plane of the first and second principal components. (b) The trained prototypes of a 10×10 grids of neurons for the same dataset are depicted on the same plane of first and second principal components.

However, the corresponding data points to each center, especially in the center of the spiral, do not follow the form of the spiral. The mean shift clustering, in contrast, has clustered the spiral perfectly in groups along the spiral form. Even so, the cluster centers in the mean shift clustering result are not good representatives of the dataset.

Balakrishnan et al. [1994] compared SOM with k-means clustering. They also observed that the SOM algorithm with a small number of neurons behaves like k-means clustering. Nattkemper and Wismüller [2005] used SOM to distinguish between benign and malignant features by training the prototypes to DCE-MRI datasets of different tumor types. They reported about different prototypes that are more specific for benign than malignant and *vice versa*.

In Figure 3.18, the trained prototypes of a 10×10 network of neurons are compared with the original DCE-MRI data points. For visualization, the data points of the DCE-MRI and the prototypes are depicted in the plane of the two principal components corresponding to the two largest eigenvalues. The 100 prototypes are successfully arranged along the data points' distribution.

As mentioned before, the SOM with a few numbers of neurons is closely related to

k-means algorithm. The randomly chosen weights can be considered as cluster centers in k-means. Due to the few number of neurons and after a few iterations, the update of weights is only done for the winning neuron. That means, the weight corresponding to the winning neuron moves toward the randomly selected data point. There is an alternative implementation of k-means clustering with the so called "Dog Rabbit Strategy" which is quite similar to the SOM updating. In this context, the SOM clustering suffers from the same problems as k-means clustering. If the number of neurons is large enough, in each iteration, weights of a set of neighbored neurons are updated. The trained prototypes are ordered by the neurons' topology. At the end, a network topology of the dataset is achieved. In this context, the SOM behaves like a nonlinear principal component analysis [Yin, 2008]. Nevertheless, the main challenge in using SOM remains the proper choice of the learning rate, the range of the kernel and the speed in which these values change in each iteration.

3.5 Number of clusters

In clustering methods, the number of clusters is either a direct parameter, or it might be controlled by other parameters of the method. Estimating the proper number of clusters is important in finding the correct clusters as well as validating the results. As mentioned before, number of clusters is a fixed parameter by some well-known clustering methods such as k-means. Therefore, it is of interest to determine this parameter directly from the dataset itself. Among several approaches, mostly based on finding well-separated compact clusters, the Gap Statistic is quite promising [Tibshirani et al., 2001].

3.5.1 Gap Statistic

The Gap Statistic is one of the most popular techniques to determine the optimal number of clusters. The idea of the Gap Statistic is to compare the within-cluster dispersion to its expectation under an appropriate null reference distribution [Tibshirani et al., 2001]. It outperforms many other methods, including the method by Kaufman and Rousseeuw [1990], the Calinski and Harabasz index [Calinski and Harabasz, 1974], the Krzanowski and Lai method [Krzanowski and Lai, 1988], and the Hartigan statistic [Hartigan, 1975, Tibshirani et al., 2001]. Therefore, the Gap Statistic is frequently used in a variety of applications, from image segmentation [Zheng-Jun and Yao-Qin, 2009], image edge detec-

tion [Qiuxia et al., 2009] to genome clustering [Wendl and Yang, 2004].

However, there are few works investigating the method itself. The tendency of the Gap Statistic to overestimate the number of clusters was reported by Dudoit and Fridlyand [2002]. It is also known that the Gap Statistic may not work correctly in cases where data are derived from exponential distributions [Sugar and James, 2003]. The Weighted Gap Statistic, proposed by Yan and Ye [2007], is an improvement, for example, in the case of the mixtures of exponential distribution. Yin et al. [2008] pointed out that in situations where a dataset contains clusters of different densities the Gap Statistic might fail. They suggested to use reference datasets sampled from normal distribution rather than uniform distribution.

The original Gap Statistic is based on some empirical choices, such as the "one standard error" style rule for simulation error, and using the logarithm of the within-cluster dispersion W_k. However, few studies have focused on analyzing the effect of these choices. It has been shown that using the logarithm of W_k is actually disadvantageous in finding the number of clusters in datasets [Mohajer et al., 2011]. Especially in cases where clustering data are sampled from multi-dimensional uniform distributions with large differences in the variances of the different clusters, it is better to use W_k instead of $\log(W_k)$ [Mohajer et al., 2011].

Gap Statistic definition

Let $\{x_{ij}\}$ be observations with $i = 1, 2, ..., n$, $j = 1, 2, ..., p$, p features measured on n independent samples, clustered into k clusters $C_1, C_2, ..., C_k$, where C_r denotes the indexes of samples in cluster r, and $n_r = |C_r|$. Let $d_{ii'}$ be the distance between samples i and i'. For example, this distance might be the squared Euclidean distance $d_{ii'} = \sum_j (x_{ij} - x_{i'j})^2$. The sum of the pairwise distances D_r for all points in cluster r is:

$$D_r = \sum_{i,i' \in C_r} d_{ii'}. \tag{3.40}$$

Let define

$$W_k := \sum_{r=1}^{k} \frac{1}{2n_r} D_r. \tag{3.41}$$

If d is the squared Euclidean distance, then W_k is the within-cluster sum of squared distances from the cluster means. W_k decreases monotonically as the number of clusters k

increases. For the calculation of the Gap function, Tibshirani et al. [2001] proposed to use the difference of the expected value of $\log(W_k^*)$ of an appropriate null reference and the $\log(W_k)$ of the dataset,

$$Gap_n(k) := E_n^* \log(W_k^*) - \log(W_k). \tag{3.42}$$

Then, the proper number of clusters for the given dataset is the smallest k such that

$$Gap_n(k) \geq Gap_n(k+1) - s_{k+1}, \tag{3.43}$$

where s_k is the simulation error calculated from the standard deviation $sd(k)$ of B Monte Carlo replicates $\log(W_k^*)$ according to the equation $s_k = \sqrt{1+1/B}sd(k)$. The expected value $E_n^*\log(W_k^*)$ of within-dispersion measures W_{kb}^* is determined as

$$E_n^* \log(W_k^*) = \frac{1}{B}\sum_b \log(W_{kb}^*), \tag{3.44}$$

where W_{kb}^* are given by clustering the B reference datasets. The sum of $\log(W_{kb}^*)$ can be written as

$$E_n^* \log(W_k^*) = \frac{1}{B}\log\left(\prod W_{kb}^*\right). \tag{3.45}$$

Therefore the Gap function from Eqn. 3.42 can be re-written;

$$Gap_n(k) = \log\left(\frac{(\prod W_{kb}^*)^{1/B}}{W_k}\right). \tag{3.46}$$

The number $(\prod W_{kb}^*)^{1/B}$ is the geometric mean of W_{kb}^*. Thus, the Gap Statistic is the logarithm of the ratio of the geometric mean of W_{kb}^* to W_k. In next section, Gap Statistic is compared to a new definition of Gap, which uses the arithmetic mean of W_{kb}^* and W_k.

Gap Statistic without logarithm function

Let's consider an alternative definition of the Gap function, $Gap_n^*(k)$, by using W_k instead of $\log(W_k)$ in Eqn. 3.42. We have

$$Gap_n^*(k) = E_n^*(W_k^*) - W_k, \tag{3.47}$$

where

$$E_n^*(W_k^*) = \frac{1}{B}\sum_b W_{kb}^*. \tag{3.48}$$

The proposed alternative Gap Statistic defined by using W_k directly is referred to as Gap_n^*; the original Gap calculated by using the logarithm of W_k is referred to as Gap_n. Scott and Symons [1971] and Tibshirani et al. [2001] note that, in case of a special Gaussian mixture model, $\log(W_k)$ has interpretation as log-likelihood. In maximum likelihood inference, it is usually more convenient to work with the log-likelihood function than with the likelihood function, in order to have sums instead of products. However, using $\log(W_k)$ has no computational advantage versus using W_k directly in the definition of the Gap Statistic.

It can be shown that an answer in the original Gap_n is a sufficient condition for the proposed Gap_n^* Statistic, but not *vice versa*. Let $A = \prod {W_{kb}^*}^{1/B}$, $B = \prod {W_{k+1b}^*}^{1/B}$, $C = \frac{1}{B}\sum_b W_{kb}^*$, $D = \frac{1}{B}\sum_b W_{k+1b}^*$, $d_1 = W_k$, and $d_2 = W_{k+1}$.

Proposition 1. *For* $\forall d_1, d_2 > 0$, $d_1 > d_2$, $A > B$, $C > D$, $A, C > d_1$ *and* $B, D > d_2$, *if*

$$log\left(\frac{A}{d_1}\right) \geq log\left(\frac{B}{d_2}\right),$$

then

$$C - d_1 \geq D - d_2.$$

Proposition 2. $\exists d_1, d_2 > 0$, $d_1 > d_2$, $A > B$, $C > D$, $A, C > d_1$ *and* $B, D > d_2$ *so that if*

$$C - d_1 \geq D - d_2,$$

then

$$log\left(\frac{A}{d_1}\right) < log\left(\frac{B}{d_2}\right).$$

Proofs are given in Appendix A.1.

Hence, if there is a possible candidate in Gap_n at point k, there is also a possible candidate in Gap_n^*. On the other hand, it is possible that there is no such k in Gap_n function while the Gap_n^* function indicates a possible candidate at point k. In the sections 6.4 and 6.5 there are examples from real and simulated data, in which the original Gap_n function is a strictly increasing function. Thus, there is no k that fulfills the condition in Eqn. 3.43. However, the proposed Gap_n^* function may be able to suggest a number of clusters for these datasets.

Weighted Gap Statistic

In Eqn. 3.41, W_k is the pooled within-cluster sum of squares. This implies, considering a point far away from the cluster mean, the large distance of this point to the cluster

center has a higher impact compared to points with small distances from the cluster mean. To this end, Yan and Ye [2007] suggested to compute $W_k^{'}$ as the average of all pairwise distances for all points in a cluster,

$$W_k^{'} = \sum_{r=1}^{k} \frac{2}{n_r(n_r - 1)} D_r. \tag{3.49}$$

This approach is called "Weighted Gap function".

Similar to the original Gap function, the weighted Gap function can also be computed with or without logarithm. However, W_k in Eqn. 3.41 is monotonically decreasing in k if the distance $d_{ii'}$ is the Euclidean distance. On the other hand, $W_k^{'}$ in Eqn. 3.49 is neither a decreasing nor an increasing function in k. Therefore, the propositions given in section 3.5.1 are not valid for the weighted Gap function. The results from the original and the weighted Gap function on two historical datasets are compared in section 6.1.

3.6 Similarity measures

A fundamental part of each clustering method is the question how to compare the samples with each other. The so called similarity measure decides which kind of patterns we are looking for in a dataset. In the most density estimate clustering methods, the Euclidean distance is the only choice. It is a distance measure that is closer to our understanding of the distance between two points in 2D and 3D vector spaces, what we measure in real life with a ruler. However, in a high dimensional vector space, where samples are described with a large number of variables, the Euclidean distance loses its meaning. On one side, it should be taken into account that the variables can have different domains, and therefore, a proper normalization method should be performed. On the other side, we are confronted with the course of dimensionality and its related problems, which are briefly discussed in section 6.4. One way to overcome these problems is the feature extraction or dimension reduction as discussed in section 3.1. The alternative is to look for new similarity measures, which may describe the dataset better. In the following section, some well-used similarity measures are discussed, and a new similarity measure is introduced.

In mathematics, a metric or distance function is a function which defines a distance between elements of a set. A set with a metric is called a metric space. A metric on a set

X is a function or distance [Fedorchuk et al., 1990]:

$$d : X \times X \rightarrow \mathbb{R}, \tag{3.50}$$

where $\mathbb{R}$ is the set of real numbers. For all x, y, z in X, this function is required to satisfy the following conditions:

1. non-negative $$d(x, y) \geq 0 \tag{3.51}$$
2. Identity $$d(x, y) = 0 \;\; if \; and \; only \; if \; x = y \tag{3.52}$$
3. Symmetry $$d(x, y) = d(y, x) \tag{3.53}$$
4. Triangle inequality $$d(x, z) \leq d(x, y) + d(y, z) \tag{3.54}$$

A semi-metric on X is a function $d : X \times X \rightarrow \mathbb{R}$ that satisfies the first three axioms, but not necessarily the triangle inequality.

3.6.1 Euclidean distance

Euclidean distance of two points in two dimensional space, for example on a paper, is the distance measured by a ruler. In general, for two vectors $v = \{v_1, ..., v_d\}$ and $u = \{u_1, ..., u_d\}$ both in $\mathbb{R}^d$, the Euclidean distance can be written as a dot product:

$$\|v, u\| = \sqrt{\sum_{i=1}^{d} (v_i - u_i)^2} = \sqrt{(v - u) \cdot (v - u)}. \tag{3.55}$$

Euclidean distance is a metric and satisfies the metric conditions in Eqn. 3.51 to Eqn. 3.54.

3.6.2 Cosine measure

Cosine similarity is a measure of similarity between two vectors by measuring the cosine of the angle between them. The cosine of the angle θ between two vectors determines whether two vectors are pointing in the same direction. For two vectors a and b, cosine similarity is equal to:

$$\cos(\theta) = \frac{a \cdot b}{\|a\| \, \|b\|}, \tag{3.56}$$

where $\cdot$ is the dot product between two vectors and $\|.\|$ is the norm of a vector. Cosine measure is a semi-metric due to the violation of triangle inequality condition in Eqn. 3.54.

3.6.3 Correlation coefficient

Correlation coefficient, also called Pearson's correlation coefficient and normally denoted by r, is a measure of linear dependency between two random variables X and Y:

$$r = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \tag{3.57}$$

where μ and σ are the mean and covariance of the corresponding random variable and $E[.]$ is the expected value. Correlation coefficient as a similarity measure between two vectors $v = (v_1, ..., v_d)$ and $u = (u_1, ..., u_d)$, is the same as the cosine measure of the centered vectors v' and u', with $v'_i = v_i - \sum_{i=1}^{d} v_i$ and $u'_i = u_i - \sum_{i=1}^{d} u_i$. The main achievement of the centering is that, if all variables of a vector are shifted by a value b, so that $v' = (v_1 + b, ..., v_d + b)$ then cosine similarity changes but correlation coefficient stays the same. That is

$$cosine(v, u) \neq cosine(v', u) \tag{3.58}$$

but

$$r(v, u) = r(v', u). \tag{3.59}$$

3.6.4 Parallelism measure

Selecting a proper distance measure is an important factor in many clustering methods. In almost all previous works in clustering of DCE-MRI of breast tumors, Euclidean distance is the measure of dissimilarity [Castellani et al., 2006, Daducci et al., 2009, Lee et al., 2010, Leinsinger et al., 2006, Meyer-Baese et al., 2009, Stoutjesdijk et al., 2007, Twellmann et al., 2008, Varini et al., 2006, Wismüller et al., 2006].

Euclidean distance measures point wise distances in two signal curves and does not take the pattern of the signal curves into account. It is possible that two curves have a larger Euclidean distance, but the shapes of the curves are more similar than other curves with less similar shapes but a smaller Euclidean distance. An example of this situation is demonstrated in Figure 3.19.

The relative intensity value at each time point is related to the amount of the contrast agent at that time point in the tissue of interest. At the same time, the form of the signal curves represents pharmacokinetic properties of the tissue.

In this work, a new similarity distance is introduced to measure the amount of parallelism in the washout phases of the different signal curves. The washout phase is the part

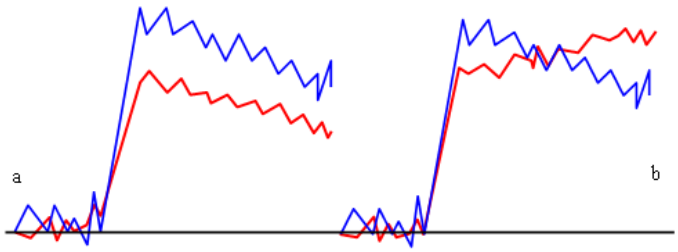


Figure 3.19: The Euclidean distance between the two curves on the left (a) is larger, but the patterns of the two curves are more similar than the curves on the right (b) with smaller Euclidean distance.

of the signal curve after the end of rapid enhancement. The rapid enhancement phase and the beginning and the end points of this phase are illustrated in an example in Figure 3.20.

The time point t_e at which the rapid enhancement phase ends, is defined by a method inspired by Cheong et al. [2003]. A piecewise linear regression is performed on the smoothed mean curve of the selected voxels. The mean curve is the average of the curves of all voxels. The mean curve is smoothed by a simple boxcar average smoothing algorithm. The piecewise regression is done for a selected size of window w, where a line $at + b$ is fitted to the partial part of the signal curve from time point t to $t + w$. The line with the largest slope is selected as the line going through the rapid enhancement phase. The rapid enhancement phase is selected as the time interval between two time points t_1 and t_2, where the mean squared errors of this interval to the fitted line is smaller than the mean squared errors of the curve to its boxcar smoothed curve.

After the starting time point of washout phase is found from previous step, the minimum of the washout part of the curve is set to zero. The similarity measure between two curves is determined as follows:

1. The distance at each time point of the min-zero washout curves are calculated. For a washout part with n time points, we obtain n distances.

2. In the next step, the n distances are sorted in decreasing order.

3. The similarity measure is:

$$sm_{ii'} = \sqrt{\sum_{j=1}^{n/2} \left(d_j^{ii'} - d_{(n-j)}^{ii'} \right)}, \tag{3.60}$$

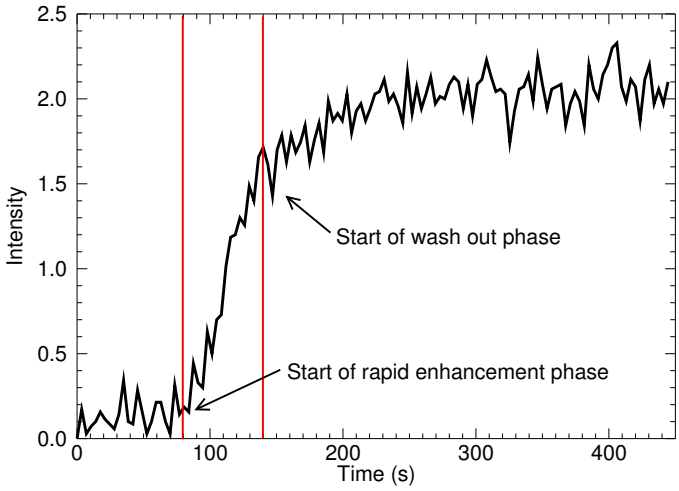


Figure 3.20: Illustration of start and end points of the rapid enhancement phase in a signal curve.

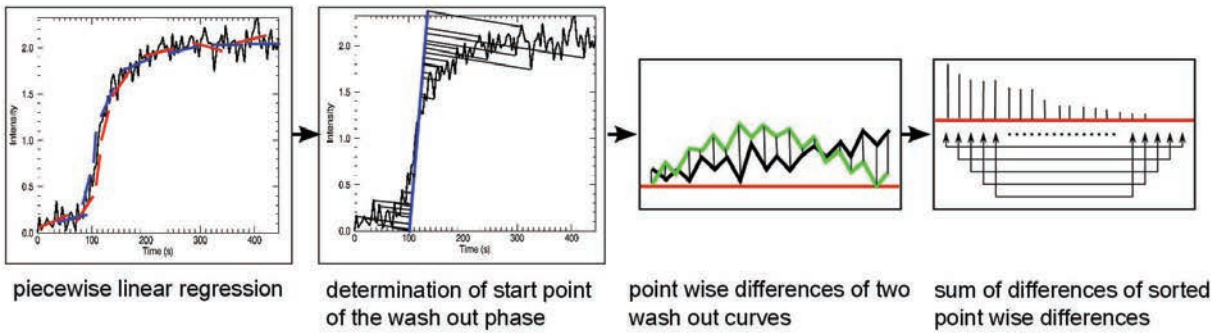


Figure 3.21: Schematic illustration of PM measure calculation procedure.

where $d^{ii'}$ is the sorted list of the distances at time points of washout curve i and curve i'.

The calculation process of the similarity measure is depicted schematically in Figure 3.21. It must be noted that this similarity measure is a semi-metric, due to the violation of the triangle inequality condition in Eqn. 3.54. The idea behind this similarity measure is that ideally two parallel curves should have equal distances at corresponding points. Taking the existence of the noise in the dataset into account, means that the distances would not be totally equal. On the other hand, the distances at different time points of non-parallel signal curves vary significantly. The introduced similarity measure tries to measure the varieties of the point wise distances between two signal curves during the washout phase. This similarity measure is referred to as the parallelism measure (PM) in this work.

3.7 Summary

In this chapter, the main topics of clustering such as dimension reduction, clustering algorithm, defining the number of clusters and similarity measure have been discussed.

Advantages and drawbacks of different clustering methods are demonstrated by means of experiments on real or simulated data. The clustering methods have been divided into two sub-categories. Clustering methods which need a specified number of clusters as an input parameter and those methods where number of clusters is controlled by other parameters. For the first category estimation of the number of clusters is requisite.

The Gap Statistic is discussed as a method to estimate the number of clusters in a dataset. A modified version of Gap Statistic (Gap*) has been proposed. Gap* Statistic does not use logarithm function in original definition of Gap Statistic. We have shown that it is possible that Gap* Statistic estimates a possible k as the number of clusters for a dataset while Gap Statistic is an increasing function and therefore is not able to estimate the cluster numbers.

As the last topic, the similarity measures are discussed. We have introduced a new similarity measure, specified for DCE-MRI signal curves. The new similarity measure (PM measure) is a measure for parallelism of two signal curves during their washout phases. In Chapter 4 and 5, we introduce a stepwise clustering method, which uses a combination of

two similarity measures, namely Euclidean distance and PM measure.

Chapter 4

Cluster analysis of breast datasets

In this chapter, we will discuss the clustering of signal curves in perfusion MRI of breast tumors. We will show different clustering scenarios performed on segmented tumors from 15 datasets. For segmentation of the tumors we introduce a fully automated segmentation method. Clustering methods have been performed both in the data space and in the feature space of the data. Besides existing clustering methods, a new two-steps clustering method has been introduced. The two-steps method combines the new PM measure (introduced in section 3.6.4) with the Euclidean distance. Furthermore, the PM measure has been compared with three well-known similarity measures. In addition, the relationship between clustering using the Euclidean distance or PM measure and the parameters from pharmacokinetic models has been studied. Finally, results have been evaluated and discussed.

4.1 Material

4.1.1 Patients

In this work, fifteen DCE-MRI datasets of female patients with breast cancer, between the ages of 30-87 years, and the mean age 55.23 years, are used for the experiments. Breast lesions were proven by histology following breast biopsy or surgery. The datasets are separated into two groups of M-series and P-series. The M-series consists of nine datasets from the German Cancer Research Center [DKFZ, 2004]. The P-series is from a study started in 2011 in cooperation with the Radiology Institute of Klinikum rechts der

Isar [Radiology, 2011]. For the study, an "Ethics Committee" approved the performance of sixty high temporal resolution DCE-MRI on volunteer patients. Table C.1 in Appendix, gives a summary of datasets used in this work.

4.1.2 MR Imaging

Data acquisition was performed with a 1.5 T MR system. The patients lay prone on the breast coil with their arms extended above the head (see Figure 2.4).

Nine M-series datasets were obtained with the following parameters: T_1-weighted FLASH image (Siemens MR), $TR = 12$ ms, $TE = 5$ ms, flip angle = 35°, slice thickness $TH = 4$ mm, matrix size = 256×256, field of view (FOV) = 320 mm; two transaxial section were defined (FOV = 320 mm, $TH = 6$ mm) crossing the lesion and the aorta for further evaluation. Strongly, T_1-weighted images ($TR = 10$ ms, $TE = 4.1$ ms, $fa = 12°$, matrix size = 256×256) were acquired from the same cross before, during, and after intravenous administration of 0.1 mmol Gd-DTPA (MAGNEVIST; Schering AG, Berlin, Germany) per kg body weight at a constant rate over 30 s with an infusion pump (CAI 626P/Tomojet; Doltron AG, Uster, Switzerland). A total of 128 acquisitions were performed over a period of 6.9 min, with one acquisition each 3.25 s [Brix et al., 2004].

The P-series data acquisition was performed with the following parameters: T_1-weighted FLASH image (Phillips MR), $TR = 8.0$ ms, $TE = 4$ ms, flip angle = 25°, slice thickness $TH = 3$ mm, matrix size = 512×512, FOV = 368.64 mm; five transversal sections were defined (FOV = 368.64 mm, $TH = 3$ mm) crossing the lesion. T_1-weighted images ($TR = 8$ ms, $TE = 4$ ms, $fa = 25°$, matrix size = 512×512) were acquired from the selected sections before, during, and after intravenous administration of 0.1 mmol Gd-DTPA per kg body weight. A total of 100 acquisitions were performed over a period of 6.6 min, with one acquisition each 4.0 s.

4.2 Methods

Before starting the cluster analysis of the signal curves, the voxels of the tumor region need to be segmented from the rest of the image. For this, a fully automated tumor segmentation has been developed. After the segmentation of the tumor, different clustering procedures were performed on the signal curves corresponding to the selected tumor voxels.

The clustering was done either in the data space of the signal curves or in the space of extracted features from signal curves.

4.2.1 Fully automated tumor segmentation

A DCE-MRI dataset is considered to consist of three types of voxels, the voxels outside the scanned body, i.e., background, the voxels without micro-vascular structure and thus without significant changes in intensity after the bolus injection, as non-enhanced tissue and the voxels with significant changes after bolus injection, i.e., enhanced tissue. For the separation of background and non-enhanced tissue, the advantage of extra information in DCE-MRI delivered by signal curves is used. The segmentation is carried out in four steps.

First step: Separation of tissue from the background

For separation of tissue from the background, a reduced dataset A^1 is calculated by the projection of the dataset along the eigenvector of the largest eigenvalue of the correlation matrix. Consider the voxels of DCE-MRI volume as samples defined by vectors, where the vector is the signal curve S_i of each voxel i. We define the centered dataset A as a matrix of size $M \times N$, with M the number of time points in S_i and N the number of voxels in MRI volume. The vectors S_i' in A (rows of matrix A) are $S_i' = S_i - \bar{s}_i$, where $\bar{s}_i$ is the mean intensity value of the dataset. For the dataset A, the $M \times M$ covariance matrix is calculated. The reduced dataset A^k is $k \times M$, built from the subspace of first k principal components U^{*k} and reduced dataset A^{*k} in principal components' space:

$$A^* = UA \quad \& \quad A^k = (U^{*k})^T A^{*k}, \tag{4.1}$$

where $(U^{*k})^T$ is the transpose matrix of U^{*k}. For more details on the principal component analysis see section 3.1.1. Along the principal component with the largest eigenvalue the data has the maximum variance. The intensity differences are maximal, and the noise is suppressed. A base curve, i.e. the eigenvector of the largest eigenvalue, is represented in all voxels of A^1. We observed that the mean values of the signal curves from the tissue are larger than the maximum values of the signal curves of the background. This observation is the base of a threshold ρ for separating the body from the background. The smallest minimum values of signal curves in A^1 are often from background signals. Thus, the threshold ρ is the mean value of the maximum values of background signal curves and is

found as follows:

$$\rho = mean(max(S_i))_{S_i \in A^1 \& min(S_i) = min(A^1)}. \tag{4.2}$$

Body voxels are voxels with signal curves S_i where $mean(S_i) > \rho$.

Second step: Separation of enhanced tissue from non-enhanced tissue

For separation of enhanced tissue from non-enhanced tissue, only the voxels selected in the previous step are used. First, the signal curves are normalized according to their pre-enhanced intensity values:

$$S_i^r = \frac{S_i - S_0}{S_0} \tag{4.3}$$

with $S_0 = \frac{1}{m} \sum_{k=0}^{m-1} S_i$, the average of the first m pre-enhanced values. Second, the average enhancement En of the post enhanced time points in a signal curve is calculated as:

$$En = \frac{1}{n_{post}} \sum_{i=post}^{n} S_i^r, \tag{4.4}$$

where n_{post} is the number of the post enhanced time points and $post$ is the starting time point of post enhanced phase. The time point $post$ is estimated by $post = pre + inj$, with pre the number of time points before injection and inj is the estimated injection time. Enhanced body voxels are then selected as voxels with $0.5 < En < 4$. This threshold is based on the results published by Mussurakis et al. [1997], who have reported that the peaks of relative intensity curves of benign and malignant tumors in various DCE-MRI breast datasets varies between 0.5 and 4.

Third step: Elimination of the heart voxels and other strongly noisy voxels

As before, only the selected voxels from the previous steps are used. In acquisition of the breast, two main sources of noise besides the environmental and random caused noise exist. These two sources of distortion are motion artifacts caused by the heart and breathing. The voxels from heart region are strongly enhanced so that they are normally selected through the previous segmentation step. To eliminate heart curves from the suspected tumor curves, we observed that signal curves of the tumor region comprised less noise than median or average noise of the dataset. The noise of a single signal curve S_i is calculated as follows:

$$noise = \frac{1}{n} \sum_{j=0}^{n-1} \left| S_{ij}^w - S_{ij}^r \right|, \tag{4.5}$$

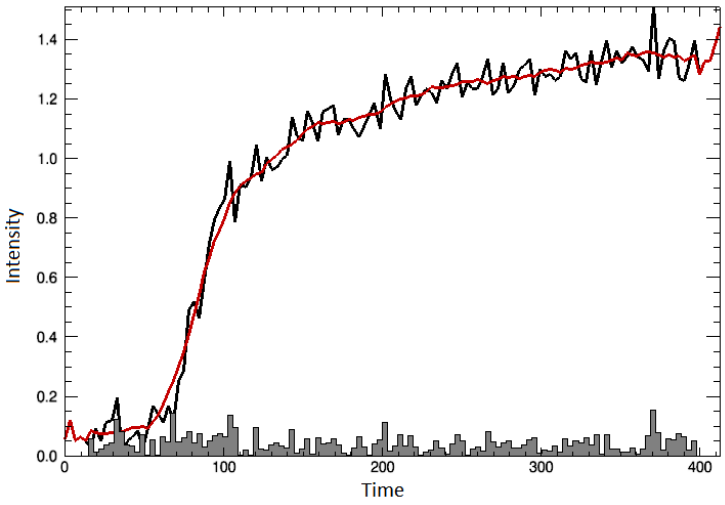


Figure 4.1: A signal curve and its corresponding smoothed curve (red curve) are depicted. The distances between signal curve and the smoothed red curve at each time point are illustrated as gray bars.

where S_i^w is the smoothed signal curve of S_i^r with w the width of smoothing window:

$$S_{ij}^w = \frac{1}{w} \sum_{k=0}^{w-1} S_{ij+k}^r. \tag{4.6}$$

Figure 4.1 depicts a signal curve and its corresponding smoothed curve (red curve) for $w = 3$. The distances, between the signal curve and the red line at each time point are depicted by gray bars in Figure 4.1. The average ς and the median τ of noise values of all signal curves in a dataset are calculated. The $min(\varsigma, \tau)$ is then used as a threshold for elimination of noisy signal curves. The curves which have average noise values greater than the threshold are eliminated.

Forth step: Post-processing of segmented area

Normally, as the result of previous steps, the voxels of the tumor area are segmented from the background tissue. However, enhanced voxels of arteries are also segmented simultaneously. In addition, it is possible that small islands of voxels inside the tumor area are not included in the segmentation. This can happen during thresholding in the second and third steps. To select a single solid segmented tumor, the following steps are performed on segmented voxels:

1. A binary image is produced where the segmented voxels are assigned the value 1 and all other voxels are set to zero.

2. In this binary image, we look for connected components of segmented voxels and numerate the components.

3. The connected components with 10 voxels or less are eliminated by setting the values of these voxels to zero in the binary image.

4. For each of remaining components, the outer border of the component is determined and all voxels inside this border are set to 1.

The tumors in nine datasets from the M-series are successfully segmented with the presented method. However, the method is less successful in the case of P-series datasets, due to the strong noise and patient movement during acquisition. Therefore, the method was extended to a semi-automated version. In the semi-automated version, the steps 2 and 4 are performed on a manually selected window around the tumor area. Step 3 is eliminated in the semi-automated version, due to the strong noise in P-series datasets. The results of segmentation for all 15 datasets are demonstrated in Appendix C.

4.2.2 Similarity measure

In DCE-MRI, the patterns of enhancement are as important as the amplitude of enhancement. The signal curves are usually compared by means of Euclidean distance. However, the Euclidean distance does not take the patterns of the signal curves into account. Hence, the Euclidean distance is not a proper choice in the cluster analysis of DCE-MRI signal curves. It is necessary to define a proper similarity measure or consider a combination of several measures. A new similarity measure was introduced and discussed in section 3.6. The PM measure is a measure for the amount of parallelism in the washout phase of two signal curves. PM similarity measure and Euclidean distance are used as similarity measures in different steps of a new two-steps clustering. Furthermore, the discussed similarity measures in section 3.6 have been compared with each other and the results are presented.

The clustering methods such as k-means clustering or mean shift are optimized for metric measures such as Euclidean distance. In contrast, hierarchical clustering methods are similarity based methods and thus they are qualified for usage by different similarity

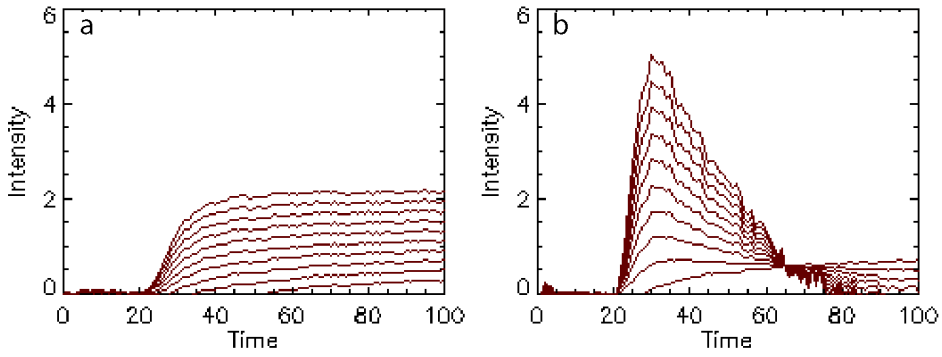


Figure 4.2: Signal curves corresponding to the selected points along the first (a) and second (b) principal components of a dataset.

measures, even non-metric measures, such as cosine or PM measure (for more details see section 3.6). Therefore, the hierarchical clustering is the method of choice in comparison of different similarity measures.

4.2.3 Dimension reduction

The goal of clustering is to divide the segmented tumor area in subgroups of similar enhancement patterns. For this, different clustering methods were performed either directly on m dimensional vector space of signal curves or on a reduced feature space. Three different feature spaces, which are the feature spaces of principal components and parameter spaces of Tofts and Brix models, were developed for the datasets listed in Table C.1 in Appendix. In the following sections, these three feature spaces are summarized.

Principal component feature space

Here, the dataset is reduced on the plane of two first principal components. This kind of dimension reduction has been explained in details, in section 3.1. It has been observed that along the first principal axis, the washout phases of the signal curves are parallel but with different intensities. Along the second principal component, variations in the shape of the washout phases can be observed. These observations are depicted for a dataset in Figure 4.2. The Figure demonstrates signal curves corresponding to the two-dimensional data points selected from the plane of two first principal components of a dataset. The selected data points can be divided into two groups:

1. Ten equally distanced data points are selected from the line parallel to the first principal component and through the zero value of the second principal component. Signal curves of these data points are depicted in Figure 4.2a.

2. Ten equally distanced data points are selected from the line parallel to the second principal component and through the zero point of the first principal component. Signal curves of these data points are depicted in Figure 4.2b.

In the case of strongly noisy datasets, the noise is present in the second group of signal curves. These two groups of data points from the plane of first and second principal components are depicted for all 15 datasets in Figures C.12 and C.13 in Appendix.

Tofts and Brix parameter spaces

For the Tofts feature space each signal curve is described by two features K_{trans} and K_{ep} from Tofts model described in section 2.4.4. Similarly, in Brix feature space, each signal curve is described by three features A, K_{21}, and K_{out} from Brix early model described in section 2.4.3.

For each dataset, the outliers in each parameter are sorted out. The outliers are found as values outside the interval:

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]\,, \tag{4.7}$$

where Q_1 and Q_3 are the lower and upper quartiles and k is a factor and was set to 1.5. After elimination of outliers, each parameter is normalized. The normalization is necessary due to the large differences between domains of the parameters. The normalization of a parameter is done as follows: Let v_i be a vector described by m parameters v_{ij}, with $j = 1, ..., m$. The normalized value $n(v_{ij})$ for parameter j of the vector v_i is:

$$n(v_{ij}) = \frac{v_{ij} - \bar{v}_j}{\sigma_j}, \tag{4.8}$$

where $\bar{v}_j = \frac{1}{n}\sum_{i=1}^{n} v_{ij}$ is the mean value of j-th parameter for all vectors in the dataset and σ_j is the standard deviation of that parameter calculated as $\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}(v_{ij} - \bar{v}_j)^2$.

Figure 4.3 compares the quality of fitting between Tofts and Brix models. The comparison is done based on average of χ^2 values of the fitting for all curves in a dataset.

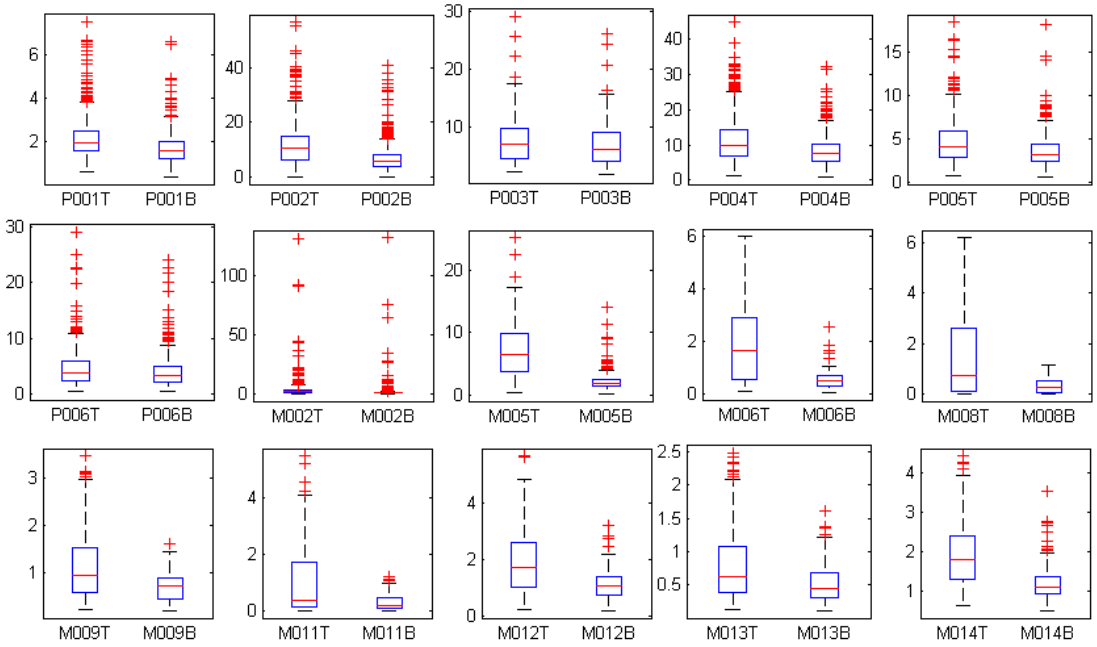


Figure 4.3: Comparison of *Chi* values of fitting using Tofts model versus Brix model for each dataset. The *Chi* value for Tofts model is marked with T and for Brix model with B.

Pearson's chi-squared test [Greenwood and Nikulin, 1996] is a measure of goodness of fit and is given by:

$$\chi^2 = \sum_{i=0}^{n-1} \frac{(S_i - E_i)^2}{E_i}, \tag{4.9}$$

where S_i is the measured value of signal point at time point i and E_i is the calculated fit value at that time point. In all datasets, the Brix model results in smaller χ^2 values than Tofts model. The fitting in the Tofts model is very unstable and the results strongly depend on given starting values. In contrast, fitting to the Brix early model is quite stable. For the calculation of fitting parameters, the Levenberg—Marquardt algorithm [Pressa et al., 2007] was used.

4.2.4 Two-steps clustering

A new clustering method is introduced based on the advantages of the PM similarity measure in combination with the Euclidean distance. The clustering method is performed in two steps.

- First step: Hierarchical clustering with complete linkage method is performed on the datasets using the PM similarity measure. Hierarchical clustering has the advantage of representing the whole structure of a dataset presented by a selected similarity measure. Therefore, it is independent of starting points or other parameters and for a selected number of clusters, the result is deterministic.

  The result of hierarchical clustering can be represented in a dendrogram. A dendrogram is a binary tree and each level of the tree demonstrates the amount of dissimilarity in which two clusters are merged together. Thus, it is sufficient to define the maximal allowed dissimilarity between the curves inside a cluster and then cut the dendrogram at that specific level. Accordingly, the number of clusters is automatically determined.

  In the two-steps clustering, the threshold is estimated from the upper bound of PM measure for two curves according to the average noise of the dataset (defined in section 4.2.1). Let m be the number of time points in washout phase and ς is the average noise of the dataset. With assumption that enhanced tumor curves have a noise less than average noise, the worst case PM for two parallel curves is estimated as follows: The point wise distances $d^{ii'}$ for two noisy parallel signal curves S_i and $S_{i'}$ are at half of the m time points equal to $2\ \varsigma$ and in the other half are equal to zero. Consequently, the worst-case PM measure for two parallel curves is $\sqrt{2m}\ \varsigma$. Using the hierarchical clustering with complete linkage method, we guarantee that each level of dendrogram represents the maximal dissimilarity inside the clusters at that level. To allow a small amount of variation and compensate the cases where noise of the data is very small, 0.5 as minimum threshold value is added to the $\sqrt{2m}\ \varsigma$. The number 0.5 is selected experimentally. To achieve the desired clustering is enough to cut the dendrogram at a level $\leq 0.5 + \sqrt{2m}\ \varsigma$.

- Second step: Each cluster from the previous step is further divided by hierarchical clustering with complete linkage method and the Euclidean distance as the distance measure. The granularity of clustering is controlled by a threshold.

  The threshold is defined in a similar way as in previous step. Suppose l is the shortest distance between two parallel curves. For two signal curves with parallel washout phases, the maximal accepted distance l for all curves in a cluster is the basis of threshold selection. The upper bound of Euclidean distance, for signal curves in a

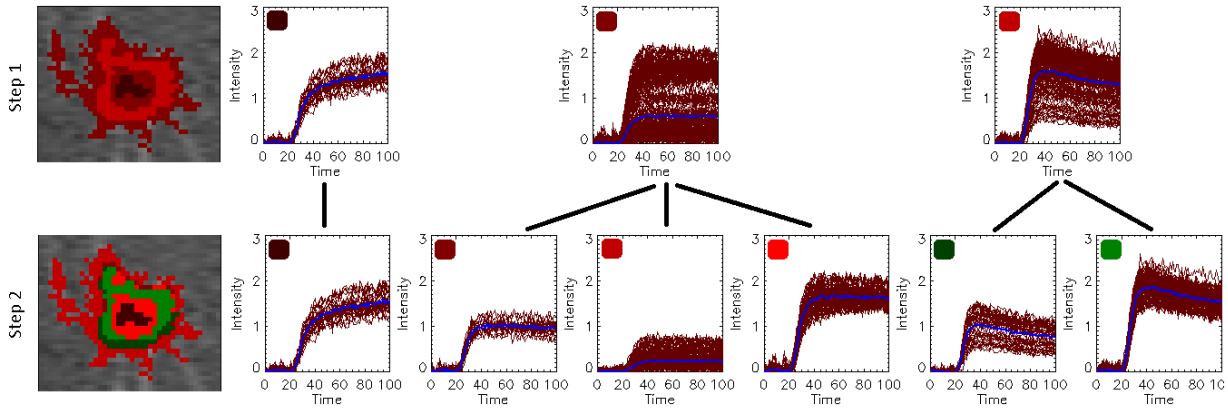


Figure 4.4: The result of two-steps clustering for the tumor region of a dataset. In the first step, the tumor is divided into three sub-regions. The curves inside each cluster are depicted on the left. In the second step, two of the three clusters are divided further.

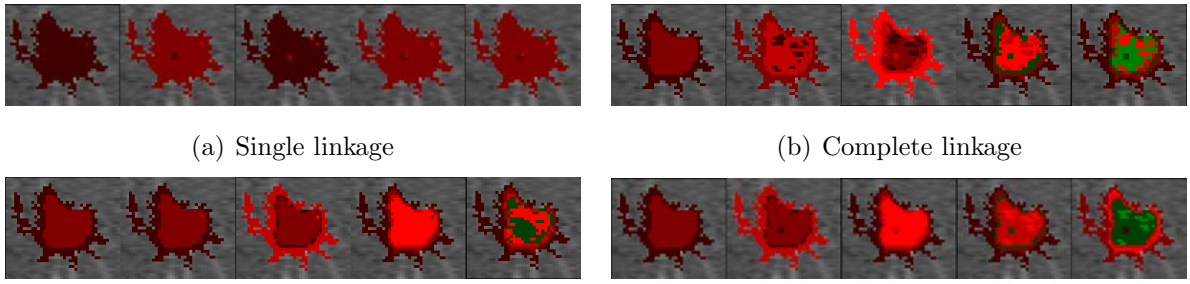
cluster can be estimated as $(l + 2\varsigma)\sqrt{n}$, with n is the number of time points in the signal curve and ς is the average noise of the dataset. This upper bound is estimated for two signal curves with point wise distances equal to $l + 2\varsigma$. In the experiments here, $l = 1.0$. That means parallel signal curves with a peak difference 1.0 or less in relative intensities, will be clustered together.

Figure 4.4 demonstrates the result of the two-steps clustering for a dataset. The results of two-steps clustering for all datasets can be found in Appendix C. The tumor region is clustered into three clusters in the first step of the clustering. The three clusters can be described by types of washout patterns such as, slow sustained enhancement, stable late enhancement, and decreasing late enhancement. In the second step of clustering, two clusters are divided further into new clusters.

The introduced two-steps clustering is compared with other clustering methods, described in section 3.2, such as: k-means, fuzzy c-means, mean shift, and hierarchical clustering. The results of these clusterings have been evaluated in the following section.

4.3 Experiments and results

This section presents the results of the experiments on 15 DCE-MRI breast datasets summarized in Table C.1 in Appendix. Different experiments have been designed to compare and evaluate a variety of clustering methods and similarity measures. The clustering has



(a) Single linkage (b) Complete linkage

(c) Average linkage (d) Centroid linkage

Figure 4.5: Clustering results of four linkage methods with Euclidean distance as the similarity measure for a dataset with different number of clusters (2-6).

been performed only on the segmented tumor area in each dataset, either on the signal curves directly or on the feature vectors of the signal curves.

4.3.1 Comparison of different linkage methods

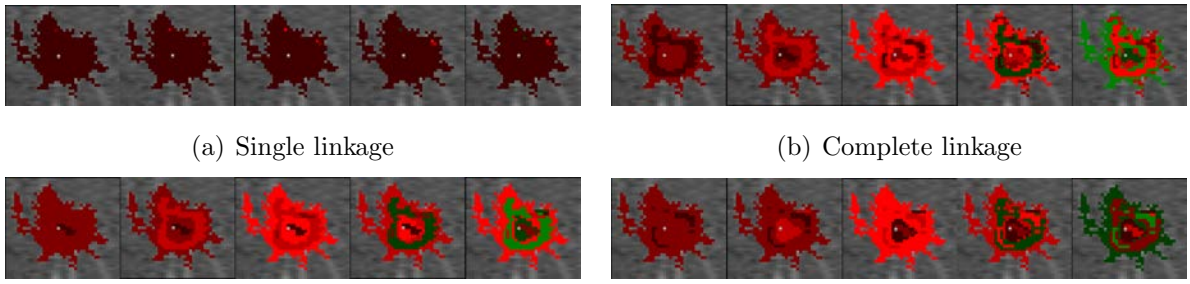
In the first group of experiments, the hierarchical clustering with four different linkage methods, single linkage, complete linkage, average linkage, and centroid linkage (for more details on linkage methods see section 3.4.1), has been performed on a dataset. Each linkage method has been applied five times with a different number of clusters from two to six. Euclidean distance has been used as the distance measure. Figure 4.5 demonstrates the results obtained from these experiments.

As expected, the single linkage method is not able to find structures inside the tumor. The single linkage is sensitive to noise. For not well-separated clusters, single linkage results in one cluster that is a chain of all not well-separated clusters.

The average linkage and centroid linkage show quite a similar behavior in the cases of two, three, and four clusters. In general, the three methods, complete, average, and centroid linkage agree on three regions of the tumor. The outer ring structure of the tumor, the central part and a border between the outer and the inner part are recognized by all these three methods.

Figure 4.6 shows the results obtained from similar experiments, but this time with the PM measure. Similar to previous experiments, the single linkage method results in a large cluster and several singleton clusters.

Figure 4.7 shows the signal curves inside the clusters for the clustering with complete,



(a) Single linkage (b) Complete linkage

(c) Average linkage (d) Centroid linkage

Figure 4.6: Clustering results of four linkage methods with PM measure as the similarity measure for a dataset with different number of clusters (2-6).

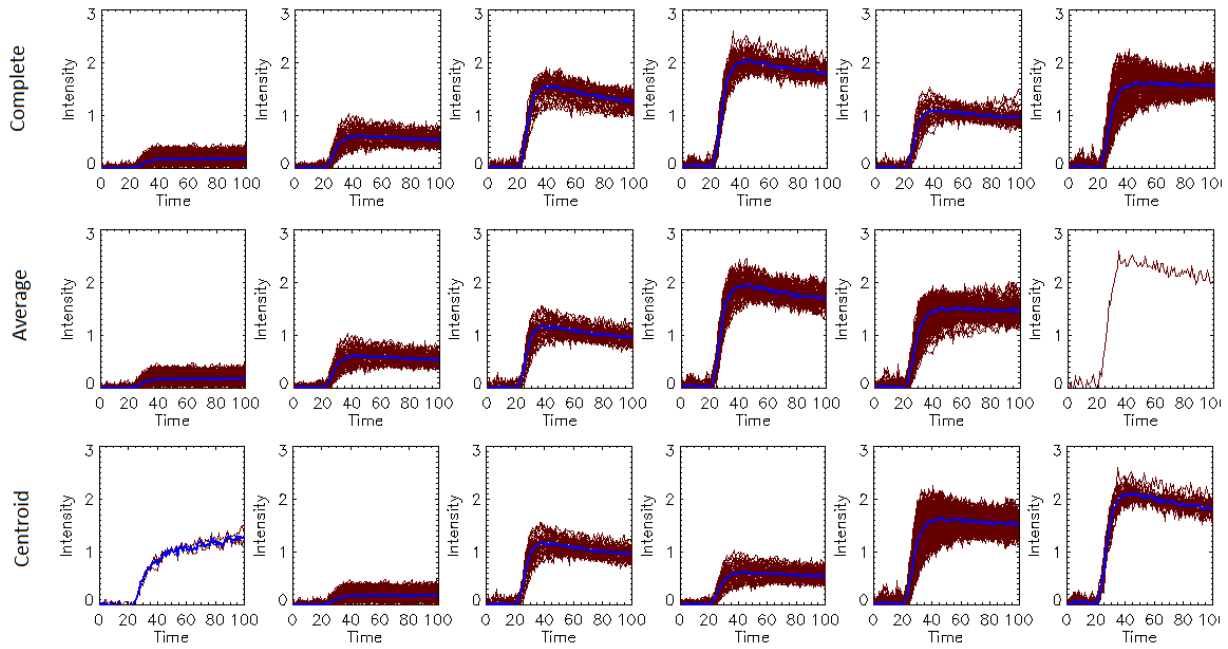
average, and centroid linkage with Euclidean distance on the top (a), and with the PM measure on the bottom (b). The number of clusters is six. For each similarity measure, the results obtained from different linkage methods are very similar. Only the order in which each cluster is found varies between different linkage methods. In contrast, the results of clustering using different similarity measures are distinguishably different.

In complete linkage method, it is possible to control the maximal dissimilarity between all data points in a cluster. This property can be very useful in the clustering of DCE-MRI datasets. It is enough to determine a threshold as allowed dissimilarity between two signal curves. For this reason, complete linkage is the clustering method used in the two-steps clustering introduced in the present work.

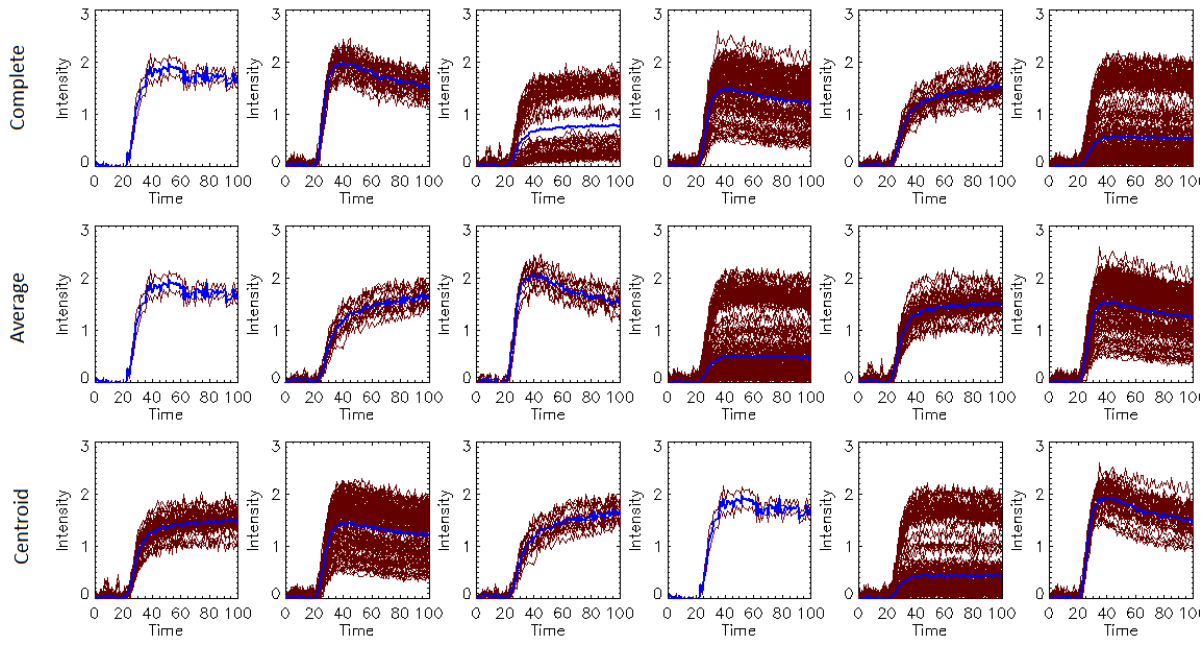
4.3.2 Mean shift clustering

As mentioned in section 3.4.2, mean shift clustering is a mode seeking method. This clustering method looks for the points with high-density neighborhood. The search is based on slightly shifting the center of a neighborhood toward the higher density location. This mechanism works only in a convex neighborhood where the distance between the data points in the neighborhood is based on a metric [Comaniciu and Meer, 2002]. The triangle inequality of metric space guarantees that the distances between the points inside the neighborhood are smaller or equal to the diameter of the neighborhood kernel.

The introduced PM measure violates the triangle inequality, and therefore, is not a proper measure for mean shift clustering. An example is shown in Figure 4.8 obtained from the mean shift clustering using the PM measure for a DCE-MRI dataset. The clustering



(a) Euclidean distance



(b) PM measure

Figure 4.7: Comparison of signal curves in each cluster for results of clustering using Euclidean distance and the PM measure.

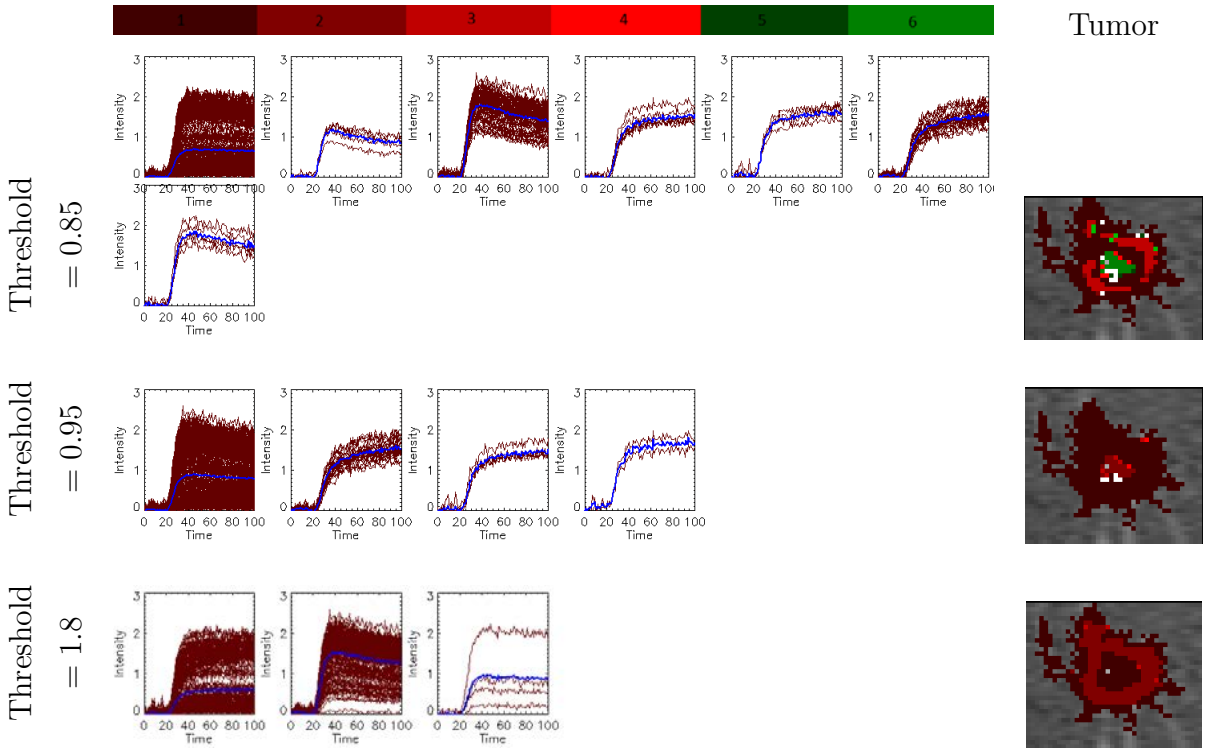


Figure 4.8: Mean shift clustering with PM measure for a DCE-MRI dataset with three different neighborhood radii.

was performed three times with different radii for the neighborhood function. The dataset, which is used in these experiments, consists of three major washout patterns, up going, down going and constant. The example shows that mean shift clustering using PM measure have difficulty finding these three groups. In clustering examples with the radius of 1.8 and 0.95 only two major categories could be found. Only Mean shift clustering with the radius of 0.85 was able to find all three major categories. This example also demonstrates the problematic of selecting the proper radius for the mean shift.

In additional experiments, the mean shift with the Euclidean distance has been performed in the space of the first two principal components and on the feature space of Tofts parameters for the same dataset. The difficulty to choose the proper radius remains the main problem. Figure 4.9 depicts the results of clustering in both cases for different numbers of neighborhood radii. As it can be observed, the results vary with different choices of the neighborhood radii. Due to this variation in results, and difficulty to select the proper radius, it is difficult to compare the mean shift clustering systematically with other clus-

tering methods. For this reason and also because the mean shift method is not suitable for semi-metric spaces, it has been evaluated as improper for clustering of DCE-MRI datasets.

4.3.3 Comparison of different clustering methods

In this section, the results obtained from thirteen different clustering scenarios on the breast datasets described in Table C.1 or on the corresponding feature-spaces are compared. The clustering experiments can be divided into four main groups. The first group consists of the clustering experiments performed directly on breast datasets. The second group consists of experiments performed on the reduced datasets on the plane of the first two principal components. The third and fourth groups consist of the results of clustering performed on the feature datasets. In the third group, the features are the parameters calculated from fitting the signal curves with Tofts model from section 2.4.4. In the last group, the features are obtained from fitting parameters using early Brix model from section 2.4.3. Each group consists of results of three clustering methods: k-means, c-means and average linkage method. In addition to these twelve clusterings, the introduced two-steps clustering from section 4.2.4 is performed on real datasets.

In both, Tofts feature space, and Brix feature space, the outliers were eliminated from the feature dataset as described in section 4.2.3. In the next step, features without outliers are normalized as explained in section 4.2.3. The three different clustering methods were performed on the normalized feature vectors.

For the comparison, two properties of the resulted clusters are evaluated. The first property is γ the average of the Euclidean distances of curves inside the clusters to the mean curve of that cluster. Suppose a dataset X with n signal curves S_i, and $i \in 0, ..., n-1$ is clustered into k clusters C_j, $j \in 0, ..., k-1$. The mean curve of the cluster C_j is:

$$\mu_j = \frac{1}{n_j} \sum_{S_i \in C_j} S_i, \tag{4.10}$$

where $n_j = |C_j|$, is the number of signal curves in the cluster C_j. Then the parameter γ for a clustering result is calculated as:

$$\gamma = \frac{1}{n} \sum_{\substack{j=0 \\ \forall S_i \in C_j}}^{k-1} \|S_i - \mu_j\|^2, \tag{4.11}$$

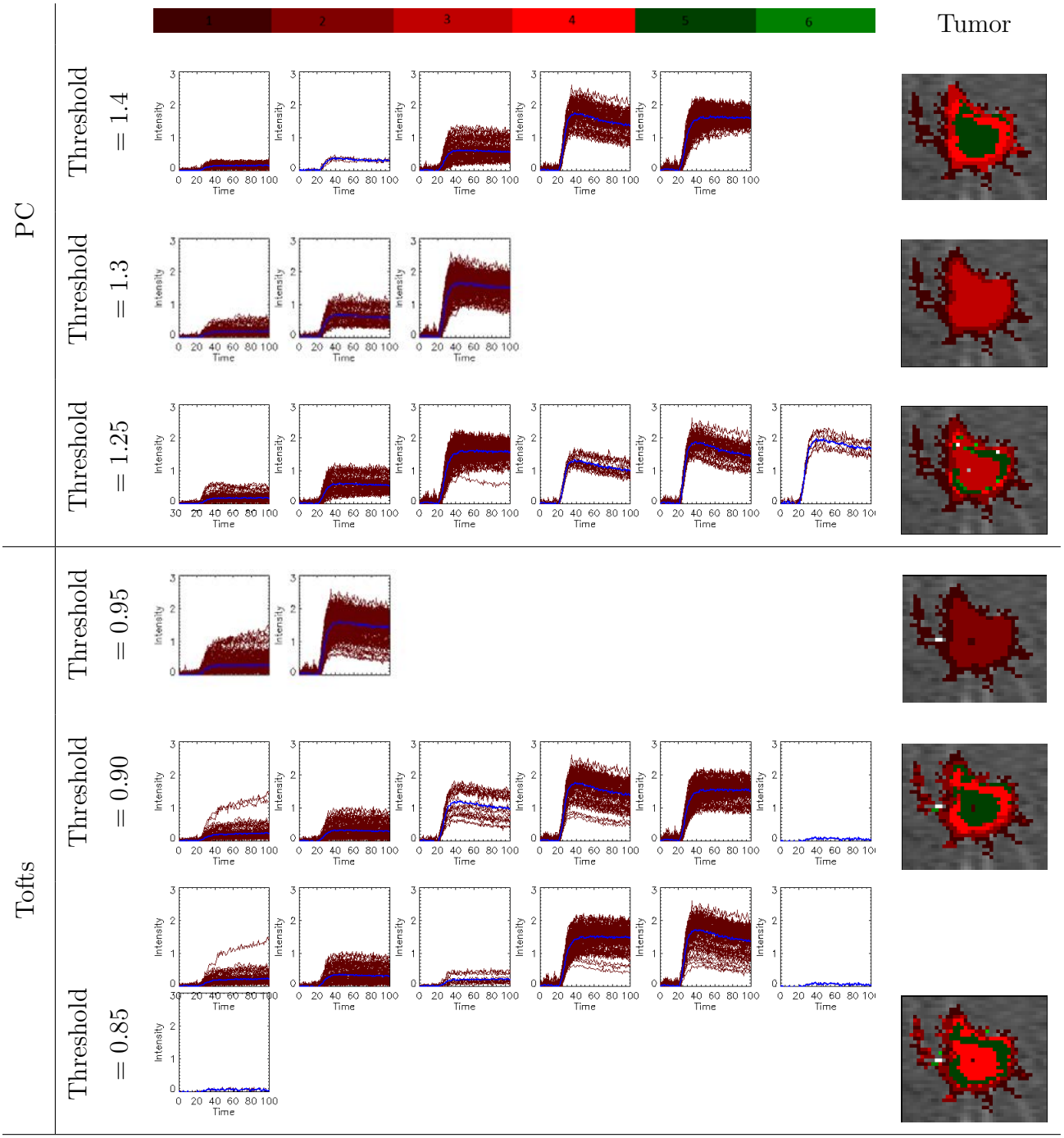


Figure 4.9: Mean shift clustering performed on principal components' space and Tofts feature space of a DCE-MRI dataset with three different neighborhood radii.

Datasets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nr. of clusters	10	31	3	52	9	10	14	10	3	6	2	3	4	5	4

Table 4.1: Number of clusters determined by two-steps clustering

where $\|S_i - \mu_j\|$ is the Euclidean distance between the signal curve S_i and the mean curve μ_j. The parameter γ can be considered as a kind of squared standard deviation of the clustering result.

The other property is the averaged differences ξ between the slopes of the signal curves inside a cluster and the average slope of the cluster. For a signal curve S_i, the slope A_i of the signal curve is determined as follows:

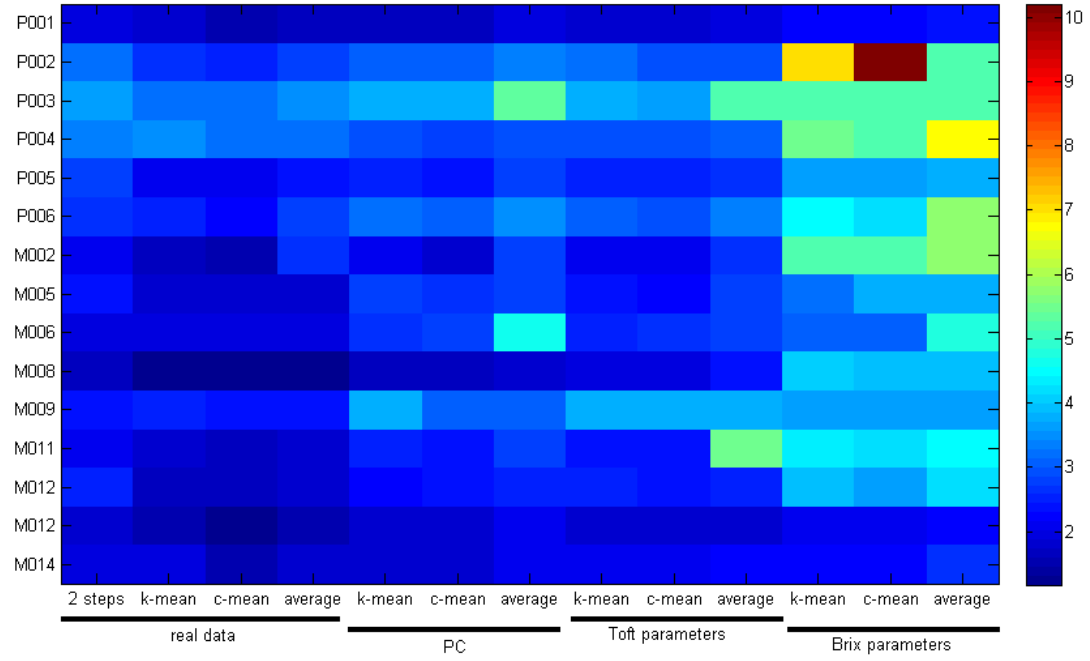
- As described in section 3.6.4, the start point t_w of the washout phase is determined for the average curve of all signal curves in dataset X.
- In the second step, a line $A_i x + B_i$ is fitted to the washout time points S_i^t, with $t = t_w, ..., n-1$.
- The property ξ is determined from the equation:

$$\xi = \frac{1}{n} \sum_{\substack{j=0 \\ \forall S_i \in C_j}}^{k-1} \left\| A_i - \bar{A}_j \right\|^2, \tag{4.12}$$

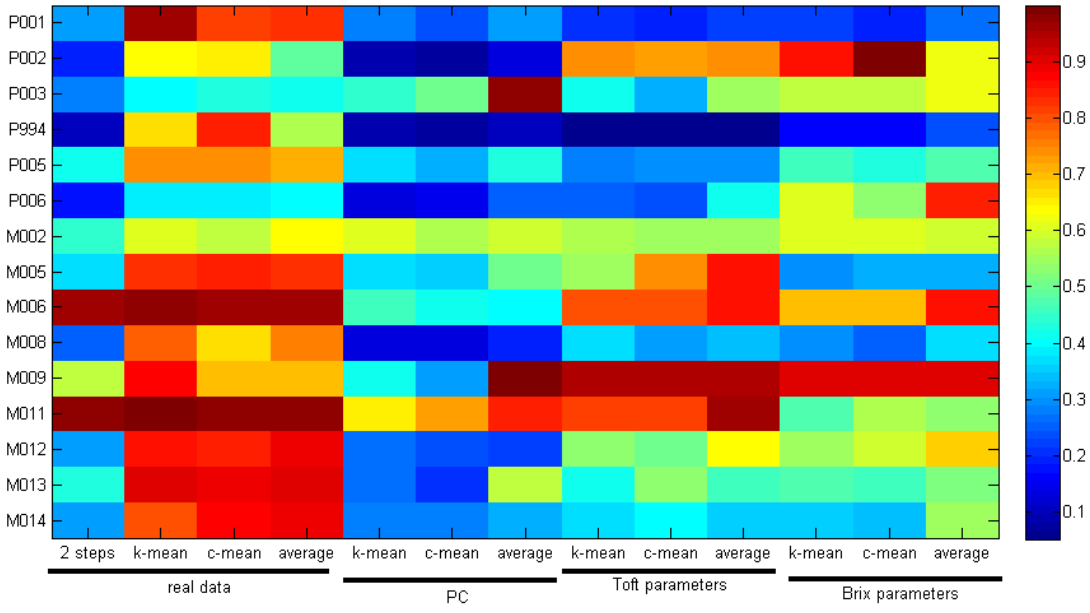
where A_i is the slope of the signal curve S_i, $\bar{A}_j$ is the average of slopes in cluster C_j, and k and n are number of clusters and number of data points, respectively.

It must be taken into account that before the evaluation, all the singleton clusters have been eliminated so that the clusters used for the calculation of γ and ξ have at least two members. All the clustering methods have been performed with the same number of clusters for a dataset. The number of clusters for a dataset is determined by the two-steps clustering. The estimated numbers of clusters from two-steps clustering on these 15 datasets are given in Table 4.1.

The values of γ and ξ for the 15 datasets are depicted as different RGB colors in Figure 4.10. A better performance is achieved by smaller values of γ and ξ. As it can be observed from Figures 4.10(a) and 4.10(b) in the case of γ values, k-means and c-means clustering methods on real datasets have a better overall performance. In the case of ξ



(a) γ



(b) ξ

Figure 4.10: The values γ and ξ for each dataset and different clustering methods are depicted as colors.

values, the k-means and c-means on reduced datasets on the plane of principal components show a better performance.

The introduced two-steps clustering exhibits in the case of γ values a comparable performance to the clustering results of principal components plane and Tofts parameters. In the case of ξ values, the two-steps clustering is slightly behind principal components and Brix parameters results. Summarizing the results together, two-steps clustering and clustering on the plane of the first two principal components have the best overall performance.

For a better understanding of differences between clustering methods, the results obtained from the thirteen clustering experiments are depicted for the same dataset used in sections 4.3.1 and 4.3.2. In the selected dataset three different patterns of the washout phase are distinguishable: up going, constant, and rapid clearance. The white pixels of the tumor area are the data points that are not considered in the clustering because they were selected as outliers in the feature space, or they were clustered to singleton clusters and therefore, eliminated from the results.

From Figure 4.11 can be observed that k-means and fuzzy c-means methods performed directly on vectors of signal curves are not able to consider the washout patterns of the curves. This observation is also confirmed by the calculated ξ values depicted in Figure 4.10(b). At the same time, k-means and fuzzy c-means methods performed on data space show significantly smaller differences between maximal enhancement (γ values), which is directly related to the Euclidean distance between two curves.

As demonstrated in section 4.2.3, fitting to the Brix model has smaller Chi values than the Tofts model. However, it is difficult to relate the parameters in the Brix model directly to peak or amplitude of the signal curves. For this reason, clustering, in feature space of Brix parameters, results in larger γ values. In contrast, different patterns in washout phase are successfully grouped as depicted in Figure 4.12.

4.3.4 Comparison of similarity measures

The introduced similarity measure PM in section 3.6.4 is compared to three widely used similarity measures: Euclidean distance, correlation coefficient and cosine measure. For the comparison, the same values γ and ξ from section 4.3.3 have been used. The complete linkage method is used as the clustering method. The clustering has been performed four times each time with a different similarity measure. The 15 breast datasets have undergone

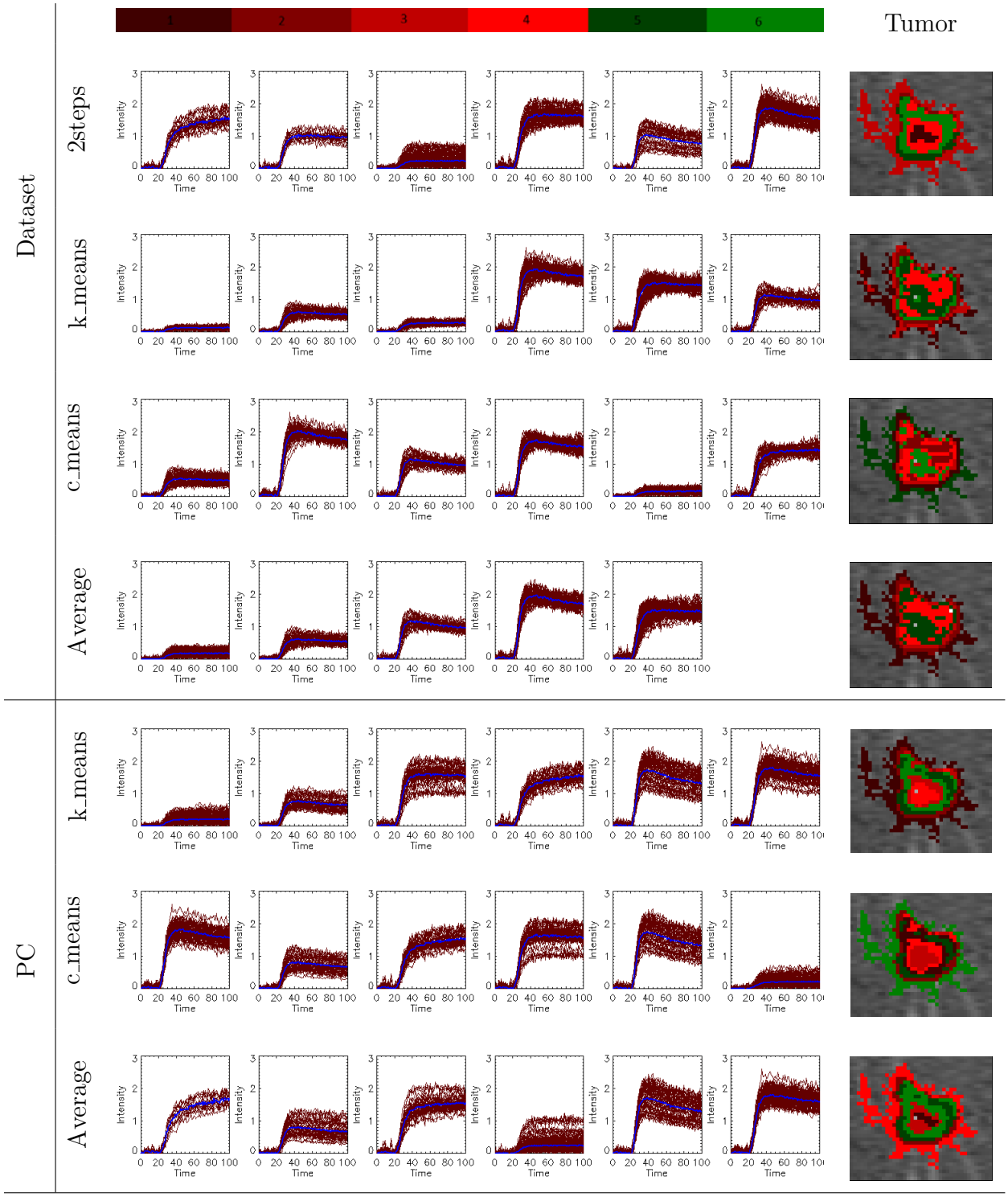


Figure 4.11: Result of different clustering methods on a dataset, performed in data space and in the plane of principal components.

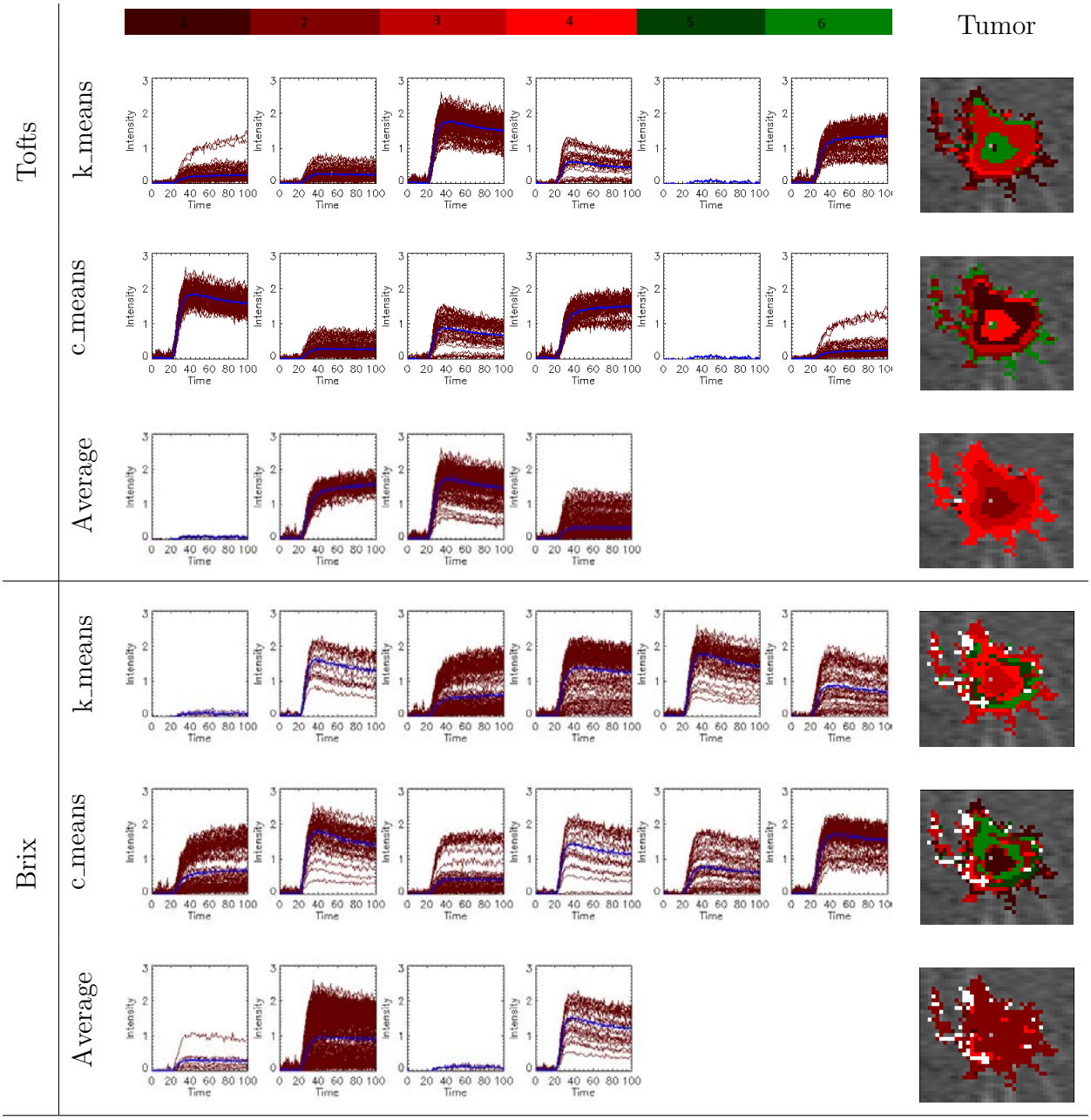
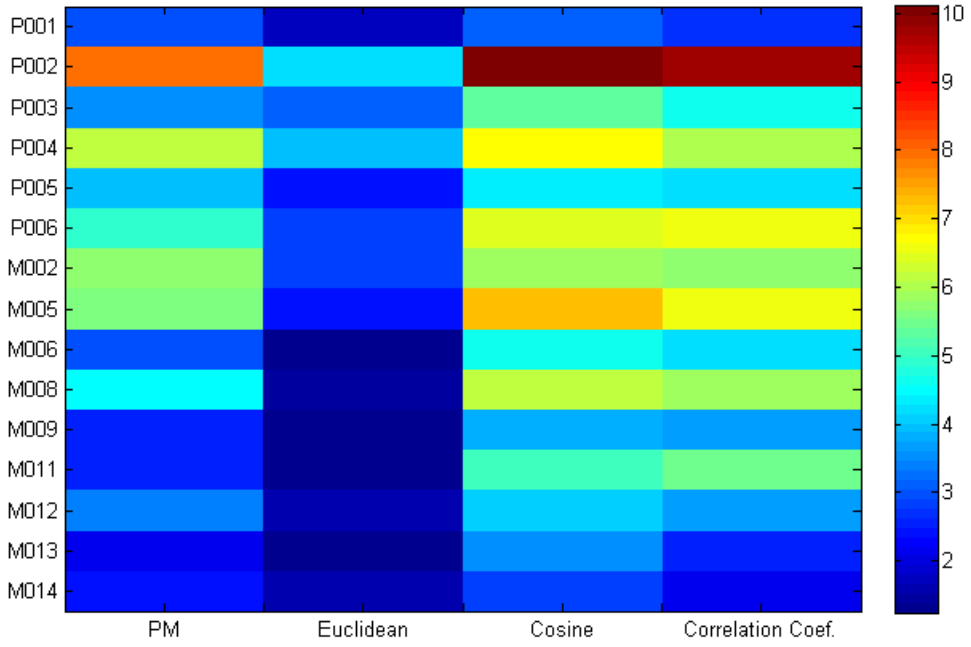
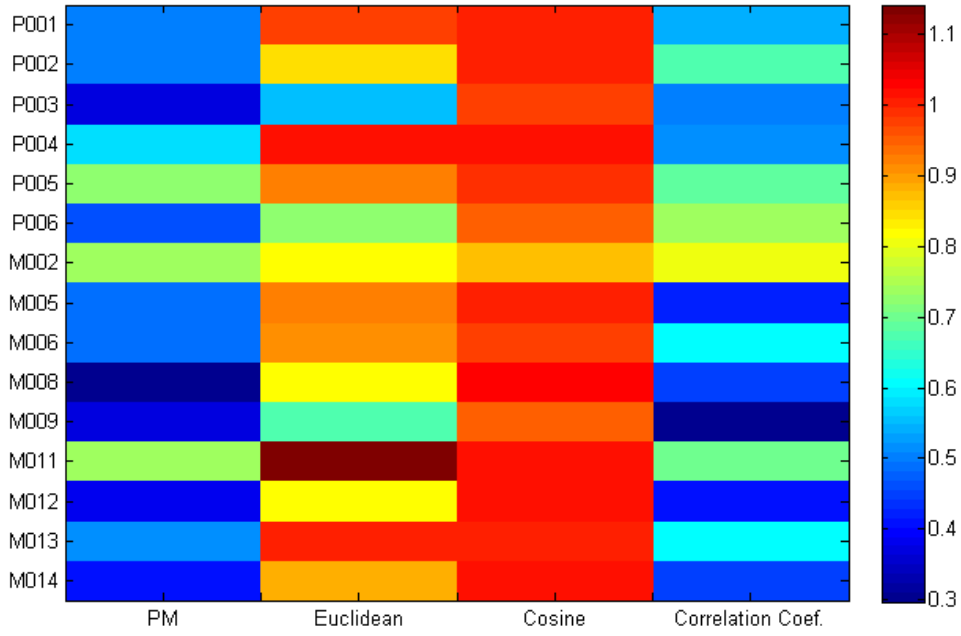


Figure 4.12: Result of different clustering methods on a dataset, performed in Tofts and Brix parameter spaces.

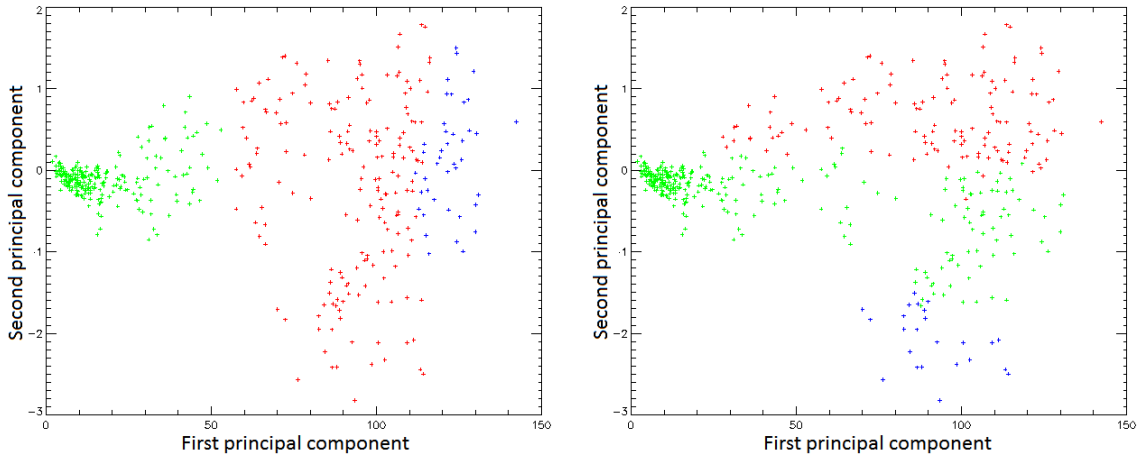


(a) γ



(b) ξ

Figure 4.13: Result of clustering with different similarity measures.



(a) Complete linkage using Euclidean distance (b) Complete linkage using PM measure

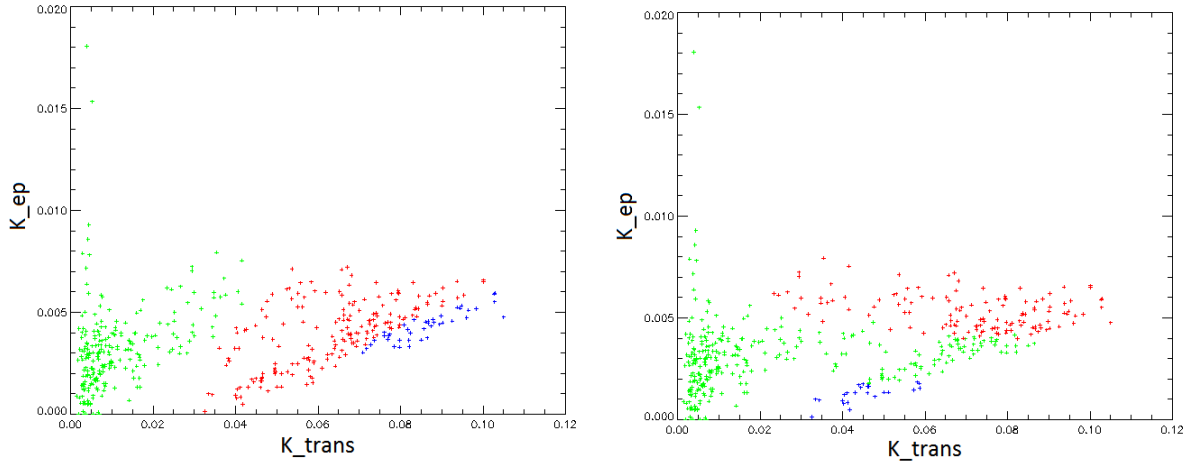
Figure 4.14: Three clusters of clustering results, using Euclidean distance and PM measure, are displayed in the plane of first and second principal components.

the experiment with the same number of clusters $k = 5$.

The results of the experiments are depicted in Figure 4.13. As expected, clustering with Euclidean distance results in smaller γ values. Nevertheless, clustering with cosine or Euclidean distance is not able to take the shapes of the signal curves into account and results in large ξ values. A large ξ value means more variety among the slopes of signal curves inside a cluster. The PM measure performs slightly better than the correlation coefficient for both values.

Figures 4.14 depicts the relationship of the Euclidean distance and the PM measure to the two first principal components for the selected dataset from previous section. The dataset is clustered by the complete linkage method directly performed on the data space with three as the number of clusters. Figure 4.14(a) shows the results of clustering with complete linkage method using Euclidean distance projected on the plane of first two principal components, while Figure 4.14(b) shows the same clustering using PM measure as the similarity measure. As expected clustering with Euclidean distance clusters the dataset along the first principal components. Using PM as the similarity measure, the dataset is clustered along the second component.

Figure 4.15 shows the clustering results in Tofts parameters' space. Figure 4.15(a) shows the result of clustering with the Euclidean distance while Figure 4.15(b) shows clustering using PM measure as the similarity measure. As expected the Euclidean distance is more



(a) Complete linkage using Euclidean distance

(b) Complete linkage using PM measure

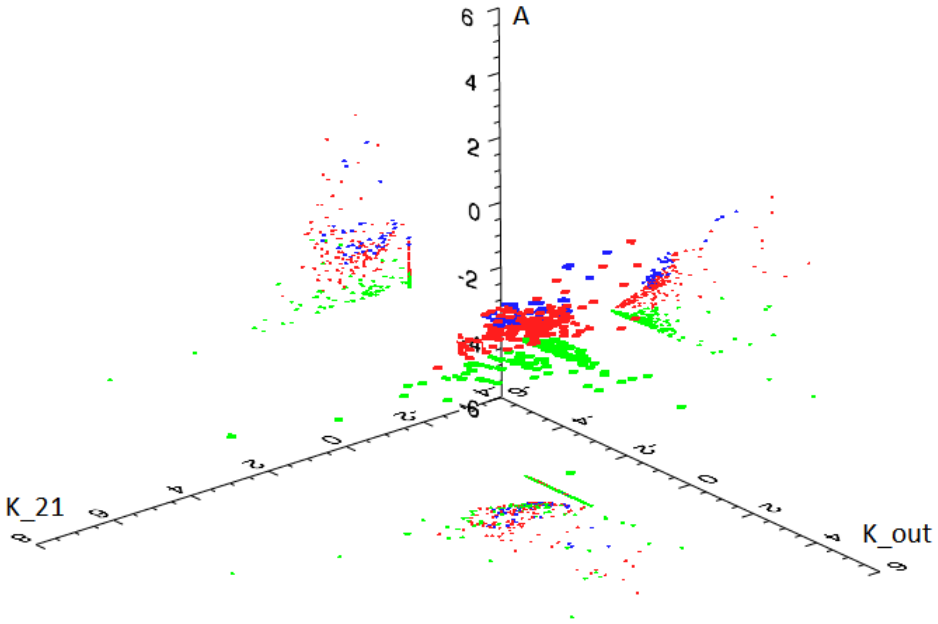
Figure 4.15: Three clusters of clustering results, using Euclidean distance and PM measure, are displayed in the Tofts parameters' space.

related to K_{trans} parameter and PM measure is more related to K_{ep}. Figure 4.16 shows the same results in Brix parameters' space. In Brix case, a direct relationship between the Euclidean distance and the PM measure with Brix parameters does not exist. However, the clusters are well separated in the plane of parameters A and k_{trans} in Figure 4.16.

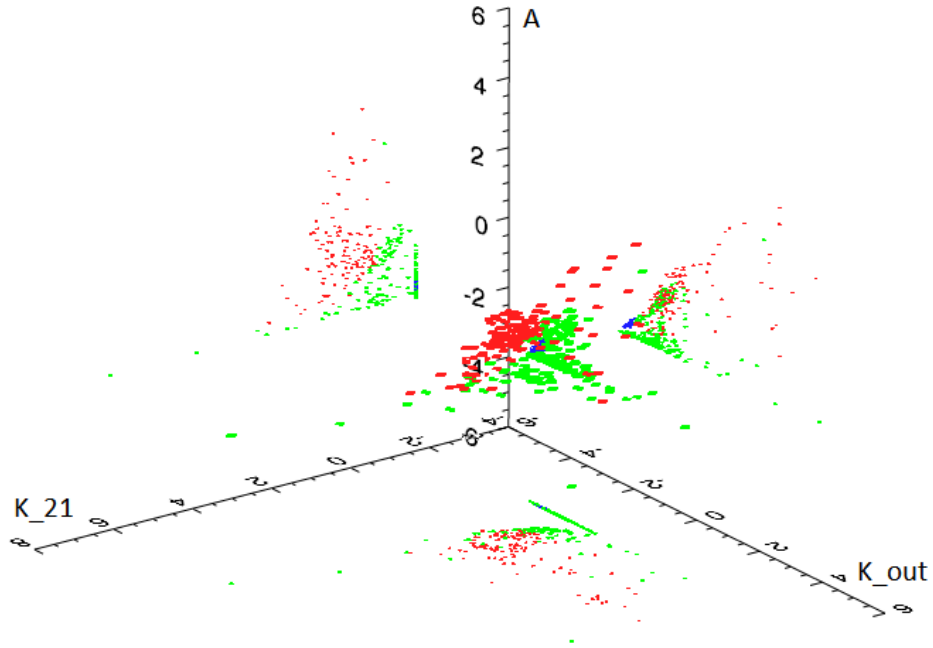
4.4 Discussion

The introduced similarity measure and the corresponding two-steps clustering were performed on 15 datasets of breast DCE-MRI. The clustering was performed on signal curves of the tumor region in each patient. The tumor's voxels were segmented according to the steps described in section 4.2.1. The two-steps clustering from section 4.2.4 was compared against 12 other clustering scenarios. Also, the PM measure was compared against three other similarity measures.

To evaluate the quality of a clustering method two properties γ and ξ are introduced. First measure γ is responsible for compactness of the clusters with respect to the Euclidean distance. The other value ξ measures the variety of washout slopes inside a cluster. The evaluation is based on the fact that the goal of clustering is to achieve compact clusters with low variation of washout shapes, which means smaller values for γ and ξ.



(a) Complete linkage using Euclidean distance



(b) Complete linkage using PM measure

Figure 4.16: Three clusters of clustering results, using Euclidean distance and PM measure, are displayed in the Brix parameters space.

From the results it is evident that using the Euclidean distance as the measure of dissimilarities between signal curves, neglects the shape of the curves. Thus, clustering with Euclidean distance results in larger ξ values as depicted in Figure 4.10(b). This behavior is also observed in the experiments comparing four similarity measures. In comparison of γ values of the three clustering methods, k-means, c-means, and average linkage, one can say that c-means has the best performance. Whiles k-means and average linkage have similar behavior.

In the other nine scenarios, different clustering methods are used on transformed datasets in the feature spaces. The clustering methods are again k-means, c-means, and average linkage. The distance between the data points in the feature space is measured by the Euclidean distance. From Figure 4.14 it is evident that Euclidean distance is related to the first principal component whereas the PM measure is related to the second principal component. A similar relationship can be observed between these two similarity measures and Tofts' parameters in Figure 4.15. For this reason, the results of clustering in principal components space and Tofts' parameters space, are quite comparable with two-steps clustering. In both feature spaces, principal components and Tofts' parameters, the c-means clustering has the best performance. The clustering in principal components space has slightly better performance than two-steps clustering and the clustering in Tofts' parameters space.

In Figure 2.8 it can be observed that the three clusters obtained using the Euclidean distance or the PM measure are not completely separable in Brix' parameters space. This observation is confirmed in clustering results. Again c-means clustering has the best performance follows by k-means clustering.

There are several aspects that must be taken into account about mentioned results. In general, it is difficult to compare the results of two clusterings. For two clustering results with the same number of clusters, the number of data points in each cluster varies. That is why the two evaluation values, γ and ξ, only depend on the number of all data points. On the other hand, singleton clusters had to get eliminated from clustering results before evaluation, since they are undesired. For this reason, it might happen that the number of data points differs from one result to the other. Furthermore, in a parameter space, it might happen that the data points should be eliminated because the corresponding calculated parameter is an outlier.

The other aspect is that the calculated γ and ξ values for various clustering results of

a dataset might be comparable. However, they are hard to compare with other datasets. Therefore, in the present work, a joint evaluation of the γ and ξ values is not given.

Even so, from the results, the success of clustering using two-steps clustering on real datasets, and c-means clustering on reduced datasets is clear. Despite its poorer performance compared to the results obtained in the principal components space, the two-steps clustering has the advantage of determining the number of clusters as part of clustering process. The number of clusters is *a priori* unknown. Fuzzy c-means clustering needs the number of clusters as a parameter.

The number of clusters is determined by means of thresholding in every step of the two-steps clustering. Thresholds are defined according to the noise in the signal curves and maximal allowed variation in a cluster.

The two-steps clustering is better than c-means in principal components space in regard to γ values. However, it is poorer with respect to ξ values. It is left to the user which property is more important. Finally, besides the introduced evaluation methods, in a further step, the results must be compared with gross pathology cuts of the tumors.

Chapter 5

Cluster analysis of liver datasets

In this chapter, we will discuss clustering of signal curves in perfusion MRI of liver datasets. The introduced method for breast tumors has been adopted to be used in the case of DCE-MRI of the liver. Two major issues in adopting breast clustering methods have to be considered. These consist of strong noises and motion artifacts in the liver datasets and a larger number of enhanced voxels. In DCE-MRI of the liver, the whole liver tissue is enhanced. Therefore, segmenting the tumor from the surrounding tissue in the same way that is done for breast tumors is impossible. Not only the tumor, but also the whole liver tissue is quite heterogeneous. For this reason, in addition to the tumor, also the clustering of the liver tissue itself is of interest.

5.1 Material

5.1.1 Patients

Nine datasets of patients with hepatocellular carcinoma (HCC) from the Radiology Institute of Klinikum rechts der Isar [Radiology, 2011] are studied in this work. The DCE-MRI was performed after starting the treatment of the cancer. To reduce the respiratory caused motions, the patients were asked to follow a volunteer breath holding protocol. For this, before starting with acquisition, the patients were trained to breathe in 4-5 seconds intervals. The volume acquisition was only performed during the idle time of breathing protocol. Each volume acquisition was performed manually by human.

5.1.2 MRI Imaging

MRI measurements were performed on a clinical 1.5T scanner (Magnetom Avanto; Siemens) using a transmit body coil. Images were acquired in trans-axial orientation, applying a 3D (Volume Interpolated Breath-hold Examination) VIBE dynamic with manual triggering in expiration to further reduce motion artifacts. The delay time between the start of the VIBE dynamic and the intravenous Gd-DTPA contrast agent injection at a standard dose of 0.1 $mmol\Delta kg^{-1}$ was approximately 25 seconds. Images were acquired for a minimum time period of 5 minutes at 3 - 5 second intervals. Slice thickness was 5 mm with no intersection gap. Images were obtained with a TR/TE of 3.9/0.96 ms, a flip angle of 15°, a FOV of 375×375 mm^2 and a matrix size of 128×128.

5.2 Method

The improvement in the MRI technique and the shortening of acquisition time introduce new possibilities to consider DCE-MRI as a method to diagnose and characterize tumors in the liver tissue [Bartolozzi et al., 1999, Rummeny and Marchal, 1997, Totman et al., 2005, Tsushima et al., 2001, Van Beers et al., 1997]. Although, the introduced methods in chapter 4 in the case of breast tumors cannot be adopted directly for the liver tissue. This is due to the strong movement of the organ and its complex vascular system. Different sources of motion are peristalsis of the stomach, breathing motion and heart beats. To reduce the noise caused by motion, instead performing clustering methods direct on the space of the signal curves, they are performed on a sub-space of the k first principal components. Furthermore, the PM similarity measure introduced for the two-steps clustering of the breast tumors is too sensitive to strong noise. Therefore, a modified version of PM measure is introduced here. However, the most significant difference in the clustering of signal curves in the case of breast tumors to liver tumors is due to the enhancement of the whole liver tissue. Therefor, whole liver is the subject of the clustering.

5.2.1 Suppressing the noise using principal component analysis

Noise and motion artifacts are more apparent in DCE-MRI of the liver compared to other organs. These artifacts make it difficult to analyze the signal curves in DCE-MRI of liver. To suppress noise and to be able to compare the most significant patterns of the signal

curves, the dataset is transformed into a new dataset in the subspace defined by the k five principal components of the data.

Considering the voxels of DCE-MRI volume as samples defined by vectors. Each vector represents the signal curve S_i of a voxel. Define the centered dataset A as a matrix of size $M \times N$, with M the number of time points in S_i and N the number of voxels in MRI volume. Consider rows of matrix A as the vectors S_i' in A with $S_i' = S_i - \bar{s}_i$, where $\bar{s}_i$ is the mean intensity value of S_i. For dataset A, the $M \times M$ covariance matrix is calculated. Using the covariance matrix, the eigenvalues and the corresponding eigenvectors are calculated. Thus, we obtain M eigenvalues and M eigenvectors U_i. The eigenvectors are the principal components of A. Considering the set of principle components as a basis for a new coordinate system, the vectors in A are transformed into the coordinate system of principal components, building a new dataset A^*:

$$A^* = UA. \tag{5.1}$$

The new dataset A^* is again a $M \times N$ matrix. The reduced dataset A^k is $k \times N$, built from the subspace of first k principal components U^{*k} and reduced dataset A^{*k} in principal components' space.

$$A^k = (U^{*k})^T A^{*k}, \tag{5.2}$$

where $(U^{*k})^T$ is the transposed matrix of U^{*k}.

Due to the observation that the most information of the signal curves is represented in the first five principal components, the value $k = 5$ is used.

The five slices of reduced datasets A^{*5} for nine liver dataset are presented in Figure D.6 in the Appendix. The corresponding five principal components are depicted in Figure D.7. Except of datasets 4 and 9, in the other seven datasets, in the second slice corresponding to the second principal component, the enhanced voxels are strongly present. The first principal component is obviously responsible for the absolute measured values. Therefore, the first slice is perfectly suitable for separating the tissue from the background.

5.2.2 Elimination of background and non-enhanced tissue

The separation of the background from tissue is similar to the introduced method in DCE-MRI of breast datasets from section 4.2.1.

After the segmentation of the tissue from the background, the enhanced tissue is separated from non-enhanced tissue. For this, a reduced dataset is calculated from the original

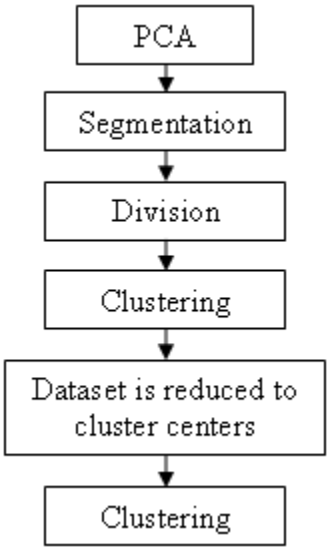


Figure 5.1: The steps of liver clustering procedure.

dataset by means of PCA. The reduced dataset A^5 is derived from projection of the original dataset in the subspace of the first five principal components. The signal curves in A^5 dataset are less noisy than the original signal curves. The number five was selected experimentally so that the valuable patterns of the signal curves are preserved while noise is suppressed. From this dataset the pre-enhanced intensity and post-enhanced intensity values for each voxel are calculated. An enhancement factor Enh_i for a voxel i is defined, which is the difference between average of pre-enhanced and average of post-enhanced intensity values for that voxel:

$$Pre_i = \frac{1}{n_{pre}} \sum_{j=0}^{a} S_{ij},$$

$$Post_i = \frac{1}{n_{post}} \sum_{j=a+1}^{T} S_{ij},$$

$$Enh_i = Post_i - Pre_i, \tag{5.3}$$

where a is the time point separating the pre-enhanced phase from the post-enhanced phase, n_{pre} is the number of pre-enhanced intensity values and n_{post} is the number of post-enhanced values for voxel i. It is observed that the enhanced areas are the voxels with an enhancement factor greater than the average enhancement factors of the all tissue voxels selected in the previous step.

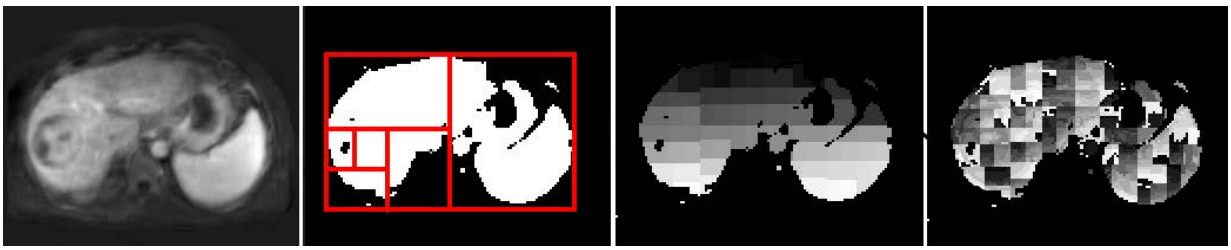


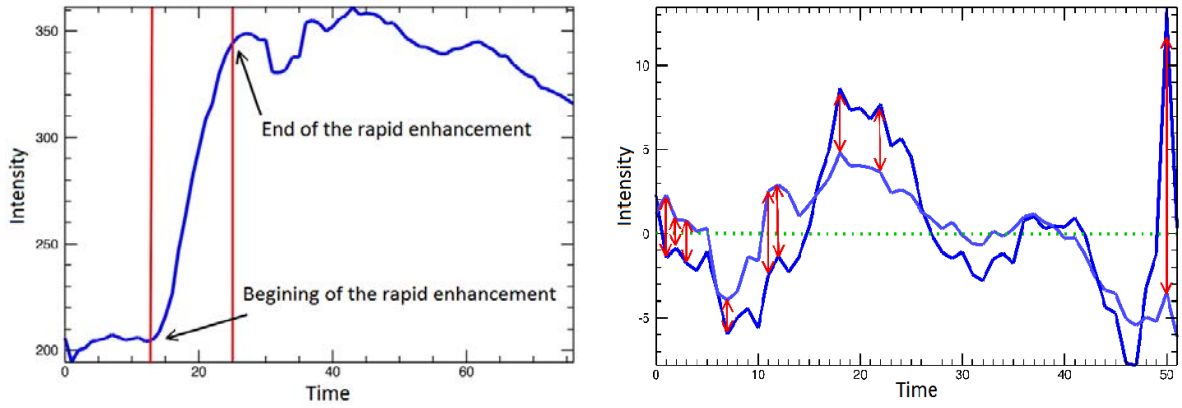
Figure 5.2: Illustration of recursively dividing of the enhanced area. From left to right, the image of the enhancement factors calculated from equation 5.3, recursively dividing of the segmented voxels, the sub regions, and the clustering inside each region.

5.2.3 Clustering

For voxel clustering, the hierarchical clustering method is used. It has the advantage of representing the whole structure of a dataset by means of a selected similarity measure. Hierarchical clustering is independent of starting points or other parameters. For a selected number of clusters, the results are deterministic. The disadvantage of the hierarchical method is that it is computationally expensive, since it needs all inter distances between vectors in the dataset. Depending on the size of the dataset, the number of distances can be very large, increasing exponentially with the size of the dataset. For a dataset with n signal curves, there are at least $n(n-1)/2$ distances to determine. The number of voxels in the segmented enhanced tissue in the liver datasets is much larger than the number of voxels in the segmented tumors in the breast datasets. In the liver datasets, beside the liver tumor, the whole tissue of liver, pancreas and spleen are enhanced as well. Thus, it is impossible to separate the tumor from surrounding enhanced tissue. Instead, the whole enhanced tissue is subject to clustering. Due to the large number of voxels a direct application of hierarchical clustering is restricted.

To overcome this problem, the dataset is recursively divided into sub-regions r_i where $i = 1, ..., k$ and k is the number of regions. The number of the voxels inside each region is less than a specific number N. The number N is the maximum number of voxels possible for the calculation of the dissimilarity matrix as restricted to machine memory.

The voxels inside each region r_i are clustered into m clusters, with $m = N/k$. Clustering is performed by hierarchical clustering (see section 3.4.1 for more details). Euclidean distance is used as the measure of similarity. After the division of sub-regions into clusters,



(a) (b)

Figure 5.3: (a) Illustration of the start and end point of rapid enhancement phase. (b) The new similarity measure calculated as the Euclidean distance of the mean zero washout parts of the curves.

a new dataset is calculated from the representative signal curve of each cluster. The representative signal curve of a cluster is the average of signal curves of all voxels in that cluster. The number of signal curves in the new dataset is less than N. The new dataset is again clustered by hierarchical clustering. A modified version of the introduced PM measure (see 5.2.4) is used as similarity measure.

The advantage of this step-wise clustering procedure is that when dividing the dataset into sub-regions, the spatial relationships between signal curves are taken into account, in the first step of clustering. The steps of the clustering schema is illustrated in a diagram in Figure 5.1 and is depicted for a dataset in Figure 5.2.

5.2.4 Similarity measure

To compare representative curves of the clusters resulting from the previous step, instead of using the Euclidean distance, a new similarity measure is used. For this similarity measure, the time point t_w and the end point of rapid enhancement phase, has to be determined. In Figure 5.3, the left image demonstrates the start and the end time points of rapid enhancement for a signal curve. The mean of later washout part of the curve and the part of the curve after the time point t_w is set to zero. The dissimilarity measure for two signal

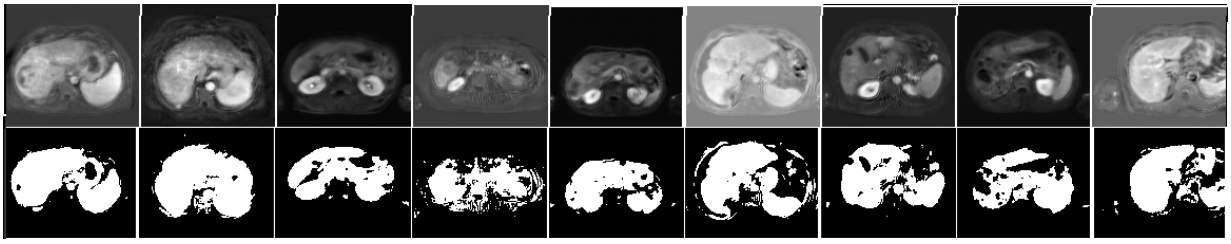


Figure 5.4: Segmentation of enhanced voxels from the rest of the image for nine datasets.

curves S_i and $S_{i'}$ is the Euclidean distance between these parts:

$$d_{ii'} = \sqrt{\sum_{j=t_w}^{d} \left(S_{ij}^w - S_{i'j}^w\right)^2}, \tag{5.4}$$

where $S_{ij}^w = S_{ij} - \frac{1}{d-t}\sum_{j=t_w}^{d} S_{ij}$ with $j = \{t_w, ..., d\}$. In the right image of the Figure 5.3, a schematic illustration of this measure for two signal curves is demonstrated. On the right, the mean values of the washout phases of two signal curves are set to zero and point wise distances are depicted with red arrows.

The introduced PM measure in section 3.6.4 is inappropriate for the liver datasets due to the strong noise of the signal curves. The introduced alternative similarity measure is more robust against the noise, at the same time less successful in measuring the parallelism in the curves.

5.3 Experiments and results

The introduced clustering method in section 5.2.3, is evaluated on nine DCE-MRI liver datasets. In the first step, the enhanced voxels are separated from the background and the non-enhanced voxels. The segmented enhanced tissue for all nine datasets is depicted in Figure 5.4. The enhanced area is clustered into subareas with similar enhancement patterns. The clustering was performed in two steps as described in section 5.2.3. In the first step, the number of voxels has been reduced by a local clustering of sub-regions using the Euclidean distance as the similarity measure. In the second step, representative curves of the clusters from the previous step are clustered further using the introduced similarity method from section 5.2.4.

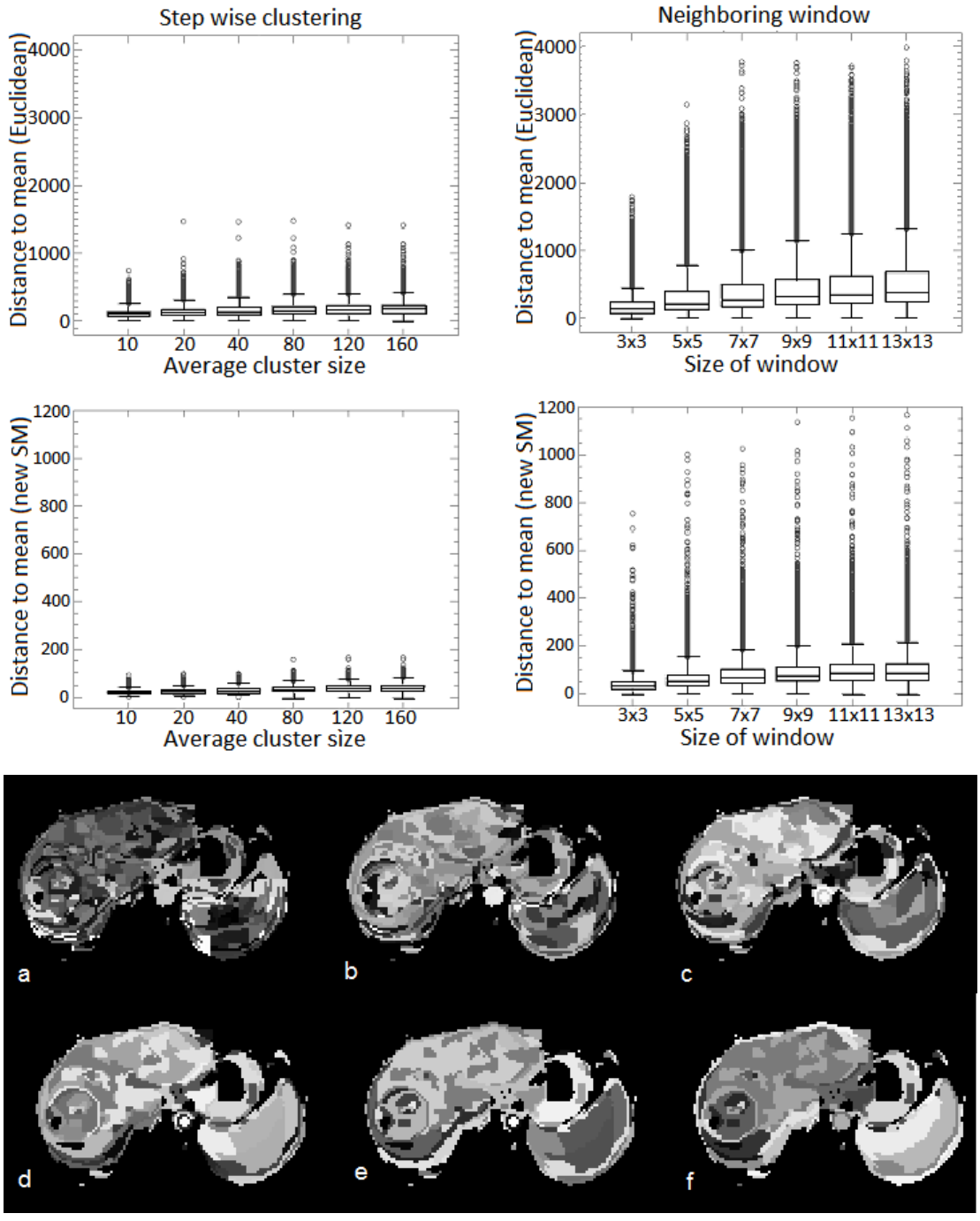


Figure 5.5: Top: Comparison of step wise clustering with neighboring windows. The distance to mean is measured by Euclidean distance in the first row, and by the introduced similarity measure in the second row. Bottom: Results of clusterings for a dataset.

The evaluation of the method is based on the fact that the distance between the signal curves of voxels belonging to a homogeneous group and the mean curve of this group is smaller than the distance of the signal curves to the mean of a heterogeneous group with the same number of voxels. The liver is a heterogeneous organ. Thus, even in a small neighborhood, it is possible that the neighbor voxels are part of several homogeneous areas.

To evaluate whether the step-wise clustering method is successful in clustering the enhanced voxels into more homogeneous areas the datasets were clustered with six different numbers of clusters. The six cluster numbers are 600, 400, 200, 100, 50, and 25 and the average number of voxels inside the clusters are $\approx$ 10, 20, 40, 80, 120, 160, respectively. The distance to the mean curves of the clusters are calculated. To evaluate how heterogeneous the neighborhood of a voxel is, a window of size $r \times r$ is considered as the neighborhood of each voxel. The voxel is in the center of the considered window. For the different values of r, the distance of a voxel to the mean curve of the voxels inside the neighborhood window is calculated. The different values of r are selected so that the numbers of the voxels inside the windows are approximately close to the average cluster size above.

The results in the case of the clustering and neighboring-window are compared in Figure 5.5 for all datasets. In the column left in Figure 5.5, the average distances to mean curve for all datasets are presented. In the boxplots on the top, the Euclidean distance is used as the distance measure to cluster-centers. In the boxplots on the bottom, the new similarity is the distance measure to cluster-centers. In the right column in Figure 5.5, the average distances to the mean curve for corresponding neighboring-window are shown.

The boxplots demonstrate that the signal curves in a neighborhood can be very dissimilar, that is an index of heterogeneity of the liver. In contrast, the boxplots of the clusters are smaller and with fewer outliers compared to the boxplots using the rectangular windows. In other words, the step-wise clustering with the new similarity measure is able to build more homogeneous groups of voxels. The results of six clusterings with different cluster numbers from the above experiment are depicted for a dataset in the bottom of Figure 5.5. The meaningful biological clusters like tumor and aorta are clearly visible in the clustering. The results of the clustering for all nine datasets are depicted in Figures D.1- D.5 in Appendix. In each figure, sub figure (a) depicts the subtraction image of a pre-enhanced slice from a post-enhanced slice. In sub figures (b, c, and d) the results of the clustering with cluster numbers 100, 50, and 25 are depicted. The five slices of reduced dataset A^{*5} on the space of the five first principal components for all nine datasets

are depicted in Figure D.6. The corresponding five principal components are depicted in Figure D.7.

5.4 Discussion

A method for heterogeneity analysis of DCE-MRI liver dataset is introduced. The experiments on nine liver datasets show that depending on the expected granularity, the tissue of liver can consist of many sub regions. The comparison of voxels inside a cluster to the voxels in a neighborhood window shows that the variation among the signal curves of the neighbor voxels can be larger than voxels with large spatial distances. The voxels inside a sub region have similar enhancement patterns, and those patterns differ from enhancement patterns of other sub regions. This method is a suitable way to reduce signal to noise value.

A challenging and interesting question is, which granularity of clustering represents the best the biological structure of the image (the minimal number of clusters, which differentiate between maximal biological structures). This question can be answered the best by comparing of different clustering results with pathological gross-cuts. However, this kind of comparison is hard to realize.

Similar to the breast clustering, the granularity can also be controlled through the acceptable dissimilarity in a cluster. However, due to the strong noise and motion artifacts in signal curves defining the appropriate threshold is difficult. To reduce the noise, the clustering was performed on the reduced dataset by means of PCA. In spite of that, the significant deformations in signal curves are present in principal components. One solution could be the registration of the dataset to a proper reference. On the other hand, registration can deform the curves in such a way that valuable kinetic information, such as peaks of artery and vein flows, could be lost. The other possibility is to look for significant irregularities in a signal curve. The irregularities are either features of the signal curve, or are undesired artifacts. The determination of whether they are undesired artifacts, can be done based on *a priories*. These irregularities can be eliminated from the measurement.

The clustering gives us information about how many significant different signal curves are present in the liver. This information helps the estimation of the number of compartments and the complexity of the system. Finally, the result of the proposed clustering technique can be used in a further non-model based classification method to distinguish between different types of tumors.

Chapter 6

Gap Statistic application to simulated and real datasets

In the section 3.5.1, the differences of the Gap functions computed with, and without logarithm was discussed. In this chapter, the original *Gap* and the proposed Gap^* Statistic are applied to simulated and real datasets, in order to evaluate the effect of the differences in both approaches.

For the experiments, agglomerative hierarchical clustering with the group average linkage method [Kaufman and Rousseeuw, 1990] is used. The average linkage method has some advantages over the widely used k-means clustering. Hierarchical clustering methods produce hierarchical representations in which the clusters at each level of the hierarchy are created by merging clusters at the next lower level. Each level of hierarchy represents a particular grouping of the data into disjoint clusters of samples. The entire hierarchy represents an ordered sequence of such groupings. Unlike k-means clustering, where the choice of different numbers of clusters can lead to totally different assignment of elements to the clusters, in hierarchical clustering the sets of clusters are nested within one another.

6.1 Two historical datasets

Two historical datasets are frequently used when discussing clustering; "Fisher's Iris dataset" [Fisher, 1963] and Wolbergs "Breast Cancer Wisconsin dataset"[Wolberg et al., 1993]. The four different definitions of the Gap Statistic have been applied to these two famous historical datasets. Fisher's Iris dataset consists of 50 samples from three species of Iris flowers.

Gap function	number of clusters	
	Iris	Breast
Gap	3^+	2^+
Gap*	3^+	2^+
weighted_Gap	2	1
weighted_Gap*	7	1

Table 6.1: Results of standard and weighted *Gap* and *Gap** functions on Iris and Breast Cancer datasets. "+" indicates the correct number of clusters for that dataset.

Four variables were measured for each sample. For the "Breast Cancer Wisconsin dataset", samples arrived periodically as Dr. Wolberg reports his clinical cases. The dataset consists of 699 samples. Each sample is described by nine variables. The whole dataset has two main groups, consisting of 458 benign and 241 malignant tumors. Table 6.1 lists the estimated number of clusters for both the iris and the breast datasets using the original Gap Statistic "*Gap*" from Eqn. 3.42 and the proposed Gap Statistic without logarithm "*Gap**" as defined in Eqn. 3.47. These two Gap functions are compared with the results of the weighted *Gap* as described in section 3.5.1 and the weighted *Gap**, *i.e.*, the weighted Gap using W_k instead of $\log(W_k)$.

In contrast to the result from k-means clustering reported in [Yan and Ye, 2007], when using average linkage clustering, the Gap Statistic with the original W_k, Eqn. 3.41, estimates the number of clusters for both datasets correctly. Figures 6.1 and 6.2 show the calculated Gap functions for the two datasets. Both, the iris and the breast cancer datasets represent their natural clusters in average linkage clustering. Thus, *Gap* and *Gap** show similar behavior. It can be observed that in the case of iris data, the weighted *Gap* suggests number 2 as the proper number of clusters but weighted *Gap** suggest 7 as the cluster number. According to the discussion in section 3.5.1, whenever a number fulfills the inequality 3.43, this number fulfills the inequality for the proposed *Gap**. However, this statement is not valid for weighted *Gap* due to the fact that $W_k^{'}$ from Eqn. 3.49 is not monotonically decreasing.

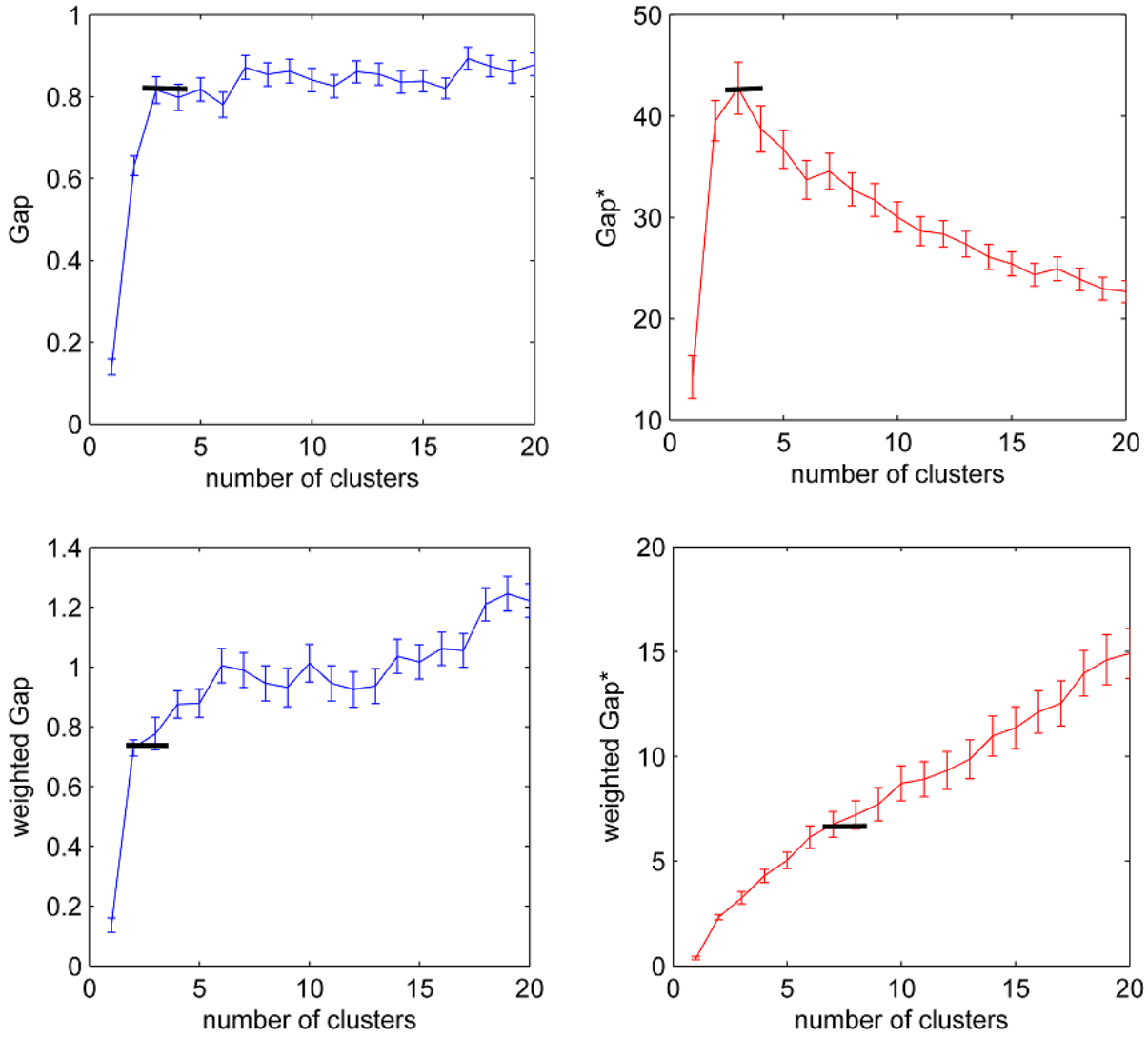


Figure 6.1: Standard and weighted-*Gap* and *Gap** functions for Iris dataset.

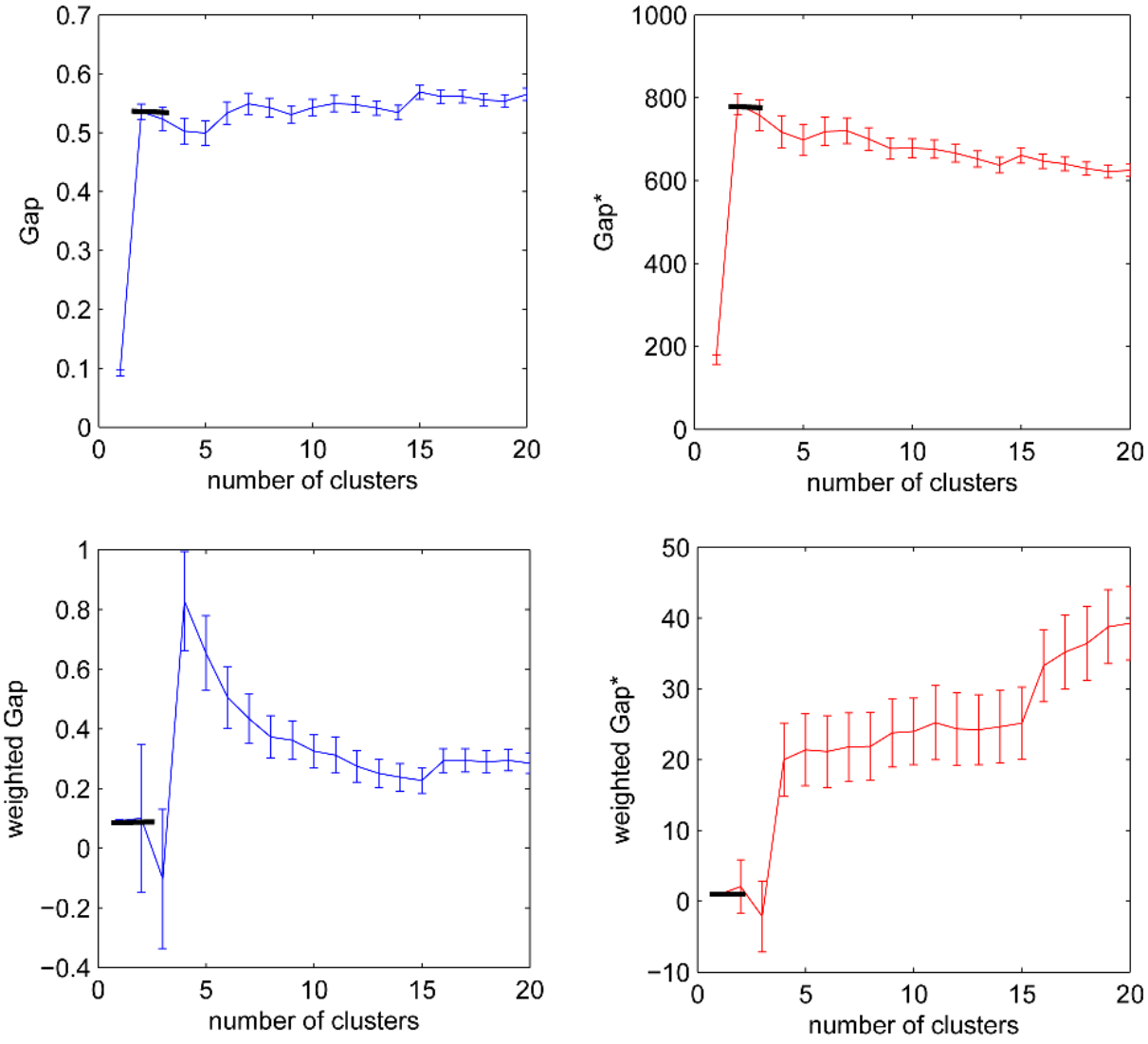


Figure 6.2: Standard and weighted *Gap* and *Gap** functions for Breast Cancer Wisconsin dataset.

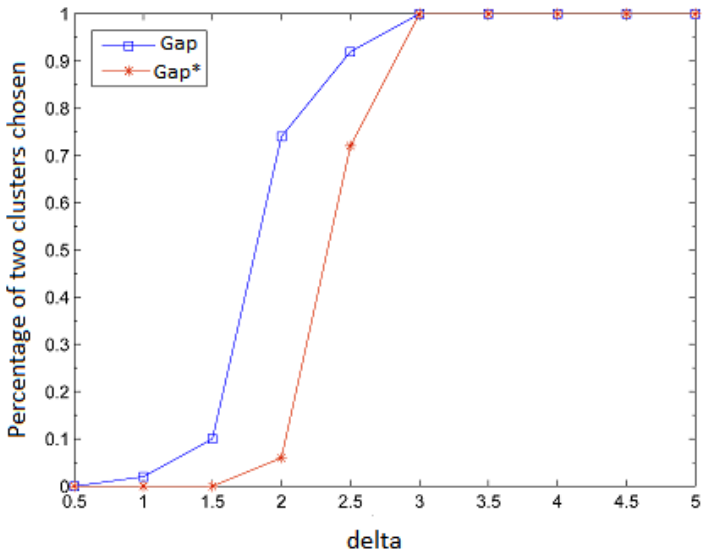


Figure 6.3: *Gap* function from Eqn. 3.42 and *Gap** function from Eqn. 3.47 are compared for 10 datasets with two clusters. Two clusters have different portion of overlapping area in each dataset.

6.2 Not well-separated clusters

Now let assume clusters, which are not well-separated. 1000 datasets are simulated with two clusters each, with different proportions of overlapping. Each cluster had 50 observations with two variables. Both variables were drawn independently from Gaussian distributions; for observations from the first cluster, both variables had expected values 0 and standard deviation 1. For observations from the second cluster, both variables were again randomly drawn from Gaussian distribution with expected value Δ and standard deviation 1. As a result, there are two clusters, where the distance between the means of the two clusters decreases with decreasing value of Δ. The values $\Delta = 0.5, 1, 1.5, \ldots, 5.0$ were used. For each of the ten unique values of Δ, 100 datasets were generated, and original *Gap* and proposed *Gap** functions were calculated for these datasets. Figure 6.3 shows the percentage of finding two as the number of clusters for each type of dataset. It can be observed that the original *Gap* was better in estimating the proper number of clusters in overlapped clusters than *Gap**. These results were expected due to the tendency of the *Gap* to overestimate the number of clusters, which has been reported by Dudoit and Fridlyand [2002].

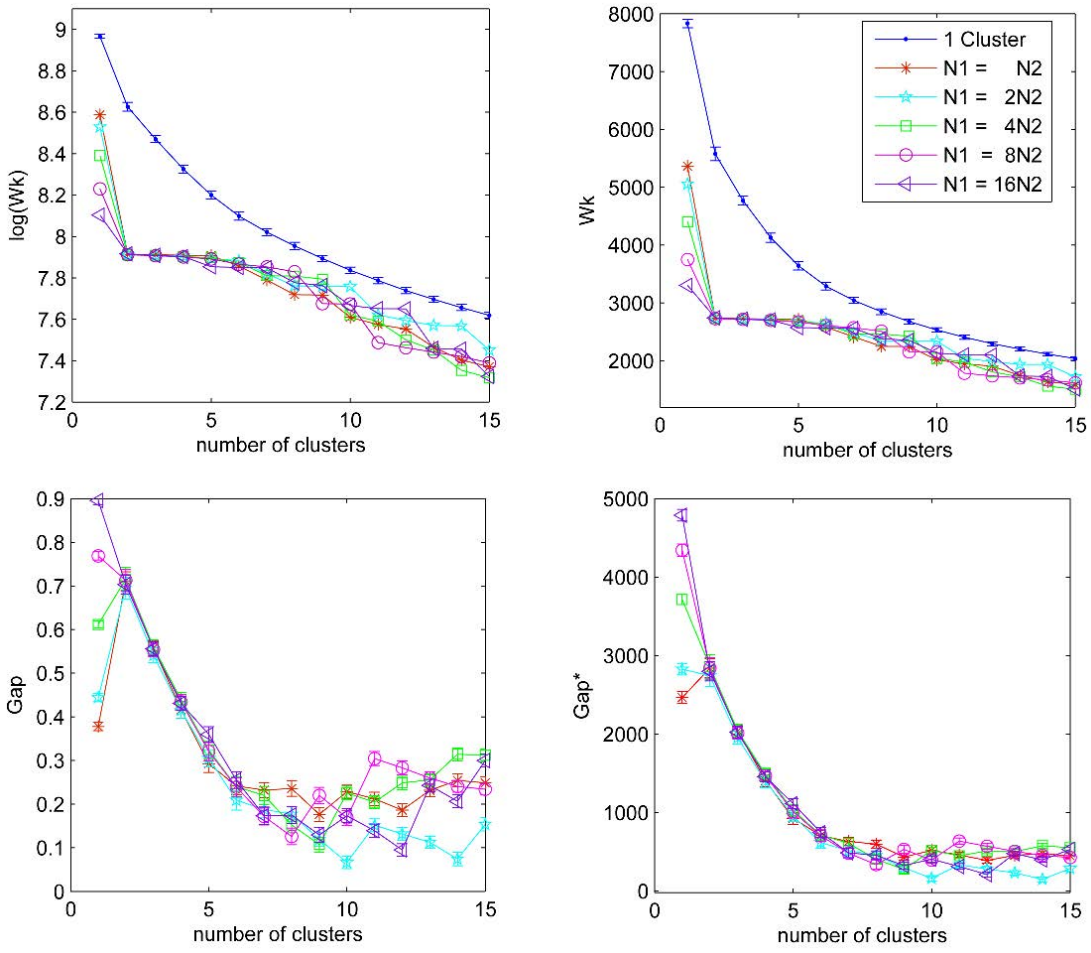


Figure 6.4: Top: $\log(W_k)$ (left) and W_k (right) five simulated datasets with two clusters each, where $N1$ and $N2$ are the numbers of samples in the first and second cluster, respectively. Bottom: *Gap* (left) and *Gap** (right) for these datasets.

6.3 Unequally sized clusters

Yin et al. [2008] report that whenever the number of observations in one cluster is more than six-fold the number of observations in the other clusters, the Gap Statistic is not able to estimate the number of clusters accurately. This effect depends not only on the number difference between clusters but also on the distance between clusters. This effect has been studied in the special case of two clusters sampled from two 2D normal distributions $N(\boldsymbol{\mu}, \boldsymbol{I})$ and $N(\boldsymbol{\mu}', \boldsymbol{I})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are two different expected values and $\mathbf{I}$ is the identity matrix. Details of this study are given in Appendix B. Suppose N_1 is the number

simulation	N_1	N_2	$m = N_1/N_2$
1	765	765	1
2	1020	510	2
3	1224	306	4
4	1360	170	8
5	1440	90	16

Table 6.2: Five simulated datasets with two clusters with N_1 and N_2 number of samples in first and second cluster respectively.

of samples in the first cluster and N_2 is the number of samples in the second cluster and $N_1 = m \cdot N_2$ and $n = N_1 + N_2$. For a fixed total number of samples n, by increasing m, the value of W_1 decreases. Thus, Gap_1 increases while Gap_2 is almost unchanged. When m becomes large enough, Gap_1 will be greater than Gap_2, and the estimated cluster number will be one. The possible numbers of m for which Gap and Gap^* can still estimate two as the proper number of clusters, can be estimated from the following two inequalities (see Appendix B inequalities Eqns. B.6 and B.7):

1. for Gap

$$\frac{md}{(m+1)^2} \geq \frac{E(d_1)}{E(d_2)} - 1, \tag{6.1}$$

2. for Gap^*

$$\frac{2md}{(m+1)^2} \geq E(d_1) - E(d_2), \tag{6.2}$$

where d is the average distance between the points in first cluster to the points in second cluster, $E(d_1)$ is the expected distance of two points from a rectangular uniform distribution with sides a and b and $E(d_2)$ is the expected distance of two points from a rectangular uniform distribution with sides $\frac{a}{2}$ and b.

These results are illustrated in an example in Figure 6.4. In this example, we compared five datasets with two clusters of different observation sizes. The total number of observations is the same in all five datasets. However, the ratio of observations is varied. Table 6.2 summarizes the size of the clusters in each dataset and the ratio between the number of observations in the two clusters. In the first dataset, the number of observations in the

first cluster (N_1) and in the second cluster (N_2) are equal. In the other four datasets, N_1 increases and N_2 decreases as given in table 6.2.

Samples were drawn as follows:

1. Select N_1^{max} as the maximum number of samples in the first cluster in all five datasets.

2. Select N_2^{max} as the maximum number of samples in the second cluster in all five datasets.

3. Draw N_1^{max} samples from a bivariate normal distribution with parameters $(\boldsymbol{\mu}, \boldsymbol{I})$, where $\boldsymbol{\mu} = (0, 0)$.

4. Draw N_2^{max} samples from a bivariate normal distribution with parameters $(\boldsymbol{\mu}', \boldsymbol{I})$, where $\boldsymbol{\mu}' = (5, 0)$.

5. For each dataset, select the first N_1 samples from the N_1^{max} sample points according to the number N_1 given for this dataset in table 6.2.

6. For each dataset, select the first N_2 samples from the N_2^{max} sample points according to the number N_2 given for this dataset in table 6.2.

According to the estimations in Appendix B and the inequalities 6.1 and 6.2, in this example, $E(d_1) \approx 4.53$, $E(d_2) \approx 2.99$, and $d \approx 3.48$. As a result only for $m < 6$ for the original Gap, and $m < 2$ for the proposed Gap^*, the Gap Statistic determines two as the proper number of clusters. Figure 6.4 shows $\log(W_k)$ and W_k for all five simulated datasets. The blue dotted line is the expected $\log(W_k)$ on the left top and expected W_k on the right top of the null reference distribution. As demonstrated in Figure 6.4, by increasing the number of samples in first cluster against the second cluster, the within-cluster dispersion W_2 remains the same but W_1 decreases. Depending on how far apart the two clusters are, increasing the ratio of observations in both clusters increases the $Gap(1)$ value. Figure 6.4 demonstrates the original Gap function (bottom left) and the proposed Gap^* function (bottom right) for these five datasets. The estimated m from the inequalities 6.1 and 6.2 is confirmed by the results illustrated in Figure 6.4.

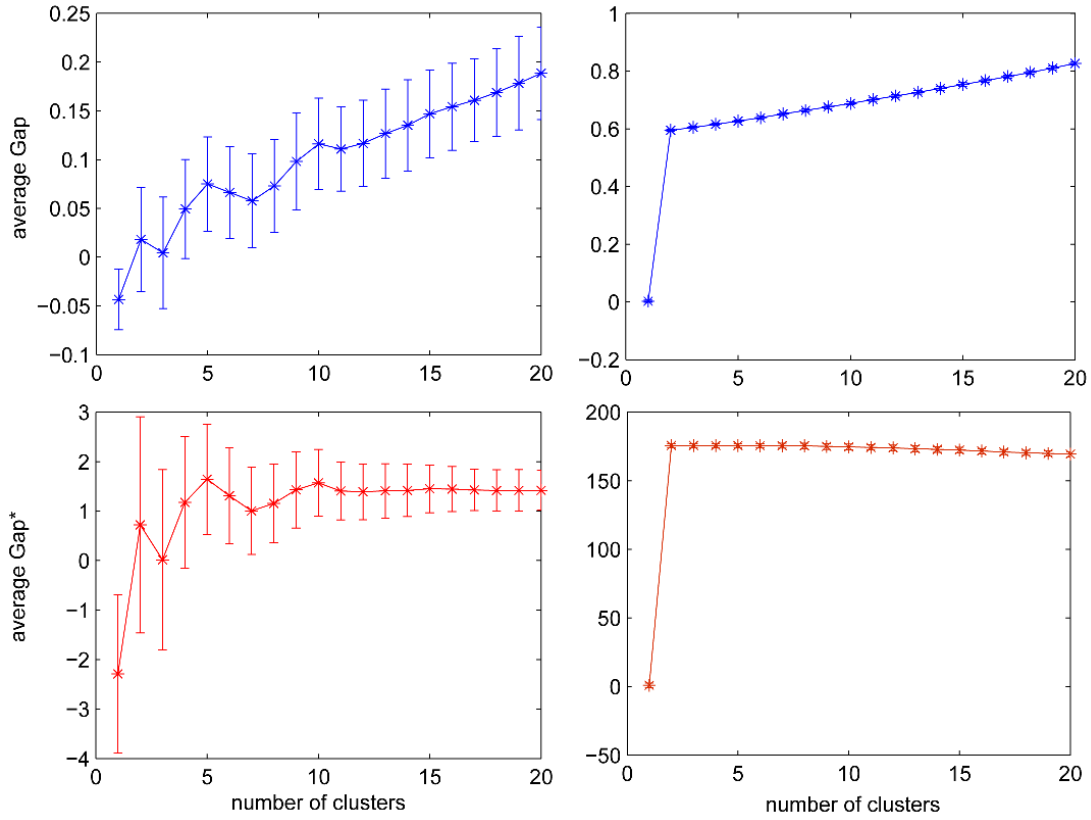


Figure 6.5: Average *Gap* and *Gap** for simulated 2D (left) and 100D (right) datasets from experiments 6.4.

6.4 Simulated data with increasing Gap function

In this experiment, data were simulated such that the calculated Gap function (*Gap* from Eqn. 3.42) is a strictly increasing function. A dataset was simulated 2000 times and for each simulated dataset the original *Gap* and the proposed *Gap** Statistic was calculated. The simulated dataset consists of two clusters. Each cluster contains 50 observations from an n-dimensional variable space. In the first cluster, each feature was sampled from a uniform distribution on the interval $[0, 10]$ at random. For the second cluster only the first variable was sampled from the same uniform distribution. All other variables of observations in the second cluster were set to zero. Half of the datasets were simulated in a 100-dimensional variable space while the other half were simulated in a 2-dimensional variable space.

Figure 6.5 depicts the average *Gap* and the average *Gap** functions for both the 2D datasets and the 100D datasets. For the 2D datasets, both Gap functions suggest two

	Method	Estimate of number of clusters									
		1	2	3	4	5	6	7	8	9	≥ 10
2D	*Gap*	368	489	143	0	0	0	0	0	0	0
	Gap*	270	567	162	1	0	0	0	0	0	0
100D	*Gap*	0	0	0	0	0	0	1	3	1	995
	Gap*	0	1000	0	0	0	0	0	0	0	0

Table 6.3: Number of times that different cluster numbers are estimated by *Gap* and *Gap** for 2D and 100D datasets.

as the proper number of clusters. However, it can be seen that the *Gap* function for the 100D datasets is a strictly increasing function. This is indeed expected due to the 'curse of dimensionality' [Bellman, 1961]. Beyer et al. [1999] have shown that the minimum and the maximum occurring distances become indiscernible, as the difference of the minimum and maximum value compared to the minimum value converges to 0 as the dimensionality m goes to infinity.

$$\lim_{d\to\infty} \frac{dist_{max} - dist_{min}}{dist_{min}} \to 0. \tag{6.3}$$

Consequently, all the distances $d_{ii'}$ from Eqn. 3.40 can be considered to be equal in a high dimensional space. Consider n observations from a 100 dimensional uniform distribution and suppose these samples are divided into k clusters $C_1, C_2, ..., C_k$, where $|C_1| = |C_2| = ... = |C_k| = \frac{n}{k}$. Consider all $d_{ii'} = dist$, thus, W_k is equal to:

$$W_k = \left(\frac{n}{2} - \frac{k}{2}\right) dist. \tag{6.4}$$

By increasing the number of clusters k, W_k in Eqn. 6.4 decreases linearly. The slope of this line is the same for all datasets sampled from the same high dimensional uniform population even with different numbers of samples. Here, in the case of a 100D dataset for all $k > 2$ only the first cluster will be divided further, due to the large distances of the samples in this cluster compared to the second cluster. Hence, W_k will be linear for $k > 2$ and parallel to $E^*(W_k^*)$. The difference $E^*(W_k^*) - W_k$ remains constant as $E^*(W_k^*)$ and W_k decrease. Therefore, the *Gap* function is strictly increasing. On the other hand, whenever the difference $E^*(W_k^*) - W_k$ remains constant, $Gap^*(k)$ and $Gap^*(k+1)$ will be

dataset	number of voxels	*Gap*	*Gap**
1	1260	7	7
2	207	9	5
3	116	9	5
4	262	*nd*	5
5	141	11	5
6	277	*nd*	5
7	151	13	4

Table 6.4: Results for all seven DCI-MRI datasets analyzed with the *Gap* and the *Gap** Statistic. "*nd*" stands for not defined.

equal. Thereby, due to the Gap condition Eqn. 3.43, k will be suggested as the proper number of clusters by the proposed *Gap** Statistic.

Table 6.3 lists the number of clusters found with the original *Gap* and the proposed *Gap** Statistic for 1000 simulations of 2D and 100D datasets, respectively. While for the 2D simulation, both original *Gap* and the proposed *Gap** Statistic perform similarly, the original *Gap* fails in finding the true number of clusters for all the 1000 simulated 100D datasets. The proposed *Gap** Statistic, however, is able to determine the true number of clusters for these simulations.

6.5 Real datasets with increasing Gap function

Both Gap functions are evaluated further on seven real datasets from DCE-MRI of breast tumors [DKFZ, 2004]. One of the main challenges on clustering DCE-MRI datasets is to determine the number of underlying patterns in the signal curves. To this end, the Gap Statistic has been applied on DCE-MRI data. As before, the average linkage clustering method with Euclidean distance as the measure of dissimilarity was used. The samples are the signal curves of voxels of which each is described by 128 features, i.e., time points.

Table 6.4 gives the number of clusters found with the original *Gap* and the proposed *Gap** for seven DCE-MRI datasets. The tumors in all of these images have the same type. Using the proposed *Gap** Statistic, the number of five clusters was found in five of the seven images, whereas with the original *Gap* Statistic, no consistent number of clusters,

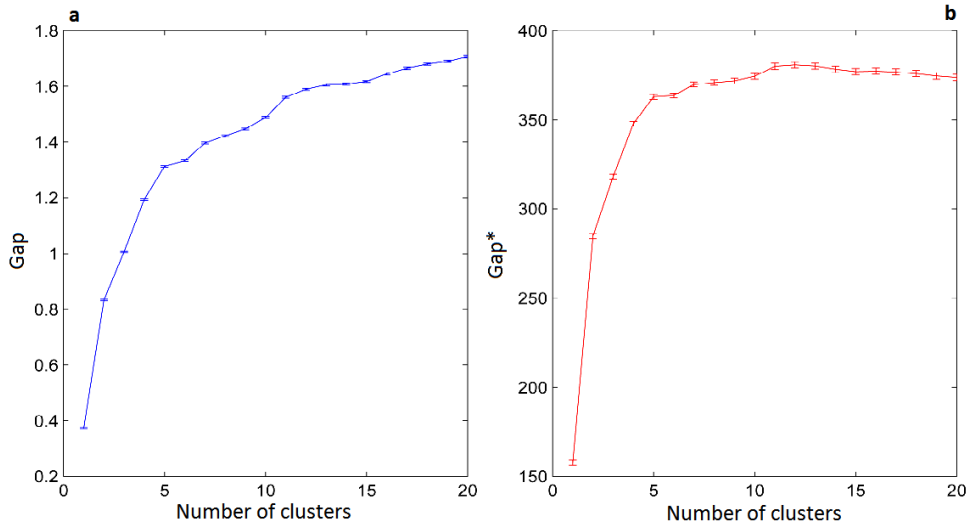


Figure 6.6: Gap functions *Gap* and *Gap** for DCE-MRI dataset of a breast tumor.

i.e., regions, was found.

Figure 6.6 shows the resulting *Gap* and *Gap** functions for one of the DCE-MR dataset (dataset 4 in Table 6.4). Similar to the simulated dataset in 6.4, the *Gap* function is a strictly increasing function, whereas the *Gap** function is not strictly increasing and suggests five as the number of clusters for this dataset. In the Figure 6.7(a) the dataset in first and second principal components plane is depicted, and the five identified clusters are shown in different colors and with different symbols. The intensity curves for voxels in a cluster are shown in Figure 6.7(c); the mean curve of each cluster is depicted in red. Figure 6.7(b) depicts the tumor image with voxel colored according to their cluster with the same colors as in sub-figure (a). A ring-shaped ordering of the five clusters can be observed in this image. This ordering is in agreement with enhancement patterns reported in medicine such as, circumferential, centripetal and peripheral ring contrast [Buadu et al., 1997]. However, so far there is no information on the number of regions.

6.6 Discussion

The *Gap* Statistic is one of the most popular methods for estimating the number of clusters in a dataset. It is rather simple to implement and is used in many, diverse applications. As reported by Tibshirani et al. [2001], it outperforms many other methods. However, the *Gap* Statistic is not able to suggest the correct number of clusters in some cases. Yin et al.

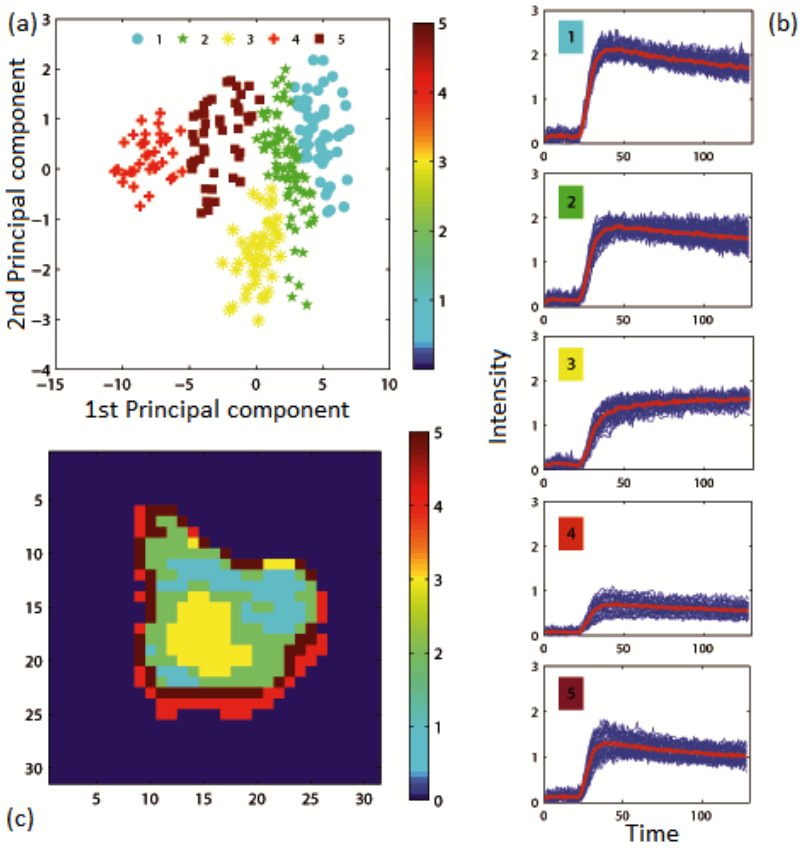


Figure 6.7: Five clusters of the DCE-MRI breast tumor found with average linkage clustering. (a) First and second principal component of the DCE-MRI signals per voxel. Voxels are colored according to their cluster affiliation. (b) Segmentation map of the tumor. Voxels are colored similar to subfigure (a). (c) Signal curves for each voxel in the five respective clusters along with the mean curve (bold red line).

[2008] have reported that in cases where the ratio of observation sizes between clusters is over than six-fold, the *Gap* Statistic does not work accurately. Dudoit and Fridlyand [2002] have mentioned the overestimation of *Gap* Statistic in some applications. Sugar and James [2003] have reported the failure of the *Gap* Statistic in the case that data were derived from exponential distributions.

Using $\log(W_k)$ instead of W_k in the calculation of the *Gap* function can be one cause of overestimation of number of clusters in the *Gap* Statistic. Theoretically, there is no feasible reason to choose Eqn. 3.42 over Eqn. 3.47 for the definition of the *Gap* Statistic. Indeed, using the logarithm function in the definition of the *Gap* Statistic has a fundamental effect

on the results of the *Gap* Statistic. This is due to a property of the logarithm function described in the following example: Consider four positive numbers a, b, c, and d, with logarithm of all of them greater than 1. Let be $a > c$ and $b > d$ and $a - b = c - d > 0$, then we will have $\log(a) - \log(b) < \log(c) - \log(d)$. As a result, by increasing the number of clusters the within cluster dispersion W_k decreases. Consequently the *Gap* function increases even when the distance between W_k^* and W_k remains the same.

Estimation of the number of clusters depends on many factors. The choice of clustering method is one of these factors. The *Gap* Statistic is designed to be applicable to any clustering method. In general, the results and discussions given in this work are not restricted to any clustering method. However, the choice of the clustering method influences the result of *Gap* Statistic. Different clustering methods look for different structures in data. The average linkage method, used for the *Gap* calculation in section 6.1, was able to find the real cluster number for both the "Iris" [Fisher, 1963] and the "Breast Cancer Wisconsin" datasets in contrast to the *Gap* function with k-means clustering reported in [Yan and Ye, 2007]. This is exactly one of the demerits of *Gap* Statistic. In the cases, that few *a priori* information about the dataset exits, it is quite difficult to say which clustering method is expedient.

Comparing the original *Gap* and proposed Gap^* Statistic, the original *Gap* Statistic has a better performance in the case of overlapped clusters than Gap^* due to the tendency of the *Gap* in overestimating the number of clusters. For real application, it is however up to the user to decide whether two clusters with overlapping area should be considered as one cluster or two. In previous studies [Dudoit and Fridlyand, 2002, Sugar and James, 2003, Tibshirani et al., 2001, Yan and Ye, 2007, Yin et al., 2008], it was reported that a null reference data generated from a uniform distribution aligned with the principal components of the data causes a better performance of *Gap* Statistic. The *Gap* function calculated from such null reference data is referred to as Gap_{pc}. It would be interesting to compare Gap_{pc} and Gap_{pc}^* in further studies.

The Gap^* compares the expected values of W_k^* with W_k. Thus, it reflects exactly the changes in the within cluster dispersion of the real data against the expected W_k^* of the null reference dataset. Whenever the original *Gap* results in a k as proper number of clusters, this k is also a possible answer with the proposed Gap^*. In contrast, there are situations where proposed Gap^* function is able to offer a number as a proper number of clusters while the original *Gap* has no answer. Our experiments suggest that such data are possibly

from multi dimensional feature space, with different variances in the different feature axes. Evaluations in section 3 verify this idea. In sections 6.4 and 6.5, the original *Gap* function is a strictly increasing function, hence it cannot find any cluster number. On the other hand, Gap^* is not strictly increasing and therefore is able to suggest a cluster number for the data. For the simulated data in section 6.4 the suggested number is equal to real number of clusters. For the real dataset in section 6.5, however, we have no reference to decide if the number suggested by the proposed Gap^* Statistic is the proper number of clusters. Further experiments are necessary on real data with known cluster number to verify the accuracy of the proposed Gap^* Statistic in cases where the original *Gap* is a strictly increasing function.

The dissimilarity between signal curves, in the experiment on real DCE-MRI dataset, is measured by Euclidean distance. The *Gap* Statistic is based on W_k, which is the within-cluster dispersion if the distance measure is Euclidean. That means, in this case, *Gap* Statistic looks for a cluster number, which maximizes the compactness in clusters and at the same time maximizes the inter cluster distances. However, using another similarity measure such as the correlation coefficient or PM measure changes this meaning. In calculation of W_k, the sum of distances in each cluster is weighted by the number of elements in that cluster. It is the same as the sum of distances to the mean of the cluster, only if the distance measure is Euclidean distance. Otherwise, such a W_k has no specific meaning. Despite the fact that the changes in weighting of the sum of intra cluster distances affect significantly the results, the meaning of *Gap* Statistic regarding other similarity measures is not clear.

If we consider weighted W_k from Eqn. 3.49 instead of W_k, we could say weighted W_k is the sum of the average distances of the clusters. Nevertheless, in a reference dataset, the sum of average distances depends on the sampled datasets and not on the number of clusters. For this reason, usage of weighted Gap is questionable.

Irregular noisy data points in a dataset are another problem by estimating the number of clusters via *Gap* Statistic. Normally, the distances of outliers and the rest of dataset are so large that each outlier builds a singleton cluster. Thus, *Gap* Statistic counts the outliers as the number of clusters instead of real patterns in the data. The problem of outliers can be overcome by eliminating the outliers before applying the *Gap* Statistic. In contrast, other obscurities such as, choice of the clustering method, and finding the proper W_k for other similarity measures are more difficult to handle.

Chapter 7

Conclusion

7.1 General view

In this work, several clustering methods are described, each of them having their own advantages and disadvantages. The focus of the work is the cluster analysis of breast tumors in DCE-MRI datasets. However, the possibility is studied to see whether similar methods can be performed on liver datasets.

The study, besides searching for the proper clustering method, follows two other major topics, the choice of an appropriate similarity measure, and determining the number of clusters. These three subjects are connected to each other in such a way that success in one direction will help solve the other problems. In the case of signal curves, if we could define, what the best criterion for similarity between two curves is, determining the number of clusters becomes dispensable. Technically, we could define a threshold for the acceptable amount of dissimilarity in a cluster. At the same time, being able to choose the number of clusters makes it easier to find the proper clustering method.

Many clustering methods are designed or optimized for a metric space. In these cases, the Euclidean distance is the common choice. Present work demonstrates that the Euclidean distance as a similarity measure for comparing the signal curves is not sufficient. Therefore, either a combination of several measures must be used for the clustering or the dataset must be transformed into a feature space. The distances of data points in feature space are measured with the Euclidean distance. However, the definition of the new similarity measures and the choice of proper features are challenging.

7.2 Similarity of the curves

This work presents a new similarity measure for comparing the parallelism in the washout phase of the signal curves. A two-steps clustering is developed to combine the new measure with the Euclidean distance. At the same time, clustering in three feature spaces is studied. One of the feature spaces is the plane of the first and second principal components. The transformed data points on this plane have the maximal variance along the principal components. The two other feature spaces are the space of parameters from Tofts model and Brix model, respectively. Each of these feature spaces has its own advantages and disadvantages.

From Experiments is evident that Euclidean distance and PM measure are related to the first and second principal components and Tofts' parameters. That means two signal curves, which have a large Euclidean distance, are also far away along the first principal component or K_{trans} axis in Tofts parameter space and *vice versa*. Similarly, two signal curves with a large PM distance have also larger distances along second principal component or K_{ep} axis. On these grounds, the results of two-steps clustering, clustering in the space of principal components and in Tofts parameter space should be similar. Indeed, the results of two-steps clustering and the clustering in principal components plane are quite analogous. However, the results of Tofts parameter space differ, due to the problematic and instability of the fitting procedure.

In contrast to the Tofts model, fitting the signal curves to Brix model is quite stable and is independent of initial values. The *Chi* values of the fitting in compare to fitting to the Tofts models are significantly smaller. That means the curves can be described more exactly with three parameters of the Brix model than the two parameters of the Tofts model. Even so, experiments show no clear evidence of a relationship to the Brix parameters and the Euclidean distance or the PM measure. In this sense, adjacencies in the Brix parameter space indicate another kind of similarity, which differ from Euclidean distance or PM measure. From a philosophical point of view, it is impossible to describe the "true" similarity of the curves. Obviously, it depends on context and interpretation. However, from biological or medical point of view, we can look for those similarities with a relation to underlying biology.

Such a relationship can be, for example, finding the correlation between clusters in the tissue of interest and the underlying cellular or molecular structure of the tissue. This can

be done, for example, by the comparison of clustering to pathological gross cuts of the tumor. By finding such a correlation, it becomes possible to understand which kind of curves is related to which property of the tissue. This will provide a basis for the modeling of the tissue according to the signal curves. Alternatively, it can be the correlation of cluster patterns with patterns from another modality, such as PET (Positron emission tomography). In both cases, either the pathology cuts or PET, the most challenging part is to define the correspondence between DCE-MRI clusters and those modalities. Newly developed hybrid device of PET-MRI could open new possibilities in this area. Indeed, to this time point, a simultaneous acquisition of PET, and DCE-MRI is questionable.

7.3 Number of clusters

Results of the two-steps clustering and clustering in the plane of the first and second principal components look quite similar. Even so, there is a major difference between two methods. In two-steps clustering, the number of clusters is defined through the chosen thresholds. The threshold in each step of clustering is related to the noise of the signal curves and the desired maximal dissimilarity inside each cluster. In contrast, clustering in the plane of the first two principal components requires the number of clusters predefined. In a tumor, it is not *a priori* known how many different kinds of curves exist there. The measure of distance between points in the plane of principal components is Euclidean distance. Therefore, the Gap statistic is an appropriate candidate for estimating the number of clusters in the plane of principal components. However, Gap Statistic has difficulty in estimating the correct number of clusters if data consists of outliers and irregularities. The other issue is that the Gap Statistic procedure is time consuming. For these reasons, the developed two-steps clustering is advantageous over clustering in the plane of principal components. A comparison of the two methods based on affiliation of elements to clusters can give a more accurate understanding of each method.

7.4 Final remarks

Cluster analysis of a data, extracts, exhibits and visualizes the relevant information of the data to the user. Considering the extended information production during the last years, the importance of cluster analysis becomes clearer. During the last decades, plenty

of clustering methods have been designed and developed for different problems. Each of these methods has been optimized for the purposed task. Studies in fields of biology and medicine are not exempt. Here also, an exact definition of application and purpose of the clustering is fundamental. Indeed, in the cluster analysis of DCE-MRI signal curves, an exact understanding of the shape variations in the curves in relation to the underlying biology of the tissue is very important. This can be done, for example, in a bottom-up manner. By looking to the histology of the tumor, the sub-regions of the tumor can be selected according to the underlying structure of the tissue. In the next step, the signal curves of DCE-MRI inside each sub-region of the tumor should be searched and analyzed for their properties. This way, the relationship between signal curves can be understood better. Furthermore, this understanding can be used to optimize the clustering method and design the appropriate similarity measure or feature space.

Appendix A

Proofs

A.1 Proof of proposition (1):

$$log(\frac{A}{d_1}) \geq log(\frac{B}{d_2}) \Rightarrow \frac{A}{B} \geq \frac{d_1}{d_2}$$

$\Rightarrow$

$$\frac{A}{B} \geq 1 \text{ and } \frac{d_1}{d_2} \geq 1 \Rightarrow \frac{A}{B} - 1 \geq \frac{d_1}{d_2} - 1$$

Proof by contradiction: If $C - d_1 \geq D - d_2$ is not true, then we have:

$$C - d_1 < D - d_2$$

$$C - D < d_1 - d_2 \Rightarrow \frac{C - D}{d_2} < \frac{d_1}{d_2} - 1$$

$\Rightarrow$

$$\frac{C - D}{d_2} < \frac{A}{B} - 1$$

$$\frac{1}{B}\sum_b W^*_{kb} - \frac{1}{B}\sum_b W^*_{k+1b} < d_2 \frac{\prod {W^*_{kb}}^{1/B}}{\prod {W^*_{k+1b}}^{1/B}} - d_2. \tag{A.1}$$

Geometric to arithmetic mean relationship says:

$$\left(\prod \frac{W^*_{kb}}{W^*_{k+1b}}\right)^{1/B} \leq \frac{1}{B}\sum_b \frac{W^*_{kb}}{W^*_{k+1b}},$$

so we can rewrite the Eqn. A.1 as follows:

$$\frac{1}{d_2}\sum_b W^*_{kb} - \frac{1}{d_2}\sum_b W^*_{k+1b} < \sum_b \frac{W^*_{kb}}{W^*_{k+1b}} - B \tag{A.2}$$

$\Rightarrow$

$$\sum_b \frac{W^*_{kb}}{d_2} - \sum_b \frac{W^*_{kb}}{W^*_{k+1b}} < \sum_b \frac{W^*_{k+1b}}{d_2} - B$$

$\Rightarrow$

$$\sum_b \frac{W^*_{kb}}{W^*_{k+1b}} \left(\frac{W^*_{k+1b} - d_2}{d_2} \right) < \sum_b \frac{W^*_{k+1b} - d_2}{d_2}.$$

For all values of b, $\frac{W^*_{kb}}{W^*_{k+1b}} \geq 1$, by setting $\frac{W^*_{kb}}{W^*_{k+1b}}$ to its minimum value 1, then we have:

$$\sum_b \frac{W^*_{k+1b} - d_2}{d_2} < \sum_b \frac{W^*_{k+1b} - d_2}{d_2}.$$

A.2 Proof of proposition (2):

From the previous proof we have:

$$\frac{A}{B} - 1 < \frac{C - D}{d_2}$$

Thus, in the case of $d_1 - d_2 = C - D$ we will have:

$$\frac{A}{B} < \frac{d_1}{d_2}$$

□

Appendix B

Case study: Unequally sized clusters

In the following case study the effect of number difference between clusters on Gap Statistic is studied. The case study considers datasets with each consisting of two clusters sampled from two 2D normal distributions $N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $N(\boldsymbol{\mu}', \sigma^2\mathbf{I})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ are expected values, $\mathbf{I}$ is the identity matrix, and $\sigma^2 > 0$ is a positive real number. According to standard score 99.7% of samples will be inside a circle with radius $3 \cdot \sigma$ [Glenberg and Andrzejewski, 2008]. Here, the uniform distribution rectangle, from which the null references are sampled, was estimated as a rectangle with sides $6 \cdot \sigma + \Delta$ and $6 \cdot \sigma$ as illustrated in Fig. B.1, where $\Delta = \|\boldsymbol{\mu} - \boldsymbol{\mu}'\|$. Let N_1 be the number of samples in the first cluster and N_2 be the number of samples in the second cluster, while $N_1 = m \cdot N_2$ and $n = N_1 + N_2$. In section 6.2, we observed that in the case of $N_1 = N_2$, for $\Delta \geq 5\sigma$ both Gap functions estimate two as the proper number of clusters. In this study, we want to show how changes in m affect the result of the Gap Statistic. Let $\Delta \geq 5\sigma$ and n be fixed. In Gap Statistic in order to be able to choose $k = 2$ as the proper number of clusters, it is necessary to have $Gap_n(1) \leq Gap_n(2) - s_2$, otherwise it suggests $k = 1$. We ignore s_2 and consider the inequality $Gap_n(1) \leq Gap_n(2)$. The two next inequalities follows from the Eqns. 3.42 to 3.47 for Gap and Gap^*, respectively:

1. Gap

$$\left(\prod \frac{W^*_{1b}}{W^*_{2b}}\right)^{\frac{1}{B}} \leq \frac{W_1}{W_2} \tag{B.1}$$

2. Gap^*

$$\frac{1}{B}\sum (W^*_{1b} - W^*_{2b}) \leq W_1 - W_2 \tag{B.2}$$

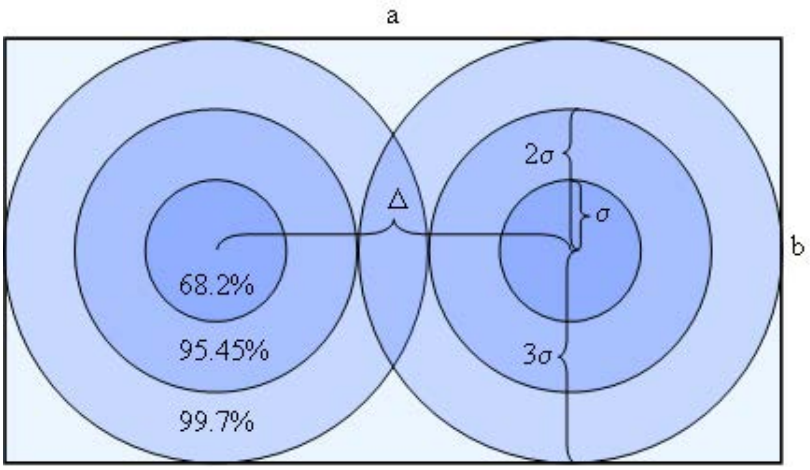


Figure B.1: Two 2D distributions from the case study in Appendix B. Each distribution is depicted with three areas of 68.2%, 95.45%, and 99.7% percentage of sample occurrence inside each area.

Each W^*_{1b} can be estimated as $nE(d_1)$, where $E(d_1)$ is the expected distance between two random points from a rectangular uniform distribution with sides $6\sigma + \Delta$ and 6σ. In a similar way W^*_{2b} can be estimated as $nE(d_2)$, where $E(d_2)$ is the expected distance between two random points from a rectangular uniform distribution with sides $\frac{6\sigma+\Delta}{2}$ and 6σ. The expected distance of two random points sampled from a rectangular uniform distribution with sides a and b with $a \geq b$ is given by [Santalo, 1976]:

$$E(d) = \frac{1}{15}\left[\frac{a^3}{b^2} + \frac{b^3}{a^2} + d\left(3 - \frac{a^2}{b^2} - \frac{b^2}{a^2}\right) + \frac{5}{2}\left(\frac{b^2}{a}\log\frac{a+d}{b} + \frac{a^2}{b}log\frac{b+d}{a}\right)\right], \quad \text{(B.3)}$$

where $d = \sqrt{a^2 + b^2}$. Using these estimations and Eqns. B.1 and B.2, we gain for

1. *Gap*:

$$\frac{E(d_1)}{E(d_2)} \leq \frac{W_1}{W_2}, \quad \text{(B.4)}$$

2. *Gap**:

$$n\left(E(d_1) - E(d_2)\right) \leq W_1 - W_2. \quad \text{(B.5)}$$

Furthermore, we can take into account that W_1 includes the inter-cluster distances between the first and second clusters in addition to all distances which are used in calculation of W_2. Therefore, W_1 can be written as $W_2 + \frac{2N_1N_2d_\Delta}{n}$, where d_Δ is the average inter-cluster distances. Consequently, inequalities (B.4) and (B.5) can be rewritten as:

1. *Gap*

$$\frac{E(d_1)}{E(d_2)} - 1 \leq \frac{md_\Delta}{\sigma(m+1)^2}, \tag{B.6}$$

2. *Gap**

$$E(d_1) - E(d_2) \leq \frac{2md_\Delta}{(m+1)^2}. \tag{B.7}$$

Appendix C

Supplementary information for breast datasets

Data description

Fifteen DCE-MRI datasets of female patients with breast cancer (ages 30-87 years, mean age 55.23 years) are used for experiments in this work. Breast lesions were proven by histology following breast biopsy or surgery. The datasets are separated into two groups of M-series and P-series. The M-series consists of nine datasets from German Cancer Research Center [DKFZ, 2004]. The P-series are from a study started in 2011 in cooperation with Radiology Institute of Klinikum rechts der Isar [Radiology, 2011]. Name of datasets from M-series starts with M and in P-series with P. Table C.1 summarizes the following information for 15 DCE-MRI breast datasets:

1. Nr. of voxels: Number of segmented voxels, which undergone the clustering.
2. Tumor size: The approximated size of the tumors in mm.
3. Ave. noise: Average noise calculated for tumor curves in each dataset by the method from section 2.3.3.
4. Diagnosis: ductal invasive carcinoma = d. inv. Carcinoma, lobular carcinoma = lob. Carcinoma

It can be observed that the M-series datasets have smaller signal to noise values in comparison with the P-series. Number of clusters in first and second step of two-steps clustering

index	dataset	Nr. voxels	Tumor size	Average noise	Diagnosis
1	M002	1633	68	0.049	d. inv. carcinoma
2	M005	192	35	0.081	fibroadenoma
3	M006	115	20	0.052	d. inv. carcinoma
4	M008	455	37	0.043	d. inv. carcinoma
5	M009	123	22	0.053	lob. carcinoma
6	M011	340	33	0.040	d. inv. carcinoma
7	M012	86	20	0.065	d. inv. carcinoma
8	M013	497	48	0.049	d. inv. carcinoma
9	M014	139	29	0.064	d. inv. carcinoma
10	P001	297	26	0.125	lob. carcinoma
11	P002	560	30	0.169	fibroadenoma
12	P003	173	23	0.240	lob. carcinoma
13	P004	836	44	0.194	carcinoma
14	P005	370	37	0.124	carcinoma
15	P006	461	56	0.153	fibroadenoma

Table C.1: Breast datasets summary.

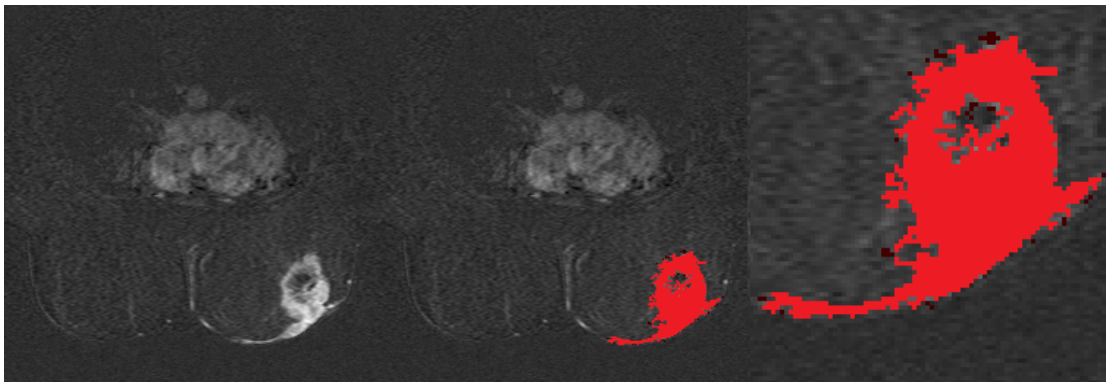
is given in Table C.2. In Figures C.1 to C.5, for each dataset, a subtraction slice of a post-enhanced, a pre-enhanced slice of that dataset, the segmented tumor and a magnified image of the segmented tumor are depicted.

C.1 Experiments' results

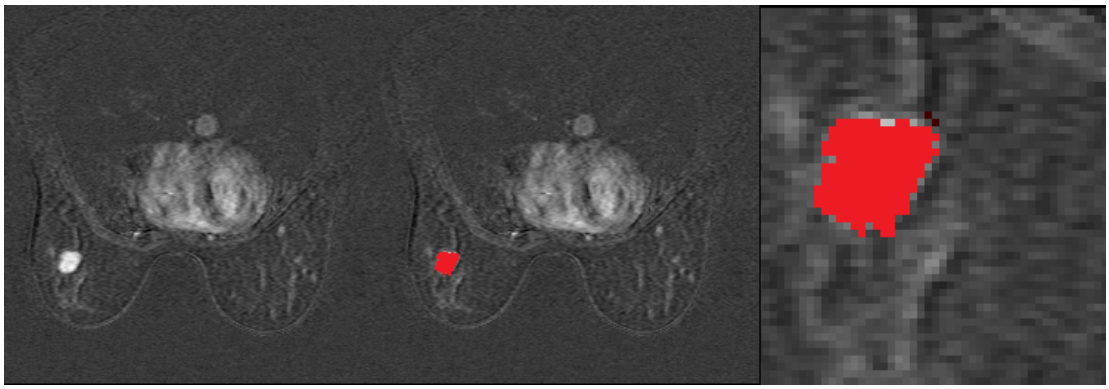
The results of clustering with two-steps clustering for each dataset are depicted in Figures C.6 to C.11. The clusters and the corresponding colors are given in a look up table

Datasets	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nr. of clusters in 1st step	5	3	1	2	1	1	2	2	2	5	10	2	21	5	3
Total Nr. of clusters	14	10	3	6	2	3	4	5	4	10	31	3	52	9	10

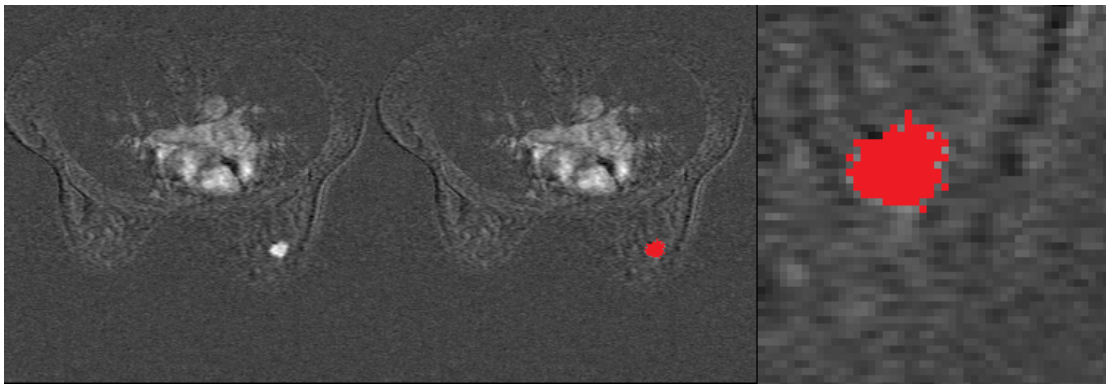
Table C.2: Number of clusters determined by two-steps clustering



(a) M002

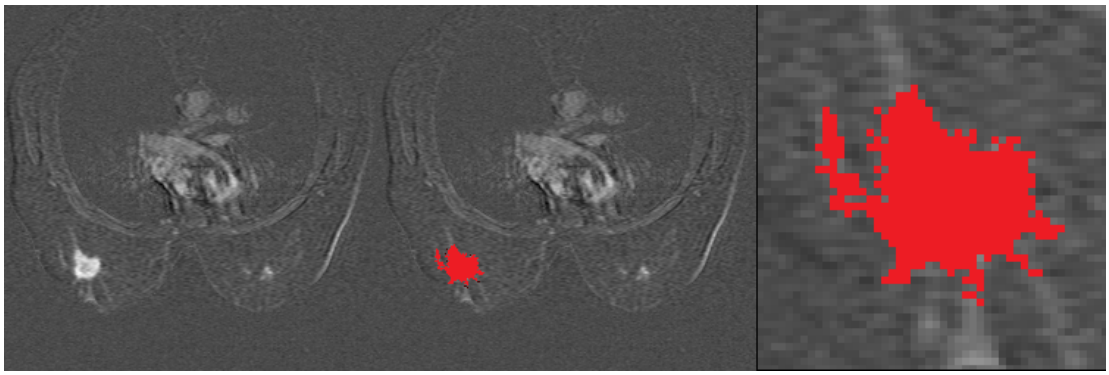


(b) M005

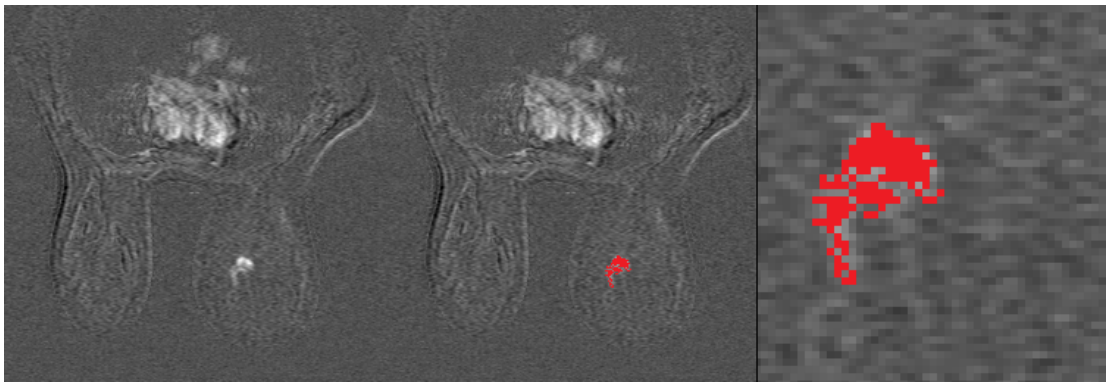


(c) M006

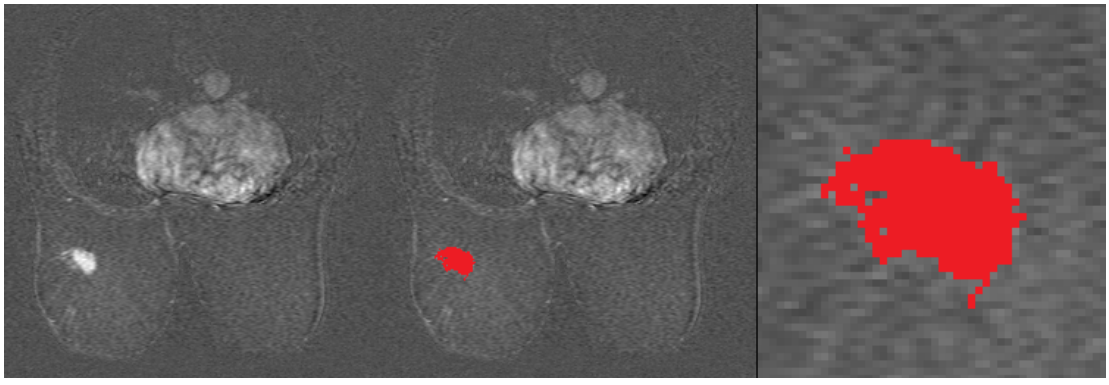
Figure C.1: Segmentation of M-series datasets



(a) M008

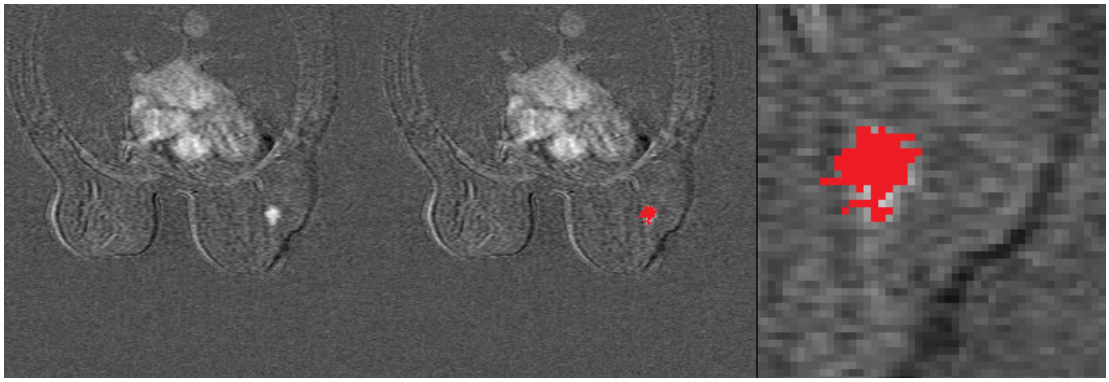


(b) M009

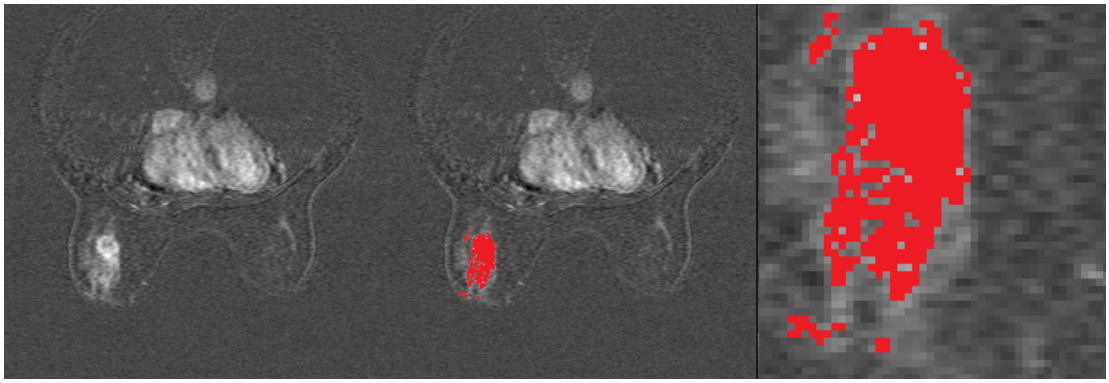


(c) M011

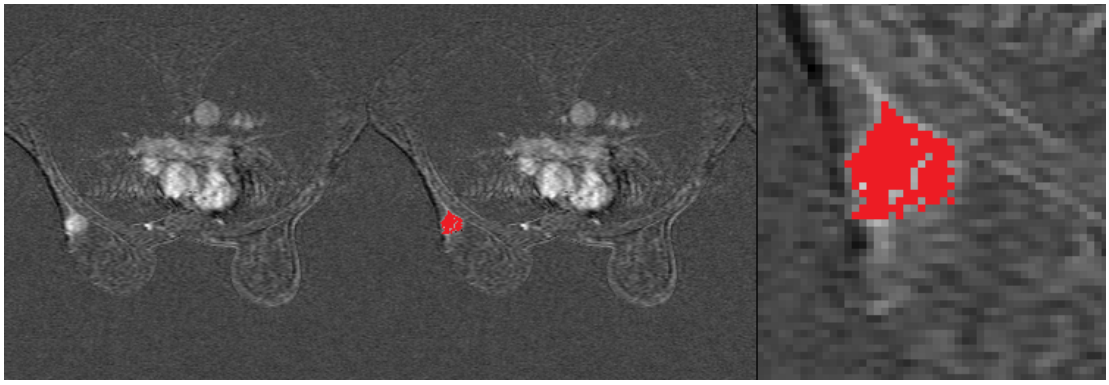
Figure C.2: Segmentation of M-series datasets



(a) M012

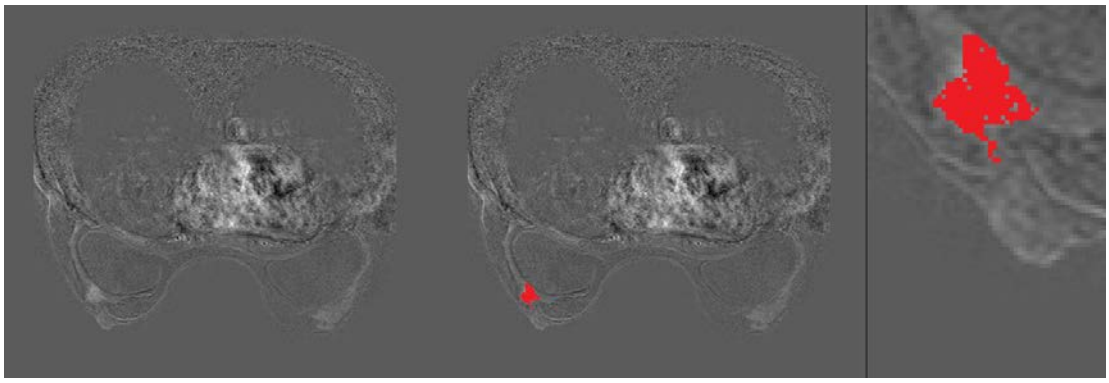


(b) M013

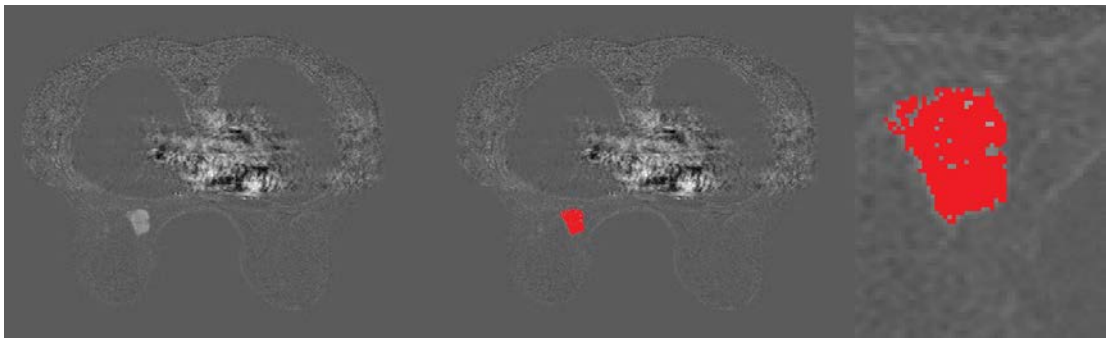


(c) M014

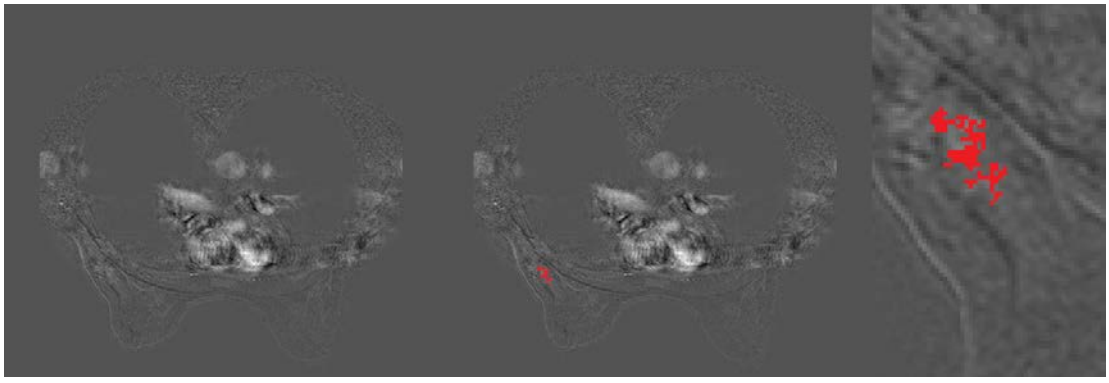
Figure C.3: Segmentation of M-series datasets



(a) P001

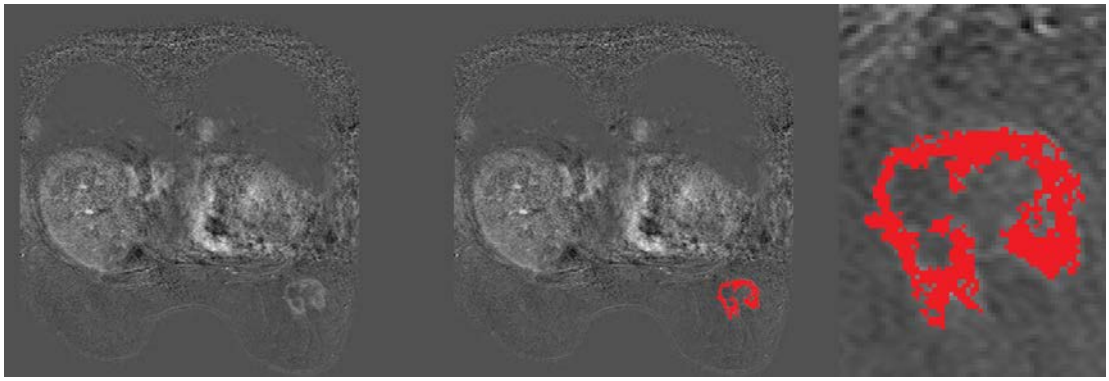


(b) P002

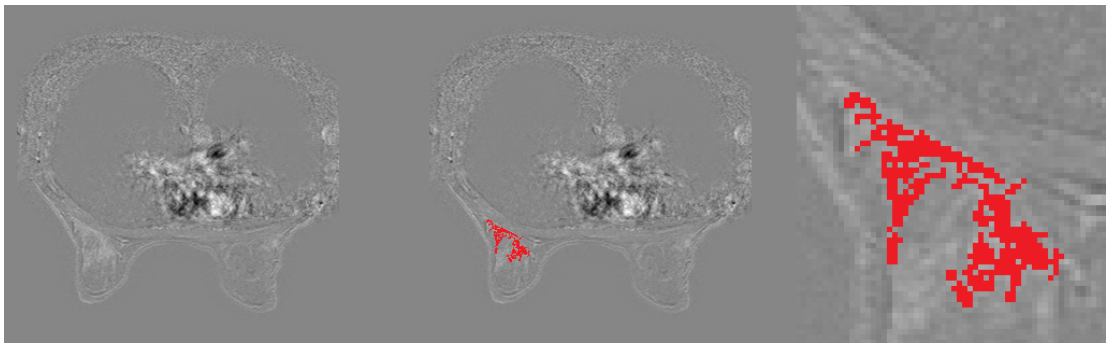


(c) P003

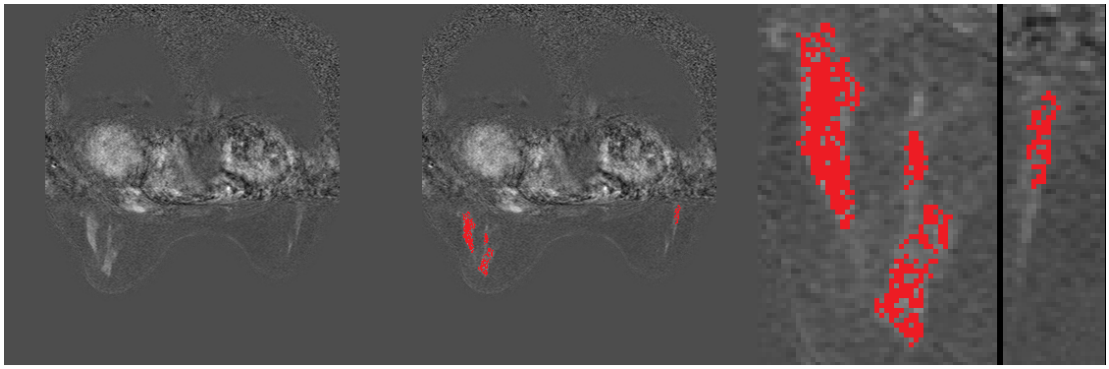
Figure C.4: Segmentation of P-series datasets



(a) P004



(b) P005

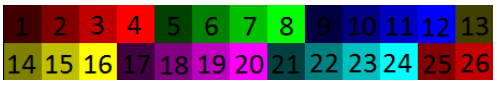


(c) P006

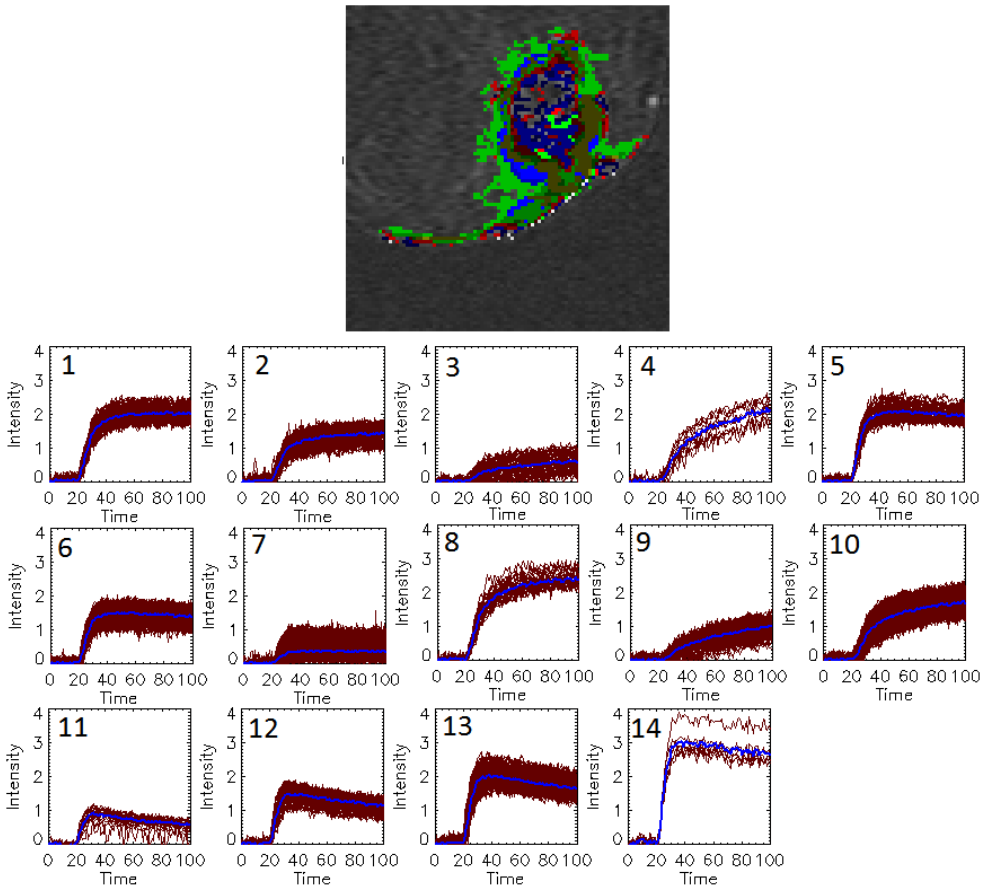
Figure C.5: Segmentation of P-series datasets

for each clustering result. In addition, the signal curves inside each cluster are depicted for each clustering result.

Signal curves corresponding to the data points selected from the plane of the two eigenvectors corresponding to the first two largest eigenvalues of the correlation matrix for the M-series and P-series are shown in Figures C.12 and C.13. The values of γ and ξ calculated from the experiments with 13 clustering methods for each dataset are given in Tables C.3 and C.4. The clustering methods, two-steps clustering, k-means, c-means, and average linkage method are marked as number 0, 1, 2, and 3 respectively. The values of γ and ξ for experiments with different similarity measures are given in Table C.5.



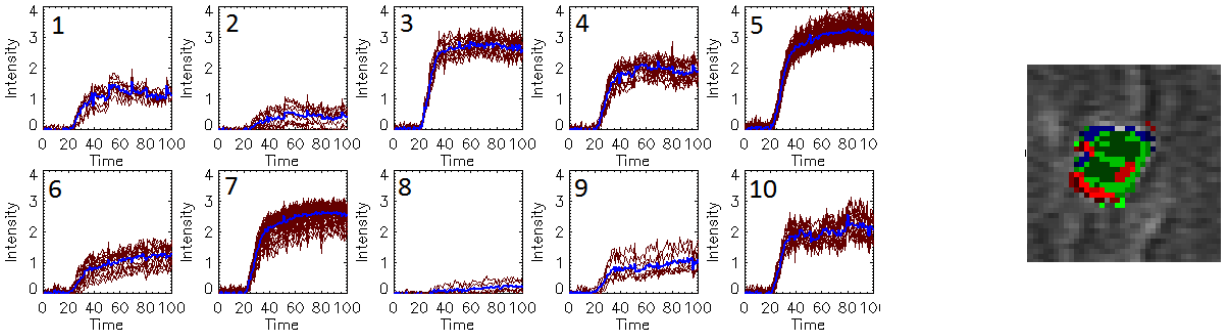
(a) Look up table for the clusters



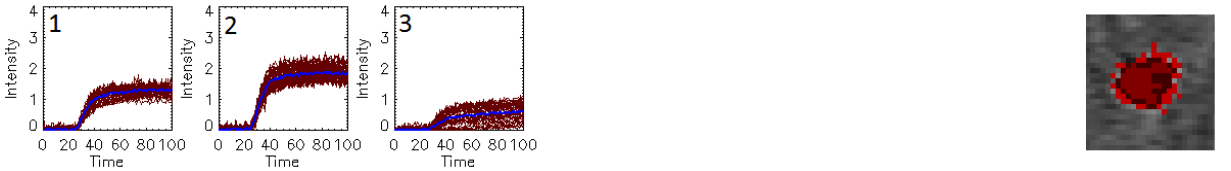
(b) M002

Figure C.6: Two-steps clustering of breast datasets

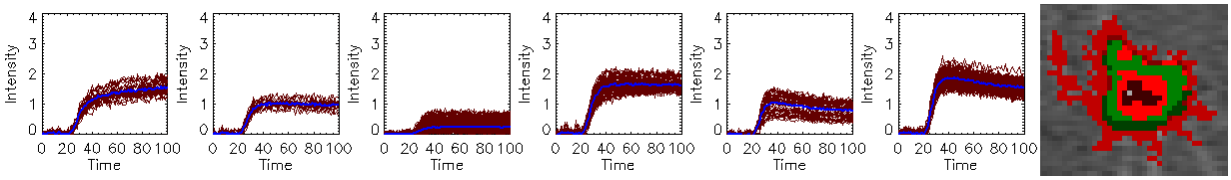
(a) Look up table for the clusters



(b) M005

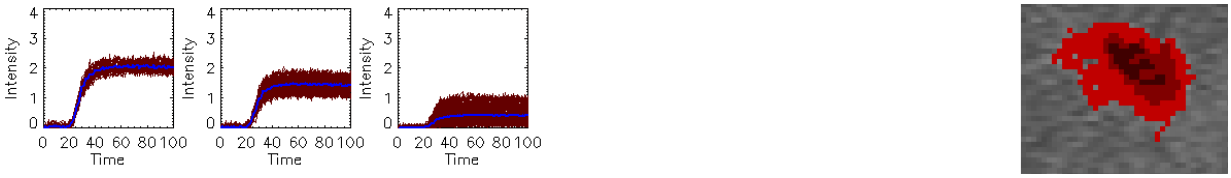


(c) M006



(d) M008

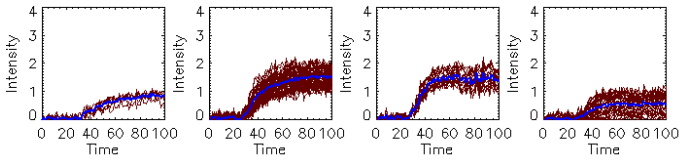
(e) M009

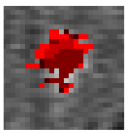


(f) M011

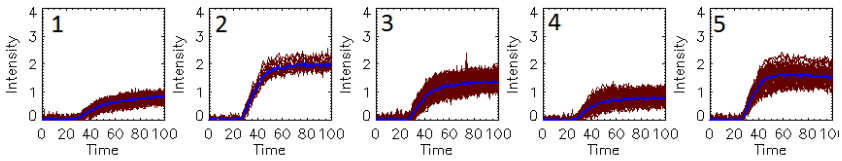
Figure C.7: Two-steps clustering of breast datasets

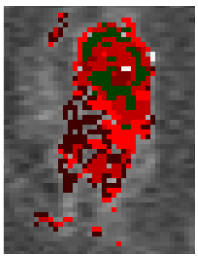
(a) Look up table for the clusters



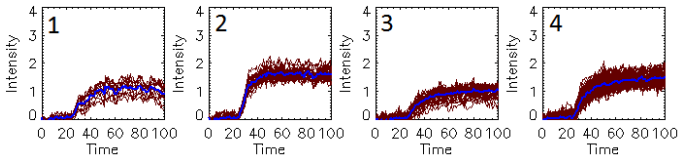


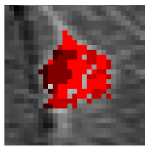
(b) M012



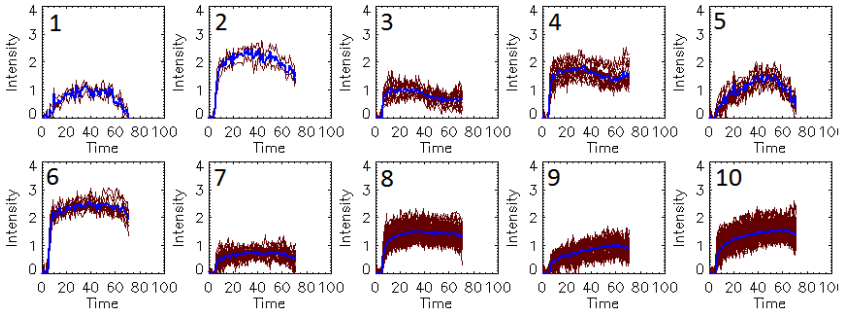


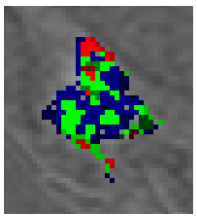
(c) M013



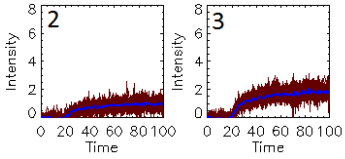


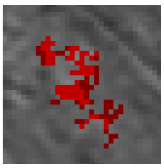
(d) M014





(e) P001

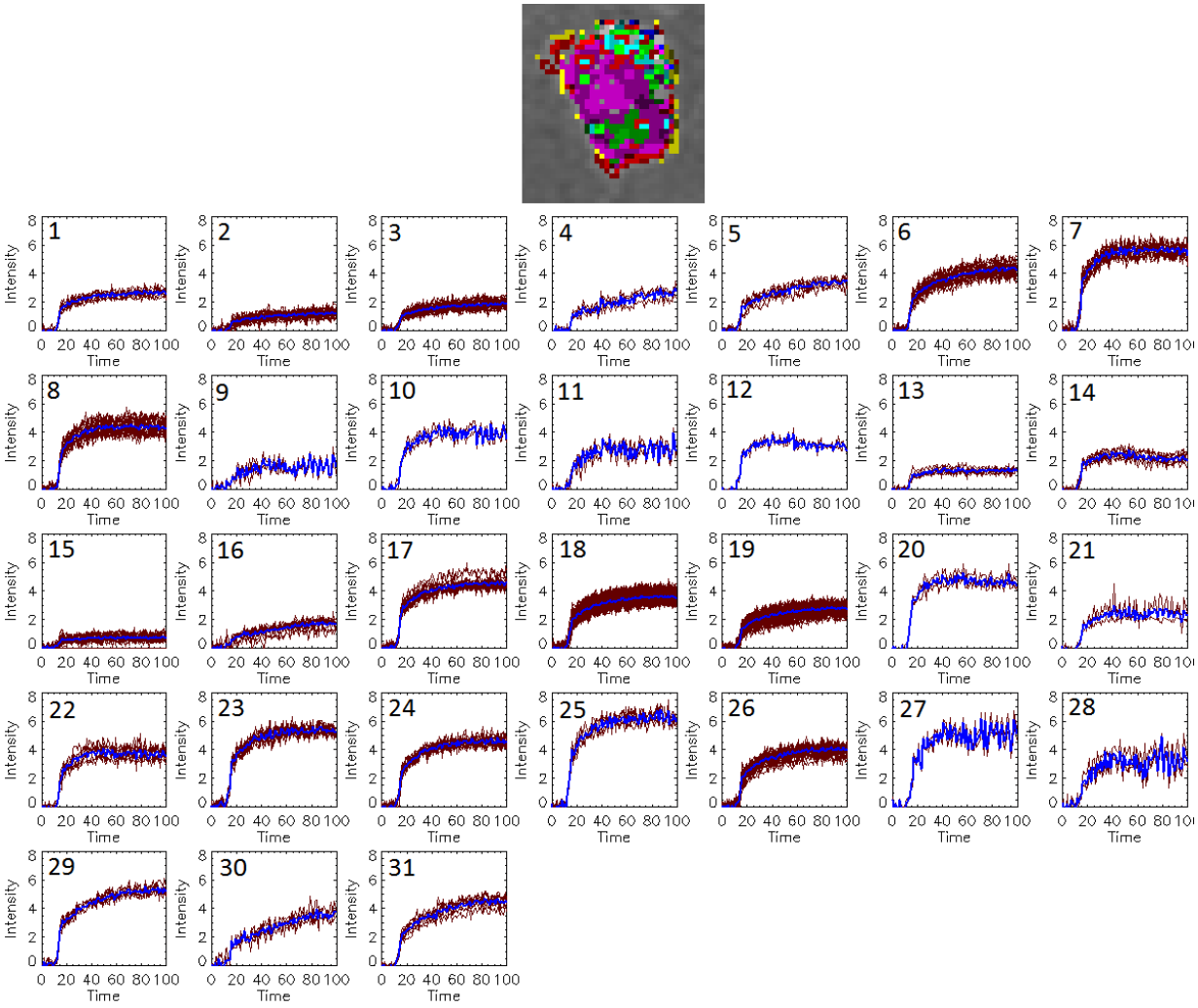




(f) P003

Figure C.8: Two-steps clustering of breast datasets

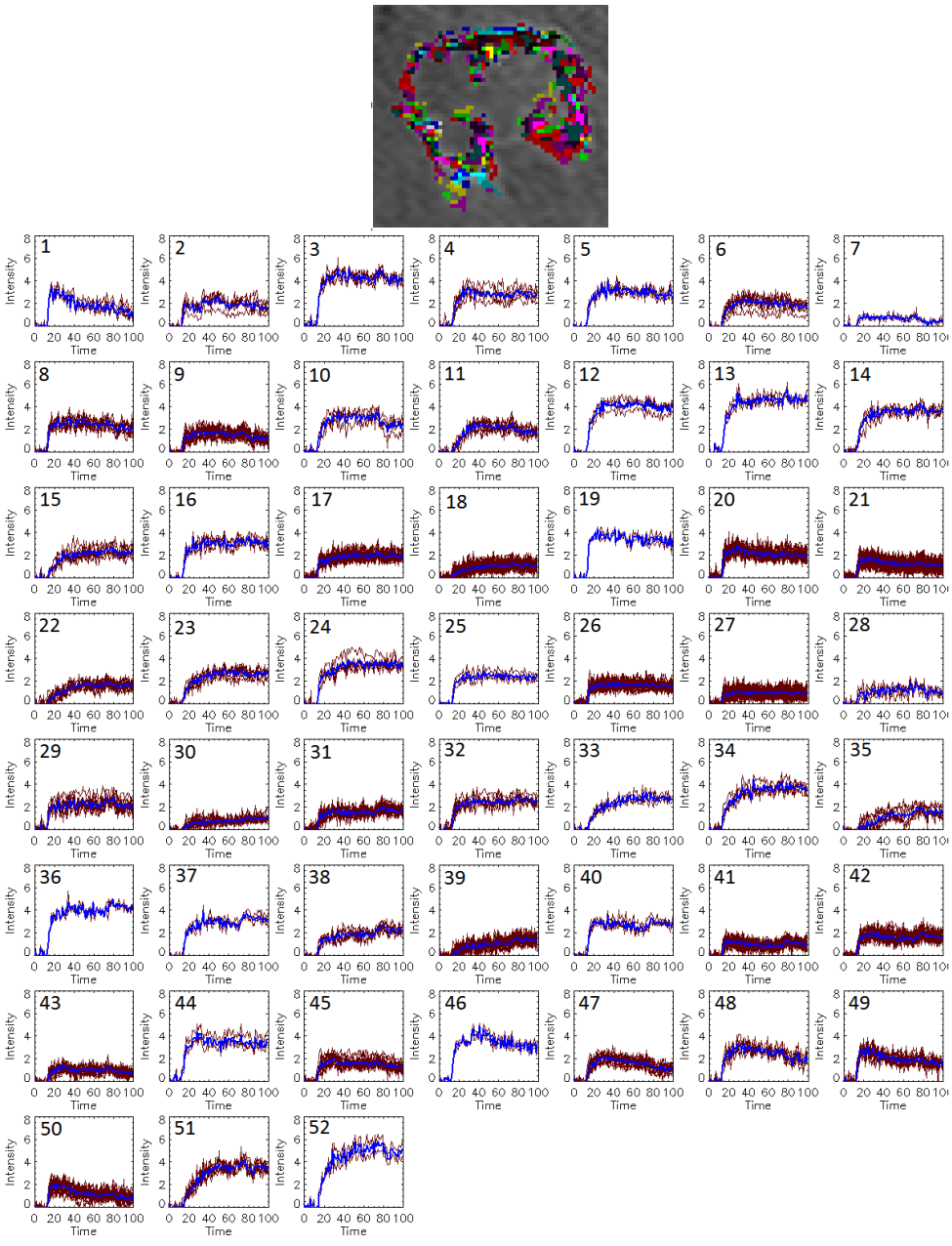
(a) Look up table for the clusters



(b) P002

Figure C.9: Two-steps clustering of breast datasets

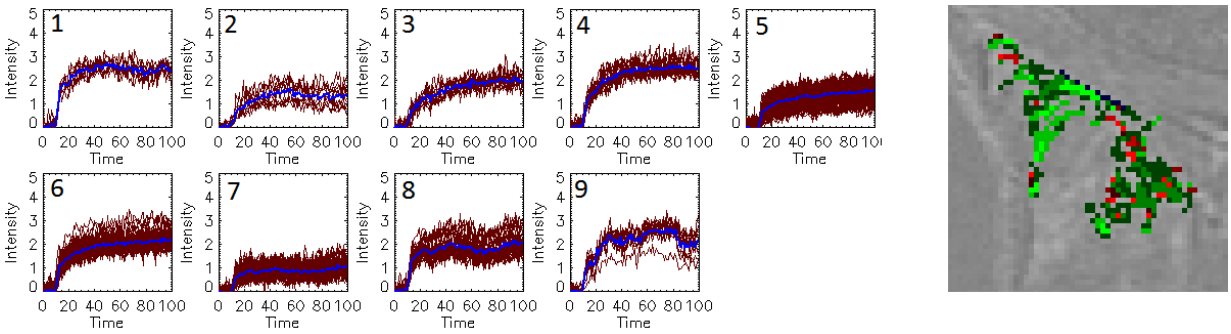
(a) Look up table for the clusters



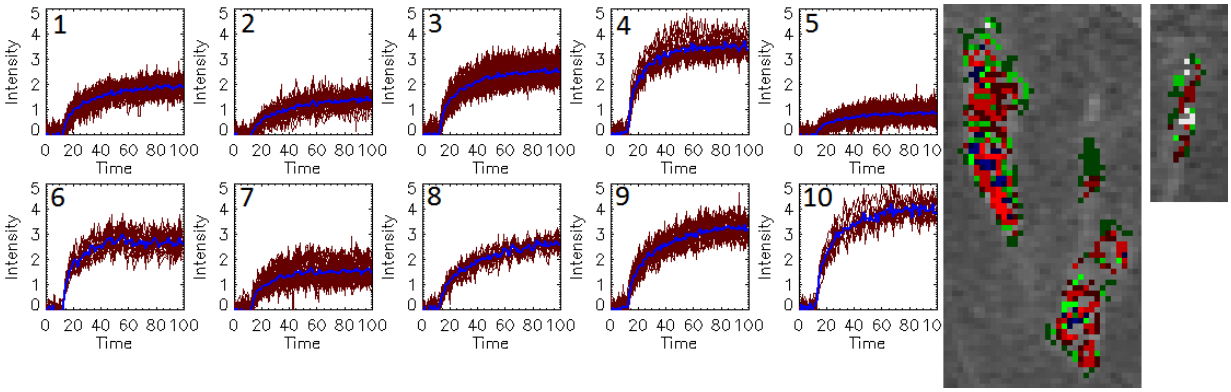
(b) P004

Figure C.10: Two-steps clustering of breast datasets

(a) Look up table for the clusters



(b) P005



(c) P006

Figure C.11: Two-steps clustering of breast datasets

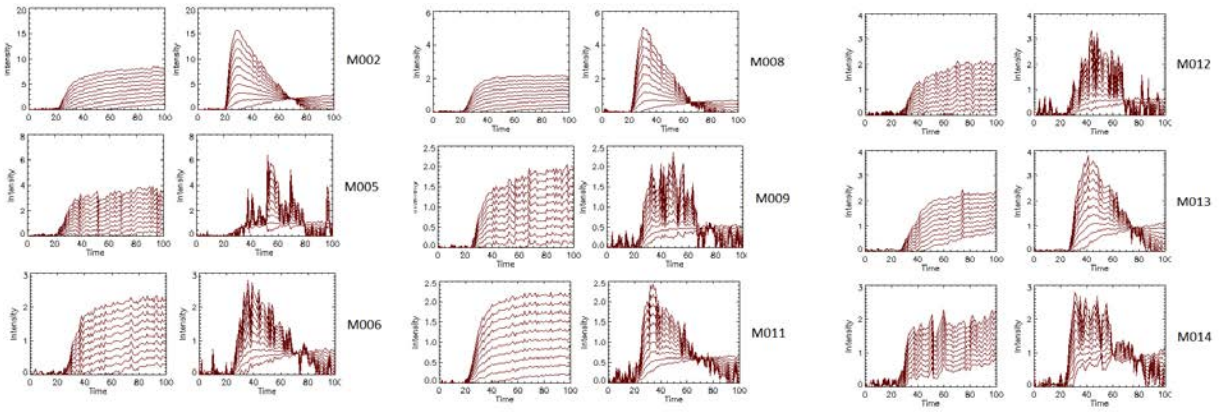


Figure C.12: Signal curves corresponding to the data points selected from the plane of the two eigenvectors corresponding to the first two largest eigenvalues of the correlation matrix for M-series datasets.

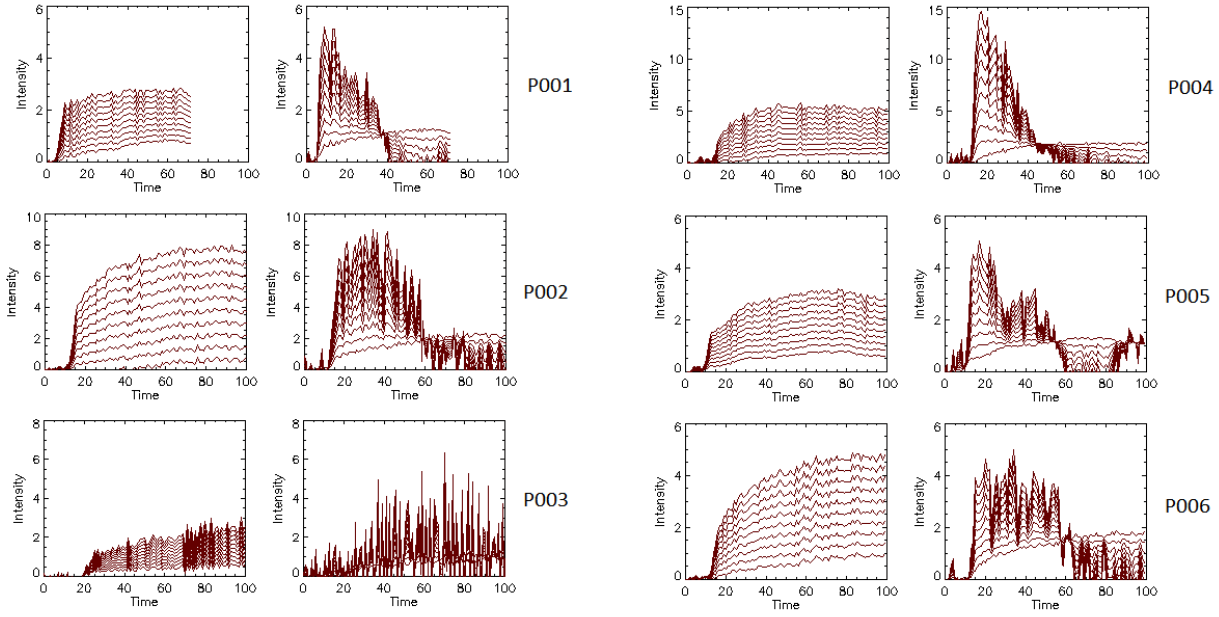


Figure C.13: Signal curves corresponding to the data points selected from the plane of the two eigenvectors corresponding to the first two largest eigenvalues of the correlation matrix for P-series datasets.

	Real				PC			Toft			Brix		
	0	1	2	3	1	2	3	1	2	3	1	2	3
P001	1.92	1.84	1.50	1.65	1.72	1.63	1.97	1.79	1.82	1.99	2.23	2.20	2.39
P002	3.22	2.70	2.44	2.73	3.03	3.10	3.39	3.17	3.00	2.99	7.02	10.21	5.20
P003	3.62	3.23	3.22	3.51	3.79	3.75	5.33	3.80	3.65	5.13	5.14	5.17	5.20
P004	3.38	3.49	3.17	3.19	2.87	2.86	2.95	2.99	2.97	3.11	5.41	5.21	6.70
P005	2.83	2.15	2.04	2.33	2.49	2.33	2.77	2.46	2.49	2.66	3.60	3.68	3.84
P006	2.69	2.49	2.23	2.73	3.15	3.10	3.56	3.09	2.95	3.43	4.52	4.22	5.76
M002	2.12	1.61	1.50	2.65	2.09	1.82	2.74	2.15	2.03	2.60	5.21	5.22	5.81
M005	2.44	1.79	1.76	1.82	2.84	2.69	2.85	2.37	2.28	2.77	3.21	3.83	3.83
M006	1.98	1.97	1.98	1.98	2.70	2.73	4.65	2.57	2.60	2.81	3.12	3.12	4.76
M008	1.63	1.17	1.18	1.28	1.65	1.64	1.78	2.01	1.88	2.31	4.13	3.91	3.88
M009	2.43	2.47	2.34	2.34	3.77	3.08	3.10	3.83	3.83	3.83	3.57	3.57	3.57
M011	2.04	1.74	1.73	1.80	2.54	2.43	2.85	2.38	2.39	5.43	4.39	4.25	4.42
M012	2.57	1.73	1.72	1.79	2.29	2.43	2.56	2.54	2.35	2.56	3.87	3.68	4.18
M013	1.74	1.52	1.23	1.50	1.76	1.78	2.04	1.80	1.80	1.85	2.14	2.11	2.26
M014	1.90	1.97	1.55	1.82	1.77	1.76	2.13	2.04	2.07	2.24	2.29	2.21	2.60

Table C.3: Distance to mean curve (γ values).

	Real				PC			Toft			Brix		
data	0	1	2	3	1	2	3	1	2	3	1	2	3
P001	0.30	0.96	0.82	0.83	0.28	0.24	0.31	0.21	0.20	0.22	0.22	0.19	0.26
P002	0.19	0.64	0.66	0.49	0.09	0.08	0.13	0.73	0.72	0.74	0.86	1.00	0.63
P003	0.28	0.40	0.42	0.41	0.45	0.50	0.97	0.41	0.33	0.55	0.58	0.58	0.63
P004	0.10	0.67	0.85	0.57	0.09	0.08	0.11	0.06	0.05	0.06	0.17	0.16	0.24
P005	0.41	0.74	0.73	0.71	0.36	0.33	0.43	0.28	0.29	0.29	0.47	0.42	0.47
P006	0.17	0.39	0.38	0.41	0.13	0.15	0.25	0.25	0.23	0.41	0.61	0.54	0.85
M002	0.45	0.60	0.58	0.63	0.60	0.56	0.59	0.57	0.54	0.55	0.60	0.60	0.60
M005	0.38	0.83	0.85	0.83	0.37	0.35	0.50	0.54	0.73	0.86	0.30	0.32	0.33
M006	0.97	0.98	0.96	0.97	0.46	0.41	0.41	0.80	0.79	0.86	0.70	0.70	0.85
M008	0.25	0.78	0.67	0.76	0.13	0.13	0.19	0.36	0.31	0.33	0.30	0.25	0.37
M009	0.57	0.88	0.70	0.70	0.42	0.32	0.99	0.94	0.94	0.94	0.90	0.90	0.90
M011	0.97	0.99	0.98	0.98	0.65	0.72	0.84	0.81	0.81	0.96	0.47	0.56	0.53
M012	0.30	0.86	0.84	0.89	0.27	0.24	0.22	0.54	0.51	0.64	0.55	0.59	0.68
M013	0.42	0.90	0.89	0.91	0.26	0.20	0.58	0.41	0.54	0.46	0.47	0.45	0.52
M014	0.32	0.80	0.87	0.88	0.29	0.29	0.33	0.37	0.39	0.36	0.35	0.34	0.55

Table C.4: Average slopes (ξ values).

	γ values				ξ values			
data	PM	Euclidean	Cosine	Correl.	PM	Euclidean	Cosine	Correl.
P004	6.225	3.961	6.747	6.069	0.580	1.015	1.014	0.519
P005	3.894	2.400	4.410	4.222	0.721	0.922	0.986	0.692
P006	4.875	2.825	6.389	6.591	0.458	0.720	0.955	0.735
M002	5.719	2.893	5.915	5.707	0.743	0.820	0.870	0.806
M005	5.606	2.461	7.211	6.530	0.486	0.928	0.998	0.424
M006	2.936	1.382	4.653	4.261	0.494	0.917	0.982	0.604
M008	4.459	1.403	6.100	5.873	0.305	0.817	1.032	0.444
M009	2.562	1.331	3.837	3.638	0.366	0.674	0.946	0.297
M011	2.539	1.247	5.067	5.439	0.741	1.142	1.011	0.705
M012	3.422	1.590	4.136	3.725	0.388	0.815	1.019	0.414
M013	2.082	1.319	3.524	2.603	0.521	0.998	0.998	0.603
M014	2.357	1.646	2.802	2.206	0.406	0.885	1.017	0.447

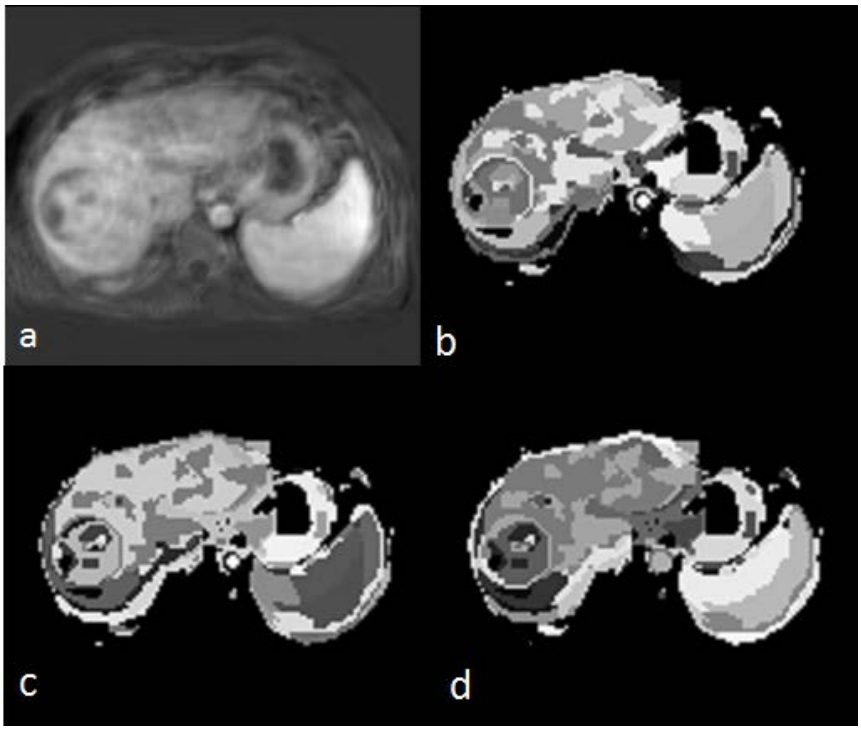
Table C.5: Similarity measures experiment

Appendix D

Supplementary information for liver datasets

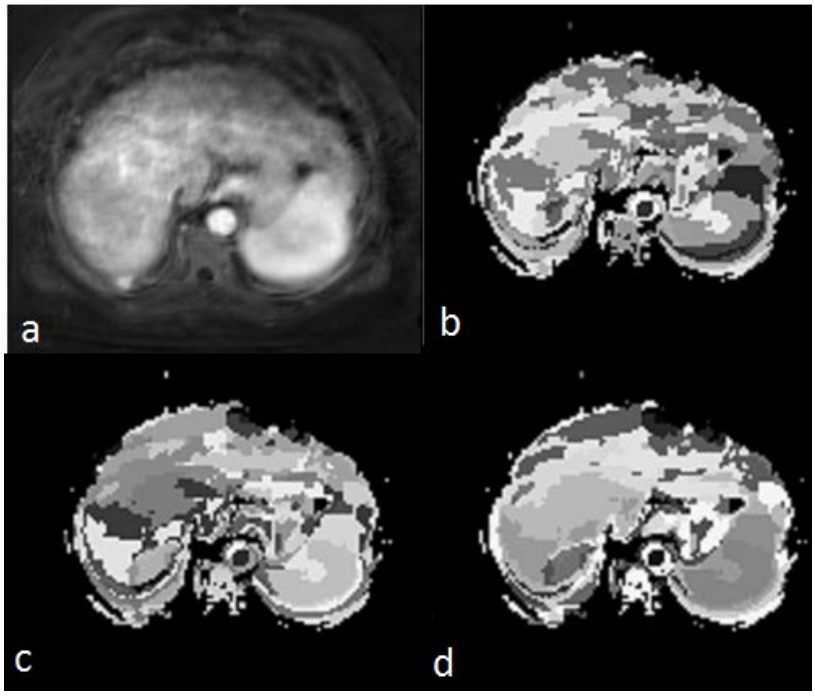
D.1 Experiments' results

Nine datasets of patients with hepatocellular carcinoma (HCC) are studied in this work. Figures D.1 to D.5 show the results of the clustering method introduced in section 5.2.3, for these nine datasets. In each Figure, sub-figure (a) depicts the subtraction image of a pre-enhanced slice from post-enhanced slice. In sub-figures (b, c, and d) the results of clustering with cluster numbers 100, 50, and 25 are depicted. The five slices of the reduced dataset A^{*5} on the space of the five first principal components for all nine datasets are depicted in Figure D.6. The corresponding five principal components are depicted in Figure D.7.

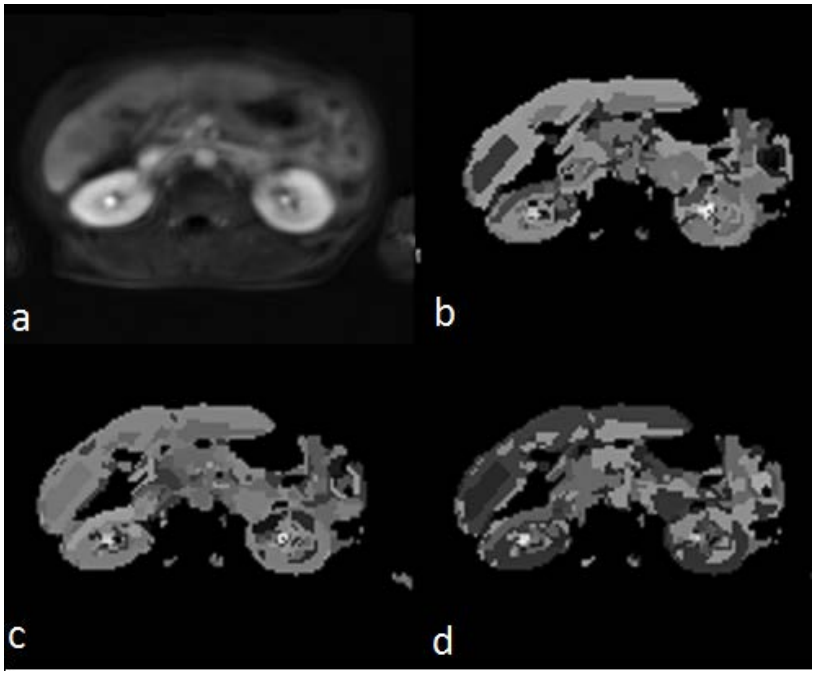


(a) Dataset 1

Figure D.1: Clustering of liver datasets

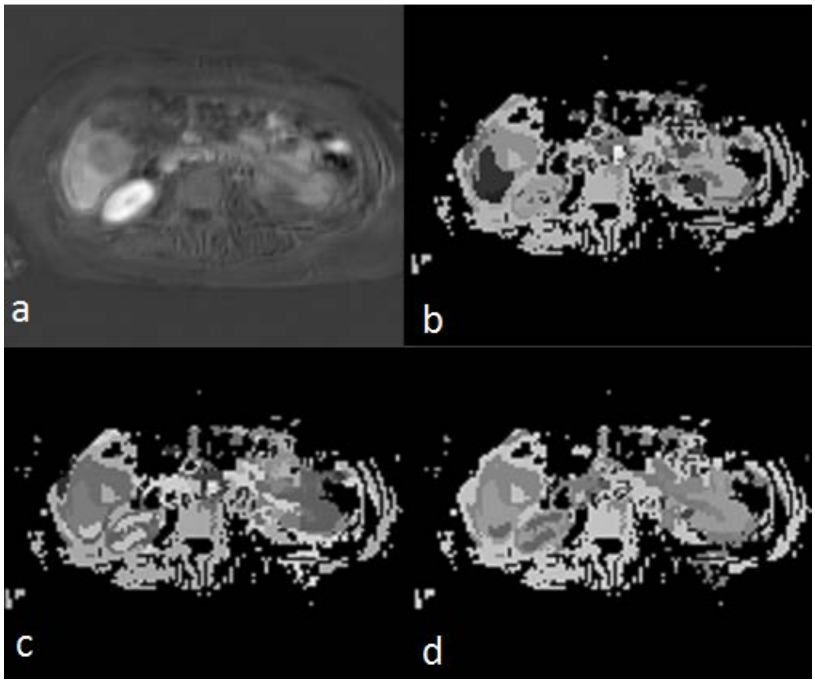


(a) Dataset 2

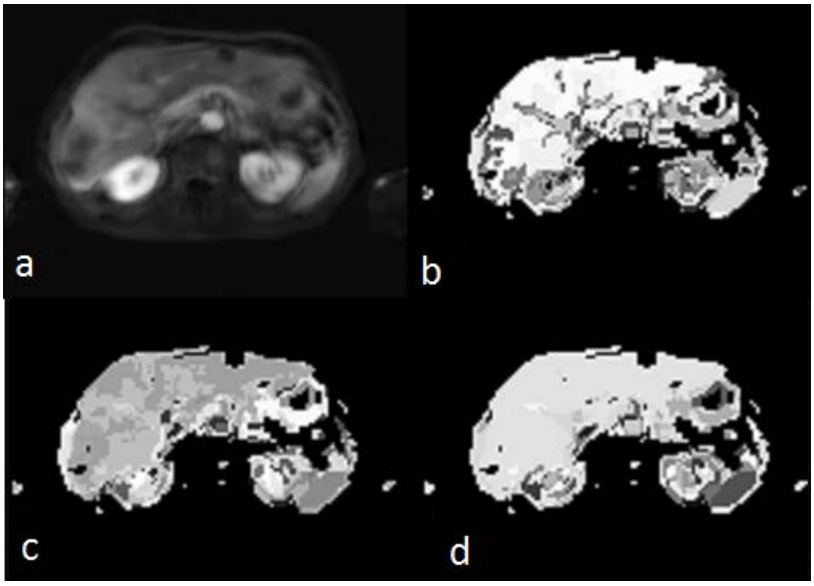


(b) Dataset 3

Figure D.2: Clustering of liver datasets

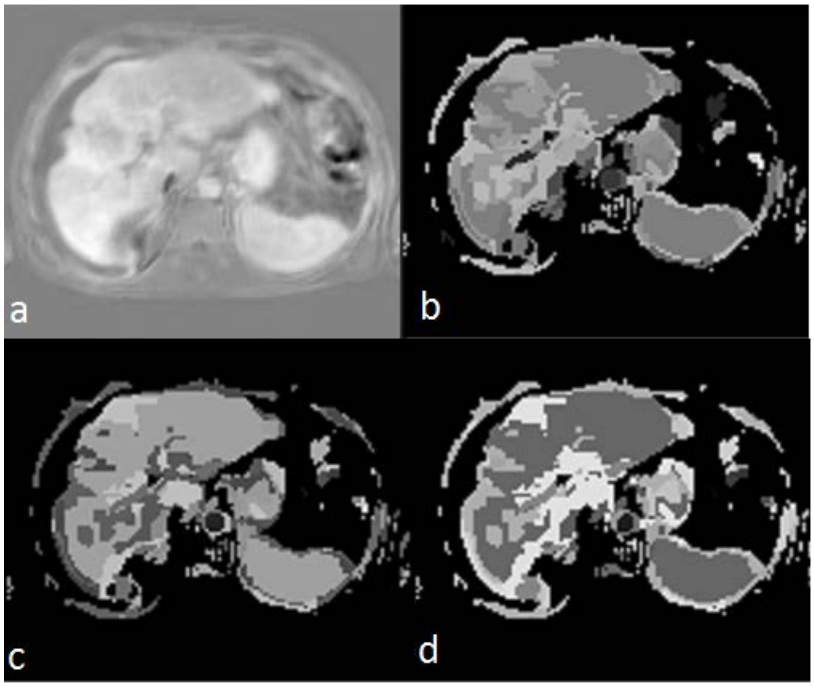


(a) Dataset 4

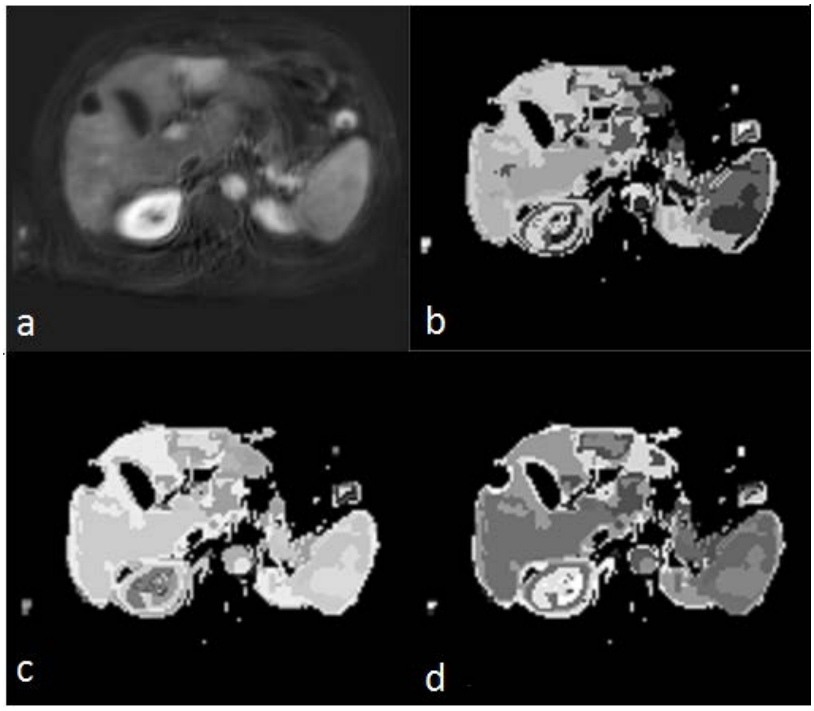


(b) Dataset 5

Figure D.3: Clustering of liver datasets

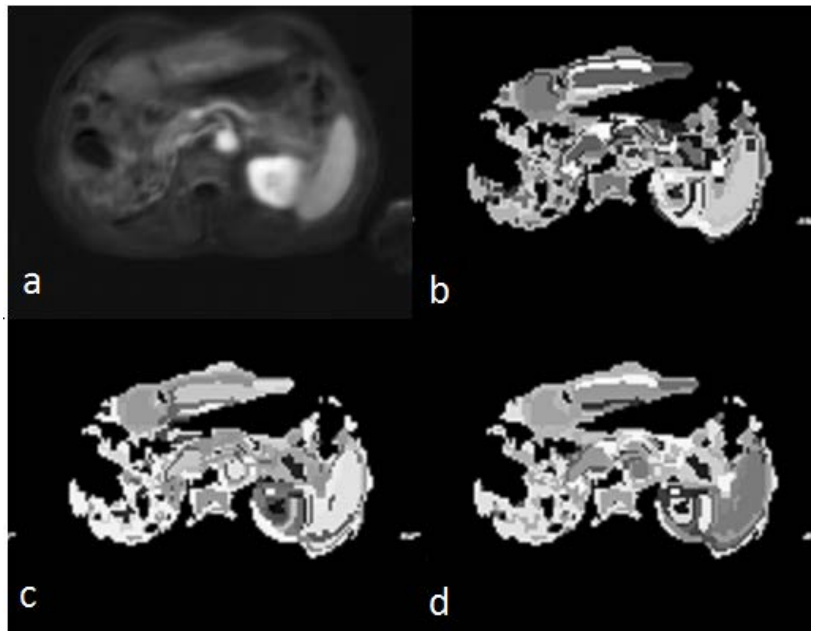


(a) Dataset 6

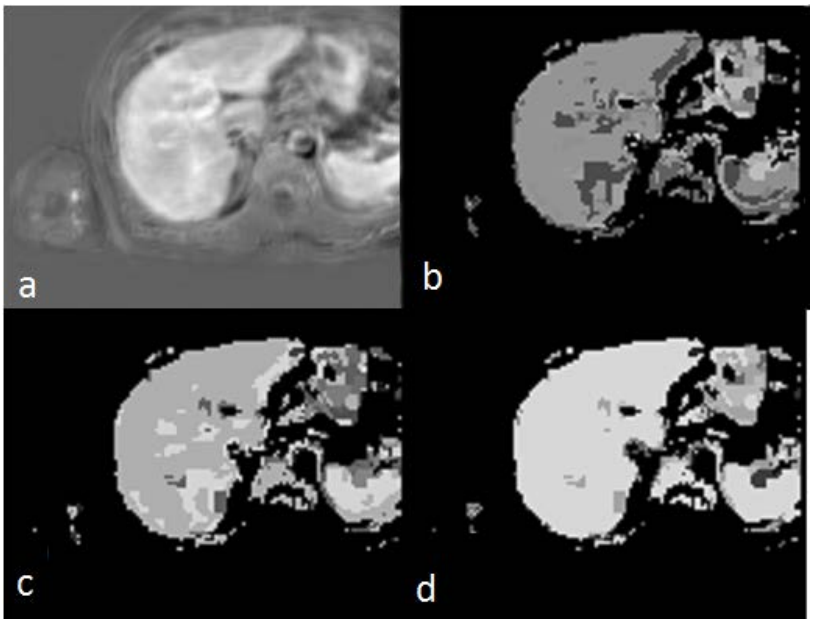


(b) Dataset 7

Figure D.4: Clustering of liver datasets



(a) Dataset 8



(b) Dataset 9

Figure D.5: Clustering of liver datasets

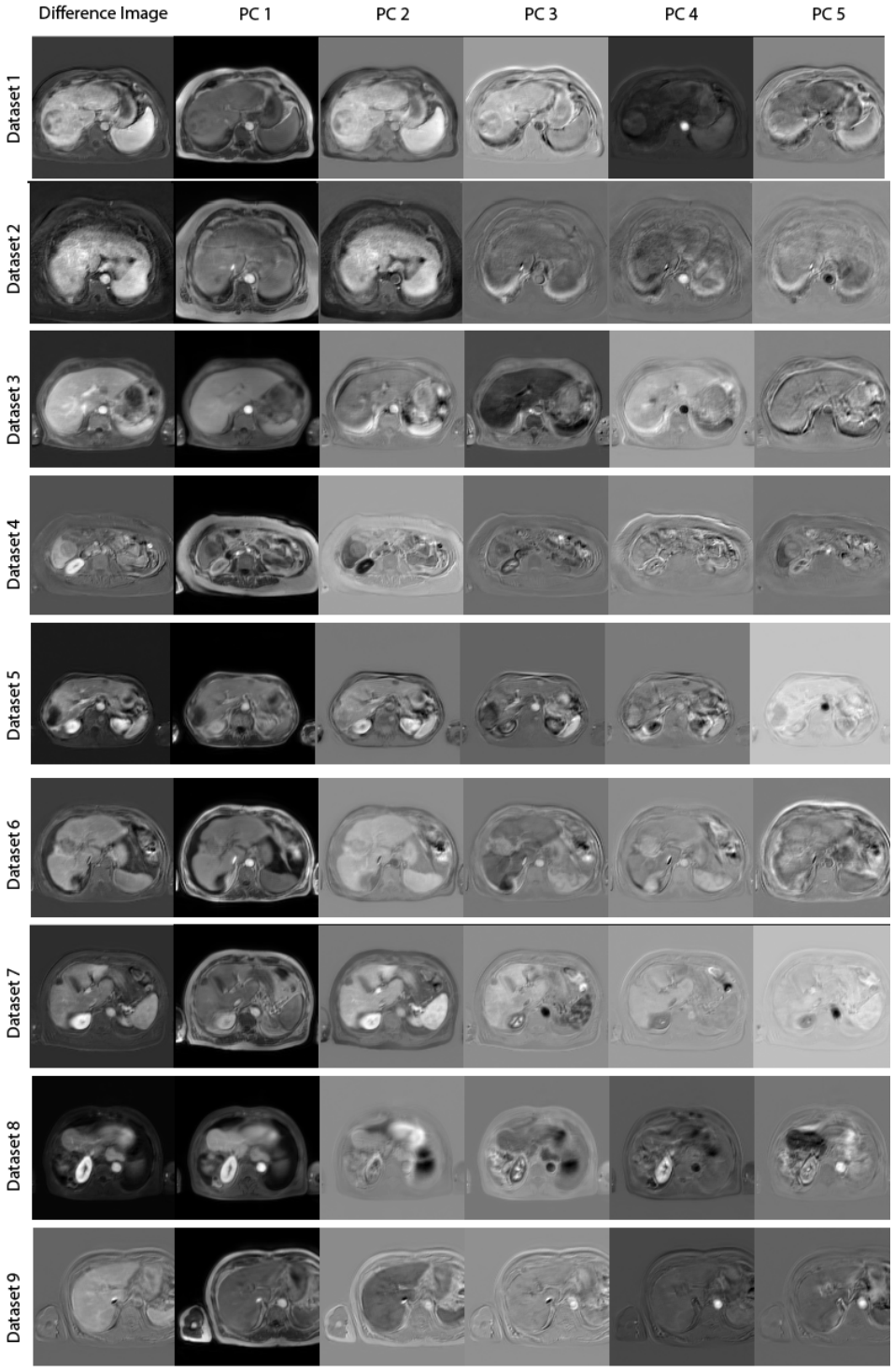


Figure D.6: Nine datasets in the sub-space of five principal components with the five largest eigenvalues.

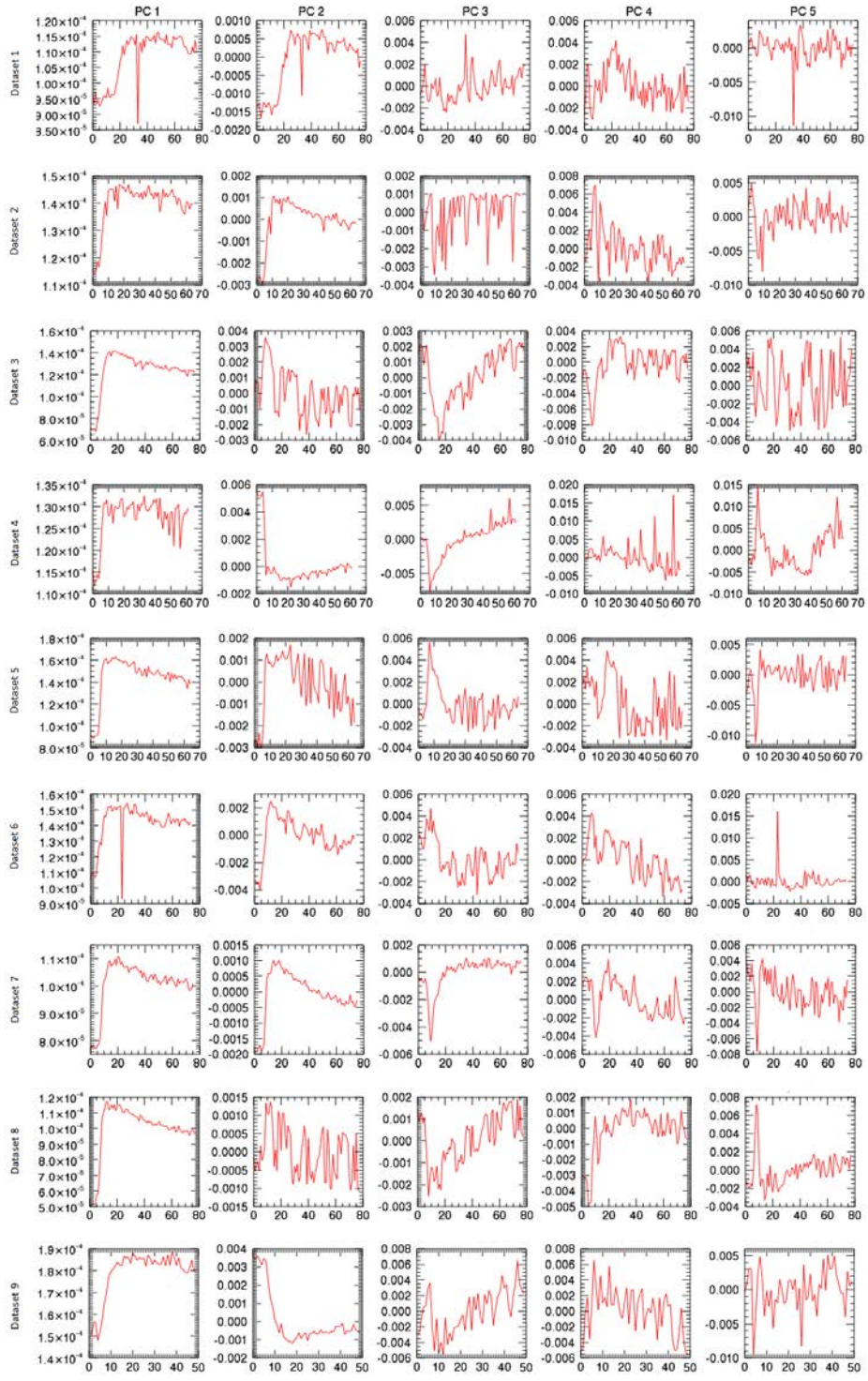


Figure D.7: Five principal components corresponding to the five largest eigenvalues of the correlation matrix.

Bibliography

H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, volume 1, pages 267–281. Springer Verlag, 1973.

F.R. Bach and M.I. Jordan. Tree-dependent component analysis. In *Uncertainty in Artificial Intelligence: Proceedings of the Eighteenth Conference), San Mateo, CA, 2002b. Morgan Kaufmann*, 2002.

P.V. Balakrishnan, M.C. Cooper, V.S. Jacob, and P.A. Lewis. A study of the classification capabilities of neural networks using unsupervised learning: A comparison withk-means clustering. *Psychometrika*, 59(4):509–525, 1994.

C. Bartolozzi, R. Lencioni, F. Donati, and D. Cioni. Abdominal MR: liver and pancreas. *European radiology*, 9(8):1496–1512, 1999.

C. Baumgartner, K. Gautsch, C. Böhm, and S. Felber. Functional cluster analysis of CT perfusion maps: a new tool for diagnosis of acute stroke? *Journal of Digital Imaging*, 18(3):219–226, 2005. ISSN 0897-1889.

R.E. Bellman. *Adaptive control processes: a guided tour*. A Rand Corporation Research Study Series. Princeton University Press, 1961.

K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is nearest neighbor meaningful? In *Database Theory*, volume 1540 of *Lecture Notes in Computer Science*, pages 217–235. Springer Berlin / Heidelberg, 1999.

C.M. Bishop. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

R. Brasch and K. Turetschek. MRI characterization of tumors and grading angiogenesis using macromolecular contrast media: status report. *European journal of radiology*, 34 (3):148–155, 2000. ISSN 0720-048X.

G. Brix, W. Semmler, R. Port, L.R. Schad, G. Layer, and W.J. Lorenz. Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. *Journal of computer assisted tomography*, 15(4):621, 1991.

G. Brix, M.L. Bahner, U. Hoffmann, A. Horvath, and W. Schreiber. Regional blood flow, capillary permeability, and compartmental volumes: measurement with dynamic ct—initial experience. *Radiology*, 210(1):269, 1999.

G. Brix, F. Kiessling, R. Lucht, S. Darai, K. Wasser, S. Delorme, and J. Griebel. Microcirculation and microvasculature in breast tumors: pharmacokinetic analysis of dynamic MR image series. *Magnetic Resonance in Medicine*, 52(2):420–429, 2004.

G. Brix, S. Zwick, F. Kiessling, and J. Griebel. Pharmacokinetic analysis of tissue microcirculation using nested models: multimodel inference and parameter identifiability. *Medical physics*, 36:2923, 2009.

G. Brix, J. Griebel, F. Kiessling, and F. Wenz. Tracer kinetic modelling of tumour angiogenesis based on dynamic contrast-enhanced CT and MRI measurements. *European journal of nuclear medicine and molecular imaging*, 37:30–51, 2010.

L.D. Buadu, J. Murakami, S. Murayama, N. Hashiguchi, S. Sakai, S. Toyoshima, K. Masuda, S. Kuroki, and S. Ohno. Patterns of peripheral enhancement in breast masses: correlation of findings on contrast medium enhanced MRI with histologic features and tumor angiogenesis. *Journal of computer assisted tomography*, 21(3):421, 1997.

F. Calamante, M. Morup, 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.

L. Caldeira, I. Silva, and J. Sanches. Automatic liver tumor diagnosis with Dynamic-Contrast Enhanced MRI. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 2256–2259. IEEE, 2008.

T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, 3(1):1–27, 1974. ISSN 0361-0926.

U. Castellani, M. Cristani, P. Marzola, V. Murino, E. Rossato, and A. Sbarbati. Cancer area characterization by non-parametric clustering. In *Workshop on Intelligent Data Analysis in bioMedicine And Pharmacology*, pages 25–30, 2006.

L.H. Cheong, T.S. Koh, and Z. Hou. An automatic approach for estimating bolus arrival time in dynamic contrast MRI using piecewise continuous regression models. *Physics in Medicine and Biology*, 48:N83, 2003.

P.L. Choyke, A.J. Dwyer, and M.V. Knopp. Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging. *Journal of Magnetic Resonance Imaging*, 17(5):509–520, 2003. ISSN 1522-2586.

D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.

A.P. Crawley, J. Poublanc, P. Ferrari, and T.P.L. Roberts. Basics of diffusion and perfusionMRI. *Applied Radiology*, 32(4):13–24, 2003.

C. Crone. The permeability of brain capillaries to non-electrolytes. *Acta Physiologica Scandinavica*, 64(4):407–417, 1965.

A. Daducci, U. Castellani, M. Cristani, P. Farace, P. Marzola, A. Sbarbati, and V. Murino. Learning approach to analyze tumour heterogeneity in DCE-MRI data during anti-cancer treatment. *Artificial Intelligence in Medicine*, pages 385–389, 2009.

B.L. Daniel, Y.F. Yen, G.H. Glover, D.M. Ikeda, R.L. Birdwell, A.M. Sawyer-Glover, J.W. Black, S.K. Plevritis, S.S. Jeffrey, and R.J. Herfkens. Breast disease: dynamic spiral MR imaging. *Radiology*, 209(2):499, 1998.

DKFZ. *Research program "Innovative Diagnosis and Therapy"*. German Cancer Research Center (DKFZ), Heidelberg, Germany, 2004.

S. Dudoit and J. Fridlyand. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome biology*, 3(7), 2002.

E. Eyal, D. Badikhi, E. Furman-Haran, F. Kelcz, K.J. Kirshenbaum, and H. Degani. Principal component analysis of breast DCE-MRI adjusted with a model-based method. *Journal of Magnetic Resonance Imaging*, 30(5):989–998, 2009.

FastICA. The FastICA package for MATLAB, 2005. URL `http://research.ics.tkk.fi/ica/fastica/`.

V.V. Fedorchuk, A.V. Arkhangelskiui, and L.S. Pontriagin. *General topology I*, volume 1. Springer, 1990.

R.A. Fisher. UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, 1963. http://archive.ics.uci.edu/ml.

K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1): 32–40, 1975.

D. Gianfelice, A. Khiat, M. Amara, A. Belblidia, and Y. Boulanger. MR imaging-guided focused ultrasound surgery of breast cancer: correlation of dynamic contrast-enhanced MRI with histopathologic findings. *Breast cancer research and treatment*, 82(2):93–101, 2003.

A.M. Glenberg and M.E. Andrzejewski. *Learning from data: An introduction to statistical reasoning*. Taylor & Francis Group, LLC, 2008.

P. Gowland, P. Mansfield, P. Bullock, M. Stehling, B. Worthington, and J. Firth. Dynamic studies of gadolinium uptake in brain tumors using inversion-recovery echo-planar imaging. *Magnetic resonance in medicine*, 26(2):241–258, 1992.

P.E. Greenwood and M.S. Nikulin. *A guide to chi-squared testing*, volume 280. Wiley-Interscience, 1996.

J. Griebel, N.A. Mayr, A. De Vries, M.V. Knopp, T. Gneiting, C. Kremser, M. Essig, H. Hawighorst, P.H. Lukas, and W.T.C. Yuh. Assessment of tumor microcirculation: a new role of dynamic contrast MR imaging. *Journal of Magnetic Resonance Imaging*, 7 (1):111–119, 1997.

J.Y. Guo and W.E. Reddick. DCE-MRI pixel-by-pixel quantitative curve pattern analysis and its application to osteosarcoma. *Journal of Magnetic Resonance Imaging*, 30(1): 177–184, 2009.

D. Hanahan and J. Folkman. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*, 86(3):353–364, 1996. ISSN 0092-8674.

J.A. Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc. New York, NY, USA, 1975. ISBN 047135645X.

R.H. Hashemi and W.G. Bradley. *MRI: the basics*. Williams & Wilkins Baltimore, Md., 2004.

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2005.

M.A. Hayat. *Cancer imaging: Lung and breast carcinomas*, volume 1, pages 83–88. Academic Press, 2007.

D.F. Hayes, T.M. Walker, B. Singh, E.S. Vitetta, J.W. Uhr, S. Gross, C. Rao, GV. Doyle, L. Terstappen, et al. Monitoring expression of HER-2 on circulating epithelial cells in patients with advanced breast cancer. *International journal of oncology*, 21(5):1111–1118, 2002.

S. Haykin. *Neural networks and learning machines*. Canada Prentice Hall, 2008.

E. Henderson, B.K. Rutt, and T.Y. Lee. Temporal sampling requirements for the tracer kinetics modeling of breast disease. *Magnetic resonance imaging*, 16(9):1057–1073, 1998.

E. Henderson, J. Sykes, D. Drost, H.J. Weinmann, B.K. Rutt, and T.Y. Lee. Simultaneous mri measurement of blood flow, blood volume, and capillary permeability in mammary tumors using two different contrast agents. *Journal of Magnetic Resonance Imaging*, 12 (6):991–1003, 2000.

J.P. Hornak. *The basics of MRI*. 2000.

G. Husmann, P. Kaatsch, A. Katalinic, J. Bertz, J. Haberland, K. Kraywinkel, and U. Wolf. Krebs in deutschland 2005/2006: Häufigkeiten und trends. *Berlin: Robert Koch-Institut*, 2010.

A. Hyvärinen, P.O. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558, 2001a.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*, volume 26. Wiley-interscience, 2001b.

A. Jackson, H. Haroon, XP Zhu, KL Li, NA Thacker, and G. Jayson. Breath-hold perfusion and permeability mapping of hepatic malignancies using magnetic resonance imaging and a first-pass leakage profile model. *NMR in Biomedicine*, 15(2):164–173, 2002.

A. Jackson, D. Buckley, and G.J.M. Parker. *Dynamic contrast-enhanced magnetic resonance imaging in oncology*. Springer Verlag, 2005.

A. Jackson, J.P.B. O'Connor, and et al. Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging. *Clinical Cancer Research*, 13(12):3449, 2007. ISSN 1078-0432.

R.K. Jain. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science*, 307(5706):58, 2005.

S.R Kannan, A. Sathya, and S. Ramathilagam. Effective fuzzy clustering techniques for segmentation of breast MRI. *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 15(3):483–491, 2011.

L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data. An Introduction to Cluster Analysis*. Wiley Interscience, New York, 1990.

S.S. Kety. The theory and applications of the exchange of inert gas at the lungs and tissues. *Pharmacological Reviews*, 3(1):1–41, 1951.

F. Kiessling, M. Jugold, E.C. Woenne, and G. Brix. Non-invasive assessment of vessel morphology and function in tumors by magnetic resonance imaging. *European Radiology*, 17(8):2136–2148, 2007.

T. Kimura, Y. Hirokawa, Y. Murakami, M. Tsujimura, Nakashima, et al. Reproducibility of organ position using voluntary breath-hold method with spirometer for extracranial stereotactic radiotherapy. *International journal of radiation oncology, biology, physics*, 60(4):1307–1313, 2004. ISSN 0360-3016.

M.V. Knopp, G. Brix, H.J. Junkermann, H.P. Sinn, et al. MR mammography with pharmacokinetic mapping for monitoring of breast cancer treatment during neoadjuvant therapy. *Magnetic resonance imaging clinics of North America*, 2(4):633, 1994.

T.S. Koh, C.H. Thng, J.T.S. Ho, P.H. Tan, H. Rumpel, and J.B.K. Khoo. Independent component analysis of dynamic contrast-enhanced magnetic resonance images of breast carcinoma: A feasibility study. *Journal of Magnetic Resonance Imaging*, 28(1):271–277, 2008.

W.J. Krzanowski and Y.T. Lai. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44(1):23–34, 1988. ISSN 0006-341X.

O.A. Kubassova, R.D. Boyle, and A. Radjenovic. Quantitative analysis of dynamic contrast-enhanced MRI datasets of the metacarpophalangeal joints. *Academic radiology*, 14(10):1189–1200, 2007.

C.K. Kuhl, P. Mielcareck, S. Klaschik, C. Leutner, E. Wardelmann, J. Gieseke, and H.H. Schild. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology*, 211(1):101, 1999.

C. Lavini, M.C. de Jonge, M.G.H. van de Sande, P.P. Tak, and Maas M. Nederveen, A.J. Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system. *Magnetic resonance imaging*, 25(5):604–612, 2007.

S.H. Lee, J.H. Kim, K.G. Kim, S.J. Park, and W.K. Moon. K-means clustering approach for kinetic pattern analysis of dynamic contrast enhancement breast MRI. 2006.

S.H. Lee, J.H. Kim, N. Cho, J.S. Park, Z. Yang, Y.S. Jung, and W.K. Moon. Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. *Medical physics*, 37:3940, 2010.

G. Leinsinger, T. Schlossbauer, M. Scherr, O. Lange, M. Reiser, and A. Wismüller. Cluster analysis of signal-intensity time course in dynamic breast MRI: does unsupervised vector quantization help to evaluate small mammographic lesions? *European radiology*, 16(5): 1138–1146, 2006. ISSN 0938-7994.

R.E.A. Lucht, M.V. Knopp, and G. Brix. Classification of signal-time curves from dynamic MR mammography by neural networks. *Magnetic resonance imaging*, 19(1):51–57, 2001.

H. Mehrabian, I. Pang, C. Chandrana, R. Chopra, and A.L. Martel. Automatic mask generation using independent component analysis in dynamic contrast enhanced-MRI. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1657–1661. IEEE, 2011.

A. Melbourne, D. Atkinson, MJ White, D. Collins, M. Leach, and D. Hawkes. Registration of dynamic contrast-enhanced MRI using a progressive principal component registration (PPCR). *Physics in medicine and biology*, 52:5147, 2007.

A. Melbourne, D. Atkinson, and D. Hawkes. Influence of Organ Motion and Contrast Enhancement on Image Registration. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 948–955, 2008.

M. Mescam, M. Kretowski, and et al. Multiscale Model of Liver DCE-MRI Towards a Better Understanding of Tumor Complexity. *Medical Imaging, IEEE Transactions on*, 29(3):699–707, 2010. ISSN 0278-0062.

A. Meyer-Baese, O. Lange, T. Schlossbauer, and A. Wismüller. Computer-aided diagnosis and visualization based on clustering and independent component analysis for breast MRI. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 3000–3003. IEEE, 2008.

A. Meyer-Baese, T. Schlossbauer, O. Lange, and A. Wismueller. Small lesions evaluation based on unsupervised cluster analysis of signal-intensity time courses in dynamic breast MRI. *Journal of Biomedical Imaging*, 2009:31, 2009.

M. Mohajer, G. Brix, and K.H. Englmeier. A novel and fast method for cluster analysis of DCE-MR image series of breast tumors. In *Proceedings of SPIE*, volume 7626, 2010.

M. Mohajer, K.H. Englmeier, and V.J. Schmid. A comparison of gap statistic definitions with and without logarithm function. *Arxiv preprint arXiv:1103.4767*, 2011.

S. Mussurakis, D.L. Buckley, P.J. Drew, J.N. Fox, P.J. Carleton, L.W. Turnbull, and A. Horsman. Dynamic MR imaging of the breast combined with analysis of contrast

agent kinetics in the differentiation of primary breast tumours. *Clinical radiology*, 52(7): 516–526, 1997.

T.W. Nattkemper and A. Wismüller. Tumor feature visualization with unsupervised learning. *Medical Image Analysis*, 9(4):344–351, 2005.

T.W. Nattkemper, B. Arnrich, O. Lichte, W. Timm, A. Degenhard, L. Pointon, C. Hayes, and M.O. Leach. Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. *Artificial Intelligence in Medicine*, 34 (2):129–139, 2005. ISSN 0933-3657.

M. Neeman. Preclinical MRI experience in imaging angiogenesis. *Cancer and Metastasis Reviews*, 19(1):39–43, 2000. ISSN 0167-7659.

P.V. Pandharipande, G.A. Krinsky, H. Rusinek, and V.S. Lee. Perfusion imaging of the liver: current challenges and future goals. *Radiology*, 234(3):661, 2005.

E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

W.H. Pressa, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, The Edinburgh Building, Cambridge, UK, 2007.

Protocol. Breast mass protocol, 1999. URL `http://www.mrprotocols.com/oldsite/MRI/Chest/breast_mass_mri.htm`.

Y. Qiuxia, T. Liangrui, D. Wenting, and S. Yi. Image edge detecting based on gap statistic model and relative entropy. In *FSKD (5)*, pages 384–387. IEEE Computer Society, 2009. URL `http://dblp.uni-trier.de/db/conf/fskd/fskd2009-5.html#YangTDS09`.

Radiology. Institute for radiology, 2011. Klinikum rechts der Isar, Technische Universität München, Ismaninger Str. 22, D-81675 Munich.

E.M. Renkin. Transport of potassium-42 from blood to tissue in isolated mammalian skeletal muscles. *American Journal of Physiology–Legacy Content*, 197(6):1205, 1959.

T. Rohlfing, C.R. Maurer, W.G. O'Dell, and J. Zhong. Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Medical Physics*, 31:427, 2004.

E.J. Rummeny and G. Marchal. Liver imaging. *Acta Radiologica*, 38(5):626–630, 1997.

A. Saalbach, O. Lange, T. Nattkemper, and A. Meyer-Baese. On the application of (topographic) independent and tree-dependent component analysis for the examination of DCE-MRI data. *Biomedical signal processing and control*, 4(3):247–253, 2009.

L.A. Santalo. *Integral geometry and geometric probability / Luis A. Santalo ; with a foreword by Mark Kac*, page 49. Addison-Wesley Pub. Co., Advanced Book Program, Reading, Mass. :, 1976. ISBN 0201135000.

T. Schlossbauer, G. Leinsinger, A. Wismüller, O. Lange, M. Scherr, A. Meyer-Baese, and M. Reiser. Classification of small contrast enhancing breast lesions in dynamic magnetic resonance imaging using a combination of morphological criteria and dynamic analysis based on unsupervised vector-quantization. *Investigative radiology*, 43(1):56, 2008.

A.J. Scott and M.J. Symons. Clustering methods based on likelihood ratio criteria. *Biometrics*, 27(2):387–397, 1971. ISSN 0006-341X.

J. Shlens. A tutorial on principal component analysis. *Measurement*, 51, 2005.

P. Somervuo and T. Kohonen. Self-organizing maps and learning vector quantization for feature sequences. *Neural Processing Letters*, 10(2):151–159, 1999.

J.V. Stone. *Independent component analysis: a tutorial introduction*. The MIT Press, 2004.

M.J. Stoutjesdijk, J. Veltman, H. Huisman, N. Karssemeijer, J.O. Barentsz, J.G. Blickman, and C. Boetes. Automated analysis of contrast enhancement in breast MRI lesions using mean shift clustering for ROI selection. *Journal of Magnetic Resonance Imaging*, 26(3): 606–614, 2007.

C.A. Sugar and G.M. James. Finding the Number of Clusters in a Data Set - An Information Theoretic Approach. *Journal of the American Statistical Association*, 98(463): 750–763, 2003.

C.H. Thng, S. Hartono, T. San Koh, and D.M. Koh. An Introduction to MR Perfusion Imaging of the Liver. *Proceedings of Singapore Healthcare*, 19(1):26, 2010.

H.S. Thomsen, S.K. Morcos, and P. Dawson. Is there a causal relation between the administration of gadolinium based contrast media and the development of nephrogenic systemic fibrosis (NSF)? *Clinical radiology*, 61(11):905–906, 2006.

R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001. URL `http://ideas.repec.org/a/bla/jorssb/v63y2001i2p411-423.html`.

P.S. Tofts. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *Journal of Magnetic Resonance Imaging*, 7(1):91–101, 1997.

P.S. Tofts and A.G. Kermode. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnetic Resonance in Medicine*, 17(2):357–367, 1991.

P.S. Tofts, G. Brix, D.L. Buckley, J.L. Evelhoch, E. Henderson, M.V. Knopp, H.B.W. Larsson, T.Y. Lee, N.A. Mayr, G.J.M. Parker, et al. Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusible tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10(3):223–232, 1999.

J.J. Totman, R.L. O'gorman, P.A. Kane, and J.B. Karani. Comparison of the hepatic perfusion index measured with gadolinium-enhanced volumetric MRI in controls and in patients with colorectal cancer. *British Journal of Radiology*, 78(926):105, 2005.

Y. Tsushima, M.J.K. Blomley, H. Yokoyama, S. Kusano, and K. Endo. Does the presence of distant and local malignancy alter parenchymal perfusion in apparently disease-free areas of the liver? *Digestive diseases and sciences*, 46(10):2113–2119, 2001. ISSN 0163-2116.

T. Twellmann, A. Saalbach, O. Gerstung, M.O. Leach, T.W. Nattkemper, et al. Image fusion for dynamic contrast enhanced magnetic resonance imaging. *Biomedical engineering online*, 3(1):35, 2004.

T. Twellmann, A. Meyer-Baese, O. Lange, S. Foo, and T.W. Nattkemper. Model-free visualization of suspicious lesions in breast MRI based on supervised and unsupervised learning. *Engineering applications of artificial intelligence*, 21(2):129–140, 2008.

B.E. Van Beers, B. Gallez, and J. Pringot. Contrast-enhanced MR imaging of the liver. *Radiology*, 203(2):297, 1997.

C. Varini, A. Degenhard, and T.W. Nattkemper. Visual exploratory analysis of DCE-MRI data in breast cancer by dimensional data reduction: A comparative study. *Biomedical Signal Processing and Control*, 1(1):56–63, 2006. ISSN 1746-8094.

S.S. Vasanawala, Y. Iwadate, D.G. Church, R.J. Herfkens, and A.C. Brau. Navigated abdominal T1-W MRI permits free-breathing image acquisition with less motion artifact. *Pediatric radiology*, 40(3):340–344, 2010.

H.Z. Wang, S.J. Riederer, and J.N. Lee. Optimizing the precision in T1 relaxation estimation using limited flip angles. *Magnetic resonance in medicine*, 5(5):399–416, 1987.

M.C. Wendl and S.P. Yang. Gap statistics for whole genome shotgun DNA sequencing projects. *Bioinformatics*, 20(10):1527–1534, 2004. ISSN 1367-4803.

C. Westbrook, C.K. Roth, and J. Talbot. *MRI in Practice*. Wiley-Blackwell, 2005.

B. Whitcher and V. J. Schmid. Quantitative Analysis of Dynamic Contrast-Enhanced and Diffusion-Weighted Magnetic Resonance Imaging for Oncology in R. *Journal of Statistical Software*, 44(5):1–29, 10 2011. URL `http://www.jstatsoft.org/v44/i05`.

A. Wismuller, A. Meyer-Baese, O. Lange, MF. Reiser, and G. Leinsinger. Cluster analysis of dynamic cerebral contrast-enhanced perfusion mri time-series. *Medical Imaging, IEEE Transactions on*, 25(1):62–73, 2006.

A. Wismüller, A. Meyer-Bäse, O. Lange, T. Schlossbauer, M. Kallergi, M.F. Reiser, and G. Leinsinger. Segmentation and classification of dynamic breast magnetic resonance image data. *Journal of Electronic Imaging*, 15:013020, 2006.

W.H. Wolberg, W.N. Street, and O.L. Mangasarian. UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, 1993. http://archive.ics.uci.edu/ml.

M. Yan and K. Ye. Determining the number of clusters using the weighted gap statistic. *Biometrics*, 63(4):1031–1037, 2007. ISSN 1541-0420.

H. Yin. Learning nonlinear principal manifolds by self-organising maps. *Principal Manifolds for Data Visualization and Dimension Reduction*, pages 68–95, 2008.

Z. Yin, X. Zhou, C. Bakal, F. Li, Y. Sun, N. Perrimon, and S.T.C. Wong. Using iterative cluster merging with improved gap statistics to perform online phenotype discovery in the context of high-throughput RNAi screens. *BMC bioinformatics*, 9(1):264, 2008. ISSN 1471-2105.

S.S. Yoo, B. Gil Choi, J.Y. Han, and H. Hee Kim. Independent component analysis for the examination of dynamic contrast-enhanced breast magnetic resonance imaging data: preliminary study. *Investigative radiology*, 37(12):647, 2002.

Y. Zheng, S. Baloch, S. Englander, M. Schnall, and D. Shen. Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images. *MICCAI 2007*, pages 393–401, 2007.

Z. Zheng-Jun and Z. Yao-Qin. Estimating the image segmentation number via the entropy gap statistic. In *ICIC '09: Proceedings of the 2009 Second International Conference on Information and Computing Science*, pages 14–16, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3634-7.

Acknowledgments

I would like to gratefully and sincerely thank my advisors Prof. Dr. Englmeier and Prof. Dr. Schmid who made this dissertation possible. I have been fortunate to have advisors who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. Thank you for your patience and support and most importantly, the friendship.

I would like to thank Prof. Dr. Brix for his helpful advices and guidance through my thesis.

I would like to thank Prof. Dr. Rummeny, Dr. Engels, Dr. Braren, and Dr. Noel for their support and assistance and, above all, by providing me with DCE-MRI datasets.

I would also like to thank all of the members of the IBMI group at Helmholtz center, especially Susanne Stern, Silvia Weinzierl, Rui Ma, Karin Schäfer, Christiane Ogorek and Zsuzsanna Öszi for always be there for me and their supports.

My special thanks go to my cousin Mahzad Mohajer for her patiently reading this work. I would also like to thank Wilhelm Huber for his proof read and helpful comments.

Finally, and most importantly, I would like to thank my husband Hamid for his support, encouragement, quiet patience and unwavering love and at the same time for patiently reading this work, corrections and helpful comments. I thank my parents, who taught me the curiosity and love of learning. Without their support I would never be here where I am. Thank you so much.

This study was granted by the German 'Competence Alliance on Radiation Research' (BMBF 02NUK008H).