Myocardial perfusion modelling: MR image processing and statistical inference

Yalei Yang

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Mathematics & Statistics
College of Science and Engineering
University of Glasgow

April 2022

Abstract

Coronary heart disease (CHD) is a major source of human mortality worldwide. Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is widely used as a non-invasive method to assess myocardial perfusion, which can be used to diagnose and detect myocardial ischaemia. The main aim of this thesis is to develop novel robust statistical classification models to detect the precise location of the ischaemia within the myocardium using DCE-MRI myocardial perfusion data. Firstly, the myocardial blood flow (MBF), a critical parameter adopted to quantify the degree of ischaemia within the myocardium, is estimated using semi-quantitative or quantitative methods. The MBF map is used as the input training data for the classification methods. Secondly, a Gaussian mixture model and its modifications, i.e. a mixture model incorporating spatial constraints, are applied to classify the myocardial tissues based on the MRI data. Markov random field priors are introduced to represent the spatial prior information. Thirdly, hierarchical Bayesian classification models are developed to combine the physiological model or the MRI meta data with spatial or spatio-temporal prior information to classify the myocardial tissues based on the MRI data. Furthermore, a polar projection method is developed to map the myocardium image to an annulus. The thesis concludes with an outlook on future work on how the methods developed in this PhD project can be extended to longitudinal analysis of myocardial perfusion DCE-MRI.

Contents

List of Tables

List of Figures

Acknowledgements

My PhD journey has come to an end. It is a challenging but exciting part of my whole life. Without help from several people and organizations, I could not have come here. So, I would like to express my sincerest thanks to them.

Firstly, I would like to thank my PhD supervisors Prof. Dirk Husmeier, Dr Aleksandra Radjenovic and Dr Hao Gao. You are the people who open a new door for me. Dear Dirk, you taught me so many things which are not only about knowledge but also the ways and the attitudes toward unknown things. I know you have spent a lot of time on my project and made me a better person. Dear Sasha, your patience and kindness always make me feel encouraged. Your point of view always inspires me. Dear Hao, you always stand out when I have some difficulties and encourage me to keep moving and overcome the challenges. I want to express the most sincere respect to all of you, and it is my honour to be your student.

Secondly, I would like to thank the School of Mathematics and Statistics for waiving my tuition fees and GlaxoSmithKline plc. for granting funds to me. Furthermore, the group of SoftMech provides me with an opportunity of communicating with other colleagues. I indeed learned a lot from the open environment. The people such as Prof. Xiaoyu Luo, Prof. Adrian Bowman, Prof. Duncan Lee and Dr Mu Niu also helped me a lot during the period of my PhD program, and I appreciate it. I also like to thank my colleagues and friends Debao Guan, Mihaela Paun, Alan Lazarus and Yingjie Wang for their help in my study and life.

Lastly and most importantly, I would like to thank my family. Everything I own today has credits to my family. My mother, Guangtao Shi, is the best mother in the world. Dear Mama, I was born as a part of you, and I will try to pass on all the beautiful qualities you gave me. Your intelligence and hard work will always inspire me. My father, Jinyun Yang, is the best father in the world. Dear Baba, your optimism and humour have made me grow as a strong and tough person. Your insight has saved me from detours. You made me have a happy life. My grandmother, Liefeng Song, just passed away a few months ago. I miss her deeply. I hope she is well in heaven. I would also like to thank my girlfriend, Qing li. Dear Qing, you have devoted so much time to for caring me, accompanying me and supporting me. Without your help and support, I cannot complete my PhD study. You make my life more joyful, and I will never forget the happy moments we spent together.

Declaration

I, Yalei Yang, declare that the thesis entitled "Myocardial perfusion modelling: MRI image processing and statistical inference" and the work proposed in it have been carried out by myself under the supervision of Prof. Dirk Husmeier, Dr Aleksandra Radjenovic and Dr Hao Gao. I confirm that:

- The work was done wholly while a candidate for a research degree at this University.
- The citation is always given where I quote the references from others.

Submitted Papers

The following papers have been accepted for publication:

Quantification of Myocardial Perfusion Lesions using Spatially Variant Finite Mixture Modelling of DCE-MRI. **Yalei Yang**, Hao Gao, Colin Berry, Aleksandra Radjenovic and Dirk Husmeier. Proceedings of the International Conference on Statistics: Theory and Applications (2019). The content about this paper can be found in Section 5.3.

Classification of Myocardial Blood Flow Based on Dynamic Contrast Enhanced Magnetic Resonance Imaging Using Hierarchical Bayesian Models. **Yalei Yang**, Hao Gao, Colin Berry, David Carrick, Aleksandra Radjenovic and Dirk Husmeier. Journal of the Royal Statistical Society: Series C (2022). The content about this paper can be found in Chapter 6.

Myocardial perfusion Dynamic Contrast Enhanced Magnetic Resonance Imaging classification using Markov random field constrained Gaussian mixture model. **Yalei Yang**, Hao Gao, Colin Berry, Aleksandra Radjenovic and Dirk Husmeier. Accepted by the International Conference on Statistics: Theory and Applications (2022). The content about this paper can be found in Section 5.4.

The following paper has been submitted for review:

Classification of Myocardial Perfusion Dynamic Contrast Enhanced Magnetic Resonance Imaging using Hierarchical Bayesian Model with Spatio-temporal Markov Random Fields. **Yalei Yang**, Hao Gao, Colin Berry, Aleksandra Radjenovic and Dirk Husmeier. Submitted to Artificial Intelligence in Medicine. The content about this manuscript can be found in Chapter 7.

Abbreviations

AHA . . . . . American Heart Association

AIC . . . . . Akaike information criterion

AIF . . . . . arterial input function

AMIS . . . . . adaptive multiple importance sampling

ARMA . . . . . autoregressive–moving-average

BIC . . . . . Bayesian information criterion

CER . . . . . contrast enhancement ratio

CHD . . . . . coronary heart disease

CFR . . . . . coronary flow reserve

CNR . . . . . concentration-to-noise ratio

DAG . . . . . directed acyclic graph

CABG . . . . . coronary artery bypass graft

DCE-MRI . . . . . dynamic contrast-enhanced magnetic resonance imaging

DIC . . . . . deviance information criterion

ECG . . . . . electrocardiogram

EM . . . . . expectation-maximization

ESS . . . . . effective sample size

GBCA . . . . . gadolinium-based contrast agent

GMM . . . . . Gaussian mixture model

GMM-MRF . . . . . Markov random field constrained Gaussian mixture model

HBM . . . . . hierarchical Bayesian model

ICM . . . . . iterated conditional modes

LAD . . . . . left anterior descending

LCX . . . . . left circumflex

LGE . . . . . late gadolinium enhanced

LV . . . . . left ventricle

MAP . . . . . maximum a posteriori probability

MBF . . . . . myocardial blood flow

ME . . . . . maximum enhancement

MCMC . . . . . Markov chain Monte Carlo

ML . . . . . maximum likelihood

MLE . . . . . maximum likelihood estimation

MPR . . . . . myocardial perfusion reserve

MPRI . . . . . myocardial perfusion reserve index

MRF . . . . . Markov random field

MSE . . . . . mean squared error

PDE . . . . . partial differential equation

PET . . . . . positron emission tomography

PINN . . . . . physics-informed neural network

PTCA . . . . . percutaneous transluminal coronary angioplasty

PSI . . . . . peak signal intensity

PSRF . . . . . potential scale reduction factor

RCA . . . . . right coronary artery

RF . . . . . radio frequency

ROI . . . . . region of interest

SI . . . . . signal intensity

SMC . . . . . sequential Monte Carlo

SNR . . . . . signal-to-noise ratio

SSE . . . . . sum of squared error

SVFMM . . . . . spatially variant finite mixture model

TIC . . . . . time signal intensity curve

WAIC . . . . . Watanabe-Akaike information criterion

Chapter 1

Introduction

Coronary heart disease (CHD) is one of the most common sources of human death worldwide, and 382,820 deaths in the US were caused by CHD in 2020 [7]. Inadequate blood supply to the heart muscle will cause coronary heart disease. Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is widely used as a non-invasive method to assess the differences in myocardial perfusion (microvascular or capillary blood flow) within the myocardium. The signal changes within the myocardium during the first pass of a contrast agent bolus (bolus of radio-opaque contrast media, gadolinium based) reflect the degree of the blood flow perfusion. Therefore, myocardial ischaemia, i.e. inadequate blood supply to the myocardium, can be diagnosed or detected using the first pass of the myocardial perfusion DCE-MRI. There are many methods aiming to analyze the myocardial perfusion DCE-MRI scans such as semi-quantitative and quantitative methods [8]. Most of the methods focus on quantifying the myocardial blood flow (MBF) and use the MBF estimates to detect the hypo-perfusion (ischaemia), i.e. inadequate perfusion of blood flow.

The main aim of the study in this thesis is to develop robust classification methods based on either original DCE-MRI data or estimated parameters, e.g. the MBF values. The statistical classification methods used in this study can be categorised into two types, i.e. finite mixture model and the hierarchical Bayesian model. The finite mixture model describes different classes (or groups) using mixture distributions. For example, for a set of data that contains many classes, the finite mixture model assumes different classes have different distributions, and it aims to classify different data points into their corresponding classes. In this study, the Gaussian mixture model [9], and its modifications, i.e. spatially variant finite mixture model [10] and Markov random field constrained Gaussian mixture model [11], are applied to classify the myocardial pixels within the myocardium images. The hierarchical Bayesian model is a statistical model that has multiple layers. The parameters located in different layers are connected based on dependence relationships. In this way, the perfusion parameters, e.g. the MBF and their corresponding classes, can exist in the same model with dependence relationships. In this study, MBF estimates based and data-based hierarchical Bayesian models are developed to classify the

myocardial pixels using myocardial perfusion DCE-MRI.

The study in this thesis also includes a description of a polar projection method that aims to lay the foundations for longitudinal analysis. This method can project any contoured myocardial image to an annulus (large circle denotes epicardium and small circle denotes endocardium). The projected annuluses have the same shape and size so that the images from different datasets can be compared pixelwise.

The structure of the thesis is listed as follows:

- Chapter 2 presents the statistical tools used in this thesis.
- Chapter 3 provides the background information on heart physiology and myocardial perfusion DCE-MRI.
- Chapter 4 presents semi-quantitative and quantitative methods for the myocardial blood flow.
- Chapter 5 introduces the applications of mixture models to myocardial perfusion DCE-MRI.
- Chapter 6 provides a Fermi model based hierarchical Bayesian classification model using myocardial perfusion DCE-MRI.
- Chapter 7 introduces a data-driven hierarchical Bayesian classification model based on myocardial perfusion DCE-MRI.
- Chapter 8 presents a polar projection method that projects any myocardial image to an annulus.
- Chapter 9 provides a summary of the thesis and the potential future work.

Chapter 2

Statistical Background

This chapter aims to outline different statistical methodologies from the literature that will be applied in the study carried out in this thesis.

2.1 Bayesian statistics

Bayesian statistics is a theory based on the Bayesian interpretation of probability. Specifically, in the Bayesian framework, the prior knowledge, e.g. previous experiment, affects the probability of an event. The Bayesian statistics makes use of prior probabilities over parameters to infer the probability. On the other hand, for the frequentist probability, the probability of an event is the limit of its relative frequency in many experiments [12]. The Bayesian statistics is named after Thomas Bayes, who introduced the Bayes' theorem in 1763.

2.1.1 Bayes' theorem

In statistics, Bayes' theorem is used to describe the probability of an event based on prior knowledge related to that event. The mathematical statement of Bayes' theorem is:

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A})P(\mathrm{A})}{P(\mathrm{B})} \tag{2.1}$$

where A and B are events and $P(\mathrm{B}) \neq 0$.

In the Bayesian framework, Bayes' theorem is used to update the probability of an event when more inputs of knowledge or information are available. Specifically, for an unknown parameter θ conditional on observed data y, its posterior distribution can be expressed as:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \tag{2.2}$$

where $P(\theta)$ is called the prior distribution of θ and $P(y)$ is named the marginal likelihood. The

posterior distribution $P(\theta|y)$ is the distribution of θ given y, i.e. y is known. The prior distribution $P(\theta)$ is the distribution of the parameter θ before any data has been observed. Sometimes, the prior distribution is hard to choose because no prior information about the parameter θ is known. Therefore, some non-informative prior distributions, e.g. Jeffreys prior [13], are chosen. The marginal likelihood is the distribution of the observed data y marginalised over the parameter θ. According to the definition of marginal distribution, $P(y)$ can be expressed as:

$$P(y) = \int P(\theta, y) d\theta = \int P(\theta) P(y|\theta) d\theta. \tag{2.3}$$

θ is assumed to be continuous here. If θ is discrete, the integrals in equation (2.3) will be changed to a summation. Since $P(y)$ is the factor that is independent on the parameter θ, equation (2.2) can be expressed as a form omitting $p(y)$:

$$P(\theta|y) \propto P(y|\theta) P(\theta). \tag{2.4}$$

2.1.2 Prior distribution

In Bayesian statistical inference, a prior distribution of a parameter is the probability distribution that represents the prior information before some added information, e.g. data, observations, is taken into account. The parameters of the prior distributions are named hyperparameters. For example, if a Gaussian distribution with mean μ and variance σ^2 is the prior distribution θ, then both μ and σ^2 are hyperparameters of the prior distribution.

There are two ways to choose the prior distribution of a specific parameter, informative priors and uninformative priors. If there are clear clues about the value of a parameter, informative priors can be given based on the clues. For instance, if a statistical model is established to predict the share price θ of a stock in the next year, then the average share price of this stock μ in this year could be used to be the mean of a Gaussian distribution. This Gaussian distribution can be chosen as the prior distribution of θ, and μ is the hyperparameter of the prior distribution. On the other hand, if the prior information of a specific parameter is unknown, then uninformative priors should be chosen. Meanwhile, if the parameter range is clear, e.g. $\theta > 0$ or $a > \theta > b$, the choice of the prior distribution should be limited by the support of such distribution. For example, if the parameter θ has upper bound b and lower bound a, the prior distribution of θ can not be a Gaussian or Gamma distribution. A uniform distribution, truncated normal distribution or 4-parameter beta distribution, on the other hand, could be chosen under this circumstance.

There is a concept about probability distributions called conjugate distributions. In the Bayesian framework, if the posterior distribution $P(\theta|y)$ and the prior distribution $P(\theta)$ are in the same probability distribution family, $P(\theta|y)$ and $P(\theta)$ are named conjugate distributions and $P(\theta)$ is named conjugate prior for the likelihood function $P(y|\theta)$. The conjugate prior provides a choice of prior distribution to give a closed-form expression for the posterior distribution. If the

prior distribution is not conjugate to the likelihood function, the mathematical expression of the posterior distribution might be analytically intractable. Numerical methods for the approximations of the posterior distribution might be introduced and the computational complexity would increase. Table D.2 in Appendix D shows some likelihoods that have common probability distributions and their corresponding conjugate prior distributions.

2.1.3 Posterior predictive distribution

Predictive inference is the method to make inferences about an unknown observable quantity. The marginal distribution of the unknown data y can be found in equation (2.3). $P(y)$ is also named as the prior predictive distribution. Let $\tilde{y}$ denote the unknown observable quantity and y denote the observation. The posterior predictive distribution of $\tilde{y}$ is:

$$\begin{aligned} P(\tilde{y}|y) &= \int P(\tilde{y},\theta|y)d\theta \\ &= \int P(\tilde{y}|\theta,y)P(\theta|y)d\theta. \end{aligned} \tag{2.5}$$

Generally, the unknown observable quantity $\tilde{y}$ is assumed to be conditionally independent from the observations y. Therefore, equation (2.5) can be simplified to

$$P(\tilde{y}|y) = \int P(\tilde{y}|\theta)P(\theta|y)d\theta. \tag{2.6}$$

2.1.4 Maximum a posteriori

A maximum a posteriori (MAP) probability estimate is used to estimate unknown parameters. This estimate equals the mode of the posterior distribution of the unknown parameter. The MAP mathematically corresponds to a penalised maximum likelihood (ML) estimation. Therefore, MAP estimation can be seen as a regularization of the ML estimation.

Let θ denote the unknown parameter given observations y. Then, the likelihood function of y is $P(y|\theta)$. The general ML estimate is:

$$\hat{\theta}_{\mathrm{MLE}}(y) = \arg\max_{\theta} P(y|\theta). \tag{2.7}$$

Now, given the prior distribution $P(\theta)$, the posterior distribution $P(\theta|y)$ is

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int P(\theta)P(y|\theta)d\theta}. \tag{2.8}$$

The MAP estimates θ by finding the value of θ that maximizes the posterior distribution of θ.

This can be mathematically expressed by:

$$\hat{\theta}_{\text{MAP}}(y) = \arg\max_{\theta} P(\theta|y) = \arg\max_{\theta} \frac{P(y|\theta)P(\theta)}{\int P(\theta)P(y|\theta)d\theta}. \tag{2.9}$$

There are many ways to obtain the MAP estimate. If the mode of the posterior distribution can be analytically derived, e.g. the mode of a Gaussian distribution is its mean μ, then the MAP estimate can be given as the closed form of the mode. If the MAP solution is intractable, numerical optimization methods, e.g. conjugate gradient method [14], or Newton's method [15], can be applied to search the local minima if first or second derivatives of the posterior distribution are evaluated analytically or numerically. Moreover, simulated annealing methods can also be used to obtain the global minimum [16]. In this study, an algorithm named iterated conditional modes (ICM) has been applied to obtain the MAP given Markov random field priors [17]. The details of the ICM algorithm can be found in Sections 2.2.4 and 5.4.

The limitation of the MAP estimate is that it is a point estimate of the mode of the posterior distribution. It is an optimization method, whereas Bayesian methods usually concentrate on drawing samples for the posterior distribution. In other words, the Bayesian methods regard a parameter to be a random variable, not a fixed value. So, not the point estimate but the specific distribution of a parameter is interesting in the Bayesian framework.

2.2 Hierarchical Bayesian models and graphical models

2.2.1 Introduction

For many statistical applications, multiple parameters are considered related or connected. Therefore, the joint probability of these parameters should reveal their dependencies. For example, there is a study related to the average age of people (the specific study can be found in [12] (Chapter 5)). A person i from city j has age $y_{i,j}$. So many people $i = 1,2,..,n$ from city j have age $\{y_{i,j}|i = 1,2,...,n\}$ (the set of $y_{i,j}$ given all $i = 1,2,...,n$), and there are $j = 1,...,J$ cities. Let θ_j denote the average age of people in city j. It is reasonable to assume the average age of people in all cities $\{1,2,...,J\}$, $\{\theta_j|j = 1,2,...,J\}$, follows a certain distribution controlled by hyperparameters ϕ. In the Bayesian framework, θ_j can therefore be considered as a sample of the prior distribution $P(\boldsymbol{\theta}|\phi)$ which is controlled by hyperparmeters ϕ. This hierarchical structure can be expressed as a directed acyclic graph (DAG) (see Figure 2.1). A nice feature of the hierarchical modelling is that the observations $\{y_{i,j}|i = 1,2,...,n\}$ can quantify θ_j though θ_j is not observed.

Some data is featured to have a "hierarchical structure" during the data collection. For the hierarchical data, the simple non-hierarchical models are inappropriate for the data in practice [12] (Chapter 5). For those simple non-hierarchical models, if the number of parameters is

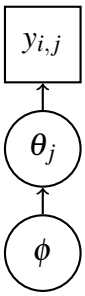


Figure 2.1: This DAG shows the dependencies of multiple parameters. The node $y_{i,j}$ denotes the age of person i in city j. The node θ_j denotes the average age of people in city j. The node ϕ denotes the hyperparmeters of $P(\theta|\phi)$.

small, they cannot fit large datasets accurately. On the other side, if the number of parameters is big, they tend to "overfit" the data. However, the hierarchical models can use large numbers of parameters and accurately fit the data. Meanwhile, a prior distribution can structure the dependencies between parameters to avoid the "overfit" issue [12] (Chapter 5).

2.2.2 Statistical inference of hierarchical Bayesian models

For the hierarchical Bayesian model (HBM) example in Figure 2.1, an important property, exchangeability, holds. Specifically, the quantity θ_j only depends on the data $\{y_{i,j}|i = 1,2,...,n\}$, and the parameters $\{\theta_j|j = 1,2,...,J\}$ have no ordering or grouping properties. The property "symmetry" can be assumed for the parameters in their prior distribution. Let $\{\pi_1,\pi_2,...,\pi_J\}$ denote any permutation of $\{1,2,...,J\}$. The "symmetry" property can be mathematically expressed:

$$P(\theta_1,\theta_2,...,\theta_J) = P(\theta_{\pi_1},\theta_{\pi_2},...,\theta_{\pi_J}). \tag{2.10}$$

Therefore, let $\boldsymbol{\theta} = \{\theta_1,\theta_2,...,\theta_J\}$, the simplest form of an exchangeable distribution has the parameter θ_j to be an independent sample from a prior (population) distribution:

$$P(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} P(\theta_j|\phi). \tag{2.11}$$

A critical feature of an HBM is that an HBM contains one or more unobserved parameters (latent variable). For the HBM in Figure 2.1, θ_j is a latent variable. The quantities $\boldsymbol{\theta}$ have their own prior distribution $P(\phi)$. The joint prior distribution is:

$$P(\boldsymbol{\theta},\phi) = P(\phi)P(\boldsymbol{\theta}|\phi). \tag{2.12}$$

Given $\boldsymbol{y} = \{y_{i,j}|i = 1,2,...,n; j = 1,2,...,J\}$, the joint posterior distribution is

$$P(\boldsymbol{\theta},\phi|\boldsymbol{y}) \propto P(\boldsymbol{\theta},\phi)P(\boldsymbol{y}|\boldsymbol{\theta},\phi). \tag{2.13}$$

Since the distribution of the data $P(\boldsymbol{y}|\boldsymbol{\theta},\phi)$ only depends on $\boldsymbol{\theta}$, the simplification

$$P(\boldsymbol{y}|\boldsymbol{\theta},\phi) = P(\boldsymbol{y}|\boldsymbol{\theta}) \tag{2.14}$$

holds. The term $P(\phi)$ is a part of the joint posterior distribution. This term, called hyperprior distribution, is the prior distribution of the hyperparameter ϕ. Theoretically, the hyperparameters of this hyperprior distribution could be either constants or hyperparameters [12].

There are three quantities that are analytically interesting, and the example in Figure 2.1 is used to illustrate the derivation of these three quantities. The first quantity is the joint posterior distribution which can be written as an unnormalised form:

$$P(\phi,\boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\phi)P(\phi). \tag{2.15}$$

The three terms in equation (2.15) are the likelihood $P(\boldsymbol{y}|\boldsymbol{\theta})$, the population (prior) distribution $P(\boldsymbol{\theta}|\phi)$ and the hyperprior distribution $P(\phi)$.

The second quantity is the conditional posterior density $P(\boldsymbol{\theta}|\phi,\boldsymbol{y})$. Since the observations $\boldsymbol{y}$ are fixed values, this quantity is a function of ϕ. According to the exchangeability in equation (2.11), the conditional posterior density $P(\theta|\phi,y)$ can be expressed as:

$$P(\boldsymbol{\theta}|\phi,\boldsymbol{y}) = \prod_{j=1}^{J} P(\theta_j|\phi,\{y_{i,j}|i=1,2,...,n\}). \tag{2.16}$$

For each $P(\theta_j|\phi,\{y_{i,j}|i=1,2,...,n\})$, it can be expressed as:

$$P(\theta_j|\phi,\{y_{i,j}|i=1,2,...,n\}) \propto \left(\prod_{i=1}^{n} P(y_{i,j}|\theta_j)\right) P(\theta_j|\phi). \tag{2.17}$$

This transformation is based on Bayes' theorem in equation (2.1). If the prior distribution $P(\theta_j|\phi)$ is a conjugate prior for the likelihood $P(y_{i,j}|\theta_j)$, the analytical form of the conditional posterior density $P(\boldsymbol{\theta}|\phi,\boldsymbol{y})$ will be given easily (the details of the conjugate prior can be found in Section 2.1.2).

The third quantity is the marginal distribution of hyperparameters ϕ, $P(\phi|\boldsymbol{y})$. This quantity can be calculated by a brute force way, that is, to integrate the joint posterior distribution over $\boldsymbol{\theta}$:

$$P(\phi|\boldsymbol{y}) = \int P(\boldsymbol{\theta},\phi|\boldsymbol{y})d\boldsymbol{\theta}. \tag{2.18}$$

Moreover, the marginal distribution can be obtained algebraically using the conditional probability formula:

$$P(\phi|\boldsymbol{y}) = \frac{P(\boldsymbol{\theta},\phi|\boldsymbol{y})}{P(\boldsymbol{\theta}|\phi,\boldsymbol{y})} \tag{2.19}$$

where the numerator is the joint posterior distribution in equation (2.15) and the denominator is

the conditional posterior distribution in equation (2.16).

Now, for the simple HBM illustrated in this section, the drawing of simulations of the parameters from the joint posterior distribution can be implemented to quantify the uncertainty of such parameters. It can be separated as three steps:

1. Draw the hyperparameters ϕ from its marginal posterior distribution $P(\phi|\boldsymbol{y})$ based on equations (2.18) or (2.19).

2. Given ϕ, draw the parameter $\boldsymbol{\theta}$ from its conditional posterior distribution $P(\boldsymbol{\theta}|\phi, y)$ based on equations (2.16) and (2.17).

3. Given $\boldsymbol{\theta}$, draw predictive values $\tilde{\boldsymbol{y}}$ from the posterior predictive distribution. The details of the posterior predictive distribution can be found in Section 2.1.3

The example illustrated in this section is simple. Specifically, the number of parameters is small, and the relationships between data, parameters and hyperparameters are straightforward. However, the applications of the HBM in the real world are usually more challenging. There are more parameters and the relationships between these parameters are usually complicated. The complicated HBM can be described by a Bayesian network [18] which will be introduced in the next section.

2.2.3 Directed acyclic graph

Generally, hierarchical Bayesian models can be solved or formulated purely by algebraic manipulation (see details in Section 2.2.2). However, diagrammatic representations of probability distributions, named probabilistic graphical models, have many advantages to tackle complicated hierarchical Bayesian models. The benefits of probabilistic graphical models are listed as follows:

- They provide the visualization of the structure of the HBM. Moreover, the design and motivation of the statistical models can be visually revealed.

- The properties of the HBM can be directly observed. Specifically, many properties, e.g. conditional independence, can be inspected from the graph visually.

- Graphical manipulations can express complex computations. For example, the relationships between different parameters for a complicated HBM can be expressed implicitly using graphical language to avoid complicated mathematical expressions.

A probabilistic graphical model contains two types of elements, nodes (or vertices) and links (or edges). The nodes denote random variables or observations, and the links denote the probabilistic relationships between nodes. There are two types of links, i.e. directed and undirected.

The difference between directed and undirected links is distinguished by their names. Specifically, the directed links have particular directionalities based on the direction of arrows, whereas there is no such directionality for the undirected links. Directed graphs can be applied to express causal relationships (dependencies) between random variables, and undirected graphs are usually applied to express soft constrains, e.g. spatial relationships, between random variables. An HBM can be expressed as a directed acyclic graph (DAG). The "acyclic" denotes that it cannot go back to itself through the directed links for any starting node.

Figure 2.1 is a very simple DAG. By application of the product rule of probability, the joint distribution of it can be expressed as follows:

$$P(y_{i,j}, \theta_j, \phi) = P(y_{i,j}|\theta_j, \phi)P(\theta_j, \phi). \tag{2.20}$$

Since the observation $y_{i,j}$ only depends on θ_j, it can be simplified as:

$$P(y_{i,j}, \theta_j, \phi) = P(y_{i,j}|\theta_j)P(\theta_j, \phi). \tag{2.21}$$

By applying the product rule again, the joint distribution becomes:

$$P(y_{i,j}, \theta_j, \phi) = P(y_{i,j}|\theta_j)P(\theta_j|\phi)P(\phi). \tag{2.22}$$

Given Figure 2.1 and equation (2.22), the higher layer is dependent on the lower layer, and the dependency can be expressed by a conditional distribution. Since the product rule can be applied repeatedly, a general expression of a DAG shows as follows [18] (Chapter 8):

$$P(\boldsymbol{\theta}) = \prod_{q=1}^{K} P(\theta_q|pa_q) \tag{2.23}$$

where K is the total number of nodes and pa_q is the set of parents of θ_q. For a DAG, if there is a directed link from node A to node B, then node A is the parent of node B. In other words, node B is the child of node A. Meanwhile, the child node is conditionally dependent on the parent node. For example, in Figure 2.1, ϕ is the parent node of θ_j and θ_j is the child node of ϕ.

Figure 2.2 shows an example of a complicated HBM (DAG). According to equation (2.23), the joint probability of this DAG is

$$P(A,B,C,D,E,F,G,H,I) = P(A|B,C)P(B|D,E,H)P(C|F)P(D|G,H)P(H|I)P(E)P(I). \tag{2.24}$$

Conditional independence [19] is a vital concept in a probabilistic graphical model. The judgement for two nodes, whether they are conditionally independent, will ultimately determine the conditional posterior distribution of a given node. For any three random variables a, b and c,

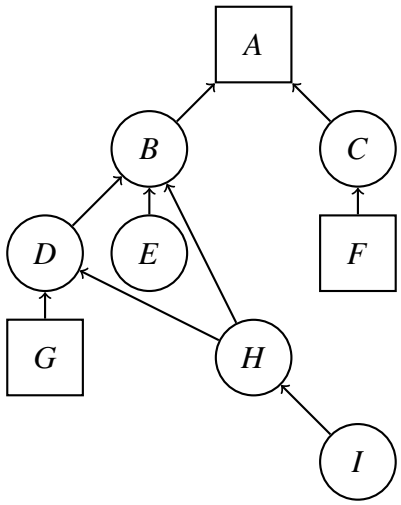


Figure 2.2: An example of a complex HBM (DAG). In this graph, the higher layers are conditionally dependent on the lower layers. The circle nodes denote variables and the rectangle nodes denote hyperparameters (G,F) or observations (A).

if

$$P(a|b,c) = P(a|c) \tag{2.25}$$

is satisfied, then it is said that a is conditionally independent of b. Another form of the conditional independence considers the joint distribution of a and b conditional on c, it gives

$$\begin{aligned} P(a,b|c) &= P(a|b,c)P(b|c) \\ &= P(a|c)P(b|c). \end{aligned} \tag{2.26}$$

It is worth mentioning that the conditional independence needs equations (2.25) or (2.26) holds for every possible value of c.

Theoretically, applying the product rule of probability repeatedly, the conditional independence property can be tested for any three nodes. However, it is pretty time-consuming in practice if the graphical models are complicated. A general method to test the conditional independence property is named d-separation where "d" indicates "directed". We consider all paths from node A to node B. A set of nodes C blocks the path from A to B if there is a node such that either

- The arrows on the path at a node is either head-to-tail or tail-to-tail and this node is in the set of C.
- The arrows on the path at a node is head-to-head and neither this node nor its children nodes are in the set of C.

If all paths are blocked, A is d-separated to B by C, and A is conditionally independent of B given C, see Section 8.2 in [18]. The definition of the three types of node can be seen in Figure 2.3. For

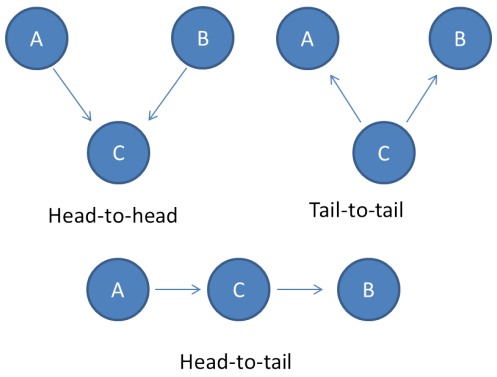


Figure 2.3: The definition of the three types of nodes, head-to-head, head-to-tail and tail-to-tail.

a DAG, it can not only be represented by a factorization of a joint distribution into a product of conditional distributions, but also represented by a set of conditional independence statements. The d-separation theorem ensures these two forms are equivalent.

In a DAG, there is another concept named Markov blanket that is related to the conditional independence property. If a DAG is very complicated, it will be inconvenient to judge whether two nodes are conditionally independent given other nodes because all paths between two nodes need to be considered . The Markov blanket can simplify this process. The Markov blanket for a node contains its parents, children and co-parents. For example, in Figure 2.2, the parents of node B are $\{D,E,H\}$. The child of node B is $\{A\}$ and the co-parent of node B is $\{C\}$. The probability of a selected target parameter conditional on all other parameters is given by the probability of the target parameter conditional on its Markov blanket [18] (Chapter 8). The Markov blanket of node A is the minimal set of nodes that can block node A from the whole graph.

2.2.4 Markov random fields

In the previous section, the concept of directed acyclic graph was reviewed. The most important property of a DAG is that it specifies a factorization of the joint distribution using a product of local conditional distributions of a set of random variables. Moreover, the directed links in a DAG define the dependencies between different nodes. There is another major type of graphical model in which the links between nodes are undirected.

A Markov random field, also named as an undirected graphical model [18] (Chapter 8), has a set of nodes with undirected links among them. The conditional independence property for a Markov random field is relatively simple. Let A, B, and C denote three sets of nodes. Considering all paths that connect nodes in A and B, if all paths contain one or more nodes in C, then all paths are "blocked", and the conditional independence

$$P(a|b,c) = P(a|c) \tag{2.27}$$

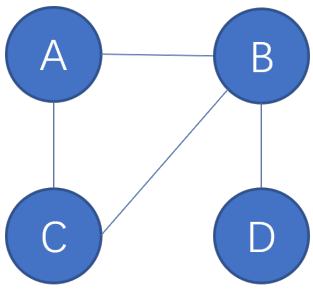


Figure 2.4: An example of a Markov random field.

holds. However, if at least one path from nodes in A to nodes in B is not blocked by nodes in C, the conditional independence property is not satisfied. An alternative way to test the conditional independence property is to remove all nodes in set C and all links connecting to the nodes in C. If there exists at least one path from nodes in A to nodes in B, the conditional independence property does not hold, and vice versa. The Markov blanket for a Markov random field is rather simple. Specifically, the Markov blanket of node A is the set of all nodes that are the neighbour of node A. Now, given the Markov blanket of the Markov random fields, the expression of conditional independence can be updated using mathematical language. Let x_i and x_j denote two nodes that are not connected by a link. Similar to equation (2.26), the conditional independence can be expressed as follows:

$$P(x_i, x_j | x_{\setminus\{i,j\}}) = P(x_i | x_{\setminus\{i,j\}}) P(x_j | x_{\setminus\{i,j\}}), \tag{2.28}$$

where $x_{\setminus\{i,j\}}$ indicates the set of all variables $\{x_k\}$ except x_i and x_j. This would lead to another concept, a clique. A clique in a Markov random field is defined as a set of nodes with links between all pairs of nodes. In other words, all nodes in a clique are connected. Furthermore, a concept named "maximal clique" is a clique such that any node in the graph cannot be added in this clique. If any node is included in a maximal clique, this clique will not be a clique anymore. In other words, if an extra node is added in a maximal clique, there must be two nodes that are not connected. Figure 2.4 shows an example of a Markov random field. In this figure, $\{A,B\}$, $\{A,C\}$, $\{B,C\}$, $\{B,D\}$ and $\{A,B,C\}$ are cliques. $\{B,D\}$ and $\{A,B,C\}$ are maximal cliques.

Let C denote a maximal clique and x_C denote the sets of variables in C, the joint probability distribution of a Markov random field can be expressed as a product of potential function $\psi_C(x_C)$ of all maximal cliques in this graph:

$$P(x) = \frac{1}{Z} \prod_C \psi_C(x_C) \tag{2.29}$$

where Z is a normalization constant

$$Z = \sum_x \prod_C \psi_C(x_C). \tag{2.30}$$

Given a strictly positive constraint to the potential function $\psi_C(x_C)$, a precise relationship between factorization, i.e. equation (2.29), and conditional independence, i.e. equation (2.28) can be made. Specifically, in [20], the Hammersley-Clifford theorem stated that the sets of distributions expressed by factorization and conditional independence are identical. Moreover, a Markov random field is equivalent to a Gibbs distribution (Boltzmann distribution) [20], so that

$$\psi_C(x_C) = \exp(-E(x_C)), \tag{2.31}$$

where $E(x_C)$ is called energy function.

Generally, the first- and second-order neighbours of the Markov random fields $\{P(x)\}$ in equation (2.29) can be specified. However, these fields have undesirable large-scale properties [17]. A scheme named "iterated conditional modes" (ICM) has been introduced to find the MAP of Markov random fields in [17]. This method can be applied to process image denoising. For this application, a hidden variable (pixel label) is usually introduced, and a Markov random field prior is usually applied to the hidden variable. Specifically, let $x = \{x_1, x_2, ..., x_n\}$ denote the hidden variable. Let $y = \{y_1, y_2, ..., y_n\}$ denote the observations. To apply the ICM algorithm, two assumptions are made. Firstly, given the hidden variable x, the observations $\{y_1, y_2, ..., y_n\}$ are conditionally independent and have same conditional density function $f(y_i|x_i)$. Therefore, the joint distribution can be described as follows:

$$P(y|x) = \prod_{i=1}^{n} f(y_i|x_i). \tag{2.32}$$

Secondly, the hidden x is a locally dependent Markov random field with Markov property. The ICM algorithm tends to iteratively update the values of the hidden variable x to achieve the local maximization of the conditional posterior distribution. Specifically, let x_i^j denote the jth iteration for element i. If there exists such a plausible choice of x_i^{j+1} that given the observations y and neighbours $x_{\setminus i}$, the conditional posterior distribution of $P(x_i|y, x_{\setminus i})$ increases, the value of x_i^{j+1} will be updated to replace x_i^j. Once this step has been through all elements $i = 1, 2, ..., n$, one cycle of iteration has been finished. After a fixed number of iterations is finished or the convergence has been achieved, the current value of x_i is the MAP estimate. Generally, the initial value of x will be chosen as the maximum likelihood estimates given by $P(y|x)$ without considering the geometrical relationship among all elements $i = 1, 2, ..., n$. The ICM algorithm ensures the conditional posterior distribution $P(x_i|y, x_{\setminus i})$ will never decrease and converge to a local maximum.

The advantage of the ICM algorithm is that this method usually converges extremely rapidly [17]. However, it has the same disadvantage: the estimates may get stuck in a local maximum, with other "greedy" optimization algorithms, e.g. gradient descent.

2.3 Markov chain Monte Carlo

2.3.1 Introduction

Markov Chain Monte Carlo (MCMC) is a general method to draw samples of a parameter θ from its approximate posterior distribution and then modify those draws to better approximate the target posterior distribution, $P(\theta|y)$. This method was originally introduced by [21,22]. The MCMC sampling is done consecutively, and the distribution of the draw is dependent on its previous draw. Therefore, all draws are from a Markov chain. A Markov chain is a sequence of random variables $\{\theta_1, \theta_2, ...\}$. For any t, the distribution of θ_t given all θ_i $(i < t)$ depends and only depends on its previous variable θ_{t-1}. Therefore, the distribution of the drawn samples is only dependent on the previous sample. It has been proved that for the MCMC sampling, the approximate distribution improves at each step and converges to the target distribution under fairly general regularity conditions called ergodicity [12].

In the applications of MCMC, independent sequences for different parameters are produced. Specifically, the sequence $\theta^1, \theta^2, \theta^3, ...$ is generated from an initial value θ^0. For any t, θ^t is generated from a transition distribution $T_t(\theta^t|\theta^{t-1})$ that is dependent on its previous value θ^{t-1}. The transition distribution $T_t(\theta^t|\theta^{t-1})$ must satisfy that the generated Markov chain converges to a unique stationary distribution, and this unique stationary distribution is the posterior distribution $P(\theta|y)$.

Generally, the MCMC methods are applied when the samples of θ are impossible (or computationally inefficient) to obtain directly from its posterior distribution $P(\theta|y)$. The MCMC methods provide an iterative sampling scheme to draw samples from a distribution that approximates the target posterior distribution $P(\theta|y)$.

Some general MCMC algorithms, e.g. Gibbs sampler, Metropolis-Hastings algorithm, are reviewed in this section. Moreover, the efficiency and convergence of the MCMC methods are also reviewed in this section. The applications of these methods will be displayed in Chapters 6 and 7.

2.3.2 Gibbs sampler

Gibbs sampler is an MCMC algorithm that is particularly useful for multidimensional sampling problems when the conditional posterior distributions have been derived. Let $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_d\}$ denote a vector of parameters. For each iteration of the sampling cycles, draw one particular parameter θ_t from $\boldsymbol{\theta}$ conditional on all other parameters in $\boldsymbol{\theta}$. Thus, there are d draws for one iteration. This process can be express as follows:

1. Initialise $\boldsymbol{\theta} = \{\theta_1^0, \theta_2^0, ..., \theta_d^0\}$.

2. Draw θ_j^t from its conditional posterior distribution $P(\theta_j^t|\boldsymbol{\theta}_{-j}^{t-1}, y)$ where $\boldsymbol{\theta}_{-j}^{t-1}$ is all elements in $\boldsymbol{\theta}$ except for θ_j with values at their current draw and y are the observations.

Mathematically, $\boldsymbol{\theta}_{-j}^{t-1}$ can be expressed as:

$$\boldsymbol{\theta}_{-j}^{t-1} = \{\theta_1^t, \theta_2^t, ..., \theta_{j-1}^t, \theta_{j+1}^{t-1}, \theta_{j+2}^{t-1}, ..., \theta_d^{t-1}\}. \tag{2.33}$$

3. Let $t = M$ where M is the length of the Markov chain. Discard the first n iterations. This process is called burn-in.

The burn-in phase is important for the MCMC samplings. The reason of burn-in is that the initial value $\boldsymbol{\theta} = \{\theta_1^0, \theta_2^0, ..., \theta_d^0\}$ might not be in the target posterior distribution. As mentioned in this section, the approximate distribution improves at each iteration and converges to the target distribution. Therefore, when the burn-in number n is large enough, the rest of the samples $\{(\theta_1^{n+1}, ..., \theta_d^{n+1}), ..., (\theta_1^M, ..., \theta_d^M)\}$ are approximately from the target distribution.

There is a strict rule for the applications of the Gibbs sampler. All conditional posterior distributions for $\boldsymbol{\theta}$ must be derived. Therefore, the choice of prior distributions $P(\boldsymbol{\theta})$ is vital. To ensure the analytical forms of the conditional posterior distributions can be derived, conjugate prior distributions (see details in Section 2.1.2) are usually adopted.

In an HBM (see Section 2.2.3), the conditional posterior distribution of a parameter from the HBM can be derived by its Markov blankets. Therefore, the prior distributions can be chosen as conjugate priors, which leads to simple simulations using the Gibbs sampler.

2.3.3 Metropolis and Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is an algorithm family of MCMC methods. In this section, both Metropolis and Metropolis-Hastings algorithms are reviewed.

The Metropolis algorithm adjusts random walk proposals with an acceptance or rejection rule. The accepted samples converge to a specified target distribution. The algorithm proceeds as follows:

1. Draw an initial value of θ_0 from a starting distribution $P_0(\theta)$. The starting distribution could be the prior distribution of θ, and the initial value could be chosen as its point estimation.

2. For $t = 1, 2, ...,$

   (a) Sample a parameter θ^* from a proposal distribution at time t, $Q_t(\theta^*|\theta^{t-1})$. $Q_t(\theta^*|\theta^{t-1})$ is the probability of moving θ^{t-1} to θ^*. For the Metropolis algorithm, the proposal distribution must be symmetric, i.e. $Q_t(\theta_a|\theta_b) = Q_t(\theta_b|\theta_a)$.

   (b) Calculate the the ratio:

$$r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}. \tag{2.34}$$

(c) Update

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r,1) \\ \theta^{t-1} & \text{otherwise.} \end{cases} \tag{2.35}$$

3. Discard the first n iterations as burn-in.

The transition probability $T_t(\theta^t|\theta^{t-1})$ is a product of the proposal distribution $Q_t(\theta^*|\theta^{t-1})$ and the acceptance distribution $A(\theta^*|\theta^{t-1})$. It can be expressed as follows:

$$T_t(\theta^t|\theta^{t-1}) = Q_t(\theta^*|\theta^{t-1})A(\theta^*|\theta^{t-1}). \tag{2.36}$$

Generally, when $\theta^t = \theta^{t-1}$, i.e. the proposal walk θ^* is rejected, it still counts one iteration in the algorithm.

There are two steps to prove that the sequences generated by the Metropolis algorithm converge to the target distributions. Firstly, each of generated sequences is proved to be a Markov chain with a unique stationary distribution. Secondly, the stationary distribution equals the target distribution. The proofs of both steps can be found in [12] (Section 11.2).

The Metropolis-Hastings algorithm is the generalization of the Metropolis algorithm. Specifically, the symmetric condition for the proposal distribution $Q_t(\theta^*|\theta^{t-1})$ is not required for the Metropolis-Hastings algorithm. Moreover, to correct the asymmetry of the proposal distribution $Q_t(\theta^*|\theta^{t-1})$, the acceptance ratio in equation (2.34) is modified to

$$r = \frac{P(\theta^*|y)Q_t(\theta^{t-1}|\theta^*)}{P(\theta^{t-1}|y)Q_t(\theta^*|\theta^{t-1})}. \tag{2.37}$$

Similarly, the proof that the sequences generated by the Metropolis-Hastings algorithm converge to the target distributions can be found in [12] (Section 11.2).

Gibbs sampler can be viewed as a special case of the Metropolis-Hastings algorithm. Specifically, the proposal sample θ_j^t from its conditional posterior distribution $P(\theta_j^t|\theta_{-j}^{t-1}, y)$ can be viewed as one Metropolis-Hastings step with acceptance ratio 1. The proof can be found in [12] (Section 11.3).

In some applications of MCMC simulations, not all closed forms of conditional posterior distributions can be derived, but its corresponding approximations can be constructed. A mixture scheme called Metropolis-within-Gibbs-sampler can therefore be used. Specifically, given the order for sampling θ, the Gibbs sampling is implemented when the corresponding conditional posterior distribution is derived. When the closed form of the conditional posterior distribution is not derived, one single step of Metropolis-Hastings algorithm will be applied to draw the sample. More details of the Metropolis-within-Gibbs-sampler can be found in [12] (Section 11.3), and [23] (Section 7.6.3).

2.3.4 Convergence and efficiency for the MCMC methods

There are two challenges for the inference of the iterative simulation. Firstly, if the number of iterations is insufficient, the simulations may not represent the target distribution. Even if the simulations converge to the target distribution eventually, the simulations in the early stage still reflect the starting distribution. This is the reason that the burn-in phase of the MCMC methods is necessary. Secondly, there are within-sequence correlations for the iterative simulation. Generally, the inference for correlated simulations are less precise than for the same number of independent simulations. The convergence and the efficiency of the MCMC methods should be considered to tackle these two challenges.

Convergence monitoring of the MCMC samples is used to provide evidence of the generated samples converging to the target distribution. Geweke's test [24] and Gelman-Rubin statistic [25] are used to diagnose the convergence of MCMC samples. It is worth mentioning that these tests are necessary conditions for convergence but not sufficient conditions.

Geweke's test compares the samples located in two different parts of the simulation chain. Specifically, if the mean values of the samples in two different parts are similar, then the samples in two different parts are assumed to be from the same distribution. Generally, the last half of the chain and a small interval at the beginning of the chain are compared.

Let θ_i denote the the i^{th} sample of the simulation chain where $i = 1, 2, ..., n$. Two sets $A = \{i, a_0 \leq i \leq a_1\}$ and $B = \{i, b_0 \leq i \leq b_1\}$ denote two different parts of the chain where $1 \leq a_0 \leq a_1 < b_0 \leq b_1 \leq n$ and a_0, a_1, b_0, b_1 are positive integer. The mean values of two sets are:

$$\bar{\theta}_A = \frac{1}{a_1 - a_0 + 1} \sum_{i \in A} \theta_i \tag{2.38}$$

and

$$\bar{\theta}_B = \frac{1}{b_1 - b_0 + 1} \sum_{i \in B} \theta_i. \tag{2.39}$$

The sample variances of two sets are:

$$\bar{s}_A^2 = \frac{1}{a_1 - a_0} \sum_{i \in A} (\theta_i - \bar{\theta}_A)^2 \tag{2.40}$$

and

$$\bar{s}_B^2 = \frac{1}{b_1 - b_0} \sum_{i \in B} (\theta_i - \bar{\theta}_B)^2. \tag{2.41}$$

If the simulation chain converges to the target distribution, then two different parts of the chain are both from the same distribution. Therefore, the mean values for both parts should be the

same, and Geweke's statistic has an asymptotically standard Gaussian distribution. Specifically,

$$Z = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\bar{s}_A^2 + \bar{s}_B^2}} \to N(0,1) \quad n \to \infty \tag{2.42}$$

The value Z will decline to 0 when $n \to \infty$. Therefore, if Z is high, then we believe that more iterations will help to improve the inference of the target distribution. More details of Geweke's test can be found in [24, 26].

For Gelman-Rubin statistic, let $\theta_{i,j}$ denote the the i^{th} sample of the j^{th} chain where $i = 1,2,...,n$ and $j = 1,2,...,m$. The chain mean is defined as:

$$\bar{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} \theta_{i,j} \tag{2.43}$$

and the grand mean is defined as:

$$\bar{\theta} = \frac{1}{m}\sum_{j=1}^{m} \bar{\theta}_j. \tag{2.44}$$

The between-chain variance B can be expressed as:

$$B = \frac{n}{m-1}\sum_{j=1}^{m} (\bar{\theta}_j - \bar{\theta})^2, \tag{2.45}$$

and the within-chain variance W can be expressed as:

$$W = \frac{1}{m}\sum_{j=1}^{m} \frac{1}{n-1}\sum_{i=1}^{n} (\theta_{i,j} - \bar{\theta}_j)^2. \tag{2.46}$$

The marginal posterior variance can be estimated as a weighted average of B and W:

$$\hat{\text{var}}(\theta|y) = \frac{n-1}{n}W + \frac{1}{n}B. \tag{2.47}$$

This quantity will overestimate the marginal posterior variance if the starting distribution is overdispersed. However, it is unbiased under stationarity, i.e. the starting distribution equals the target distribution.

A quantity named the potential scale reduction factor (PSRF) can be used to monitor the convergence of the iterative simulations. Specifically, the PSRF can be estimated by:

$$\hat{R} = \sqrt{\frac{\hat{\text{var}}(\theta|y)}{W}}. \tag{2.48}$$

This quantity will decline to 1 when $n \to \infty$. Therefore, if the quantity $\hat{R}$ is high, then we believe that more iterations will help to improve the inference of the target distribution. More details of

Gelman-Rubin statistic can be found in [12] (Section 11.4), and in [25].

The samples generated based on MCMC methods are inferior to the iid samples from the posterior distribution because of the Markov property. Specifically, the Markov property causes autocorrelation within samples, and therefore these samples are not independently distributed. The effective sample size (ESS) for correlated simulation draws can be defined to consider the statistical efficiency using the sample mean to estimate the posterior mean. In [12] (Section 11.5), the ESS is defined as

$$n_{\text{eff}} = \frac{mn}{1+2\sum_{t=1}^{\infty}\rho_t}. \tag{2.49}$$

where ρ_t is the autocorrelation of the sequence θ at lag t. To compute the ESS, the summation of the autocorrelation should be calculated. However, even though the series may converge, it is usually impossible to calculate it in practice because the sample correlation is too noisy when the value of t is large. Therefore, an estimate of the ESS is measured by:

$$\hat{n}_{\text{eff}} = \frac{mn}{1+2\sum_{t=1}^{T}\hat{\rho}_t}. \tag{2.50}$$

In this equation, T denotes the first odd positive integer for which $\hat{\rho}_{T+1}+\hat{\rho}_{T+2}$ is negative. This value is suggested in [12] (Section 11.5). $\hat{\rho}_t$ is defined by:

$$\hat{\rho}_t = 1 - \frac{V_t}{2\hat{\text{var}}(\theta|y)} \tag{2.51}$$

where $\hat{\text{var}}(\theta|y)$ can be found in equation (2.47) and

$$V_t = \frac{1}{m(n-t)}\sum_{j=1}^{m}\sum_{i=t+1}^{n}(\theta_{i,j}-\theta_{i-t,j})^2. \tag{2.52}$$

More details of the efficient sample size can be found in [12] (Section 11.5), and [27].

2.4 EM algorithm

2.4.1 Introduction

The expectation maximization (EM) algorithm is a general method for searching maximum likelihood (ML) or maximum a posteriori (MAP) solutions for statistical models that have latent variables [28, 29]. Typically, the ML or MAP solutions are obtained by taking the derivatives of the likelihood function with respect to all the parameters and simultaneously solving the equations. However, the closed form solution of the derivative of a statistical model with latent variables does not exist. In this section, the general approach of EM algorithm is reviewed. The application of it will be illustrated in Section 5.

2.4.2 Decomposition of likelihood function

Let X denote the observations of a statistical model, and Z denote the latent variable of this model. In other words, X denotes the incomplete data and $\{X,Z\}$ denote the complete data. θ is a set of parameters to govern the joint distribution $P(X,Z|\theta)$. The goal of the EM algorithm is to maximize the likelihood function that is expressed as follows:

$$P(X|\theta) = \sum_{Z} P(X,Z|\theta). \tag{2.53}$$

The latent variable is assumed to be discrete though the derivation is identical to the continuous case with summation replaced by integration. The optimization of the incomplete likelihood $P(X|\theta)$ is usually difficult, but the optimization of the complete likelihood $P(X,Z|\theta)$ is available in closed form for certain probability distributions (which are in the exponential family).

According to the product rule of probability, the log-likelihood of the joint distribution $P(X,Z|\theta)$ can be written as:

$$\ln P(X,Z|\theta) = \ln P(Z|X,\theta) + \ln P(X|\theta). \tag{2.54}$$

This equation can be finally rewritten to

$$\ln P(X|\theta) = L(q,\theta) + KL(q||p) \tag{2.55}$$

for any choice of $q(Z)$. In equation (2.55),

$$L(q,\theta) = \sum_{Z} q(Z) \ln \left\{ \frac{P(X,Z|\theta)}{q(Z)} \right\} \tag{2.56}$$

and

$$KL(q||p) = -\sum_{Z} q(Z) \ln \left\{ \frac{P(Z|X,\theta)}{q(Z)} \right\}. \tag{2.57}$$

The derivation from equation (2.54) to equation (2.55) can be found in Appendix A.

It can be noticed that the term $KL(q||p)$ is the Kullback–Leibler divergence between $q(Z)$ and $P(Z|X,\theta)$ [30]. According to Gibbs' inequality, the Kullback–Leibler divergence satisfies $KL(q||p) \geq 0$ and the equality holds if and only if $q(Z) = P(Z|X,\theta)$. Therefore, the term $L(q,\theta)$ satisfies $L(q,\theta) \leq \ln P(X|\theta)$. In other words, $L(q,\theta)$ is a lower bound of $\ln P(X|\theta)$.

The EM algorithm has two stages to iteratively find the optimization of the ML. The first step is called "expectation (E) step". In this step, given the fixed old value θ_{old}, the lower bound $L(q,\theta_{\text{old}})$ is pushed to the log likelihood with respect to $q(Z)$. Since $\ln P(X|\theta_{\text{old}})$ does not depend on $q(Z)$, the maximization of $L(q,\theta_{\text{old}})$ will happen when the equality of Kullback–Leibler divergence holds. At this time, $q(Z)$ equals to $P(Z|X,\theta)$. The second step is called "maximization (M) step". In this step, given the fixed distribution $q(Z)$, the lower bound $L(q,\theta)$ is maximised

with respect to θ to generate the new value θ_{new}. This will increase the lower bound L and necessarily increase the corresponding log likelihood function $\ln P(X|\theta)$. However, since $q(Z)$ is determined by the old value θ_{old} and it is kept fixed during the M step, $q(Z) \leq P(Z|X,\theta_{\text{new}})$ holds, and hence the Kullback–Leibler divergence is non-negative. This will cause the increase of the log likelihood function to be larger than the increase of the lower bound q. Specifically, if we substitute $q(Z) = P(Z|X,\theta_{\text{old}})$ into equation (2.56), it turns to

$$\begin{aligned} L(q,\theta) &= \sum_Z P(Z|X,\theta_{\text{old}}) \ln P(X,Z|\theta) - \sum_Z P(Z|X,\theta_{\text{old}}) \ln P(Z|X,\theta_{\text{old}}) \\ &= Q(\theta,\theta_{\text{old}}) + \text{constant} \end{aligned} \tag{2.58}$$

where

$$Q(\theta,\theta_{\text{old}}) = \sum_Z P(Z|X,\theta_{\text{old}}) \ln P(X,Z|\theta) \tag{2.59}$$

is the expectation of the log likelihood function of θ with respect to the conditional distribution of Z given X and old value θ_{old}. The constant in equation (2.58) is the negative entropy of distribution q that is independent of θ. Therefore, the quantity that is maximised in the M step is the expectation of the complete data log likelihood. For probability distributions in the exponential family, the M-step has a closed form solution. The E-step pushes the lower bound up so it becomes equal to the log likelihood. It is this combination of facts that is the source of the improvement. Therefore, the iterative EM cycle will increase the log likelihood function until the log likelihood does not change, and the estimates of θ that maximises the incomplete data log likelihood will be found.

2.4.3 Modification and generalization of EM algorithm

The EM algorithm can also maximise the posterior distribution under the Bayesian framework. Let $P(\theta)$ denote the prior distribution of θ. According to Bayes' theorem in equation (2.2), the posterior distribution $P(\theta|X)$ can be expressed as

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}. \tag{2.60}$$

By applying logarithm to both sides of the equation, and it turns to

$$\ln P(\theta|X) = \ln P(X|\theta) + \ln P(\theta) - \ln P(X). \tag{2.61}$$

By substituting equation (2.55) into equation (2.61), it turns to

$$\begin{aligned} \ln P(\theta|X) &= L(q,\theta) + KL(q||p) + \ln P(\theta) - \ln P(X) \\ &\geq L(q,\theta) + \ln P(\theta) - \ln P(X) \end{aligned} \tag{2.62}$$

where $\ln P(X)$ is a constant. The EM algorithm for the posterior distribution is to maximise the right hand side of equation (2.62) with respect to q and θ. Since q is only dependent on $L(q,\theta)$, the E step for the posterior distribution is exactly the same as the standard EM algorithm. However, the M step for the posterior distribution is different. Specifically, for the posterior distribution case, the maximization of $Q(\theta,\theta_{\text{old}})$ with respect to θ for the standard EM algorithm is modified to maximise $Q(\theta,\theta_{\text{old}})+\ln P(\theta)$ with respect to θ.

Loosely speaking, the EM algorithm overcomes a potential difficulty that the maximization of the log likelihood function is intractable by separating the maximization into two steps, and for each step, the maximization is easier to implement. However, the EM algorithm may be less effective and less efficient if either or both E and M steps are intractable. This potential less effectiveness leads to the modification of the EM algorithm. The generalised EM (GEM) algorithm tends to fix the intractable issue for M step. The EM algorithm aims to maximise $L(q,\theta)$ with respect to θ, but the GEM aims to search a proper update of θ to increase the value of $L(q,\theta)$. The GEM can be applied using nonlinear optimization like conjugate gradients algorithm [31]. Expectation conditional maximization (ECM) has been introduced to be another form of GEM [32]. Similar to the modification to the M step, the generalization of E step has been introduced in [33]. In this modification, partial optimization of $L(q,\theta)$ has been performed with respect to $q(Z)$ during the E step of the EM algorithm.

More details about the EM algorithm can be found in [18], Section 9.4. The applications of the EM algorithm to the medical image processing can be found in Chapter 5.

2.5 Gaussian process

2.5.1 Introduction

In statistics, a Gaussian process is a stochastic process, and any finite collection of random variables of the Gaussian process is multivariate Gaussian distributed. In other words, the Gaussian process can be viewed as an infinite-dimensional generalization of the Gaussian distribution. In this section, the Gaussian process regression is reviewed, while the application of the Gaussian process regression will be illustrated in Chapter 4. More details of the Gaussian process can be found in [12] (Chapter 21), and [34].

2.5.2 Gaussian process regression

Let y denote the observation (dependent variable) and x denote the input (independent variable). The regression analysis aims to reveal the underlying relationships between y and x. Generally, the observation y contains some unknown noise, and the regression can be mathematically expressed as

$$y = f(x)+\varepsilon \tag{2.63}$$

where the function f represents the relationship between y and x. If the mathematical expression of the function f is known, e.g. simple linear model $f(x) = ax + b$, the regression analysis would be the quantification and inference of the unknown parameter a and b. However, in reality, the mathematical expression of f might be unknown, and therefore nonparametric regression methods can be used. A Gaussian process can be assigned as the prior distribution of the function f as a flexible nonparametric regression.

A Gaussian process can be written as

$$f \sim GP(m,k) \tag{2.64}$$

where m is the mean function and k is the kernel (co-variance) function. The Gaussian process $f \sim GP(m,k)$ is a nonparametric model because the regression function f itself is an infinite object. The mean function m denotes the initial guess of the function f, and therefore is usually suggested as some convenient forms, e.g. linear model $m(\boldsymbol{x}) = \boldsymbol{x}\beta$ and constant $m(\boldsymbol{x}) = C$ (see details in [12], Chapter 21). The function k represents the kernel function between the process at any two points, x and $x^{'}$. Specifically, given $\boldsymbol{x} = [x_1, x_2, ..., x_n]^{\mathrm{T}}$ and a $n \times n$ covariance matrix K, the element (i, j) of the matrix can be expressed as $k(x_i, x_j)$, and it is usually written as a shorthand form $k(x, x^{'})$. The kernel function controls the degrees of the smoothness and fluctuation of the Gaussian process. The details of the kernel function of the Gaussian process can be found in Section 2.5.3.

2.5.3 Kernel function

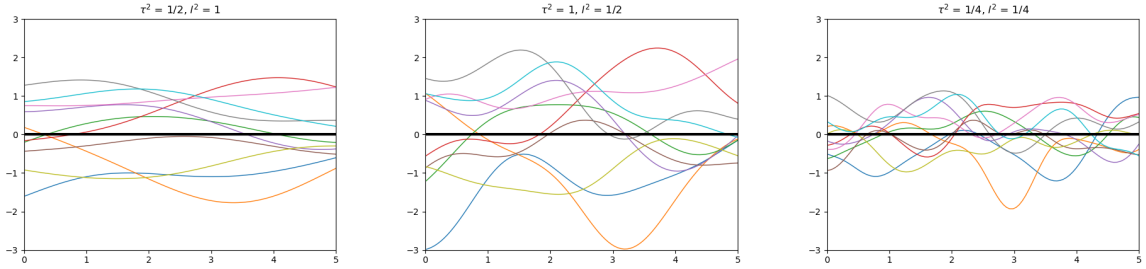
The choice of the kernel function of a Gaussian process prior will impose structural assumptions like smoothness, fluctuation, periodicity. Since the summation and the product of Gaussian processes are still Gaussian processes, their kernel functions can be easily combined. Some commonly used kernel functions are reviewed in this section.

One of the most commonly used kernel functions is called the squared exponential covariance function. Its mathematical expression is

$$k(x, x^{'}) = \tau^2 \exp\left(-\frac{|x - x^{'}|^2}{l^2}\right). \tag{2.65}$$

where τ and l are unknown parameters, and $|x - x^{'}|^2$ is the squared Euclidean distance between x and $x^{'}$. In this kernel function, the parameter τ controls the magnitude of the function, and the parameter l controls the smoothness of the function. Figure 2.5 shows the comparisons of different values of hyperparameters in the kernel function. The controlling of the values of the hyperparameters can be easily observed from these figures.

The hyperparameter l in equation (2.65) is invariant. This means that the degrees of the smoothness in all directions are the same. The invariant property is named “isotropic”. In some



(a) $\tau^2 = 0.5, l^2 = 1$ (b) $\tau^2 = 1, l^2 = 0.5$ (c) $\tau^2 = 0.25, l^2 = 0.25$

Figure 2.5: Gaussian process priors given different values of hyperparameters. Panels (a) - (c) show the Gaussian process priors with different values of hyperparameters given squared exponential covariance function and 0 mean function.

high-dimensional cases, the isotropic Gaussian process is not ideal because it cannot characterize the regression surface efficiently. On the other hand, anisotropic Gaussian processes may perform better for high-dimensional cases. For example, a modified squared exponential covariance function,

$$k(x,x') = \tau^2 \exp\left(-\sum_{j=1}^{P} \frac{|x_j - x_j'|^2}{l_j^2}\right) \tag{2.66}$$

is an anisotropic Gaussian process. The hyperparameter l_j is a length scale parameter that controls the smoothness of the Gaussian process prior in the direction of the j^{th} predictor. Model selections can be implemented to choose the values of the hyperparameters $\{l_j\}$ so that the anisotropic smoothness in different directions can be adapted given the input data.

Many commonly used kernel functions have been introduced in [2] (Chapter 4). A brief summary of these kernel functions can be found in Table D.1 in Appendix D.

2.5.4 Inference

Given the observed data, the Gaussian process prior of a function f can be updated to the posterior. In practice, the infinitely dimensional Gaussian process posterior cannot be estimated, but any finite subset of the realization of the function $\{f(x_i)\}_{i=1}^n$ is multivariate Gaussian distributed given the observed data points $\{x_i\}_{i=1}^n$. Moreover, given newly observed points $\tilde{x}$, the posterior predictive distribution at $\tilde{x}$ can be inferred. Firstly, the joint distribution for f and f^* given a zero mean Gaussian process prior $GP \sim (0,k)$ is

$$\begin{bmatrix} f \\ f^* \end{bmatrix} = N\left(0, \begin{bmatrix} K(x,x) + \sigma^2 I & K(x,\tilde{x}) \\ K(\tilde{x},x) & K(\tilde{x},\tilde{x}) \end{bmatrix}\right) \tag{2.67}$$

where noise variance σ^2 has been imposed to the diagonal of covariance of f because of the ε in equation (2.63). According to standard Gaussian conditional expressions in [18], the posterior

predictive distribution of f^* is

$$P(f^*|f) = N(\boldsymbol{\mu}, \Sigma) \tag{2.68}$$

where

$$\boldsymbol{\mu} = K(x,\tilde{x})(K(x,x) + \sigma^2 I)^{-1} f \tag{2.69}$$

and

$$\Sigma = K(\tilde{x},\tilde{x}) - K(x,\tilde{x})(K(x,x) + \sigma^2 I)^{-1} K(\tilde{x},x). \tag{2.70}$$

2.6 Model selection and cross-validation

2.6.1 Introduction

Model selection is the task to compare different candidate models given data. Although all models are usually considered to have mismatches with the data in practice, it is necessary to assess the accuracy of the prediction given different models and choose the one that is fit to the data. Moreover, the model that is fitted to the data with the least mismatches might not be favoured because the complexity of the model should also be considered. In other words, the degree of the complexity (or the number of parameters) of a model should be penalised for comparisons. Different information criteria are introduced to estimate the prediction error and select the statistical model given the penalty of the number of parameters.

Generally, for fit models, the out-of-sample predictions are less accurate than the within-sample predictions, and this phenomenon is called overfitting. Cross-validation is one type of approach that assesses the prediction accuracy between different statistical models that are generalised given independent data set. Typically, the cross-validation method can avoid the overfitting issue, but it also has some drawbacks, e.g. computationally expensive.

In this section, some commonly used information criteria for model selection are reviewed. Moreover, some cross-validation approaches are also reviewed.

2.6.2 Prediction accuracy assessment

Prediction accuracy is one way to evaluate the performances of different models. There are many different ways to define the prediction accuracy of different models. Some generic scoring functions and rules are reviewed here.

For point prediction, only a single value is predicted as the future observation. The functions used to measure the point prediction are named scoring functions. The most commonly used scoring function is called “mean squared error (MSE)”. Let $\{y_i\}_{i=1}^n$ denote the observation and

θ denote the parameters of the model. The MSE can be expressed as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \text{E}(y_i|\theta))^2, \tag{2.71}$$

or a weighted version

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \text{E}(y_i|\theta))^2}{\text{var}(y_i|\theta)}, \tag{2.72}$$

where $\text{E}(\cdot|\cdot)$ and $\text{var}(\cdot|\cdot)$ denote the conditional expectation and the conditional variance respectively. In regression analysis, the MSE is sometimes referred to as the unbiased estimate of error variance, which is calculated by the sum of squared errors (SSE) divided by the number of degrees of freedom of the model. The SSE is expressed as

$$\text{SSE} = \sum_{i=1}^{n}(y_i - f(x_i))^2 \tag{2.73}$$

where x_i is the independent variable and f is the regression function.

For probabilistic prediction, the probability of future events is given as the forecast. Specifically, in contrast to the point prediction that predicts the future observation $\tilde{y}$, the probabilistic prediction aims to obtain the inferences of $\tilde{y}$ with the full analysis of the uncertainty. Scoring rules are used to measure the probabilistic prediction [35]. The most commonly used scoring rule is called logarithmic score [36]. Given the likelihood function $p(y|\theta)$, the logarithmic score is $\log p(y|\theta)$. Rather than the log-posterior, the log-likelihood is used for the prediction assessment. The reason is that the fitness of the model to the data is used to assess the prediction accuracy, but the introduction of the prior does not aim to improve the prediction accuracy. More details about the logarithmic score can be found in [12] (Section 7.1) and [35].

The ideal measurement of the fitness of a model is the out-of-sample prediction given new data that is noise free. Let f denote the true data-generation function, y denote the observed data and $\tilde{y}_i$ denote a new point that is generated by f. The logarithmic score of the out-of-sample prediction given $\tilde{y}$ is

$$\log P_{\text{post}}(\tilde{y}_i) = \log \text{E}_{\text{post}}\left(P(\tilde{y}_i|\theta)\right) = \log \int P(\tilde{y}_i|\theta)P_{\text{post}}(\theta)d\theta \tag{2.74}$$

where the subscript post denotes the probability or the expectation over the posterior distribution of θ. Considering the true future data $\tilde{y}$ is usually hard to acquire, an alternative choice is the expected out-of-sample log predictive density (elpd) that can be expressed as

$$\text{elpd} = E_f\left(\log P_{\text{post}}(\tilde{y}_i)\right) = \int \left(\log P_{\text{post}}(\tilde{y}_i)\right) f(\tilde{y}_i)d\tilde{y}_i \tag{2.75}$$

where $f(\tilde{y}_i)$ is the realization of the future data $\tilde{y}$ that is generated by f. In practice, the data generation function f is usually unknown, and therefore the estimate of f will be inserted.

However, this will lead to the overfitting issue. Given n number of new points, equation (2.75) expands to expected log pointwise predictive density (elppd),

$$\text{elppd} = \sum_{i=1}^{n} E_f \left(\log P_{\text{post}}(\tilde{y}_i) \right). \tag{2.76}$$

Given the point estimate $\hat{\theta}(y)$, the elpd in equation (2.75) can also be expressed as

$$\text{elpd} = \text{E}_f \left(\log P(\tilde{y}|\hat{\theta}) \right). \tag{2.77}$$

In reality, the parameter θ is unknown, and the posterior distribution $P_{\text{post}}(\theta) = P(\theta|y)$ can be used to calculate the log predictive density $\log P(y|\theta)$. According to equation (2.74), the log pointwise predictive density is expressed as

$$\text{lppd} = \log \prod_{i=1}^{n} P_{\text{post}}(y_i) = \sum_{i=1}^{n} \log \int P(y_i|\theta) P_{\text{post}}(\theta) d\theta. \tag{2.78}$$

Given $s = 1, 2, ..., S$ simulations of θ, the lppd can be approximately computed by

$$\text{computed lppd} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} P(y_i|\theta^s) \right) \tag{2.79}$$

when the number of simulation S is large enough such that the simulations can explore the whole posterior distribution of θ.

2.6.3 Information criteria

Akaike information criterion (AIC)

Akaike information criterion is one type of Information criterion that is based on the point estimate $\hat{\theta}$, especially the maximum likelihood estimate (MLE) [37]. The out-of-sample prediction accuracy of the AIC is defined by equation (2.77) with corrections based on the number of the parameters. The correction is imposed to add the penalty based on the model complexity to avoid the overfitting issue. Equation (2.77) can not be calculated directly, and therefore it is estimated by the maximum log likelihood function $\log P(y|\hat{\theta})$. The expression of the AIC is

$$\text{AIC} = -2 \log P(y|\hat{\theta}) + 2k \tag{2.80}$$

where k is the number of the parameters in the model. -2 is multiplied by the estimation of the corrected elpd making this first term equivalent to the deviance in a general linear model framework.

One disadvantage of the AIC is that the correction term k is inaccurate in some circum-

stances, e.g. given informative prior distribution or the model having a hierarchical structure. For informative priors, or the model having a hierarchical structure, the effective number of parameters is less than the nominal number of parameters. Therefore, the correction term k cannot accurately reflect the penalty of the model complexity.

Bayesian information criterion (BIC)

Bayesian information criterion is a similar Information criterion to the AIC. This information criterion is firstly published in [38]. The BIC introduces the sample size of the data n into the penalty term, and therefore its expression is

$$\mathrm{BIC} = -2\log P(y|\hat{\theta}) + k\log n. \tag{2.81}$$

Therefore, for a large set of data, the penalty of the model complexity will be increased dramatically, and a simpler model is preferred. Different from other information criteria introduced here, the BIC aims to approximate the marginal probability density of the data, $P(y)$. On the other hand, the other information criteria are motivated by an estimation of predictive fit.

Deviance information criterion (DIC)

Deviance information criterion [39] can be viewed as a Bayesian version of the AIC, which is usually applied to select the model when the posterior distributions are approximately obtained by the MCMC approaches. Two changes are made to equation (2.80) to form the DIC. Firstly, the MLE $\hat{\theta}$ is replaced by the posterior mean $\hat{\theta}_{\mathrm{Bayes}} = \mathrm{E}(\theta|y)$. Secondly, the number of parameters k is replaced by a data-based bias correction. The predictive accuracy of the DIC is

$$\widehat{\mathrm{elpd}}_{\mathrm{DIC}} = \log P(y|\hat{\theta}_{\mathrm{Bayes}}) - P_{\mathrm{DIC}} \tag{2.82}$$

where

$$P_{\mathrm{DIC}} = 2\left(\log P(y|\hat{\theta}_{\mathrm{Bayes}}) - \mathrm{E}_{\mathrm{post}}\left(\log P(y|\theta)\right)\right) \tag{2.83}$$

is the effective number of parameters. Given simulations θ^s, $s = 1, 2, ..., S$, P_{DIC} can be calculated by

$$\text{computed } P_{\mathrm{DIC}} = 2\left(\log P(y|\hat{\theta}_{\mathrm{Bayes}}) - \frac{1}{S}\sum_{s=1}^{S}\log P(y|\theta^s)\right) \tag{2.84}$$

according to [40]. For a linear model given a uniform prior distribution, the effective number of parameters P_{DIC} degenerates to the number of parameters k and $P(y|\hat{\theta}_{\mathrm{Bayes}}) = P(y|\hat{\theta})$. This makes the DIC equivalent to the AIC. Similarly, the value of DIC is the predictive accuracy multiplied by -2 as follows

$$\mathrm{DIC} = -2\log P(y|\hat{\theta}_{\mathrm{Bayes}}) + 2P_{\mathrm{DIC}}. \tag{2.85}$$

Watanabe-Akaike information criterion (WAIC)

Watanabe-Akaike information criterion [41] is an asymptotic Bayesian leave-one-out cross-validation. In other words, Bayesian leave-one-out cross-validation converges asymptotically to WAIC. Two WAIC correction methods are reviewed here to adjust the log point prediction in equation (2.79).

The first method of the correction is similar to the structure of P_{DIC}. Its mathematical expression is

$$P_{\text{WAIC1}} = 2\sum_{i=1}^{n}\left(\log \mathrm{E}_{\text{post}}(y_i|\theta) - \mathrm{E}_{\text{post}}\left(\log P(y_i|\theta)\right)\right). \tag{2.86}$$

The expectation terms in equation (2.86) can be computed by approximations using averages over the posterior samples θ^s as follows:

$$\text{computed } P_{\text{WAIC1}} = 2\sum_{i=1}^{n}\left(\log\left(\frac{1}{S}\sum_{s=1}^{S}P(y_i|\theta^s)\right) - \frac{1}{S}\sum_{s=1}^{S}\log P(y_i|\theta^s)\right). \tag{2.87}$$

The second method of the correction is based on the variance of the log predictive density over n points. Specifically,

$$P_{\text{WAIC2}} = \sum_{i=1}^{v}\text{var}_{\text{post}}(\log P(y_i|\theta)) \tag{2.88}$$

where the posterior variance of the log predictive density can be approximately calculated by

$$\text{computed } P_{\text{WAIC2}} = \sum_{i=1}^{n}V_{s=1}^{S}\left(\log P(y_i|\theta^s)\right) \tag{2.89}$$

where $V_{s=1}^{S}$ denotes the sample variance,

$$V_{s=1}^{S}(a_s) = \frac{1}{S-1}(a_s - \bar{a})^2. \tag{2.90}$$

Therefore, the WAIC can be obtained by

$$\hat{\text{elppd}}_{\text{waic}} = \text{lppd} - P_{\text{WAIC}} \tag{2.91}$$

where lppd can be approximately acquired from equation (2.79) and P_{WAIC} can be calculated by either equation (2.87) or equation (2.89). Generally, P_{WAIC2} is suggested because it give a closer results to leave-one-out cross-validation and the details can be found in Section 7.2, [12]. Similarly to other criteria, the value of WAIC is computed by $\hat{\text{elppd}}_{\text{waic}}$ multipiled by -2.

Compared to the AIC and the DIC, the WAIC uses the posterior distribution of the parameters θ instead of using the point estimation of the parameters θ. This property is favoured because the WAIC evaluates the prediction accuracy that is used for new data. On the other hand, the prediction accuracies obtained according to AIC and DIC is based on the plug-in predictive

density. Therefore, in this study, the model selections in Chapters 6 and 7 are both based on the values of the WAIC. BIC could also be applied for those two chapters. However, for the consistency of the work, WAIC has been chosen to apply the model selection.

2.6.4 Cross-validation

Cross-validation [42] is a technique to use the training data to fit the model and use the testing data to evaluate the prediction accuracy. It aims to assess the ability of a model to predict new data that is not used in the training session. This technique can avoid the overfitting issue. However, cross-validation might be computationally expensive. The leave-one-out cross-validation is an extreme case that fit the model n times.

In the Bayesian framework, the data are separated into two sets y_{train} and y_{test} repeatedly. The model firstly is fitted by the training set y_{train}, and generates the posterior distribution $P(\theta|y_{\text{train}})$. Then, the evaluation of the prediction accuracy using log predictive density is carried out based on the testing set. This can be expressed as

$$\log P_{\text{train}}(y_{\text{test}}) = \log \int P_{\text{pred}}(y_{\text{test}}|\theta)P(\theta|y_{\text{train}})d\theta. \tag{2.92}$$

Similar to the contents in Section 2.6.3, the log predictive density can be approximately computed by

$$\log P_{\text{pred}}(y_{\text{test}}|\theta) = \log\left(\frac{1}{S}\sum_{s=1}^{S} P(y_{\text{test}}|\theta^s)\right). \tag{2.93}$$

where $\theta^s = \theta^1, \theta^2, ..., \theta^S$ are the S simulation draws of θ.

Leave-one-out cross-validation is specified here because leave-p-out cross-validation is rather similar to it. For the case of leave-one-out cross-validation, the data is partitioned n times. For each time, only one data point is put into the testing set, and other data points are put into the training set. Therefore, n different inferences $P_{\text{post}(-i)}$, ($P_{\text{post}(-i)}$ denotes the inference given test data (i) and training data (all data except for i)), are generated, and for each generated inference, S simulations $\{\theta^s_{-i}\}_{s=1}^{S}$ (the simulations given test data (i) and training data (all data except for i)) are drawn. The Bayesian leave-one-out estimation for out-of-sample predictive fit is

$$\text{lppd}_{\text{loo-cv}} = \sum_{i=1}^{n} \log P_{\text{post}(-i)}(y_i), \tag{2.94}$$

and it can be approximately computed by

$$\hat{\text{lppd}}_{\text{loo-cv}} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} P(y_i|\theta^s_{-i})\right). \tag{2.95}$$

Since each prediction is calculated by $n-1$ data points, underestimation of the predictive fit is

caused. If the number of data points n is large, the underestimation can be neglected. However, if n is small, a first order bias correction can be adopted. The details of the bias correction can be found in [12], Section 7.2.

The leave-p-out cross-validation partitions n data points into training set with $n - p$ points and testing set with p points. C_p^n times inferences are carried out overall.

Chapter 3

Myocardial Perfusion DCE-MRI

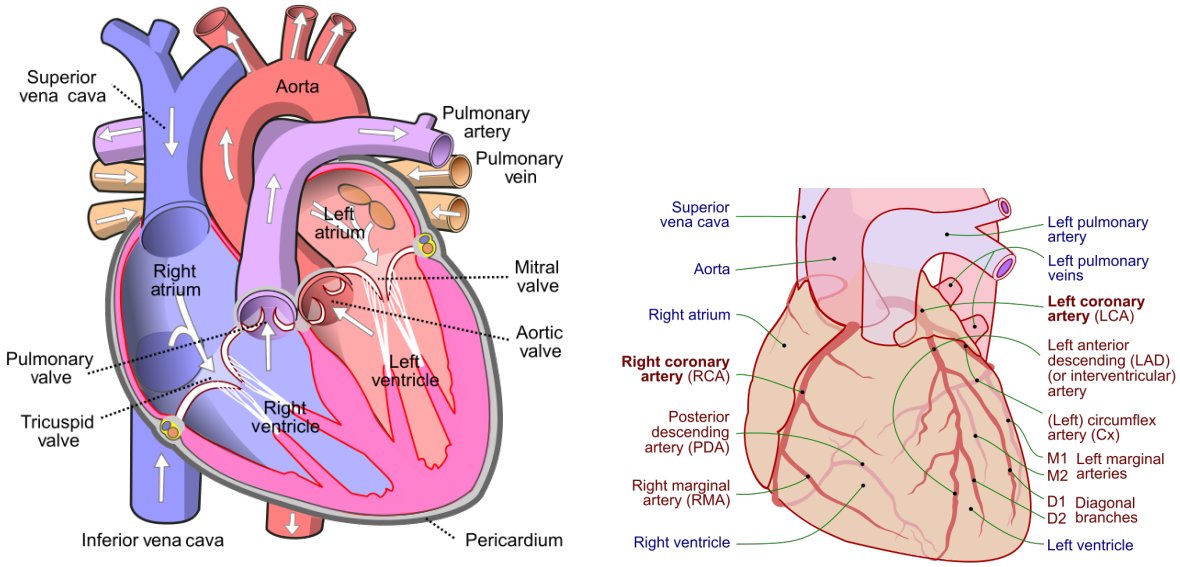
Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is widely used as a non-invasive approach to assess coronary heart disease (CHD) [8]. It can obtain multiple images of the region of interest (ROI) (e.g. left ventricular myocardium) before, during and after intravenous administration of exogenous MRI contrast agent (see Figure 3.7). The intravenously injected contrast agent alters image signal intensity so that the hypo-perfused region can be detected. Typically, three positions, i.e. apical, mid-cavity and basal slices, are acquired. The images shown in Figure 3.7 are examples of T1-weighted myocardial perfusion DCE-MRI. Gadolinium-based contrast agents (GBCA) shorten local T1 (spin-lattice or longitudinal) relaxation time in proportion to local GBCA tissue concentration. In this chapter, the background information pertaining to myocardial perfusion DCE-MRI and basic heart physiology is reviewed.

3.1 Heart physiology and coronary heart disease

3.1.1 Cardiac anatomy

The human heart consists of four chambers, the right and left atria and the right and left ventricles (see panel (a), Figure 3.1). The function of the heart can be described by two circulation systems, i.e. systemic circulation and pulmonary circulation. For systemic circulation, the heart provides oxygen-rich blood through the aorta from the left side of the heart to other organs of the human body. For pulmonary circulation, the right side of the heart receives low oxygen blood from the circulation system and pumps it into the lungs, where the blood receives oxygen and gives off carbon dioxide. The left ventricle wall is thicker than the right ventricle wall because the left side of the heart has a higher workload. More details of the heart physiology can be found in [43].

Since the nutrients cannot diffuse through the heart quickly, the coronary arteries, the blood vessels wrapping the heart, provide blood. The details of the coronary circulation are different



(a) Diagram of a human heart

(b) Coronary circulation

Figure 3.1: Diagram of human hearts. Panel (a)[1] shows the major components of a human heart which include four chambers, main blood vessels, the direction of blood flow, etc. Panel (b)[2] shows the coronary arteries of a human heart.

between individuals, but the right coronary artery (RCA), left anterior descending (LAD) and left circumflex (LCX) are the three main branches of the epicardial blood supply network (see panel (b), Figure 3.1).

3.1.2 Coronary heart disease

Coronary heart disease (CHD) is a major cause of death all over the world [44]. 382,820 deaths in the US were caused by CHD in 2020 [7] . Myocardial ischaemia (inadequate perfusion or capillary blood flow) is implicated in many forms of CHD. In particular, impaired blood supply to the myocardium occurs as a consequence of the narrowing of one or more coronary arteries. The limited supply of blood causes hypoxia (reduced oxygen supply) in the myocardium, especially under stress (either physical or pharmacological). Chronic hypoxia causes the dramatic reduction in cell function, and the patient may suffer angina pectoris. However, if the blood supply to these cells has been re-established, they will recover. On the other side, complete obstruction of the blood flow, e.g. thrombus formation on a vulnerable plaque, may cause myocardial infarction (heart attack). The myocytes in the affected region will die and are replaced by the non-contractile scar tissue. If the patient survives after the myocardial infarction, the lesion can cause the heart to have an abnormal contractile function.

[1]The source of the image is from Wikipedia by user Wapcaplet, and the links is: `https://en.wikipedia.org/wiki/Heart#/media/File:Diagram_of_the_human_heart_(cropped).svg`.

[2]The source of the image is from Wikipedia contributed by Patrick J. Lynch, and the links is: `https://en.wikipedia.org/wiki/Coronary_arteries#/media/File:Coronary_arteries.svg`.

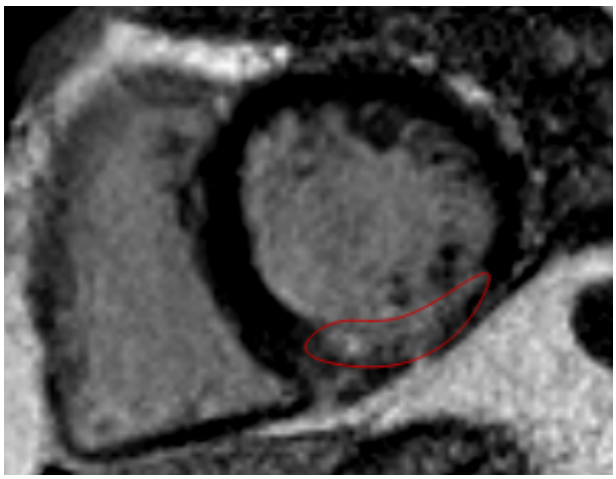


Figure 3.2: Myocardial lesion in a LGE image. The source of the image is from the datasets of the PhD program. The tissues inside the red line denote the myocardial lesion in a LGE image.

There are many medical therapies to prevent myocardial ischaemia. For example, lifestyle changes, e.g. exercises, smoking cessation, diet improvements, can slow the progress of ischaemia [45]. Cholesterol lowering agents can decrease the probability of the generation of blood clots [46]. Beta blockers can reduce oxygen demand and decrease hypoxia by lowering heart rate [46]. Nitrates can increase blood supply by vasodilation (widening of the blood vessels) [47].

Mechanical interventions, e.g. physically restoring the blood flow to the heart, can also prevent myocardial ischaemia. Some surgical approaches are applied to patients who are at a high risk of myocardial infarction. Percutaneous transluminal coronary angioplasty (PTCA) involves insertion of a catheter into an artery in the arm or leg of the patient. Under X-ray guidance, the catheter is moved to the stenotic site, and an inflated balloon is applied to squash the plaque against the artery wall. To prevent restenosis after the PTCA, a stent is deployed in the artery to keep the artery open. A coronary artery bypass graft (CABG) makes use of an artery or vein from another part of the body to connect the aorta directly to the stenosis region.

It is beneficial to re-establish the blood supply to living but hypo-perfused tissues. However, if the tissue is infarcted, i.e. non-viable, it does not help to restore the blood supply. Furthermore, the risk imposed by the arterial catheterization procedure is small but significant. Therefore, it is crucial to avoid unnecessary surgery. MRI can be used to assess the viability of myocardial tissue. The lesion shows as a bright zone on late gadolinium enhanced (LGE) images (see Figure 3.2). These images offer vital diagnostic information about whether or not the patient will experience improvement after the restoration of blood supply [48].

3.2 Myocardial perfusion DCE-MRI: artifacts

The myocardial perfusion DCE-MRI data can be affected by several different kinds of artifacts. These artifacts could cause errors in the results of quantitative analysis. Generally, the artifacts stem from two sources: the acquisition system and the patient. Both types of artifacts can cause low quality imaging and further affect the accuracy of the quantitative analysis. In this section, major artifacts and their corresponding remedies are introduced.

3.2.1 Surface coil inhomogeneity

Surface coils are applied to improve the quality of the MR images, i.e. to improve the signal-to-noise ratio (SNR). However, since the sensitivity of the coils decreases as the distance from the coils increases, the generated signals are distributed inhomogeneously [49, 50]. One straightforward method is to partially compensate for the inhomogeneity by putting two surface coils on both sides of the body. For myocardial perfusion DCE-MRI, making use of adiabatic (B1-insensitive) radio frequency (RF) inversion or saturation pulses, or composite pulses is the most common method to solve the inhomogeneity [8]. After the acquisition of the images, baseline correction can also be used to correct for inhomogeneity. Specifically, since the surface coil inhomogeneity is spatially distributed within the myocardium for all time points, this inhomogeneity can be mapped out from the pre-contrast images. Therefore, the myocardial signals can be normalised by the pre-contrast values, and the effect of surface coil inhomogeneity will be reduced [8, 49].

3.2.2 Dark rim artifacts

The dark rim artifacts describe a phenomenon that pixels close to the endocardium boundary exhibit extremely low signal intensities during the first pass of the contrast agent. These pixels might be mistaken for hypo-perfusion and sub-endocardial defects. There are two main differences between dark rim artifacts and genuine hypo-perfusion defects. Firstly, the dark rim artifacts usually only last for a few heartbeats, but the genuine hypo-perfusion defects usually show persistently low signal intensities [51]. Secondly, the dark rim artifacts can cause the signal intensities of the affected pixels to decrease below the baseline (pre-contrast) values [52]. There are multiple factors that contribute to the appearance of the dark rim artifacts, and the most prominent amongst them are listed below:

- The truncation or the non-uniformity across k-space causes the distortion of point spread function, which would cause signal variations adjacent to the boundaries showing bright or dark bands [53]. This effect becomes more substantial when a higher dose of contrast agent is used. Spatial smoothness methods, e.g. window function, are used to reduce the effect of this factor [54].

- The spill-over of signals from neighbouring pixels will also cause the dark rim artifacts [53, 54]. Adding a spill-over term to the myocardium contrast enhancement model would solve this problem [55].
- The motion of the object during the acquisition of the images will cause discontinuities in k-space, and band effect occurs at the boundary of the tissues [56].
- The increase of magnetic field distortions around the boundaries of the tissues may also cause the dark rim artifacts [52, 57].

3.2.3 Motion artifacts

During the acquisition of the myocardial perfusion DCE-MRI, the deformation or the displacement of the heart causes motion artifacts. The two main causes of the deformation or the displacement of the heart are cardiac motion and respiratory motion. The motion artifacts lead to blurring, mis-registrations and deformation of the images, and further lead to inaccurate quantitative analysis [58].

The cardiac motion is caused by the pumping action of the heart. Theoretically, this kind of artifact can be eliminated by making use of electrocardiogram (ECG) to determine at what phase of the cardiac cycle the heart images will be obtained [59]. Nevertheless, it is still possible that the trigger time has been set wrongly, and the deformation or blurring of the heart is generated.

Respiratory motion artifacts are most common patient-related artifacts in myocardial perfusion DCE-MRI. The effects of the respiratory motion have been studied in [60, 61]. Moreover, the quantification of rigid and non-rigid motion of the heart caused by respiratory motion has been studied in [62]. During the acquisition of myocardial perfusion DCE-MRI, the patients are usually asked to hold their breath during the contrast wash-in phase to reduce the effect of the respiratory motion. However, for a patient with cardiac heart disease, it is hard to hold the breath for a long period of time. This may cause the deformation or blurring of the heart images.

The effects of all of the artifacts mentioned in this section should be minimised prior to image processing. Some solutions are applied during the acquisition of the images, e.g. ask patients to hold their breath, and some solutions are applied after the acquisition of the images, e.g. normalisation of myocardial signals using baseline values. In Section 3.3, the contouring method for the segmentation of the myocardium image will be introduced to reduce the effects of artifacts further.

3.3 Myocardial perfusion DCE-MRI data

The artifacts that may affect myocardial perfusion DCE-MRI have been introduced in the previous section. In this section, a brief introduction to the myocardial perfusion DCE-MRI data and the contouring method for the myocardium image are described.

Table 3.1: The scan parameters of the myocardial perfusion DCE-MRI used in this study

TR (repetition time)	194.00ms
TE (echo time)	0.98ms
flip angle	8°
TS (saturation time)	101.00ms
Matrix size	92×128 (reconstruct to 184×256)
Slice thickness	8mm
mean FOV (Field of view)	266.43mm $\times$ 371.54mm
FOV range	223mm - 359mm $\times$ 320mm - 500mm

.

3.3.1 Introduction of the data

The cardiac MRI exams were performed with a Siemens MAGNETOM Avanto (Erlangen, Germany) 1.5-Tesla scanner with a 12-element phased array cardiac surface coil in Golden Jubilee National Hospital, Glasgow, UK at 2011. The assessment of resting myocardial perfusion was performed during intravenous administration of 0.075 mmol/kg of contrast agent (gadoterate meglumine, Dotarem, Guebert S.A.).

In this study, the myocardial perfusion DCE-MRI data were obtained in 30 patients who suffered myocardial infarction. For each patient, longitudinal data were acquired on day 1, day 2, day 7 and day 180 post myocardial infarction. Each DCE-MRI dataset was indexed by the patient number $\{i\}_{i=1}^{30}$ and the scan number $\{j\}_{j=1}^{4}$. Three DCE-MRI datasets were not available and their indexes are $(i = 13, j = 1)$, $(i = 16, j = 4)$ and $(i = 22, j = 3)$. For each DCE-MRI dataset, three slices, i.e. apical, mid-cavity and basal, have been acquired. For each slice, 21 - 60 dynamic frames were acquired. The structure of the data is illustrated in Figure 3.3. The scan parameters are shown in Table 3.1.

Since the signal-to-noise ratio (SNR) for the mid-cavity slice is usually the highest, all analyses in this study are applied to the mid-cavity slice. However, it is worth mentioning that all methods used in this study are generic to all three slices. In specific, the methods introduced in Chapter 5 are applied to a single image. The methods developed in Chapters 6 and 7 are applied to a single DCE-MRI dataset.

3.3.2 Myocardium contouring

All data used in this thesis are pre-processed by manual contouring to define the left ventricle (LV) myocardial wall. Specifically, for each DCE-MRI dataset, the endocardium and the epicardium are manually drawn by the author under supervision of an expert using a cardiac image analysis software (Mass 8.1, Medis, Leiden University, Leiden, The Netherlands). Figure 3.4 shows an example of the delineated LV myocardium. The green curve and the red curve denote the epicardial and the endocardial borders, respectively. The yellow square represents an example region of the LV blood pool. The blue cross is a reference point chosen as the infe-

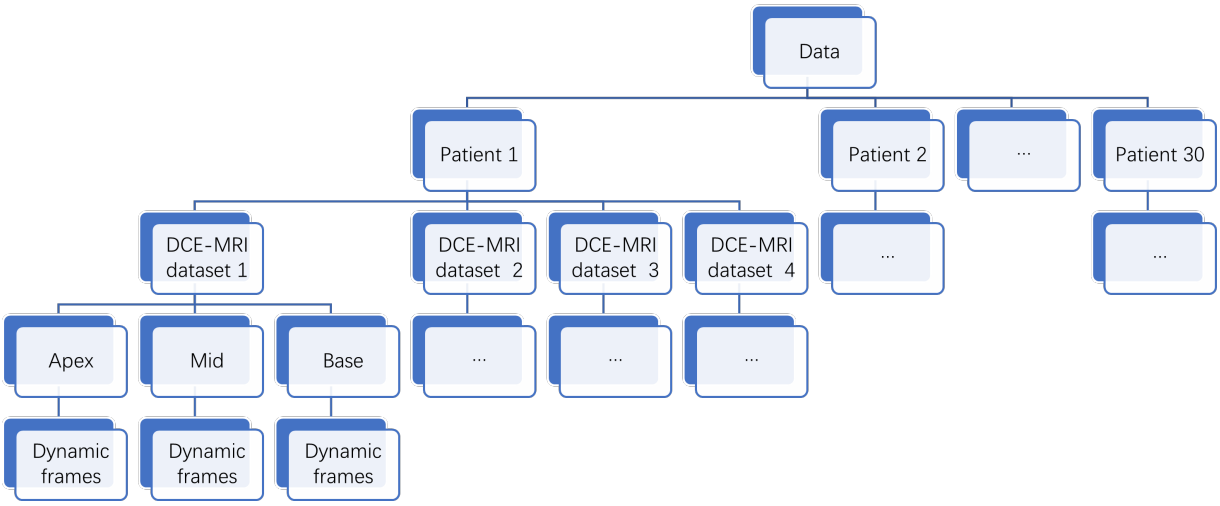


Figure 3.3: The structure of the data used in this study is shown. Specifically, there are thirty patients overall. For each patient, there are four DCE-MRI datasets. For each DCE-MRI dataset, there are three slices. For each slice there are $21-60$ dynamic frames.

rior insertion point, which is located at the junction of the right ventricle and the left ventricle of the inferior. The region between the red curve and the green curve is the LV myocardium. The five white lines and one blue line separate the myocardium into six regions (60 degrees for each segment) which is the regional segmentation suggested in [4]. The blood pool ,i.e. represented by the yellow square ROI, is used to represent the arterial input of the blood supply to the myocardium. The blue reference point is used to locate the relative position of the heart and therefore ensures to subdivide the myocardium into circumferentially equidistant regions suggested in [4].

In this study, the myocardium contouring was performed manually based on the maximum enhancement (ME) image. Then, the contours are replicated to all images in the same DCE-MRI dataset. Rigid translations are allowed to correct for motion artifacts (see Section 3.2.3) during the acquisition of the images. In other words, the motions of the myocardium during the acquisition of the images are assumed to be limited to rigid translation. However, in many cases, non-rigid deformation and rotation can be observed. This limitation may cause inaccurate analysis. However, if the contours are drawn separately within every dynamic frame, it is difficult to accurately contour the endocardium during the pre-contrast (baseline) phase. Figure 3.5 shows an example of the dynamic image frame acquired prior to the arrival of the contrast into

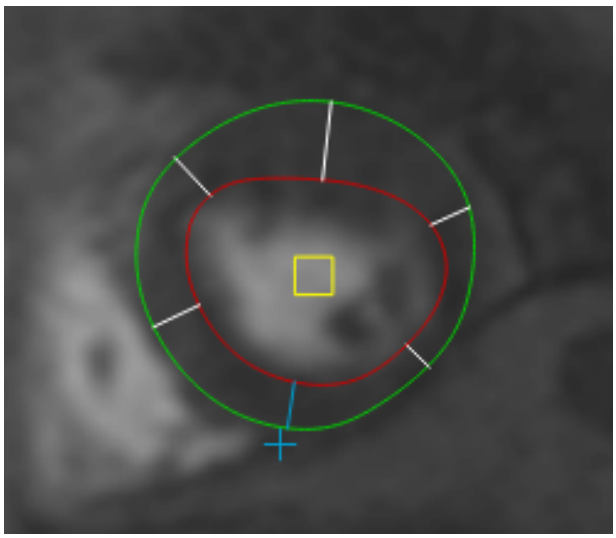


Figure 3.4: Myocardial perfusion DCE-MRI contours. The green curve denotes the LV epicardial border (including the septal wall). The red curve denotes the LV endocardial border. The tissues inside the yellow square denote the LV blood pool sample, and the blue cross denotes the reference point.

the LV cavity. It can be observed that the endocardial border is unclear, and it is hard to draw contours based on this image. Moreover, for the dynamic image frame with non-rigid contours, pixelwise analysis will be impossible. Although some projection methods, e.g. the method illustrated in Chapter 8, can be applied to tackle this problem, there still will be some information loss because the projection methods cannot reproduce all information (see details in Chapter 8). Therefore, in this study, the contours are drawn using the ME image and rigid translations are applied to compensate the motion artifacts.

Since the dark rim artifacts may cause errors in both quantitative and semi-quantitative analysis, they are avoided by excluding the dark rim region from the myocardium. Figure 3.6 shows the dark rim regions in different points. The region inside the orange contours is a suspicious dark rim region. It can be seen that the signal intensities in the dark rim region in panel (a) are visually lower than the signal intensities of other myocardial pixels. However, the signal intensity of the dark rim region in panel (b) are not visually lower than the signal intensity of other myocardial pixels. Therefore, this region is excluded from the myocardium. Although this may generate some losses to the myocardial tissues, and failure to detect some subtle subendocaridal perfusion defects, this approach minimizes the probability of false positive findings.

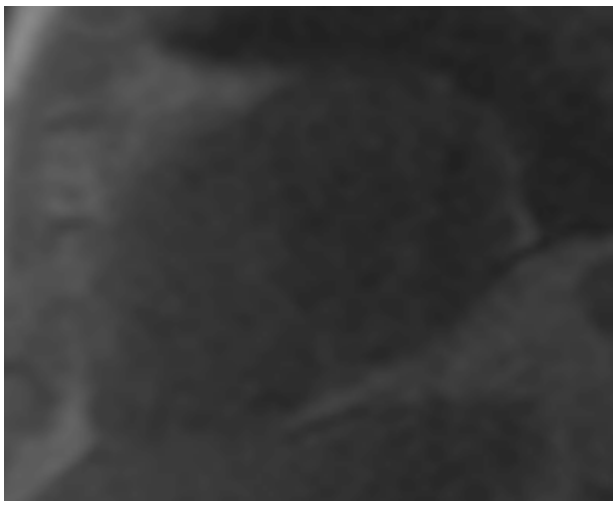


Figure 3.5: Myocardial perfusion DCE-MR image before the contrast agent flows into the heart. This image is from the same DCE-MRI data as Figure 3.4.

3.4 Diagnosing coronary heart disease using myocardial perfusion DCE-MRI

The myocardial perfusion reserve (MPR) is a criterion to diagnose coronary heart diseases. The MPR is the ratio of stress and resting myocardial blood flow (MBF) estimates.

$$\text{MPR} = \frac{\text{Stress MBF}}{\text{Resting MBF}} \tag{3.1}$$

The MBF can be obtained by quantitative or semi-quantitative approaches, which will be explicitly illustrated in Chapter 4. The rationale for expressing perfusion measures using MPR is based on the concept of coronary flow reserve (CFR). The ratio of stress and resting coronary flow measures the ability of the coronary system to maintain adequate blood flow in the presence of stenosis. Specifically, given stenosis in the coronary artery, the coronary circulation will compensate for the increased resistance to blood flow by vasodilation, and this ability is measured by the CFR. The MPR is an analogous measure to CFR and has been proved to be effective evidence of a diagnosis of CHD [63–66].

The MBF is also a criterion to diagnose coronary heart disease. To quantitatively or semi-quantitatively analyze the MBF, the time signal intensity curve (TIC) needs to be obtained. A ROI, i.e. the myocardium, is drawn on each image of the dynamic series. For each pixel located within the myocardium, its signal intensity can be plotted at each time point. In panel (f), Figure 3.7, the TICs of two specific pixels are plotted. One located in the hypo-perfused region, and the other located in the normal-perfused region, are plotted. The pixel from healthy tissue has both a larger maximum value and higher growth rate (gradient or up-slope) of signal intensity than the pixel from lesion during the first pass of the contrast agent, i.e. time 18 to

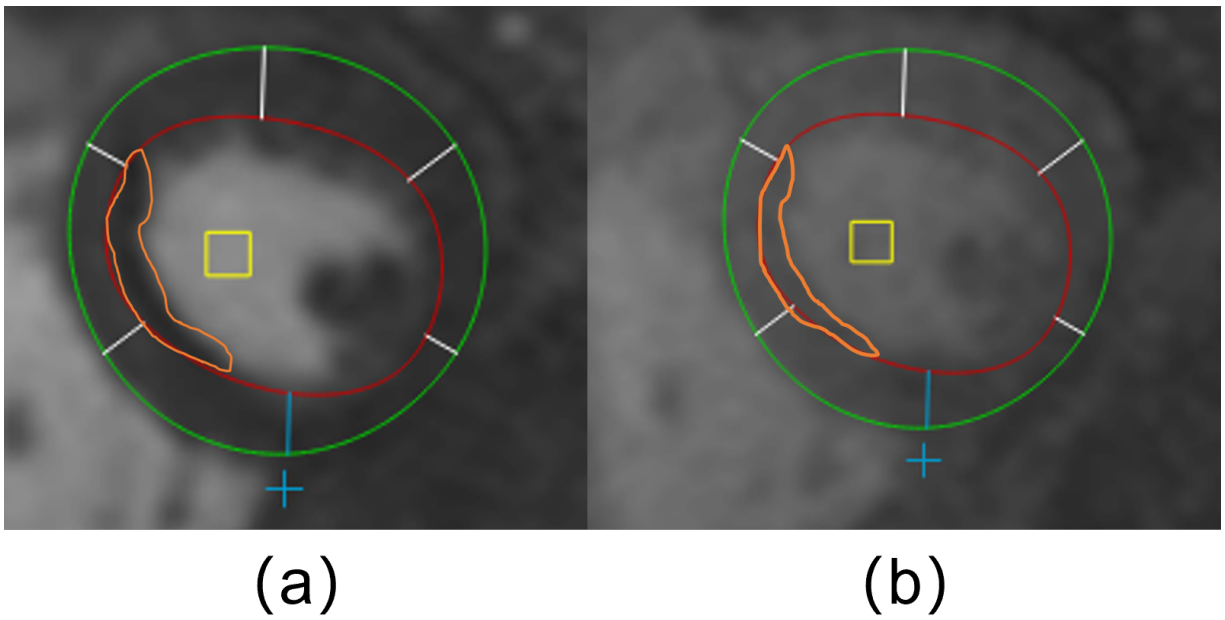


Figure 3.6: Dark rim regions in different time points
Panel (a) shows the image when the LV is enhanced. Panel (b) shows the image when the myocardium is enhanced. The pixels inside the orange contours are suspicious dark rim region.

37 in panel (f), Figure 3.7. Meanwhile, the TIC of a ROI located in the blood pool represents the contrast agent passing through the ventricular blood pool (i.e. arterial input), and the TICs for myocardial tissues represent the contrast agent remaining within the myocardium. For both myocardial TIC and blood pool TIC, there are initial dynamic frames that are named "baseline", i.e. time 0 to 12 in panel (f), Figure 3.7. This segment indicates the baseline signal intensity before the contrast agent reaches the ventricles. The models using TICs to estimate the MBF quantitatively or semi-quantitatively are introduced in Chapter 4.

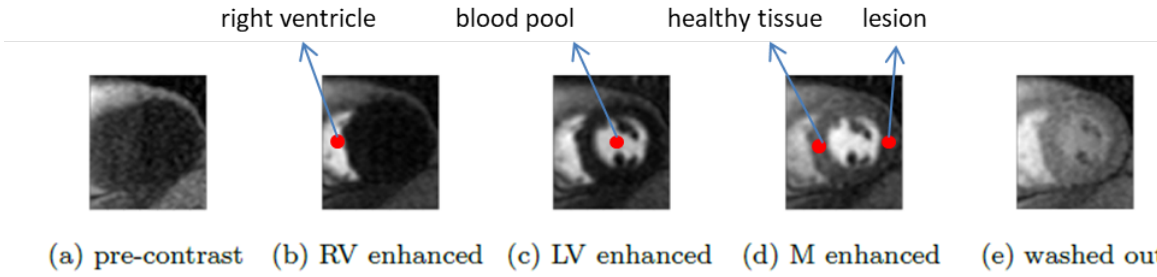
There are many different approaches to estimate the MBF. For semi-quantitative analysis, up-slope [38, 67] and MPR [68–70] are usually applied to estimate the MBF. For quantitative analysis, the Fermi-constrained method [71–73], the model independent method [74–76], the tracer-kinetic modelling [77,78], the one compartment model [79,80] and the uptake model [81] have all been used to estimate MBF. There are numerous studies reporting the results of the comparisons among these quantification or semi-quantification methods [64, 80, 82]. One study has reported that different methods are not significantly different [80], but one study has reported that different methods are significantly different under certain circumstances [82]. Specifically, for the rest data, there are no differences between the four methods. But for the stress data, the Fermi method significantly overestimate the MBF. The principal aim of this thesis was to develop methods for classification of the results of MBF estimation, regardless of their nature and particulars of implementation. The proposed classification methods are generalisable, and can be applied to TIC data (with or without some form of calibration), and also to a variety of

perfusion-related indices, such as up-slope, MPR and MBF. The classification algorithms result in a simple, dichotomous grouping of tissue elements into the healthy tissues group and the lesions set. The proposed classification results are designed to be independent on the absolute values of underlying signal intensities or derived TIC parameters, as long as their relative ranking between different pixels remains unchanged. In other words, even if a particular quantification method will overestimates the MBF systemically, e.g. Fermi method [82], the classification approaches can still generate correct results based on the systemically overestimated MBF map, because the systematic overestimation issue happens to affects to all pixels. Similarly, the non-linearity relationship between signal intensity and contrast concentration will cause an overestimation issue (see details in Section 3.5), but it will little affect the classification results.

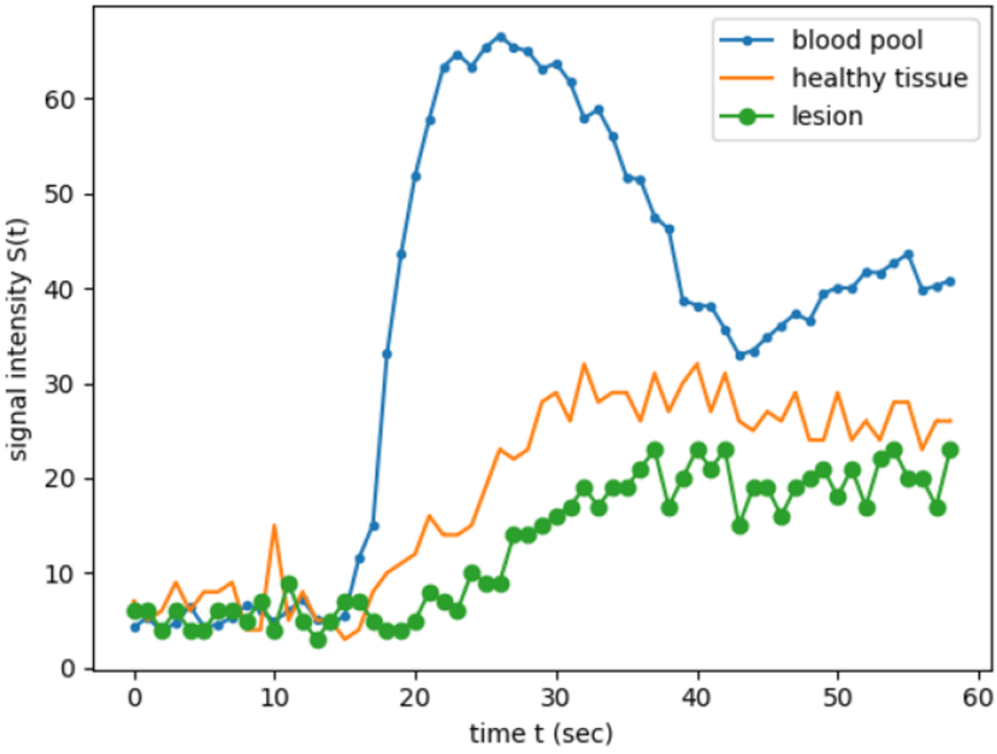
Most applications of the methods mentioned in the last paragraph are region-based using the American Heart Association (AHA) segmentation (see details in Section 3.6). However, the definition of discrete myocardial segments is subject to operator bias, and there is an inevitable loss of granularity caused by signal averaging. The pixel-based method overcomes this drawback fundamentally. Thus, methods for classification of the results of the existing quantification and semi-quantification methods described in this thesis are designed to be performed on a pixel level.

3.5 Signal intensity to contrast agent concentration

The gadolinium concentration in local tissue has a non-linear relationship with the corresponding signal intensity when the contrast agent dose is high [83,84]. The degree of the non-linearity depends on the dose and rate of the contrast agent injection. Specifically, when the dose is low, e.g. 0.01 mmol/kg, the relationship between signal intensity and contrast concentration is approximately linear [84]. However, this low dose dramatically reduces the signal-to-noise ratio (SNR) and the concentration-to-noise ratio (CNR). Since the myocardium has a lower concentration compared with the arterial input function (AIF) in the LV blood pool, the signal changes within the myocardium are highly affected by the noise. This causes inaccurate estimations of the MBF. With the dose of contrast agent getting higher, the relationship between signal intensity and contrast concentration turns to be significantly non-linear, and eventually, the blood pool signal will be saturated. This will lead to an error of the peak signal intensity and therefore cause a systematic overestimation of the MBF. However, this will not affect the relative distributions between the signal intensity and the contrast concentration since the signal intensity is a monotonously increasing function of contrast concentration over the relevant range of concentrations within the myocardial tissue. Therefore, the non-linearity issue will not cause errors systemically when the analysis aims to detect the extent of the hypo-perfusion region, i.e. classification of pixels into healthy and lesion groups, which is based on the relative distributions of



(a) pre-contrast (b) RV enhanced (c) LV enhanced (d) M enhanced (e) washed out



(f) signal intensity S(t)

Figure 3.7: Myocardial perfusion DCE-MRI. Panels (a) - (e) show 5 steps of myocardial perfusion with time. The signal intensity in right ventricle (RV), left ventricle (LV) and myocardium (M) are enhanced over time. (f) shows DCE-MRI signal intensity S(t) before, during and after administration of MRI contrast agent. The three curves represent signal intensity changes in three regions of interest: LV blood pool, lesion and contralateral "healthy" tissue.

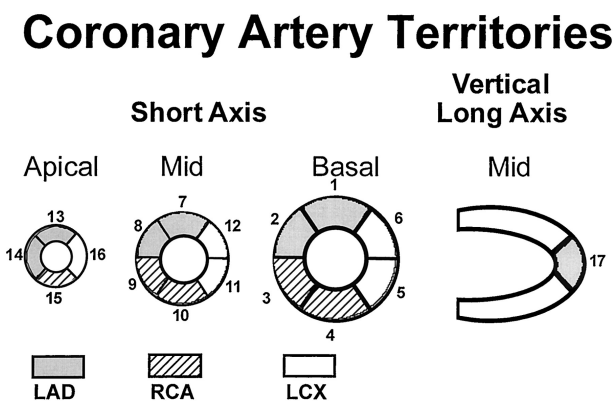


Figure 3.8: The 17 segments and their corresponding main branches of the coronary artery network. The source of this figure is from [4].

the MBF for all myocardial pixels.

All myocardial perfusion DCE-MRI datasets are obtained with the contrast dose 0.075 mmol/kg in this study. Local tissue concentration of the contrast agent at time t, $C(t)$, is approximated by the difference between the post-contrast signal $S(t)$ and the baseline signal $S(0)$:

$$C(t) \approx \Delta S(t) = S(t) - S(0). \tag{3.2}$$

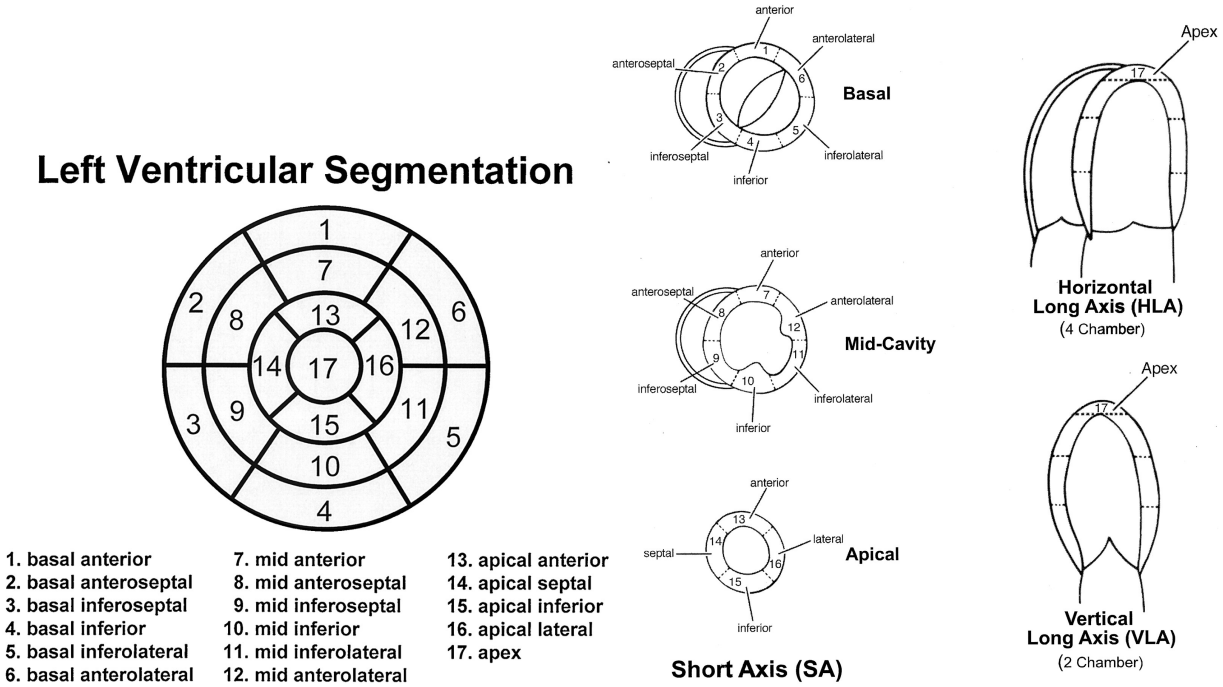
As a result of the non-linear relationship between $C(t)$ and $S(t)$ in equation (3.2), the MBF values will be systematically overestimated. While, the assumption of the linearity will affect the absolute values of MBF but will not affect their relative distribution if $S(t)$ is a monotonously increasing function over a relevant range of $C(t)$ [85] (Chapter 2).

3.6 The AHA mapping

A standardised regional segmentation method for heart images has been proposed by the American Heart Association (AHA) Writing Group, and is commonly referred to as the AHA segmentation model [4]. The AHA mapping is widely adopted in the cardiac imaging community. Although the methods illustrated in this study are pixel-based and the AHA mapping is region-based, it is still worth explaining this method because the rationale of the AHA mapping is also relevant for the interpretation of the pixel-based segmentation.

Based on the AHA mapping, the myocardium of the DCE-MRI is divided into 17 segments. The myocardium in the basal slice is separated into 6 segments and indexed as numbers $1-6$. In the mid-cavity, the myocardium is separated into 6 segments and indexed as numbers $7-12$. In the apical slice, the myocardium is separated into 4 segments and indexed as numbers $13-16$. The 17^{th} segment is located in the apical cap. Figure 3.9 shows the details of the 17-segment system. The source of Figure 3.9 is from [4].

The blood supply to the myocardium is delivered by three main branches, i.e. the right



(a) 17-segment system overall

(b) 17-segment system given different planes of the images

Figure 3.9: The AHA 17-segment system. Panel (a) shows the suggested nomenclature for tomographic imaging of the heart and the relative positions for 17 segments. Panel (b) shows the vertical long-axis, horizontal long-axis and short-axis planes. The names and locations for different segments in basal, mid-cavity and apical slices are also marked in this panel. The sources of panels (a) and (b) are from [4].

coronary artery (RCA), the left anterior descending (LAD) and the left circumflex (LCX), as mentioned in Section 3.1.1. An important aspect of the AHA mapping method is that this method can link the three main branches with the 17 segments. In other words, each segment corresponds to one of the three main branches under the AHA mapping. Figure 3.8 shows the details of the correspondences between 17 segments and the three main branches of the coronary artery.

In the AHA mapping framework, the ischaemia of the myocardium can be diagnosed by comparing the MBF values with a given threshold for different segments. Specifically, if the MBF value of a segment is lower than a given threshold (defined by the clinician), the myocardium in this segment is suspected to be affected by ischaemia. Moreover, if the hypo-perfusion is visually observed, the AHA mapping method can be used to locate the position of the hypo-perfusion area and identify the likely location of the stenosis within a particular coronary artery.

Chapter 4

Quantitative Myocardial Blood Flow

Myocardial blood flow (MBF) is an important diagnostic indicator of coronary heart disease, as described in Section 3.4. In this chapter, both semi-quantitative and quantitative methods for the estimation of MBF are introduced. Furthermore, the time-intensity curve (TIC) is also described in this chapter, as TIC is an essential component of these methods.

4.1 Time-intensity curve

A time-intensity scatter plot shows the variation of signal intensity (SI) for a specific pixel or a specific region of interest (ROI, an average of a group of pixels) over time. Various methods can be employed to fit the time-intensity points, and the fitted curve is referred to as TIC. Figure 4.1 shows the TICs fitted by different methods, i.e. polynomial regression, Gamma variate regression and Gaussian process regression, for a pixel in the myocardium and a group of pixels in the LV blood pool. The details of these regressions will be introduced in Section 4.2.

For the pixels in both the LV blood pool and the myocardium, their SIs have a period when the values remain low and fluctuate around the value we refer to as "baseline" (from time 0 to time 15 in Figure 4.1). During this time, the contrast agent has not yet reached the ROI, e.g. LV the blood pool and the myocardium. The baseline is critical because it represents the state of the ROI before the arrival of contrast agent, and therefore it can be used to normalize the SI after the contrast agent flows into the ROI. In this study, the baseline is measured by a least-squared method given a biexponential function [86], in which the signal intensity $S(t)$ is expressed as

$$S(t) = \begin{cases} C & t \leq t_0 \\ C + \dfrac{p_2 p_3}{p_2 - p_1} \times \left(e^{-p_1 t} - e^{-p_2 t}\right) & t > t_0 \end{cases} \tag{4.1}$$

where C is the baseline and t_0 is the end time point of the baseline. C, t_0, p_1, p_2 and p_3 can be estimated by the least-square regression. The SSE in equation (2.73) is used to select the best combination of the parameter estimates.

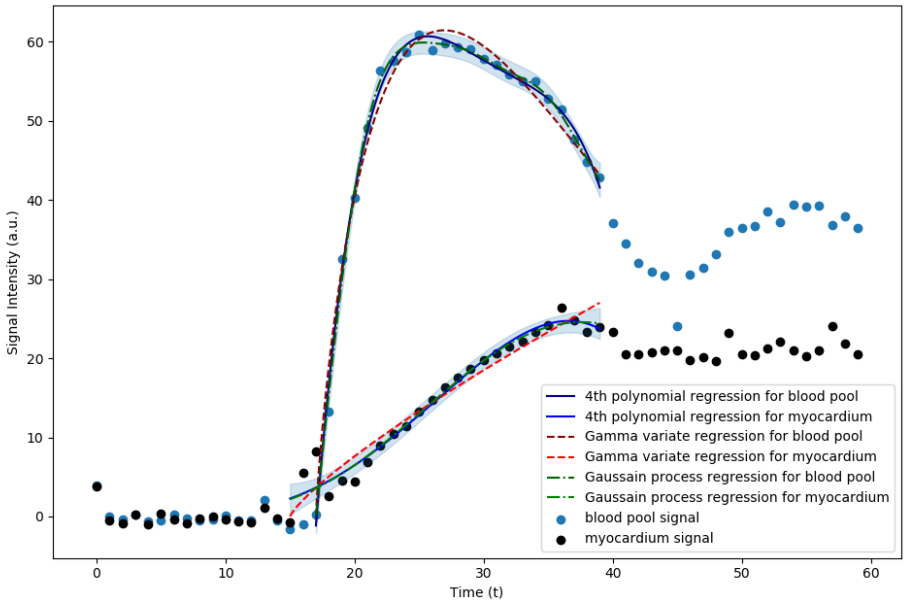
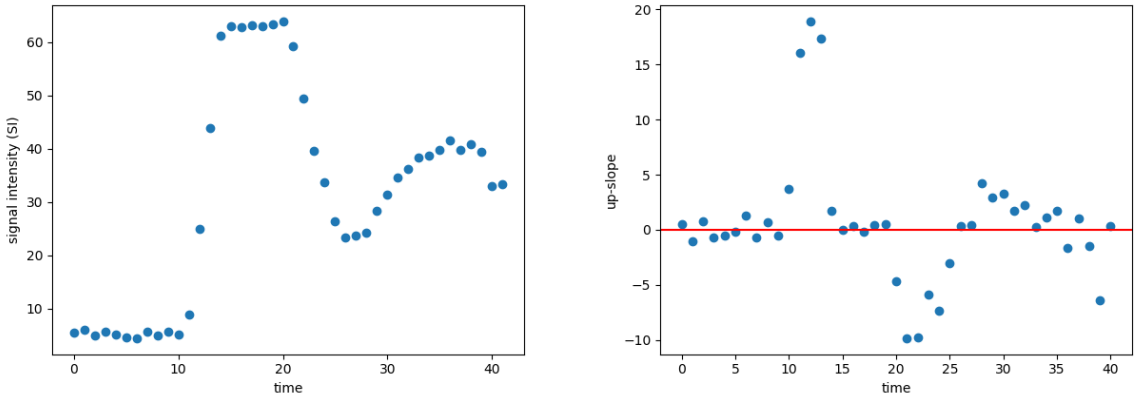


Figure 4.1: Time-intensity curve (TIC) fitted by polynomial regression, Gamma variate regression and Gaussian process regression. The shadow area is the 95% confidence interval predicted by the Gaussian process regression. The x-axis shows the time point of the SI. Only the first pass phases of the curves have been fit.

Generally, the perfusion process in the blood pool can be separated into two phases. The first phase is the pre-contrast phase, and the SIs fluctuate around the baseline. Next comes the first-pass of the contrast agent through the blood pool. However, the data may also contain the second-pass of the contrast agent through the blood pool. If the second-pass data is included for some specific methods, e.g. Fermi-constrained deconvolution (see Section 4.3.3), poor fitting and inaccurate estimations may occur. Therefore, it is necessary to detect the end time point of the first-pass and discard the second-pass data. In this study, the end time point of the first-pass is detected by up-slope values. The up-slope is the first derivative of the TIC. The end time point of the first-pass is the time point that the first positive up-slope value appears after the minimal negative up-slope appears. Specifically, the first positive up-slope time point after the minimal negative up-slope time point represents the time that the SI stops decreasing and it increases. This means the second-pass starts. Figure 4.2 shows an example of the end time point detection. In panel (a), Figure 4.2, the $S(t)$ from the blood pool is shown. The up-slope in panel (b), Figure 4.2, is obtained in a simple way: the value of up-slope at time point i in panel (b) is calculated by the differences of SI between time point $i+1$ and i in panel (a). There are many more precise method to obtain the values of up-slope, e.g. moving average filter. However, the sign of the up-slope is interested here, and therefore the simple proposed method is enough to detect the sign of the up-slope. Visually, the minimal up-slope in panel (b) can be



(a) time-intensity points (b) up-slope

Figure 4.2: time-intensity points and up-slope for the blood pool. Panel (a) shows time-intensity points for the LV blood pool, and panel (b) shows the corresponding up-slope estimates. The red line in panel (b) shows the value 0.

observed at time point 21. The first positive value of up-slope after the minimal up-slope time is at time point 26. This means that the end time point of the first-pass is at time 26 in panel (a), Figure 4.2. In practice, the value of up-slope can be more accurately estimated using curve fitting method [87, 88]. However, to identify the end time point of the first-pass, the simple method mentioned above is sufficient because not the value but the sign of the up-slope determines the end time point of the first-pass.

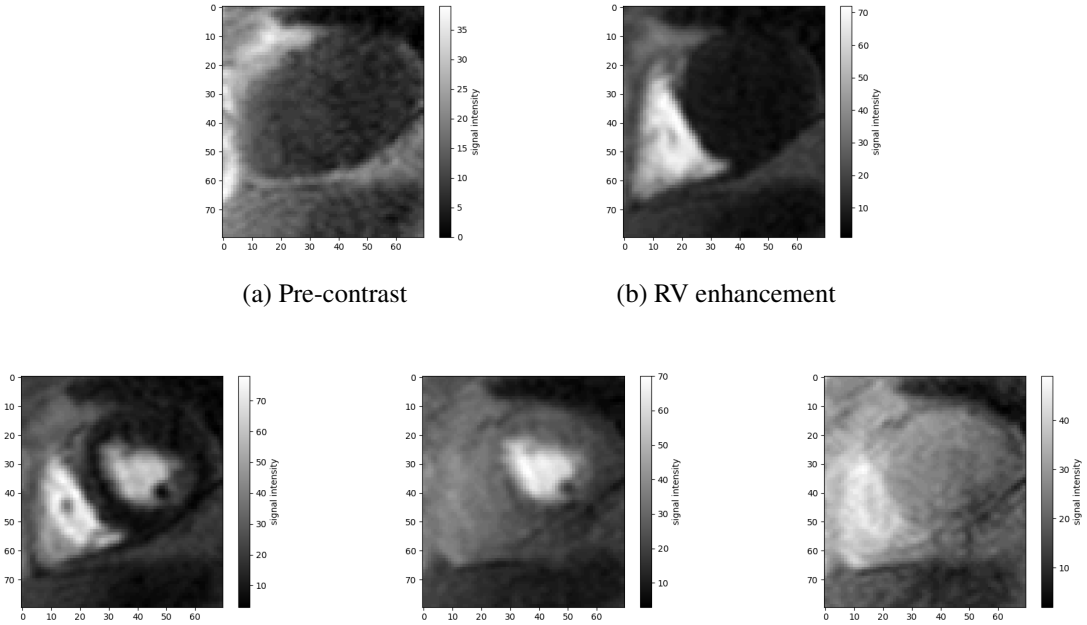
Given both end time points of the baseline and the first-pass, the complete signal variation of the first-pass of the contrast agent through the LV blood pool can be extracted from the TIC. Semi-quantitative and quantitative analysis methods are based on analysis of TIC over this limited time segment, and these methods will be explicitly illustrated in Sections 4.2 and 4.3.

4.2 Semi-quantitative analysis

Semi-quantitative analysis of myocardial perfusion has been applied to detect significant perfusion defects using DCE-MRI [49,89]. This analysis is performed with the perfusion features that are derived from the TIC (see Figure 4.1) or the time-signal points (see panel (a), Figure 4.2). Some of the most commonly used perfusion features are listed. Figure 4.3 shows the example data used in this chapter.

- Peak signal intensity (SI_{peak}, PSI) and contrast enhancement ratio (CER): the PSI is the peak value of the TIC during the first pass of the contrast agent. Generally, this feature is standardised by the baseline. Specifically, the standardised quantity can be expressed as

$$\mathrm{CER} = \frac{\mathrm{SI}_{\mathrm{peak}} - \mathrm{SI}_{\mathrm{baseline}}}{\mathrm{SI}_{\mathrm{baseline}}}, \tag{4.2}$$



(a) Pre-contrast (b) RV enhancement

(c) LV enhancement (d) Myocardium enhancement (e) Wash out

Figure 4.3: Five stages, pre-contrast, right ventricle enhancement, left ventricle enhancement, myocardium enhancement and wash out, of the myocardial perfusion DCE-MRI. All results shown in Chapter 4 are obtained from this set of data. It is emphasised that all results in this chapter are just examples for visualisation, and the methods proposed are generic to all myocardial perfusion DCE-MRI data.

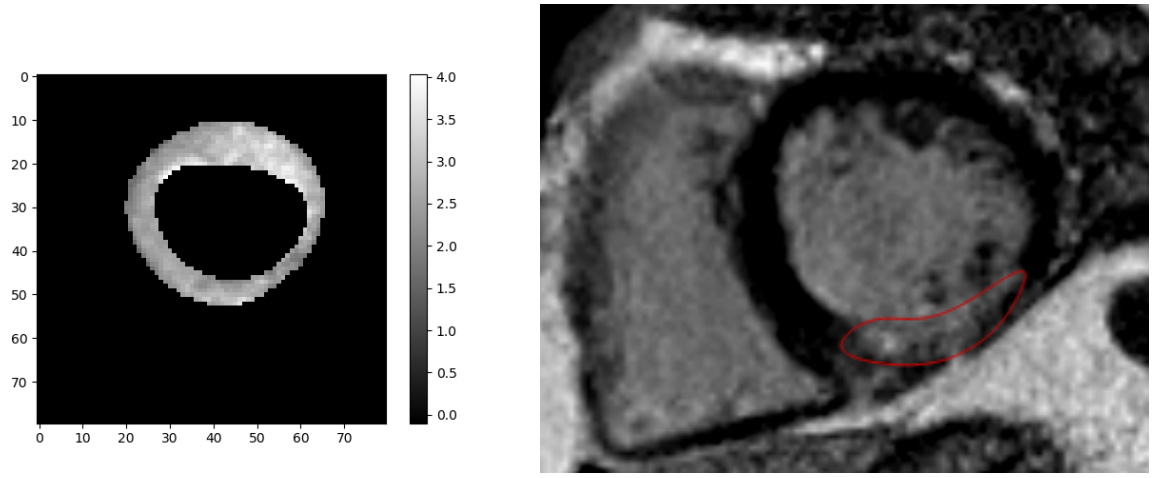
and it is named contrast enhancement ratio (CER) [90, 91] or peak contrast enhancement (pSI) [92].

- Up-slope: the first derivative of the TIC during the initial increase of the signal intensity for the ROI [65, 66]. It can be calculated by a "sliding window" method. Specifically, given a time interval d, the straight line gradient calculated from t_0 to t_d is recorded. Then, repeat this procedure to time points t_1 to t_{1+d}. the index 0 represents the end time point of the baseline. The maximum of the measured gradient is taken as the up-slope (see Section 4.2, [85]). The up-slope can also be estimated by a fitted curve. In this study, the estimations of up-slope using different regression methods are compared.

- Time to peak (T_{peak}): the time from the end of baseline to the time achieving peak signal intensity.

- Mean transit time (T_{mean}): the average time duration for a contrast agent to pass through the ROI [93].

- Myocardial perfusion reserve index (MPRI): the MPRI is estimated as the ratio of up-slopes for rest and stress myocardial perfusion DCE-MRI [49, 94]. The MPRI can be



(a) CER (b) LGE

Figure 4.4: Contrast enhancement ratio. Panel (a) shows the CER maps in the myocardium, and panel (b) shows the corresponding LGE image. The area inside the red curve in panel (b) shows the myocardial defect visually.

utilised to detect significant coronary artery stenoses with high diagnostic accuracy [94]. In this study, only resting DCE-MRI datasets were available, and therefore MPRI could not be estimated.

It is worth mentioning that the up-slope and MPRI are considered as the most critical semi-quantitative features for coronary heart diseases detection. In this study, because of the limitation of data (lack of stress data), the MPRI cannot be estimated. Therefore, the CER and up-slope are explicitly illustrated in the latter sections.

4.2.1 Contrast enhancement ratio

Contrast enhancement ratio (CER) has been extracted from Gamma variate TIC fit to derive a vital feature, mean transit time, T_{mean}, [93]. The mean transit time is highly related to the myocardial blood flow F based on indicator-dilution theory [95]:

$$F = \frac{V}{T_{mean}}, \tag{4.3}$$

where V is the distribution volume of the indicator (the contrast agent).

Figure 4.4 shows the CER map and its corresponding late gadolinium enhanced (LGE) image. In this study, the LGE image is used to be a reference of the location of the myocardial defect. From panel (b), Figure 4.4, the myocardial defect region can be visually observed in the right bottom of the myocardium in the image (the region inside the red curve). However, in the CER map, the CER values in the right bottom of the myocardium are not clearly lower than the values in other regions. According to the visual inspection, the CER values cannot truly reflect

Table 4.1: Sum of square error (SSE) for different orders of polynomial regression

	0 order	1 order	2 order	3 order	4 order
Blood Pool	5125	4162	820	194	34
Myocardium	1602	97	81	53	52

the degree of hypo-perfusion in this DCE-MRI dataset.

4.2.2 Up-slope

The up-slope can be calculated using the "sliding window" method. It can also be obtained by curve fitting [88]. Figure 4.1 shows the fitted TIC for the time-intensity points using three different methods.

Gamma variate regression is a commonly used method to fit the blood pool and myocardial signal intensity for the up-slope phase [87, 88]. Equation (4.4) shows the fitting formula:

$$S(t) = \begin{cases} 0 & t \leq t_0 \\ A(t-t_0)^{\alpha} \exp \dfrac{-(t-t_0)}{\tau} & t > t_0 \end{cases} \tag{4.4}$$

where t_0 denotes the end time point of the baseline phase. A, α and τ are regression coefficients. Generally, a least-squares fitting based on equation (4.4) is performed to approximate the shape of TIC [88].

Polynomial regression has also been used to fit the TIC [83]. The polynomial regression can be expressed as:

$$S(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_n t^n \quad t > t_0. \tag{4.5}$$

In this study, $n = \{0, 1, 2, 3, 4\}$ have been applied to compare the performances between different orders as summarised in Table 4.1. Sum of squared errors (SSE) (see Section 2.6) is used to compare different orders of the polynomial regression. Specifically, the SSE in this circumstance can be expressed as:

$$\text{SSE} = \sum_{t=1}^{m} (S(t) - S_{\text{obs}}(t))^2 \tag{4.6}$$

where $S(t)$ can be found in equation (4.5) and $S_{\text{obs}}(t)$ is the observed signal intensity. Since the 4^{th} order of polynomial regression performs the best with the least SSE, the 4^{th} order polynomial regression is chosen to compare with other regression models.

Gaussian process is a state of the art method to apply the regression (see details in Sections 2.5). The Gaussian process regression model is:

$$S(t) = f(t) + \varepsilon_t \tag{4.7}$$

Table 4.2: Average sum of square errors (SSE) with leave-one-out cross-validation for different methods

	Polynomial	Gamma Variate	Gaussian Process
Blood Pool	34.3	86.4	25.7
Myocardium	53.5	92.6	59.0
Sum	87.8	179.0	84.7

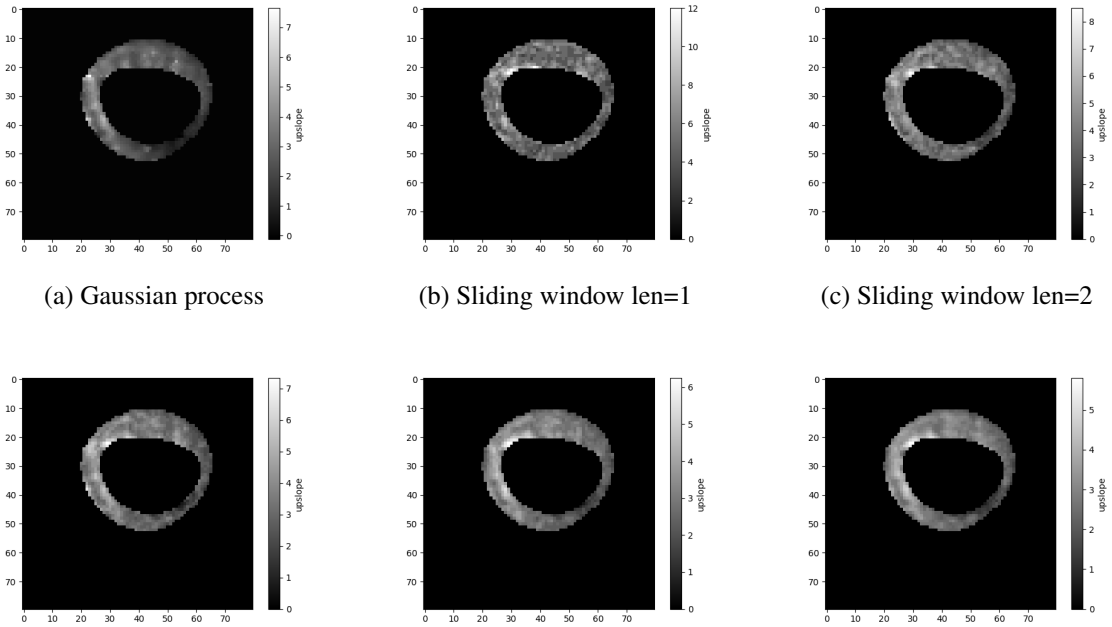
where $f(t) \sim \mathrm{GP}(\mu(\cdot), k(\cdot,\cdot^{'}))$ and $\varepsilon_t \sim \mathrm{N}(0, \gamma^2\mathrm{I})$. In this case, $\mu(\cdot) = 0$ and

$$k(t,t^{'}) = \tau^2 \exp\left(-\frac{|t-t^{'}|^2}{l^2}\right) \tag{4.8}$$

are chosen. γ^2, τ^2 and l^2 are unknown parameters.

A leave-one-out cross-validation (see details in Section 2.6) is used to compare the Gamma variate, Polynomial and Gaussian process regressions, and the results can be found in Table 4.2. It can be seen that the Gaussian process regression outperforms among three methods because it has the least SSE. All fitted curves are illustrated in Figure 4.1.

In this study, we also compare the values of up-slope generated by Gaussian process regression and the "sliding window" method. Figure 4.5 shows the comparisons between Gaussian process regression and sliding window methods given different window lengths len. Based on the visual inspection, the larger the window length len is, the smoother the up-slope estimations are. However, the degree of the smoothness cannot directly infer the goodness of the estimation. The LGE image is used as the reference of the location of the myocardial defect. According to panel (b), Figure 4.4, the myocardial defect region visually occupies 60° of the myocardium in the right bottom of the image (the region inside the red curve). This means that the up-slope values in this region should be the lowest in the up-slope map. When len $= 1$, the up-slope map is too noisy to find the lowest value region. When len $= 2,3,4,5$, the lowest value region has different degrees of shrinkage (underestimation of the myocardial defect region). For the best performance one, $n = 5$, the lowest value region visually occupies 30° of the myocardium. On the other hand, the lowest value region for the up-slope map generated by the Gaussian process regression is basically consistent with the myocardial defect region in the LGE image. This qualitative justification shows that the up-slope map generated by Gaussian process regression performs better than the up-slope map generated by the "sliding window" method in this DCE-MRI dataset. However, the visual inspection is inaccurate and subjective, and quantification methods such as classification models could be applied to compare these two methods. In fact, in Chapters 5, 6 and 7, different classification models are applied to the myocardial perfusion DCE-MRI data.



(a) Gaussian process (b) Sliding window len=1 (c) Sliding window len=2

(d) Sliding window len=3 (e) Sliding window len=4 (f) Sliding window len=5

Figure 4.5: Up-slope estimations using different methods and parameters. Panel (a) shows the estimated up-slope map using Gaussian process regression. Panels (b) - (f) show the estimated up-slope maps using sliding window method given window length len $= 1, 2, 3, 4, 5$ respectively.

4.3 Quantitative analysis

4.3.1 Introduction

The methods for quantifying the myocardial blood flow (MBF) can be broadly separated into two categories: model-based and model-independent [8]. The model-based methods focus on specifying a functional model of the tissues. Specifically, the functional model tries to describe how the tracer moves in the tissues. On the other hand, the model-independent methods do not specify a functional model.

The model-independent methods are based on the central volume principle [96]. This principle is originally derived from Fick's principle [8]. After derivations of this principle, the MBF can be obtained by deconvolution of an impulse response function. Many methods have been introduced to solve this deconvolution problem. Specifically, in [71,72], a widely applied Fermi deconvolution method has been reported. In 2002, a model-independent deconvolution method was introduced in [74]. To extend the work in [74], hierarchical Bayesian modelling (HBM) has been applied in [97] to reconstruct the arterial input function (AIF). In [98], the modelling in [97] has been extended by importing spatio-temporal constraints using Gaussian Markov random fields. An exponential function constrained deconvolution approach has been announced in [75]. An autoregressive–moving-average (ARMA) constrained deconvolution method has been described in [76] to expand the application of the exponential function constrained decon-

volution approach in [75]. Moreover, a comparison of the robustness of deconvolution methods has been published in [99].

The model-based methods are usually based on compartmental analysis [49]. This analysis was originally derived by [86, 100, 101]. In [102], the standardization of quantities and symbols for the model-based methods have been introduced. A full review of model-based methods have been published in [77]. Bayesian approaches have been applied to the model-based methods to improve the estimations of the tracer kinetic parameters. A hierarchical Bayesian model has been developed to estimate kinetic parameters of DCE-MRI based on Gaussian Markov random fields priors in [103]. A comparison between Bayesian approaches and non-linear regression algorithms have been applied based on a tracer kinetic model [104]. A large-scale Bayesian spatio-temporal regression method has been applied to quantify the MBF, which has used Gaussian Markov random fields priors to express the spatio-temporal constraints in [105]. In [5], the quantification of myocardial blood flow using a hierarchical Bayesian model has been reported based on a two-compartment exchange model.

In [1], a detailed CHD detection meta-analysis for qualitative, semi-quantitative and quantitative methods is performed, and sensitivity and specificity of the results obtained in individual studies were reported. Table 4.3 summarizes different type of methods and their performances. Information presented in Table 4.3 is based on [1] with additional entries denoted by a *.

In the following section, Fermi-constrained deconvolution method will be described. A comparison between Fermi method and a Fermi-based hierarchical Bayesian model will be illustrated in Chapter 6.

4.3.2 Central volume principle

The central volume principle describes that the rate of a substance accumulation in tissues of ROI can be calculated as the differences between concentrations of tracers flowing in and out of the tissues multiplied by the flow rate:

$$\frac{dC_{\text{myo}}(t)}{dt} = F_b \cdot \big(C_{\text{in}}(t) - C_{\text{out}}(t)\big) \tag{4.9}$$

where $C_{\text{in}}(t)$ and $C_{\text{out}}(t)$ denote the arterial input and venous output of the contrast concentrations respectively. F_b denotes the blow flow and $C_{\text{myo}}(t)$ denotes the contrast concentration in the myocardium (ROI). Let $h(t)$ represent the fraction of the tracer leaving the myocardium (ROI) per unit time at time t [124], the contrast concentration of venous output $C_{\text{out}}(t)$ can be expressed as:

$$C_{\text{out}}(t) = h(t) * C_{\text{in}}(t), \tag{4.10}$$

Table 4.3: Summary for qualitative, semi-quantitative and quantitative methods. Table originally compiled in [1], with additional entries denoted by *.

author	Patient No.	type of method	sensitivity	specificity	stenosis
*Biglands et al. [106]	128	qualitative	0.81	0.86	$\geq 50\%$
Cury et al. [107]	46	qualitative	0.97	0.75	$\geq 70\%$
*Greenwood et al. [108]	752	qualitative	0.86	0.83	$\geq 70\%$
Ishida et al. [109]	104	qualitative	0.90	0.85	$\geq 70\%$
Kawase et al. [110]	50	qualitative	0.94	0.94	$> 50\%$
Klem et al. [111]	92	qualitative	0.89	0.87	$\geq 70\%$
*Klem et al. [112]	136	qualitative	0.85	0.88	$\geq 70\%$
*Paetsch et al. [113]	79	qualitative	0.91	0.62	$\geq 50\%$
*Patel et al. [114]	30	qualitative	0.79	0.83	$\geq 50\%$
Pilz et al. [115]	171	qualitative	0.96	0.83	$> 70\%$
Plein et al. [116]	68	qualitative	0.96	0.83	$\geq 70\%$
Sakuma et al. [117]	40	qualitative	0.81	0.68	$> 70\%$
*Schwitter et al. [118]	42	qualitative	0.85	0.67	$\geq 50\%$
*Schwitter et al. [119]	533	qualitative	0.67	0.61	$\geq 50\%$
Takase et al. [120]	102	qualitative	0.93	0.85	$> 50\%$
*Watkins et al. [121]	101	qualitative	0.91	0.94	$\geq 70\%$
*Wolff et al. [122]	75	qualitative	0.93	0.75	$\geq 70\%$
*Biglands et al. [80]	50	semi-quantitative	0.89	0.77	$\geq 70\%$
Doyle et al. [123]	184	semi-quantitative	0.57	0.78	$\geq 70\%$
Giang et al. [67]	44	semi-quantitative	0.93	0.75	$\geq 50\%$
*Huber et al. [64]	31	semi-quantitative	0.88	0.87	$50\% - 70\%^*$
Nagel et al. [65]	84	semi-quantitative	0.88	0.90	$\geq 75\%$
Plein et al. [69]	82	semi-quantitative	0.88	0.74	$> 50\%$
Schwitter et al. [38]	47	semi-quantitative	0.86	0.70	$\geq 50\%$
*Biglands et al. [106]	128	Fermi (quantitative)	0.88	0.85	$\geq 50\%$
*Biglands et al. [80]	50	Fermi (quantitative)	0.94	0.87	$\geq 70\%$
*Biglands et al. [80]	50	OCM** (quantitative)	0.72	0.87	$\geq 70\%$
*Biglands et al. [80]	50	MID*** (quantitative)	0.83	0.77	$\geq 70\%$
*Costa et al. [63]	37	Fermi (quantitative)	0.85	0.49	$\geq 70\%$
*Huber et al. [64]	31	MID*** (quantitative)	0.77	0.86	$50\% - 70\%^*$
*Patel et al. [114]	30	Fermi (quantitative)	0.88	0.67	$\geq 50\%$

*Summarised by the author.
*Using fractional flow reserve (FFR).
**The abbreviation “OCM" denotes one-compartment model.
***The abbreviation “MID" denotes model-independent deconvolution.
The reference standard for myocardial ischemia is based on a quantitative coronary x-ray angiogram or single photon emission computed tomography stenosis severity.
The table is ordered by the type of the method and the author name.

where the symbol $*$ represents the convolution operation. By substituting equation (4.10) into equation (4.9), we have

$$C_{\text{myo}}(t) = R_f(t) * C_{\text{in}}(t) \tag{4.11}$$

where $R_f(t)$ denotes the impulse response function, whose expression is given by

$$R_f(t) = F_b - F_b \int_0^t h(s)ds. \tag{4.12}$$

The derivation of equation (4.11) can be found in Appendix A. According to the definition of the impulse response function $R_f(t)$ in equation (4.12), the MBF F_b can be calculated by substituting $t = 0$ into equation (4.12)

$$R_f(t = 0) = F_b. \tag{4.13}$$

Moreover, the value of $R_f(t)$ decreases from its initial amplitude F_b with time because the probability that the contrast agent remains in the ROI decreases.

The estimates of contrast concentrations for the myocardium $C_{myo}(t)$ and the arterial input $C_{\text{in}}(t)$ can be acquired from the DCE-MRI series. Therefore, it is possible to solve equation (4.12) by applying deconvolution operation. However, a unique and universal form of $R_f(t)$ does not exist, and its free-form inference is an "ill-posed" problem due to its intrinsic unidentifiability [8]. Specifically, the brute force method of solving equation (4.11) using Fourier transform would generate mathematically unstable and physiologically unrealistic results [8].

4.3.3 Fermi-constrained deconvolution

In order to estimate F_{b}, a widely-used function, Fermi function, is chosen for $R_f(t)$. It has been firstly applied in a brain perfusion study by computed tomography [71]. Then, it has been introduced to the myocardial perfusion DCE-MRI to estimate the MBF [72]. The analytical form of the Fermi model is

$$R_f(t) = \frac{A}{1 + e^{(t-\omega)/\lambda}} \tag{4.14}$$

where A, ω, λ are shape parameters. In this study, the vector $\boldsymbol{\Theta}$

$$\boldsymbol{\Theta} = \{A, \omega, \lambda\} \tag{4.15}$$

is used to represent all three Fermi parameters. Given the form of impulse response function $R_f(t)$, equation (4.11) can be solved by fitting the shape parameters in equation (4.14) using the DCE-MRI series. Generally, least-squares fitting has been used to estimate the parameters in equation (4.14) [5, 72]. Specifically, this method aims to minimise the cost function,

$$\sum_t (C_{\text{myo}}(t|R_f(t)) - C_{\text{obs}}(t))^2 \tag{4.16}$$

where $C_{\text{myo}}(t|R_f(t))$ is expressed as equation (4.11), and $C_{\text{obs}}(t)$ is the observations of contrast concentrations of the myocardium from DCE-MRI series. Once the shape parameters A, ω, λ are determined by this method, the MBF can be estimated

$$R_f(t=0) = \frac{A}{1+e^{-\omega/\lambda}}. \tag{4.17}$$

However, the algorithm of this optimization method may get stuck in local minima. This issue is not unique for the Fermi method but universal for many applications, e.g. tracer kinetic model [5, 104].

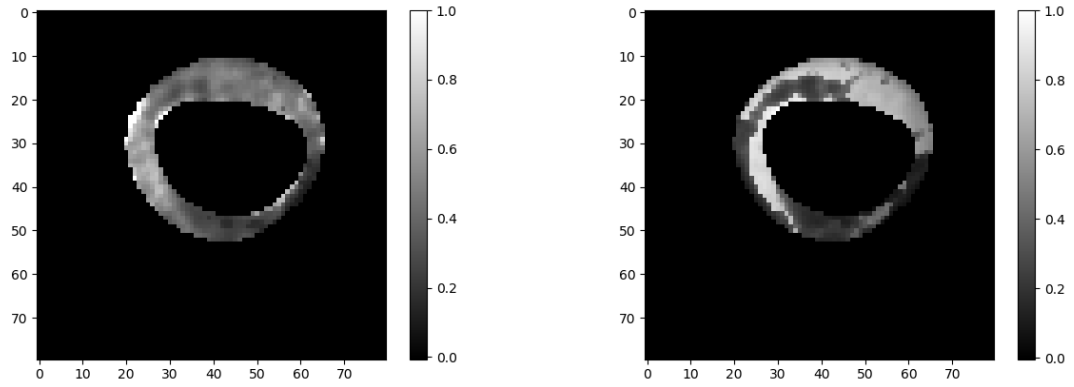
4.3.4 Fermi-constrained deconvolution application

According to Sections 4.3.2 and 4.3.3, the theoretical part of the Fermi-constrained deconvolution method has been reviewed. However, in practice, the estimated MBF will be affected by the numerical approximation significantly. In this section, the details of the application of the Fermi-constrained deconvolution are discussed.

Deconvolution based on the first-pass interval

The Fermi deconvolution should be implemented during the first-pass phase. This means that the data during the baseline phase and the second-pass phase must be discarded. Meanwhile, the data should be standardised by the baseline. In other words, for the curves of the myocardium and the blood pool, the starting points should be located around value 0. In fact, these principles should be followed by all methods based on the central volume principle.

Figure 4.6 shows the comparisons of the MBF estimations between different selected time interval for the first-pass phase. It is worth mentioning that the least-squared fitting method is the same for both cases in Figure 4.6, and the only difference between panels (a) and (b) in Figure 4.6 is the selected time interval for the first-pass phase. The location of the reference of the myocardial defect region can be found in panel (b), Figure 4.4. The myocardial defect region should be located at the right bottom of the myocardium in the image. For panel (a) in Figure 4.6, though some pixels located at the right bottom endocardium are affected by the blood pool and have high MBF values, the low MBF values region basically covers the same region as the myocardial defect region in panel (b), Figure 4.4. On the other hand, for panel (b) in Figure 4.6, many pixels in the left top of the myocardium in the image also have low MBF values. However, in the corresponding LGE image in panel (b), Figure 4.4, no myocardial defect can be observed in that area visually. This comparisons express the importance of the selected time interval for the first-pass phase. Similarly, the visual inspection is inaccurate and subjective. The methods illustrated in Chapter 5 could be applied to the MBF maps to detect the hypo-perfusion region.



(a) MBF based on accurate time interval

(b) MBF based on inaccurate time interval

Figure 4.6: MBF estimations comparisons between different time interval usage. Panel (a) shows the MBF estimations (standardised to range $[0, 1]$) using the accurate time interval of the first-pass data. Panel (b) shows the MBF estimations (standardised to range $[0, 1]$) using the inaccurate time interval of first-pass data. Specifically, the time interval used for the deconvolution method contains part of baseline (2 points) and part of second-pass (5 points).

Interpolation to the TIC

Generally, the data obtained from the MR images contain a varying degree of noise. Moreover, the artifacts mentioned in Section 3.2 also decrease the quality of the data. In practice, the fitted smooth TIC can improve the performances of the deconvolution methods. There are many ways to interpolate the time-intensity points. In this study, two types of interpolation are implemented and compared.

The first type of interpolation method is to apply linear or cubic interpolation methods directly to the time-intensity data. However, this method has potential drawbacks. Direct interpolation methods may fail to reduce the noise of the data and can even increase the noise of the data. Specifically, the direct interpolation methods will generate many new points between the outlier and normal point in the time domain. Since the outlier point is affected by high noise, all newly generated points are affected by the noise too. In practice, this will make the outliers in the estimated MBF maps to be further away from the normal values. Figure 4.7 shows the MBF estimation map using a cubic interpolation method to the TIC of the blood pool and myocardium. Some pixels located at the edge of the epicardium have extremely high values. Similarly, some pixels at the endocardium boundary also have very high values. Theoretically, this is caused by the motion artifacts and artifacts generated by the manual contouring. These abnormal values are far away from the normal range of MBF values and reduce the dynamic range for the display of the MBF values of the remaining pixels. Therefore, visual inspection and quantitative analysis of the maps affected by outliers can be very hard.

The second type of interpolation method is based on the fitted curve of the time-intensity

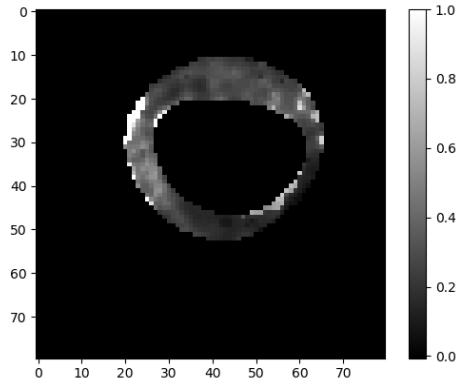


Figure 4.7: MBF estimations using cubic interpolation to the TIC of the blood pool and the myocardium.

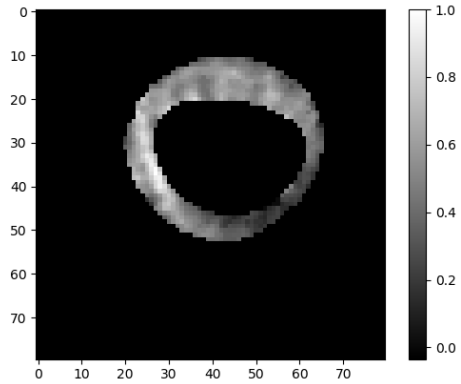


Figure 4.8: MBF map using Gaussian process regression.

data. Specifically, any number of points can be extracted from the fitted curve. These extracted points can be applied to implement the numerical deconvolution. In Section 4.2, the Gaussian process regression has the lowest average SSE compared with polynomial and Gamma variate regressions. Therefore, in this section, the MBF map generated by the Gaussian process regression is compared with the results in Figures 4.6.

Figure 4.8 shows the MBF estimations using Gaussian process regression. Compared with panel (a) in Figure 4.6, the abnormally high MBF values at endocardium and epicardium disappear in Figure 4.8. The hypo-perfusion region can be inspected visually, and this region is the same as the myocardial defect region in panel (b), Figure 4.4. Therefore, the MBF map generated using the Gaussian process fitted TIC is considered to have better performance than the one generated without using interpolation method.

4.4 Discussion and conclusion

In this chapter, the semi-quantitative and quantitative analysis methods have been reviewed. Because of the limitation of the data (no stress data), the experiments for some semi-quantitative analysis methods, e.g. MPRI, cannot be implemented. However, all methods illustrated in Section 4.2 can be directly applied to complete data. The major finding of the brief exploration of the semi-quantitative methods in this study is that the Gaussian process regression outperforms some traditional methods, e.g. Gamma variate regression and sliding window method. The Gaussian process regression has the lowest average SSE compared with Gamma variate regression and polynomial regression. Moreover, the up-slope map generated by the Gaussian process regression is more accurate than the one generated by the sliding window method based on visual inspection based on the reference LGE image.

A typical quantitative analysis method, the Fermi-constrained deconvolution, has been reviewed. This method is based on the central volume principle. The numerical approximation has been applied to obtain the estimation of the MBF. Two key factors have been discussed. First, the start and end points of the deconvolution is critical. The start point can be calculated by a parameter of a given piecewise function. The least-squared method has been applied to obtain the estimations of the piecewise function. The end point is given by a specific point that has the first positive gradient after the maximum negative (minimum) gradient. Second, the interpolation of the TIC can improve the estimation of the MBF using the Fermi-constrained method. Similarly, the MBF map generated by the Gaussian process fitted TIC outperforms the one using the original data. A critical point is emphasised that all methods based on the central volume principle are suitable to be applied using the numerical approximation.

There are also many potential improvements for the semi-quantitative and quantitative analysis. The comparisons between the parametric maps generated using the Gaussian process fitted TIC and the one generated using original data points are visually inspected. The visual inspection has the drawback that it is inaccurate and subjective. Therefore, in Chapter 5, mixture models are used to classify the myocardial tissues within the myocardium image based on different parametric maps. Moreover, the properties of the hypo-perfused tissues are different from the normal-perfused tissues. This is an apparent hierarchical structure, and the hierarchical Bayesian model introduced in Section 2.2 can be used. The details of the hierarchical Bayesian model application based on the Fermi-constrained method is introduced in Chapter 6.

Chapter 5

Mixture Models for Myocardial Perfusion DCE-MRI

5.1 Introduction

For medical images analysis, mixture models are usually applied to classify different tissues. Specifically, the observed variables (pixel signals) and latent variables (pixel status) can be modelled as a joint distribution that is more tractable [18] (Chapter 9). Gaussian mixture model (GMM) is a general method to apply segmentation and classification of medical images, e.g. DCE-MRI [125, 126], positron emission tomography (PET) [127, 128], ultrasound [129, 130]. However, the standard Gaussian mixture model classification cannot make use of spatial information, which is usually adopted for image segmentation and classification. Therefore, many spatial constrained GMM methods have been introduced to improve the pixel classification, such as spatially variant finite mixture model (SVFMM) [10, 131], Markov random field constrained Gaussian mixture model (GMM-MRF) [11, 132, 133], and other spatial constrained finite mixture models [134–136].

In this chapter, Gaussian mixture models (GMM) are applied to classify the DCE-MRI myocardial perfusion images using an expectation-maximization (EM) algorithm. Moreover, the spatial constrained mixture models, SVFMM and GMM-MRF, are also applied to improve the GMM classification method.

5.2 Gaussian mixture model classification

In practice, the distributions of some sets of data have more than one dominant clump. In this way, a simple distribution cannot capture the structure of data, but linear combinations of many simple distributions, e.g. Gaussian distribution, are able to capture the structure of data. Therefore, finite mixture models [137, 138] are introduced to formulate such probabilistic models. In this section, the most general one, the Gaussian mixture model, and its applications

on pixel classification are reviewed.

5.2.1 Statistical model

The Gaussian mixture model can be mathematically expressed by:

$$P(\boldsymbol{x}) = \sum_{j=1}^{K} \pi_j N(\boldsymbol{x}|\boldsymbol{\mu}_j, \Sigma_j), \tag{5.1}$$

where K is the total number of clusters, $N(\boldsymbol{x}|\boldsymbol{\mu}_j, \Sigma_j)$ is the multivariate Gaussian distribution of $\boldsymbol{x}$ with mean $\boldsymbol{\mu}_j$ and covariance matrix Σ_j. π_j is the mixing coefficient which will be explicitly illustrated later. Let $\boldsymbol{k} = \{k_1, k_2, ..., k_K\}$ denote a vector with dimension K which satisfies that one element $k_j = 1$ and others $k_{p\neq j} = 0$. $\boldsymbol{k}$ has a 1-of-K representation, e.g. [18] (chapter 9). In this way, k_j can be straightforwardly inferred that $k_j \in \{0, 1\}$ and $\sum_j k_j = 1$. Therefore, there are K states of $\boldsymbol{k}$ and it depends on which element is equal to 1. The variable $\boldsymbol{k}$ is a latent variable to indicate which Gaussian distribution $\boldsymbol{x}$ belongs to. In other words, a joint distribution $P(\boldsymbol{x}, \boldsymbol{k})$ is defined as a marginal distribution $P(\boldsymbol{k})$ and conditional distribution $P(\boldsymbol{x}|\boldsymbol{k})$ where $P(\boldsymbol{k})$ is specified by the mixing coefficients π_j by:

$$P(k_j = 1) = \pi_j \tag{5.2}$$

where $\{\pi_j\}$ must satisfy

$$0 \leq \pi_j \leq 1 \tag{5.3}$$

and

$$\sum_{j=1}^{K} \pi_j = 1. \tag{5.4}$$

Since the random variables $\boldsymbol{k}$ have a 1-of-K representation, the marginal probability of $\boldsymbol{k}$ can be expressed as follow:

$$P(\boldsymbol{k}) = \prod_{j=1}^{K} \pi_j^{k_j}. \tag{5.5}$$

Once the latent variable k_j is given, i.e. the data $\boldsymbol{x}$ follows the j^{th} Gaussian distribution, the conditional distribution $P(\boldsymbol{x}|k_j)$ can be expressed as:

$$P(\boldsymbol{x}|k_j = 1) = N(\boldsymbol{x}|\boldsymbol{\mu}_j, \Sigma_j). \tag{5.6}$$

Similarly, $\boldsymbol{k}$ has a 1-of-K representation, the conditional distribution can be written as:

$$P(\boldsymbol{x}|\boldsymbol{k}) = \prod_{j=1}^{K} N(\boldsymbol{x}|\boldsymbol{\mu}_j, \Sigma_j)^{k_j}. \tag{5.7}$$

The marginal distribution $P(\boldsymbol{x})$ can be written as:

$$P(\boldsymbol{x}) = \sum_{\boldsymbol{k}} P(\boldsymbol{x}, \boldsymbol{k}) = \sum_{\boldsymbol{k}} P(\boldsymbol{k}) P(\boldsymbol{x}|\boldsymbol{k}). \tag{5.8}$$

By substituting equations (5.5) and (5.7) into equation (5.8), equation (5.8) turns to be equation (5.1). The derivation of this equation can be found in Appendix A.

In the Bayesian framework, the posterior distribution $P(\boldsymbol{k}|\boldsymbol{x})$ plays an important role in the statistical inference. Generally, $\gamma(k_j)$ denotes $P(k_j = 1|\boldsymbol{x})$, and its expression can be inferred by Bayes' theorem (2.1):

$$\begin{aligned} \gamma(k_j) \equiv P(k_j = 1|\boldsymbol{x}) &= \frac{P(k_j = 1) P(\boldsymbol{x}|k_j = 1)}{P(\boldsymbol{x})} \\ &= \frac{\pi_j N(\boldsymbol{x}|\boldsymbol{\mu}_j, \Sigma_j)}{\sum_{p=1}^{K} \pi_p N(\boldsymbol{x}|\boldsymbol{\mu}_p, \Sigma_p)}. \end{aligned} \tag{5.9}$$

5.2.2 Statistical inference and EM algorithm

Let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N\}$ be the set of observations. Gaussian mixtures with K components are used to model this set of data. An $N \times D$ matrix $\boldsymbol{X}$ is used to represent this set of data, $\boldsymbol{X} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, ..., \boldsymbol{x}_N^T]$. D is the dimension of the input data $\boldsymbol{x}$. The dataset $\boldsymbol{X}$ is assumed to be i.i.d. According to equation (5.1), the likelihood function is:

$$P(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{\Sigma}) = \prod_{n=1}^{N} \sum_{j=1}^{K} \pi_j N(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \Sigma_j) \tag{5.10}$$

where $\boldsymbol{\pi} = \{\pi_1, \pi_2, ..., \pi_K\}$, $\boldsymbol{M} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, ..., \boldsymbol{\mu}_K\}$ and $\boldsymbol{\Sigma} = \{\Sigma_1, \Sigma_2, ..., \Sigma_K\}$. Therefore, the log-likelihood function can be expressed as:

$$\ln P(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left(\sum_{j=1}^{K} \pi_k N(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \Sigma_j) \right). \tag{5.11}$$

Maximum likelihood estimation (MLE) is the most straightforward method to estimate the unknown parameters $\boldsymbol{\pi}, \boldsymbol{M}$ and $\boldsymbol{\Sigma}$ because the MLE has the properties of consistency[1], asymptotical unbiasedness[2] and asymptotical efficiency[3]. However, the maximization of the log-likelihood function of Gaussian mixtures is an ill-posed problem [18] (chapter 9). The most widely applied method to this problem is called expectation-maximization (EM) algorithm (see [28, 29] and Section 2.4 in this thesis).

To maximize the log-likelihood in equation (5.11), the derivatives of $\ln P(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{\Sigma})$ with

[1] The MLE converges in probability to the value of the parameter.
[2] The expected value of the MLE converges to the value of the parameter.
[3] The MLE asymptotically achieves the Cramér–Rao lower bound.

respect to the mean $\boldsymbol{\mu}_j$ are set to be 0:

$$0 = -\sum_{n=1}^{N} \frac{\pi_j N(\boldsymbol{x}_n|\boldsymbol{\mu}_j,\Sigma_j)}{\sum_{p=1}^{K}\pi_p N(\boldsymbol{x}_n|\boldsymbol{\mu}_p,\Sigma_p)}\Sigma_j(\boldsymbol{x}_n-\boldsymbol{\mu}_j) \tag{5.12}$$

where the term $\frac{\pi_j N(\boldsymbol{x}_n|\boldsymbol{\mu}_j,\Sigma_j)}{\sum_{p=1}^{K}\pi_p N(\boldsymbol{x}_n|\boldsymbol{\mu}_p,\Sigma_p)}$ is exactly the posterior distribution in equation (5.9). Based on the assumption that Σ_j is nonsingular, both sides of equation (5.12) multiply Σ_j^{-1}, and this equation turns to be:

$$\boldsymbol{\mu}_j = \frac{\sum_{n=1}^{N}\gamma(k_{nj})\boldsymbol{x}_n}{\sum_{n=1}^{N}\gamma(k_{nj})} \tag{5.13}$$

where

$$\gamma(k_{nj}) = \frac{\pi_j N(\boldsymbol{x}_n|\boldsymbol{\mu}_j,\Sigma_j)}{\sum_{p=1}^{K}\pi_p N(\boldsymbol{x}_n|\boldsymbol{\mu}_p,\Sigma_p)}. \tag{5.14}$$

It is emphasised that equation (5.13) is intractable, because the terms on the right depends on the term on the left. In other words, the term $\gamma(k_{nj})$ depends on all parameters (i.e. π_j, $\boldsymbol{\mu}_j$ and Σ_j). The essential idea of the EM algorithm is to decouple the parameters on the left of (5.14) from those on the right of (5.14). Those on the right, the so-called “old" parameters, are used to compute the weighting parameters on the left, which are then used to compute the “new" parameters, as shown below. This leads to tractable equations, which as opposed to (5.14) have to be iterated though. Let

$$N_j = \sum_{n=1}^{N}\gamma(k_{nj}). \tag{5.15}$$

The term N_j can be interpreted as the effective number of points set in the cluster k [18] (Chapter 9). Similarly, the derivatives of $\ln P(\boldsymbol{X}|\boldsymbol{\pi},\boldsymbol{M},\boldsymbol{\Sigma})$ with respect to the covariance matrix Σ_j is set to be 0. The solution of Σ_j is:

$$\Sigma_j = \frac{1}{N_j}\sum_{n=1}^{N}\gamma(k_{nj})(\boldsymbol{x}_n-\boldsymbol{\mu}_j)(\boldsymbol{x}_n-\boldsymbol{\mu}_j)^T. \tag{5.16}$$

Similarly, equation (5.16) is also intractable. Finally, the maximization of the log-likelihood $\ln P(\boldsymbol{X}|\boldsymbol{\pi},\boldsymbol{M},\boldsymbol{\Sigma})$ with respect to π_j can be obtained using a Lagrange multiplier. Since $\sum_{j=1}^{K}\pi_j = 1$, the maximization can be achieved by maximising the following quantity:

$$\ln P(\boldsymbol{X}|\boldsymbol{\pi},\boldsymbol{M},\boldsymbol{\Sigma}) + \lambda\left(\sum_{j=1}^{K}\pi_j - 1\right). \tag{5.17}$$

Similarly, set the derivatives of the quantity in equation (5.17) with respect π_j to be 0. The solution of π_j is:

$$\pi_j = \frac{N_j}{N}. \tag{5.18}$$

The derivation of this solution can be found in Appendix A. Similarly, equation (5.18) is intractable.

Now, the EM algorithm can be used to maximize the likelihood function with respect to the parameters of the Gaussian mixture model. Specifically, in the E step, given the old values of parameters π_j, $\boldsymbol{\mu}_j$ and Σ_j, the value of $\gamma(k_{nj})$ is computed by equation (5.14). Then, in the M step, the new values of parameters π_j, $\boldsymbol{\mu}_j$ and Σ_j are re-estimated using equations (5.13), (5.16) and (5.18) given the computed value $\gamma(k_{nj})$ in E step. Each step of the EM algorithm increases the log-likelihood, equation (5.11), and that the algorithm is guaranteed to converge to a zero-gradient point of the log-likelihood. This is a local optimum of the log-likelihood or, in rare cases, a saddle point. The detail of the algorithm is illustrated in Algorithm 1.

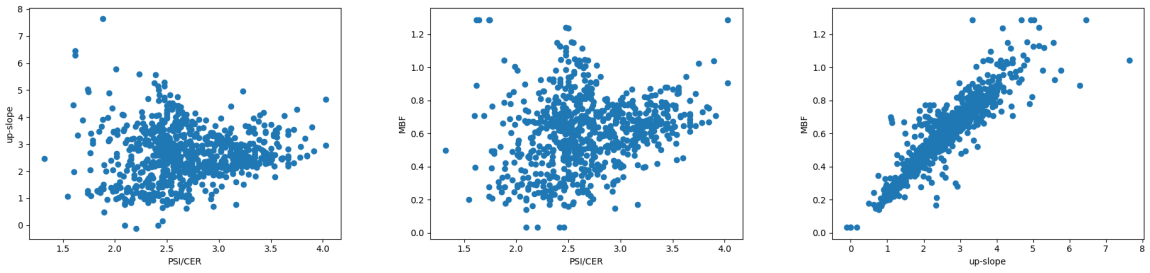
Algorithm 1: EM algorithm for Gaussian mixture model

Data: $\boldsymbol{X} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, ..., \boldsymbol{x}_N^T]$
Output: Parameters $\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{\Sigma}$
Set initial values of $\boldsymbol{\pi}, \boldsymbol{M}, \boldsymbol{\Sigma}$ and evaluate the log-likelihood according to equation (5.11);
for $k \leftarrow 1$ **to** K **do**
  E step: **for** $n \leftarrow 1$ **to** N **do**
    Calculate the quantity $\gamma(k_{nj})$ according to equation (5.14) using old values of $\boldsymbol{\mu}_j$, Σ_j and π_j;
  end
  M step: Re-estimate new parameters $\boldsymbol{\mu}_j$, Σ_j and π_j based on equation (5.13), equation (5.16) and equation (5.18) given $\gamma(k_{nj})$ calculated in E step;
end
Evaluate the log-likelihood based on equation (5.11) and check the convergence of all parameters. If the convergence criterion is not satisfied, repeat both E and M steps until the convergence criterion is satisfied;

5.2.3 Applications of Gaussian mixture model on myocardial perfusion DCE-MRI

Gaussian mixture models can be applied to the data of myocardial perfusion DCE-MRI. Specifically, the GMM can be used on both original signals and some features derived by semi-quantitative analysis (see Section 4.2) and quantitative analysis (see Section 4.3). Generally, the features derived by semi-quantitative analysis and quantitative analysis are able to reveal the degree of hypo-perfusion. Classification methods are applied to these features to separate them into different groups.

In this section, the GMM classification method is applied to several features derived by semi-quantitative analysis and quantitative analysis. Meanwhile, statistical inferences and results analysis are also introduced in this section. The GMM classification is implemented by a build-in Python package “sklearn.mixture.GaussianMixture”. The number of clusters is set to “2”,



(a) CER → up-slope (b) CER → MBF (c) up-slope → MBF

Figure 5.1: Pairwise scatter plots for CER, up-slope and MBF.

Table 5.1: Pearson correlation coefficients for CER, up-slope and MBF

	CER	up-slope	MBF
CER	1	0.10	0.27
up-slope	0.10	1	0.90
MBF	0.27	0.90	1

which means there are only two groups, 'hypo-perfused' (or lesion) and 'normal-perfused' (or healthy). The covariance type is set 'full', which means that the covariance matrices for different clusters are unconstrained and different. The initial values (means and precisions) are given by "kmeans" algorithm [139]. Figure 4.3 shows the myocardial perfusion DCE-MRI data used in this section.

In Chapter 4, the CER, up-slope and MBF maps are explicitly discussed. A naive threshold is used to distinguish the hypo-perfusion and normal-perfusion regions. This threshold is given by the intuition of the user, which means it is not robust. This disadvantage can be solved by applying the classification methods such as the GMM classification because the EM algorithm for the GMM can iteratively choose the means of different groups with the maximum log likelihood.

Figure 5.1 shows the pairwise scatter plots for CER, up-slope and MBF. Table 5.1 shows the Pearson correlation coefficients for the pairwise parameters. The CER values are linearly unrelated to both up-slope and MBF values according to these quantities. However, the up-slope and MBF values are highly linearly related with the correlation coefficient value 0.90. Meanwhile, according to panel (c) in Figure 5.1, most points are located around a straight line. Therefore, a further linear model regression is worth exploring the relationships between the up-slope and MBF. This simple linear model can be expressed as:

$$y = ax + b \tag{5.19}$$

where y is the MBF and x is the up-slope. Figure 5.2 shows the fitted simple linear model given the values of up-slope and MBF. Least-squared fitting is used to fit this model and the estimations of coefficients are $a = 0.198$ and $b = 0.089$. The R^2 is 0.81. This means the goodness of fit of

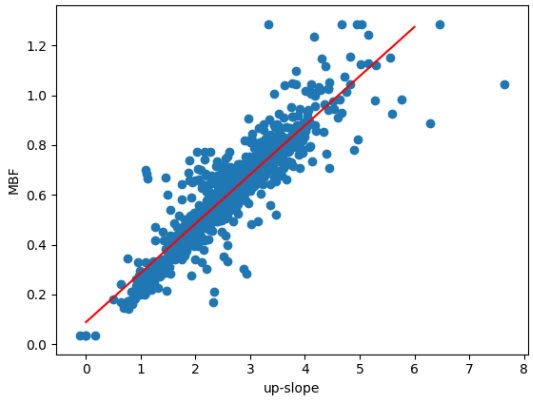
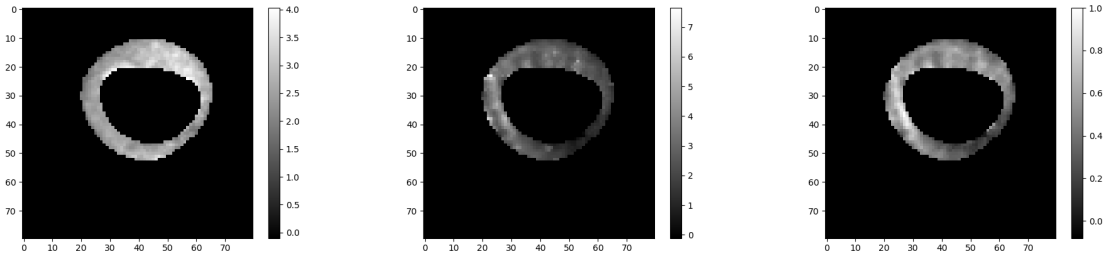


Figure 5.2: Simple linear model fitted by the values of up-slope and MBF. The blue points indicate the input data and the red line indicate the fitted simple linear model.



(a) CER (b) up-slope (c) MBF

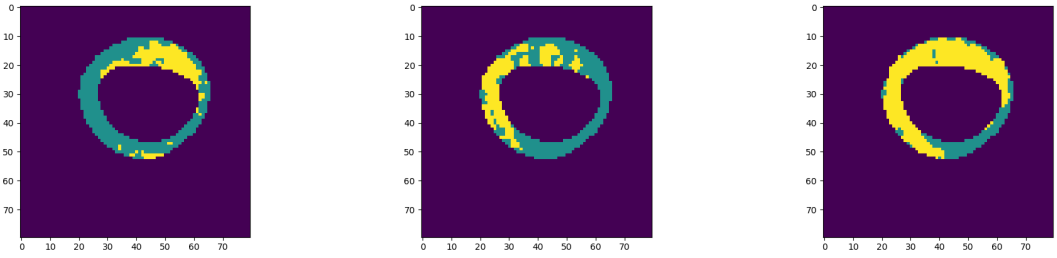
Figure 5.3: Panels (a) - (c) show the parametric maps of CER, up-slope and MBF. The values of background pixels are set to black, which can be ignored.

this simple linear model is high, especially when there are more than 750 points.

This result is consistent with the results in Chapter 4. The hypo-perfusion can be detected by the MBF and up-slope values. On the other hand, the CER is not able to detect the hypo-perfusion region because it lacks information on the growth speed of the TIC. Moreover, a new finding is that the linear relationship between the up-slope and MBF is very high in this case. This can be visually observed from panels (b) and (c), Figure 5.3.

Figure 5.3 shows the parametric maps of CER, up-slope and MBF. According to the quantities in Table 5.1, the values of up-slope and MBF are highly related. This can also be observed from the parametric maps. The right bottom of the rings in panels (b) and (c) are dark, indicating the lesion's location.

Figure 5.4 shows the classification maps of CER, up-slope and MBF based on the parametric maps. Panel (a) in Figure 5.4 shows the GMM classification using the CER map. Similar to the conclusions in Section 4.2, the CER map cannot detect the hypo-perfusion region. Panel (b) and (c) shows the GMM classification maps based on up-slope and MBF maps. According to the reference LGE image in panel (b), Figure 4.4, the hypo-perfusion region is located in the inferior

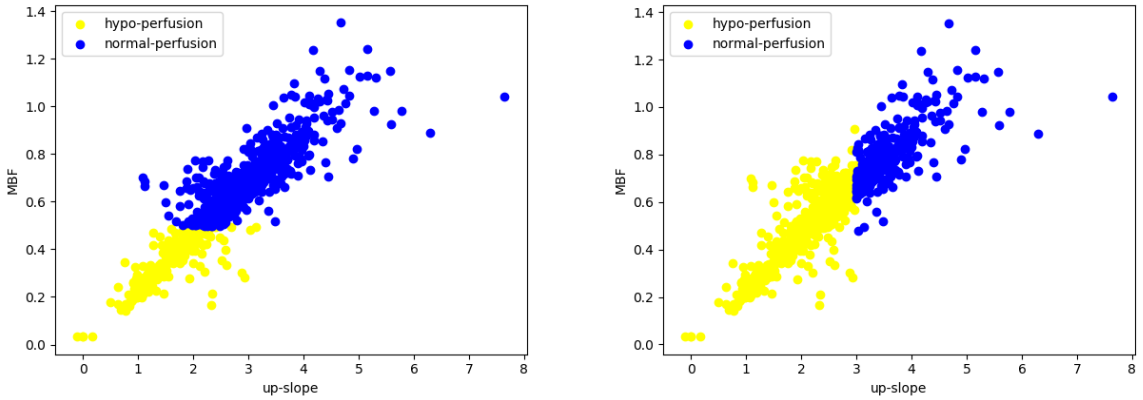


(a) CER classification (b) up-slope classification (c) MBF classification

Figure 5.4: Panels (a) - (c) show the classification maps of CER, up-slope and MBF. The yellow segments indicate healthy tissues. The dark green segments indicate lesion tissues. The black pixels indicate the background.

of the myocardium. The up-slope classification map overestimate the hypo-perfusion region, and the MBF classification map accurately estimates the hypo-perfusion region with a few small and implausible segments. Although the linear relationship between the up-slope and MBF values are high, the classification maps are quite different. This phenomenon is explainable. The linear relationship between up-slope and MBF values shows that the relative distributions of up-slope and MBF values are similar. However, the threshold values between hypo-perfusion and normal perfusion pixels for up-slope and MBF values are different. Figure 5.5 shows the different thresholds based on MBF and up-slope values, and this figure explains the different performances between MBF based and up-slope based classification maps. Moreover, the up-slope map is obtained using the information of the myocardial pixels (see Section 4.2), and the MBF map is obtained using the information of the myocardial pixels and blood pool pixels. The MBF map outperforms the up-slope map because of more information usage.

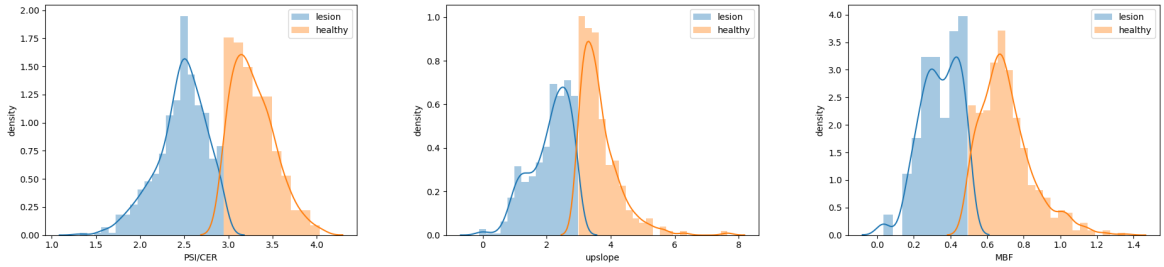
For panels (b) and (c) in Figure 5.4, the classification maps either have an unclear boundary between healthy tissues and lesions or have some single and small isolated segments. However, the boundary between lesion and healthy tissues should be clear if the noise of the myocardial perfusion DCE-MRI is low. The reason is that any specific heart wall segment is dominantly perfused by one main coronary artery, and the narrowing or blockage of that coronary artery will cause a reduced blood supply to this segment. There are three main arteries supplying blood to the myocardium (see details in Section 3.1.1), and therefore either single and small isolated segments or unclear boundaries are physiologically unrealistic, i.e. one large lesion is more likely than many isolated small lesions in myocardium [131]. The unclear boundary and single and small isolated segments are caused by the limitation of GMM classification. Specifically, the GMM classification method classifies the pixels in images based on their values, and it does not make use of the spatial relations among pixels. Figure 5.6 shows the histograms and kernel density estimates of the values of CER, up-slope and MBF, respectively. The two different groups, healthy tissues and lesions, have a clear threshold on the values of parameters. In other words, if the value of a pixel is larger than this threshold, this pixel is labelled as healthy tissue



(a) GMM classification based on MBF values

(b) GMM classification based on up-slope values

Figure 5.5: GMM classification based on MBF and up-slope values. Panels (a) and (b) show the classification scatter plots between up-slope and MBF values based on MBF and up-slope values respectively. The yellow points indicate the points in the hypo-perfusion class and the blue points indicate the points in the normal-perfusion class.



(a) CER histogram and KDE

(b) up-slope histogram and KDE

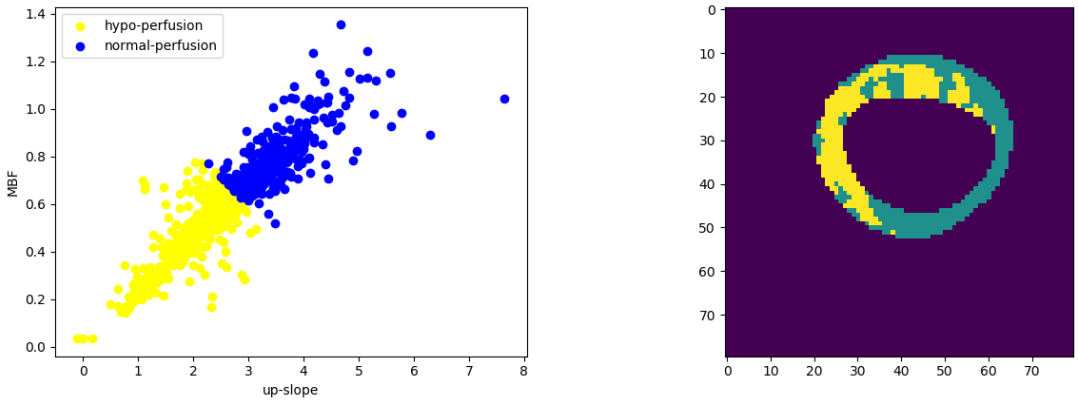
(c) MBF histogram and KDE

Figure 5.6: Panels (a) - (c) show the histograms and kernel density estimates of CER, up-slope and MBF.

and vice versa. Therefore, the GMM classification method cannot avoid unclear boundaries and single and small isolated segments issues if the parametric map is noisy.

Now, the reason of the generation of the single and small isolated segments is clear. The artifacts (information loss and errors generation during the manual contouring and motion correction) cause the low SNRs. Then the low SNRs cause inaccurate parameter estimations. Finally, The inaccurate estimations of parameters cause inaccurate classification, i.e. unclear boundary and single and small isolated segments. To address this issue, some methods can be adopted. Specifically,

- Improve the methods of parameter estimation to reduce the effects of the low SNRs.
- Improve the classification method to avoid unclear boundaries and single and small isolated segments.



(a) scatter plot

(b) classification map

Figure 5.7: GMM application to multiple parameters (up-slope and MBF). Panel (a) shows the scatter plot between the values of up-slope and MBF. The yellow points indicate the points in the lesion class and the blue points indicate the points in the healthy class. Panel (b) shows the classification map using GMM based on multiple parameters. The yellow segments indicate healthy tissues. The dark green segments indicate lesion tissues. The black pixels indicate the background.

Either or both methods can improve the classification results. A straightforward implementation is to introduce spatial information during the phase of parameter estimation and the phase of classification. The specific solutions are explicitly illustrated in Sections 5.3, 5.4 and Chapters 6 and 7.

The applications of the GMM classification method in this section are based on a single parameter. A potential extension is to apply the GMM method to multiple parameters. The parameter CER is discarded because it cannot accurately delineate the hypo-perfusion region in panel (a), Figure 5.4. Therefore, the GMM is fitted using the up-slope and MBF as inputs. Since the TIC is normalised by the baseline, the noise is assumed to be from the same distribution for all pixels in the myocardium. Therefore, the covariance matrix is set to be "tied", which means that both groups (lesion and healthy groups) share the same general covariance matrix. Figure 5.7 shows the results that are obtained by the GMM model fitted using the up-slope and MBF. The classification map in panel (b), Figure 5.7 is a mixture of panels (b) and (c) in Figure 5.4. Given the ground truth in panel (b), Figure 4.4, the classification map in panel (b), Figure 5.7 is not as accurate as the MBF classification map. The reason is that the parameter up-slope negatively affects the final result. Specifically, the estimations of the up-slope only make use of the SI of the pixels within myocardium, and the estimations of the MBF use the SI of pixels within both myocardium and LV blood pool.

5.3 Spatially variant finite mixture modelling of myocardial perfusion DCE-MRI

5.3.1 Introduction

In Section 5.2, Gaussian mixture model classification to myocardial perfusion DCE-MRI has been introduced. This application has also addressed the unclear boundary and single and small isolated segments issues. One possible solution to these issues is to improve the Gaussian mixture model classification method by importing neighbourhood information.

The spatially variant finite mixture model (SVFMM) is an extension of the Gaussian mixture model. Both models use finite mixtures to describe the data with a mixed distributions structure. However, the SVFMM improves the GMM by combining signal intensities (or other parameters) with neighbourhood information and therefore obtaining more realistic classification maps. The SVFMM method is applied to the original signals in this section. However, it is just an example of the application. In fact, this method can be directly applied to other parametric maps, e.g. the up-slope and MBF, with the replacement from signal intensity to the target parameters. It is worth clarifying that the proposed method has not been applied to other parametric maps for the reason that this method was developed in the early stage of the PhD program.

The content in this section has been published in [131]:
***Yalei Yang**, Hao Gao, Colin Berry, Aleksandra Radjenovic and Dirk Husmeier. Quantification of Myocardial Perfusion Lesions using Spatially Variant Finite Mixture Modelling of DCE-MRI. Proceedings of the International Conference on Statistics: Theory and Applications (ICSTA) (2019).*

5.3.2 Statistical model

Let x^i denote the signal intensity (or other parameter) at the i^{th} pixel of an MR image ($i = 1,...,N$), where N is the total number of pixels in this image. $K = 2$ clusters, "lesion" and "healthy tissue", are assumed. Parameters $\pi_j^i = P(j|x^i)$ denote the prior probabilities of the i^{th} pixel belonging to the j^{th} cluster ($j = 1,2$), and they have to satisfy the constrains that $0 \leq \pi_j^i \leq 1$ and $\pi_1^i + \pi_2^i = 1$. Let $\boldsymbol{\pi}_j$ denote the vector for cluster j, which is $\boldsymbol{\pi}_j = \{\pi_j^1, \pi_j^2, ..., \pi_j^N\}$, and let $\boldsymbol{\Pi}$ denote the set of probability vectors $\boldsymbol{\Pi} = \{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2\}$. Similarly, $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ denotes the set of component parameters, which are the parameters of the corresponding mixture components. The probability density function for each observation x^i is

$$f(x^i|\boldsymbol{\Pi}, \boldsymbol{\Theta}) = \pi_1^i \phi(x^i|\boldsymbol{\theta}_1) + \pi_2^i \phi(x^i|\boldsymbol{\theta}_2) \tag{5.20}$$

where $\phi(x^i|\boldsymbol{\theta}_j)$ is a Gaussian distribution with parameters $\boldsymbol{\theta}_j = \{\mu_j, \sigma_j^2\}$. The parameter μ_j is the mean of the j^{th} Gaussian component and σ_j^2 is its variance. Let $\mathbf{x} = \{x^1, x^2, \ldots, x^N\}$ denote

the full data, i.e. the set of all pixel intensities of the MR image. The log-likelihood function of the data is given by

$$l(\boldsymbol{\Theta}|\mathbf{x},\boldsymbol{\Pi}) = \sum_{i=1}^{N} \log[\pi_1^i \phi(x^i|\boldsymbol{\theta}_1) + \pi_2^i \phi(x^i|\boldsymbol{\theta}_2)]. \tag{5.21}$$

As mentioned in Section 5.2, the maximization of this so-called 'incomplete' log-likelihood function has no closed-form solution, due to the summation terms inside the logarithm. A standard procedure, therefore, is to apply the EM algorithm by incorporation of latent variables [28]. To be specific, we include an unobserved binary latent variable k_j^i [18], which indicates which component the i^{th} pixel belongs to, and define the probability function

$$f(x^i, k_j^i|\boldsymbol{\Pi},\boldsymbol{\Theta}) = (\pi_1^i \phi(x^i|\boldsymbol{\theta}_1))^{k_1^i} \times (\pi_2^i \phi(x^i|\boldsymbol{\theta}_2))^{k_2^i} \tag{5.22}$$

where the k_j^i are subject to the constraints that $k_j^i \in \{0,1\}$ and $k_1^i + k_2^i = 1$. This implies that any pixel i can only be represented by one Gaussian component. Given the set of binary latent variables k_j^i, we have the complete log-likelihood function

$$l(\boldsymbol{\Theta}|\mathbf{x},\boldsymbol{\Pi},\mathbf{K}) = \sum_{i=1}^{N} \Big[k_1^i \Big(\log(\pi_1^i) + \log \phi(x^i|\boldsymbol{\theta}_1) \Big) + k_2^i \Big(\log(\pi_2^i) + \log \phi(x^i|\boldsymbol{\theta}_2) \Big) \Big] \tag{5.23}$$

where $\mathbf{K} = \{\mathbf{k_1}, \mathbf{k_2}\}$ and $\mathbf{k_j} = \{k_j^1, k_j^2, ..., k_j^N\}$. By substituting $k_2^i = 1 - k_1^i$ and $\pi_2^i = 1 - \pi_1^i$, we can maximize equation (5.23) for given latent variables $\mathbf{K}$. To reduce the number of single and small isolated segments mentioned in Section 5.2, the prior knowledge, adjacent pixels are more likely to have the same label, is introduced. In the SVFMM, this prior knowledge is translated to that the prior probabilities $\boldsymbol{\Pi}$ of spatially related pixels are similar. Specifically, the prior distribution over the parameter $\boldsymbol{\Pi}$ [10] [133] [140] is introduced to take the spatial information into account. This prior distribution is a Markov random field prior following Gibbs function:

$$P(\boldsymbol{\Pi}) = \frac{1}{C} \exp(-U(\boldsymbol{\Pi})) \tag{5.24}$$

where

$$U(\boldsymbol{\Pi}) = \beta \sum_{i=1}^{N} V_{\mathcal{N}_i}(\boldsymbol{\Pi}) \tag{5.25}$$

and C is a normalizing constant to ensure the integral over $P(\boldsymbol{\Pi})$ is 1. The so-called potential function $U(\boldsymbol{\Pi})$ depends on a regularization parameter $\beta > 0$. Each term in the sum, $V_{\mathcal{N}_i}(\boldsymbol{\Pi})$, is a function of $\{\pi_j^m\}$ which contains the neighbourhood information of the i^{th} pixel and is defined as

$$V_{\mathcal{N}_i}(\boldsymbol{\Pi}) = \sum_{m \in \mathcal{N}_i} g(u_{i,m}) \tag{5.26}$$

where $\mathcal{N}_i$ is the set of neighbourhood pixels of the i^{th} pixel (set to the 4 nearest neighbours in the present work) and $u_{i,m}$ is the distance between two mixing coefficients π_j^i and π_j^m. Here we

use the squared Euclidean distance suggested in [141].

$$u_{i,m} = \sum_{j=1}^{2} (\pi_j^i - \pi_j^m)^2. \tag{5.27}$$

The function $g(u)$ must be non-negative and monotonically increasing [10] [133] [140]. Standard choices are $g(u) = (1+u^{-1})^{-1}$ [140] or $g(u) = u$ [10]. For the work reported in this section, we have used $g(u) = u$. Therefore, the posterior distribution given "complete" data $\{\boldsymbol{x}, \boldsymbol{K}\}$, $P(\boldsymbol{\Theta}, \boldsymbol{\Pi}|\mathbf{x}, \boldsymbol{K})$ is:

$$P(\boldsymbol{\Theta}, \boldsymbol{\Pi}|\mathbf{x}, \boldsymbol{K}) = \frac{P(\mathbf{x}, \boldsymbol{K}|\boldsymbol{\Theta}, \boldsymbol{\Pi})P(\boldsymbol{\Pi})}{P(\mathbf{x}, \boldsymbol{K})} \propto P(\mathbf{x}, \boldsymbol{K}|\boldsymbol{\Theta}, \boldsymbol{\Pi})P(\boldsymbol{\Pi}) \tag{5.28}$$

where $P(\mathbf{x}, \boldsymbol{K}|\boldsymbol{\Theta}, \boldsymbol{\Pi})$ can be found in equation (5.22) and $P(\boldsymbol{\Pi})$ can be found in equation (5.24).

5.3.3 EM algorithm for the SVFMM

Similar to the EM algorithm for Gaussian mixture model in Section 5.2.2, the EM algorithm can also be used for the SVFMM. In the Expectation step, the expected value of the latent variables $\mathbf{K}$ is calculated as

$$\begin{aligned}
\gamma_j^i = E(k_j^i|x^i, \pi_j^i, \boldsymbol{\theta}_j) &= \sum_{j=1}^{2} k_j^i f(k_j^i|x^i, \pi_j^i, \boldsymbol{\theta}_j) \\
&= 1 \times f(k_j^i = 1|x^i, \pi_j^i, \boldsymbol{\theta}_j) + 0 \times f(k_j^i = 0|x^i, \pi_j^i, \boldsymbol{\theta}_j) \quad j = 1,2 \\
&= \frac{\pi_j^i \phi(x^i|\boldsymbol{\theta}_j)}{\pi_1^i \phi(x^i|\boldsymbol{\theta}_1) + (1-\pi_1^i)\phi(x^i|\boldsymbol{\theta}_2)}. \quad j = 1,2
\end{aligned} \tag{5.29}$$

The Maximization step is different from EM algorithm for the Gaussian mixture model. Instead of maximizing the expected complete log-likelihood function, the SVFMM method maximizes the expected posterior of $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, which can be computed by

$$E[P(\boldsymbol{\Theta}, \boldsymbol{\Pi}|\mathbf{x}, \boldsymbol{K})] = E\left[\frac{P(\mathbf{x}, \boldsymbol{K}|\boldsymbol{\Theta}, \boldsymbol{\Pi})P(\boldsymbol{\Pi})}{P(\mathbf{x}, \boldsymbol{K})}\right] \propto E[P(\mathbf{x}, \boldsymbol{K}|\boldsymbol{\Theta}, \boldsymbol{\Pi})P(\boldsymbol{\Pi})]. \tag{5.30}$$

It is equivalent to maximize $E[P(\boldsymbol{\Theta}, \boldsymbol{\Pi}|\mathbf{x}, \boldsymbol{K})]$ by maximizing the logarithm of $E[P(\mathbf{x}, \boldsymbol{K}|\boldsymbol{\Theta}, \boldsymbol{\Pi})P(\boldsymbol{\Pi})]$, which is

$$\begin{aligned}
Q_{\text{MAP}}(\boldsymbol{\Pi}, \boldsymbol{\Theta}) &= \log(\prod_{i=1}^{N} f(x^i, \gamma_j^i|\boldsymbol{\Pi}, \boldsymbol{\Theta})P(\boldsymbol{\Pi})) \\
&\propto \sum_{i=1}^{N}\sum_{j=1}^{2} \gamma_j^i[\log(\pi_j^i) + \log(\phi(x^i|\theta_j^i))] - \beta \sum_{i=1}^{N}\sum_{m\in\mathcal{N}_i} g(u_{i,m}).
\end{aligned} \tag{5.31}$$

To maximize this function with respect to $\mathbf{\Pi}$ and $\mathbf{\Theta}$, it is straightforward to calculate the corresponding partial derivatives and setting them to 0:

$$\frac{\partial Q_{\text{MAP}}}{\partial \mu_j} = \sum_{i=1}^{N} \gamma_j^i \times \frac{x^i - \mu_1}{2\sigma_j^2} = 0 \tag{5.32}$$

and

$$\mu_1 = \frac{\sum_{i=1}^{N} \gamma_1^i x^i}{\sum_{i=1}^{N} \gamma_1^i} \quad \text{and} \quad \mu_2 = \frac{\sum_{i=1}^{N} \gamma_2^i x^i}{\sum_{i=1}^{N} \gamma_2^i}. \tag{5.33}$$

Similarly,

$$\frac{\partial Q_{\text{MAP}}}{\partial \sigma_j^2} = \sum_{i=1}^{N} \gamma_j^i (-\frac{1}{2\sigma_j^2} + \frac{(x^i - \mu_j)^2}{2\sigma_j^2}) = 0 \tag{5.34}$$

and

$$\sigma_1^2 = \frac{\sum_{i=1}^{N} \gamma_1^i (x^i - \mu_1)^2}{\sum_{i=1}^{N} \gamma_1^i} \quad \text{and} \quad \sigma_2^2 = \frac{\sum_{i=1}^{N} \gamma_2^i (x^i - \mu_2)^2}{\sum_{i=1}^{N} \gamma_2^i}. \tag{5.35}$$

For the parameters $\mathbf{\Pi}$, only π_1^i is under consideration because of the constraint $\pi_2^i = 1 - \pi_1^i$. Substituting $\pi_2^i = 1 - \pi_1^i$ and $\gamma_2^i = 1 - \gamma_1^i$ into equation (5.31), it turns to be

$$Q_{\text{MAP}}(\pi_1^i) \propto \gamma_1^i \log \pi_1^i + (1 - \gamma_1^i) \log(1 - \pi_1^i) - \beta \sum_{m \in \mathcal{N}_i} \sum_{j=1}^{2} (\pi_j^i - \pi_j^m)^2 \tag{5.36}$$

and

$$\frac{\partial Q_{\text{MAP}}}{\partial \pi_1^i} = \frac{\gamma_1^i}{\pi_1^i} - \frac{1 - \gamma_1^i}{1 - \pi_1^i} - 4\beta \sum_{m \in \mathcal{N}_i} (\pi_1^i - \pi_1^m) = 0. \tag{5.37}$$

Since both $\pi_1^i \neq 0$ and $1 - \pi_1^i \neq 0$, multiplying $\pi_1^i(1 - \pi_1^i)$ on both sides of equation (5.37)

$$h(\pi_1^i) = 4\beta M(\pi_1^i)^3 - 4\beta(\sum_{m \in \mathcal{N}_i} [\pi_1^m + 1])(\pi_1^i)^2 + [(4\beta \sum_{m \in \mathcal{N}_i} \pi_1^m) - 1]\pi_1^i + \gamma_1^i = 0 \tag{5.38}$$

where M is the size of the neighbourhood ($M = 4$ in the present work). The mixing parameters π_2^i can be calculated by

$$\pi_2^i = 1 - \pi_1^i. \tag{5.39}$$

Since $h(\pi_1^i) \to \gamma_1^i > 0$ when $\pi_1^i \to 0$ and $h(\pi_1^i) \to \gamma_1^i - 1 < 0$ when $\pi_1^i \to 1$, there must be at least one root located in the interval $(0, 1)$, and this root is the target value of π_1^i. In this study, the numerical solution of equation (5.38) can be obtained using a line search in the interval $(0, 1)$. Note that for $\beta = 0$, the SVFMM reduces to the standard GMM. This means that the GMM is a special case of the SVFMM. The details of the EM algorithm for the SVFMM is shown in Algorithm 2.

Algorithm 2: EM algorithm for SVFMM

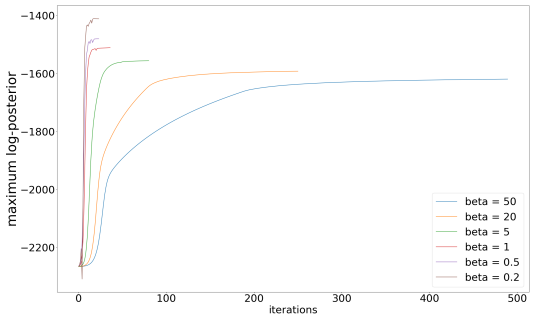
Data: $\boldsymbol{x} = \{x^1, x^2, ..., x^N\}$
Output: Parameters $\boldsymbol{\Pi}, \boldsymbol{\Theta}$
Set initial values of $\boldsymbol{\Pi}, \boldsymbol{\Theta}$. Define a fixed value for the regularization parameter β. ;
for $j \leftarrow 1$ **to** K **do**
  E step: **for** $i \leftarrow 1$ **to** N **do**
    Calculate the quantity γ_j^i according to equation (5.29) ;
  end
  M step: Re-estimate parameters μ_1, μ_2, σ_1^2, σ_2^2, π_1^i, π_2^i using equation (5.33), equation (5.35), equation (5.38) and equation (5.39) given γ_j^i calculated in E step;
end
Evaluate Q_{MAP} using equation (5.31) and check for convergence. If the convergence criterion, i.e. $|Q_{\text{MAP}_{\text{OLD}}} - Q_{\text{MAP}_{\text{NEW}}}| < \xi$, where ξ is a given positive small value, is not satisfied, return to E step;

5.3.4 Application of SVFMM on myocardial perfusion DCE-MRI

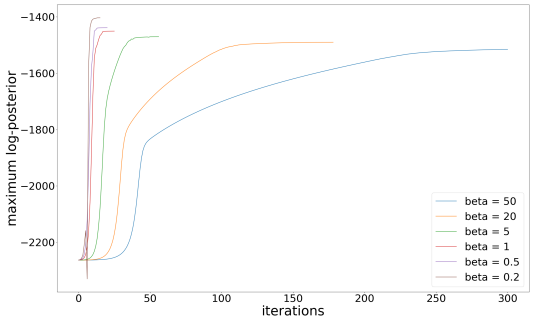
In this section, the proposed SVFMM method has been applied on both clinical and synthetic data. Gaussian mixture model classification method (see Section 5.2) has been used to be the benchmark. The original images (The ME image at time point 35) of the myocardial perfusion DCE-MRI are used as the clinical data. The results generated by the original image at the time of myocardium enhanced stage from a patient has been shown in this section. It is worth mentioning that though the original image has been used as an example for the SVFMM classification method, other parameters such as "peak signal intensity" and "MBF" are also applicable for the proposed SVFMM method by swapping signal intensity of a pixel to "peak signal intensity" or "MBF" of that pixel.

The synthetic data is used to validate the algorithm illustrated in Algorithm 2, and they are designed so as to mimic the clinical data. Specifically, the shape and the size of the synthetic myocardium is designed to be the same as for the clinical data. The whole myocardium ring is divided into two parts. The top half is designed to be "healthy tissue" and the bottom half is designed to be "lesion tissue". The signal intensity of the "healthy tissue" group is sampled from $N(\mu_h, \sigma_h^2)$, and the signal intensity of the "lesion group" is sampled from $N(\mu_l, \sigma_l^2)$. The choice of parameters $\mu_h, \mu_l, \sigma_h^2, \sigma_l^2$ is empirical. To be specific, they are chosen according to the GMM results: μ_h and σ_h^2 have been set to the mean and variance of the "healthy tissue" groups from the GMM clustering results; similarly μ_l and σ_l^2 have been obtained from the "lesion tissue" group. This design of the synthetic data mimics the same SNR (signal-to-noise ratio) as for the clinical data. The number of wrongly segmented pixels using the GMM is positively correlated with the SNR value. Therefore, the performance of the GMM method for the synthetic data reflects the SNR for the clinical data.

Figure 5.8 shows the evolution of the log-posterior as a function of the number of EM cycles



(a) clinical data

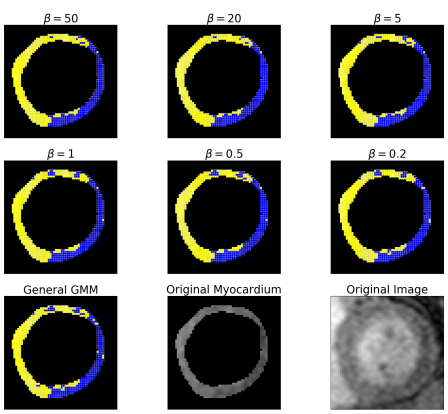


(b) synthetic data

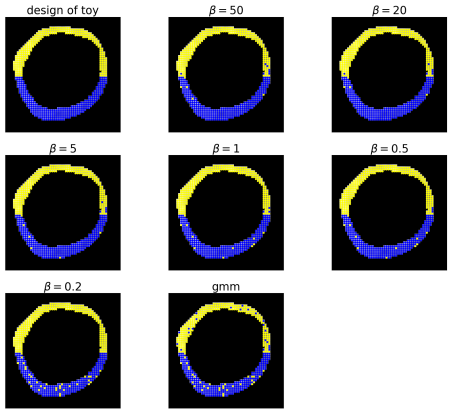
Figure 5.8: Evolution of the log-posterior for clinical (top panel) and synthetic (bottom panel) data during the optimization procedure. The horizontal axis shows the number of EM cycles.

for different choices of the regularization parameter β. It can be seen that the rate of convergence depends on β, with increasing values of β leading to a slower convergence. The regularization parameter β denotes the weight of the neighbourhood information usage. Specifically, larger β means higher usage of neighbourhood information. For the extreme case, $\beta = 0$, the usage of neighbourhood information is 0 and the SVFMM method degenerates to the GMM method. Therefore, the regularization parameter β not only controls the weight of the usage of neighbourhood information, but also controls the computational complexity of the EM algorithm for the SVFMM method.

Figure 5.9 shows a comparison of the classification results “healthy tissue” versus “lesion” between GMM and SVFMM, for different values of β. The blue areas are labelled as "lesion tissue" and the yellow areas are labelled as "healthy tissue". The original MR image in the bottom right of panel (a) shows that there is a large shadow area located in the right bottom of the ring. However, the SNR is poor. Therefore, it is challenging to determine the lesion visually. Our results demonstrate that the segmentation obtained with the SVFMM is different from that obtained with the GMM. As mentioned in Section 5.2.3, the GMM classification method has the unrealistic unclear boundary and single and small isolated segments issues. These issues also exist here because of the low SNR of the original image. Therefore, the number of isolated pixels



(a) clinical data



(b) synthetic data

Figure 5.9: The comparison of lesion detection of clinical and synthetic data. Panel (a) shows results for the clinical data, while panel (b) was obtained from the synthetic data. The blue areas are labelled as "lesion tissue" and the yellow areas are labelled as "healthy tissue". The centre bottom image in panel (a) is the segmented MRI myocardium grey map and the right bottom image in panel (a) is the original MRI left ventricle grey map. These two images are used to show the original source data.

is used as a performance criterion for the clinical data because of the pathology of coronary heart diseases illustrated in Section 5.2.3. For the synthetic data, on the other hand, the gold standard is known, as the location and size of the lesion are designed. Therefore, the number of misclassified pixels has been used as the performance criterion. There are fewer isolated pixels when applying the SVFMM method (for all values of β) to the clinical data. For the synthetic data, SVFMM leads to fewer misclassifications of pixels than GMM. This means that the SVFMM method indeed improves the classification in terms of the number of isolated pixels. Table 5.2 summarizes the numbers of isolated pixels for the clinical data and wrongly segmented (i.e. misclassified) pixels for the synthetic data. When $\beta = 20$, the number of isolated and misclassified pixels is the lowest. It can be seen that the results of $\beta = 50$ are not as good as the results of $\beta = 20$. The reason could be that the convergence for the case of $\beta = 50$ is slow, and the EM stops when the log-likelihood have not achieved maximum. Overall, the The SVFMM has fewer than half the number of isolated pixels than the GMM method for clinical data, and only about $\frac{1}{5}$ times the number of misclassified pixels than the GMM method for synthetic data. One possible reason that the proposed method performed better in the synthetic data than the clinical data is that the clinical data has spatially related noise and the synthetic data has white noise. The spatially related noise will generate jointly isolated misclassified pixels and the proposed method cannot deal with it.

By taking the neighbourhood information into account, the total number of isolated pixels decreases substantially in the clinical data, and the number of misclassified pixels in the

Table 5.2: Number of isolated pixels for the clinical data and misclassified pixels for the synthetic data.

	clinical data	synthetic data
SVFMM $\beta = 50$	27	12
SVFMM $\beta = 20$	25	10
SVFMM $\beta = 5$	32	11
SVFMM $\beta = 1$	44	14
SVFMM $\beta = 0.5$	47	14
SVFMM $\beta = 0.2$	49	33
GMM	59	55

synthetic data also decreases dramatically. However, there are still some isolated pixels, and the performance on the synthetic data appears to be noticeably better than on the clinical data. There are two main reasons for this trend. Firstly, some pixels are located close to the endocardium and the epicardium (i.e. the inner and outer boundaries of the ring). In Figure 5.9 panel (a), in the centre bottom of the rings, some pixels are labelled as "healthy" but surrounded by "lesion". The reason is that these pixels are close to the blood pool. In DCE-MRI, the blood signal increases (bright, hyperintense in the grey-scale) and the zone of impaired perfusion appears dark (hypointense) due to a relative lack of contrast agent mixed in blood. The hypointense zone spatially depicts the zone of impaired perfusion. Those pixels that are affected by blood will appear in the extreme bright range of the grey scale, and the segmentation will be affected accordingly. This effect has not been simulated in the synthetic data and constitutes the main reason why the SVFMM algorithm works better for the synthetic than the clinical data. Since the contouring has been implemented manually, the blood affection cannot be avoided ideally. Secondly, some isolated pixels are located in some areas whose SNR is very low. A consequence of this low SNR value is that too many adjacent pixels get jointly misclassified, and the correction mechanism based on neighbourhood information, on which SVFMM is based, breaks down.

In conclusion, compared to GMM classification method, although the SVFMM classification method is capable to reduce the number of unrealistic single and small isolated segments, it cannot eliminate all unrealistic single and small isolated segments. One reason is that the SVFMM imports the neighbourhood information to the mixing coefficients π^i. This is an indirect way to introduce the neighbourhood information. A more directed method is that the neighbourhood information is imported to the latent variable, the label $\boldsymbol{K}$, and use an iterative algorithm to remove all isolated segments. A further method is to use morphological operations to find the largest connected segment. Moreover, if the SNR of a certain image is very low (or the contouring of the myocardium is inaccurate), the number and the size of the jointly misclassified segments will be high. The SVFMM method is not reliable to tackle this kind of high complexity situation.

5.4 Markov random field constrained Gaussian mixture model classification

5.4.1 Introduction

In the previous Section 5.3, a spatial modification of the GMM method, SVFMM, has been introduced to classify the myocardial perfusion DCE-MRI. In practice, the SVFMM method cannot remove all isolated pixels. Therefore, in this section, a Markov random field constrained Gaussian mixture model (GMM-MRF) has been introduced to improve it. An iterative algorithm, iterated conditional modes (ICM), has been applied to obtain the maximum a posteriori probability (MAP) estimate.

The content in this section has been accepted to publish in the International Conference on Statistics: Theory and Applications (ICSTA) (2022):
***Yalei Yang**, Hao Gao, Colin Berry, Aleksandra Radjenovic and Dirk Husmeier. Myocardial Perfusion Classification Using A Markov Random Field Constrained Gaussian Mixture Model.*

5.4.2 Statistical model

Let x_i represent the signal intensity (or MBF, up-slope) of pixel i of an image (or a parametric map) where $i = 1,2,...,N$ and N is the number of pixels in this image. Any pixel i in this image has either "healthy" or " lesion" label. The pixel with the healthy label has the value $k = 0$, and on the other hand, the pixel with the lesion label has the value $k = 1$. The first assumption of this model is that all pixels that have healthy labels are from one Gaussian distribution with mean μ_0 and variance σ_0^2. Similarly, all pixels that have lesion labels are formed by Gaussian distribution with mean μ_1 and variance σ_1^2. Therefore, the statistical model can be expressed as:

$$P(x_i|\boldsymbol{\mu},\boldsymbol{\sigma}^2,k_i) = \frac{k_i}{\sqrt{2\pi}\sigma_1}\exp\left(\frac{-(x_i-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-k_i}{\sqrt{2\pi}\sigma_0}\exp\left(\frac{-(x_i-\mu_0)^2}{2\sigma_0^2}\right) \tag{5.40}$$

where $\boldsymbol{\mu} = \{\mu_0,\mu_1\}$, $\boldsymbol{\sigma}^2 = \{\sigma_0^2,\sigma_1^2\}$. Specifically, when a pixel i is in the healthy group and $k_i = 0$, its probability distribution is

$$P(x_i|\boldsymbol{\mu},\boldsymbol{\sigma}^2,k_i = 0) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(\frac{-(x_i-\mu_0)^2}{2\sigma_0^2}\right). \tag{5.41}$$

On the other hand, if a pixel is the in the lesion group, its probability distribution has a similar form to equation (5.41) with replacements of the parameters from μ_0 and σ_0^2 to μ_1 and σ_1^2. The advantage of this modelling is that given the label k_i of pixel i, the conditional distribution of x_i

is very simple. Let $X = \{x_1, x_2, ..., x_N\}$, the joint probability distribution of X is

$$P(X|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, K) = \prod_{i=1}^{N} P(x_i|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, k_i) \tag{5.42}$$

where $K = \{k_1, k_2, ..., k_N\}$. Because of the sum in equation (5.40), the log likelihood function of equation (5.42) is intractable. However, if the label K is known, the conditional likelihood function of equation (5.42) is surprisingly simple. Specifically, given a known $K = K^{'}$, the conditional negative log likelihood function can be expressed as

$$-\log P(X|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, K = K^{'}) \propto N_0 \log \sigma_0 + N_1 \log \sigma_1 + \sum_{m=1}^{N_0} \frac{(x_m - \mu_0)^2}{2\sigma_0^2} + \sum_{n=1}^{N_1} \frac{(x_n - \mu_1)^2}{2\sigma_1^2} \tag{5.43}$$

where N_0 is the number of pixels that are labelled as healthy and N_1 is the number of pixels that are labelled as lesions.

As mentioned in Section 5.3, the myocardial lesion pixels are assumed to be connected spatially because the blood supplement to the myocardium is from three main arteries (see details in Section 3.1.1). Similar to the SVFMM, the spatial information is also introduced using the Markov random field priors in this study (see details in Section 2.2.4). Specifically, the Markov random field prior for label k_i can be expressed as

$$P(k_i|k_{-i}) = \frac{1}{Q} \exp\left(-U(k_i|k_{-i})\right) \tag{5.44}$$

where k_{-i} denotes the set of all other labels except k_i. Specifically, $k_{-i} = \{k_j\}_{j \neq i}$. Q is a normalization constant of the prior distribution $P(k_i|k_{-i})$. The function U is usually called the energy function [142]. Since only the neighbours of pixel i is considered to affect the value of k_i, the energy function U is defined as

$$U(k_i|k_{-i}) = \frac{1}{T} \sum_{i \sim j} u(k_i|k_j) \tag{5.45}$$

where $i \sim j$ represents all pixels j that are the neighbours of pixel i and T is the weight parameter. In this study, the neighbours of pixel i is defined in Figure 5.10, and therefore u is defined as

$$u(k_i|k_j) = \begin{cases} -\left(\frac{1}{2}\right)^{o-1} & k_i = k_j, \\ \left(\frac{1}{2}\right)^{o-1} & k_i \neq k_j \end{cases} \tag{5.46}$$

where o is the degree of order for the neighbours. For instance, if pixel j is the first order neighbour of pixel i, $o = 1$, and if j is the second order neighbour, $o = 2$. In this work, $o \in \{1, 2\}$.

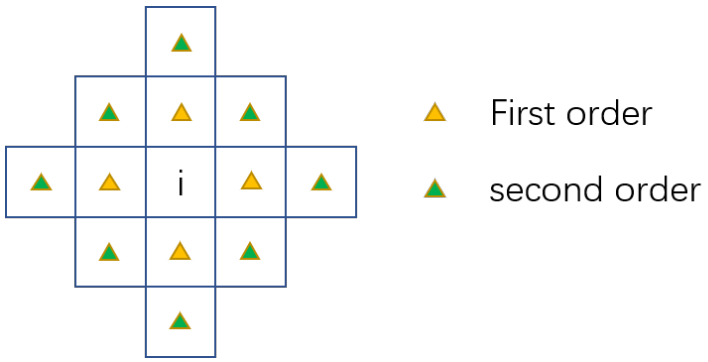


Figure 5.10: This figure shows pixel i and its neighbours.
The four pixels marked by yellow triangles are the first order neighbours. The eight pixels marked by green triangles are the second order neighbours. Four nearest neighbours are applied in this thesis because the myocardial wall is usually very thin. Therefore, the first order and the second order neighbours are enough in this work.

According to Bayes theorem in equation (2.4), the posterior distribution of k_i is

$$P(k_i|x_i,\boldsymbol{\mu},\boldsymbol{\sigma}^2,k_{-i}) \propto P(x_i|k_i,\boldsymbol{\mu},\boldsymbol{\sigma}^2)P(k_i|k_{-i}) \tag{5.47}$$

and the posterior distribution of K is

$$P(K|X,\boldsymbol{\mu},\boldsymbol{\sigma}^2,K^-) = \prod_{i=1}^{N} P(x_i|k_i,\boldsymbol{\mu},\boldsymbol{\sigma}^2)P(k_i|k_{-i}) \tag{5.48}$$

where K^- is the set of the neighbouring labels, $K^- = \{\{k_j\}_{i\sim j}\}_{i=1}^{N}$. This posterior distribution is intractable because of the sum in equation (5.40). However, an algorithm called iterated conditional modes (ICM) (see details in Sections 2.1.4 and 2.2.4) can be applied to find the Maximum a posteriori (MAP) of equation (5.48).

5.4.3 ICM algorithm

In Section 2.2.4, the ICM algorithm has been briefly reviewed. Two assumptions are satisfied in this case. Firstly, the observations X are conditionally independent given the labels K. Secondly, the labels K are locally dependent Markov random fields with Markov property. Therefore, the ICM algorithm can be applied in this case.

The negative logarithm of posterior distribution in equation (5.47) can be expressed as

$$-\log P(k_i|x_i,\boldsymbol{\mu},\boldsymbol{\sigma}^2,k_{-i}) \propto \begin{cases} \log\sigma_0 + \dfrac{(x_i-\mu_0)^2}{2\sigma_0^2} + \dfrac{1}{T}\displaystyle\sum_{i\sim j} u(k_i|k_j) & k_i = 0, \\ \log\sigma_1 + \dfrac{(x_i-\mu_1)^2}{2\sigma_1^2} + \dfrac{1}{T}\displaystyle\sum_{i\sim j} u(k_i|k_j) & k_i = 1. \end{cases} \tag{5.49}$$

This expression shows that given the value of k_i, its conditional posterior distribution is easy to obtain. Moreover, given $K = K^{'}$, the negative logarithm of conditional posterior distribution of K can be expressed as

$$-\log P(K = K^{'}|X, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, K^{-}) \propto N_0 \log \sigma_0 + N_1 \log \sigma_1 + \sum_{m=1}^{N_0} \frac{(x_m - \mu_0)^2}{2\sigma_0^2} + \sum_{n=1}^{N_1} \frac{(x_n - \mu_1)^2}{2\sigma_1^2} + \frac{1}{T} \sum_{i=1}^{N} \sum_{i \sim j} u(k_i|k_j). \tag{5.50}$$

The MAP estimation is the estimation that maximizes the posterior distribution or equivalently minimizes the negative logarithm of the posterior distribution. Although the negative logarithm of the posterior distribution is intractable, the simple form of the negative logarithm of conditional posterior distribution can be used to infer the MAP estimation using the ICM algorithm. Different from the methods illustrated in Sections 5.2 and 5.3, the MAP is obtained by the ICM algorithm in this section. Specifically, the MAP is obtained by iteratively updating the values of K to search the local minimum of equation (5.50). This also shows the disadvantage of the ICM algorithm. Since the algorithm is greedy, i.e. the updates will never move to lower posterior directions, the results are highly dependent on the initial values and may get stuck in local minima. In this case, multiple initial values are generated to deal with this issue. The difference between the EM algorithm and the ICM algorithm has been summarised in [143]. The EM algorithm aims to maximize the parameter θ ($\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ in our case) by iteratively proceeding:

1. $\theta^* = \arg\max_\theta E[\log P(\theta, k, x)]_{\gamma(k)}$
2. $\gamma(k) = P(k|\theta^*, x)$.

On the other hand, the ICM algorithm aims to maximize the parameter θ and the hidden parameter k by iteratively proceeding:

1. $\theta^* = \arg\max_\theta P(k^*, \theta, x)$
2. $k^* = \arg\max_k P(k, \theta^*, x)$.

Furthermore, the choice of the algorithm is considered as a trade-off between computational efficiency and accuracy [143]. In this study, we choose the ICM algorithm for the GMM-MRF because it has been initially adopted in [17] and the ICM algorithm is more computationally efficient. In fact, it is equivalent to the coordinate descent algorithm [144].

Moreover, the conditional posterior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ are also simple. Since the prior distribution of k_i is independent on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$, the estimations of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ can be obtained using the log conditional likelihood function. Specifically, taking derivatives with respect to $\boldsymbol{\mu}$

and $\boldsymbol{\sigma}^2$ to equation (5.43) to solve the maximum likelihood estimates (MLE), we have

$$\hat{\mu}_0 = \frac{1}{N_0} \sum_{m=1}^{N_0} x_m \tag{5.51}$$

and

$$\hat{\sigma}_0^2 = \frac{1}{N_0} \sum_{m=1}^{N_0} \left(x_m - \frac{1}{N_0} \sum_{m=1}^{N_0} x_m \right)^2 . \tag{5.52}$$

Similarly, we have

$$\hat{\mu}_1 = \frac{1}{N_1} \sum_{n=1}^{N_1} x_n \tag{5.53}$$

and

$$\hat{\sigma}_1^2 = \frac{1}{N_1} \sum_{n=1}^{N_1} \left(x_n - \frac{1}{N_1} \sum_{n=1}^{N_1} x_n \right)^2 . \tag{5.54}$$

Given the expression of the conditional posterior distribution and the MLE of the parameters, the details of MAP estimation using ICM algorithm can be found in algorithm 3.

5.4.4 Application of Markov random field constrained Gaussian mixture model

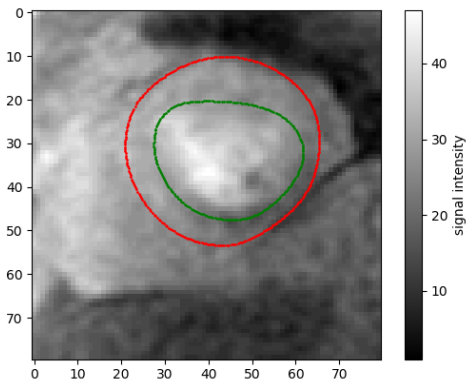
In this section, the GMM-MRF has been applied to both the original data and the parametric map to detect the hypo-perfusion region. The clinical data used in this section is the same as the data used in Section 5.2.

Figure 5.11 shows the maximum enhancement (ME) image, its corresponding GMM classification map and the LGE image of this DCE-MRI dataset. The ME image is the myocardium image obtained when the myocardium enhancement reaches the maximum. The single or small clusters mentioned in Section 5.2 exist in panel (b), Figure 5.11. The classification map shown in panel (b), Figure 5.11 is used to compare with the classification map generated by the GMM-MRF method. The LGE image shown in panel (c), Figure 5.11 is used to be the reference of the location of the myocardial defect. The GMM-MRF method introduced in this section aims to remove the small or single clusters in panel (b), Figure 5.11.

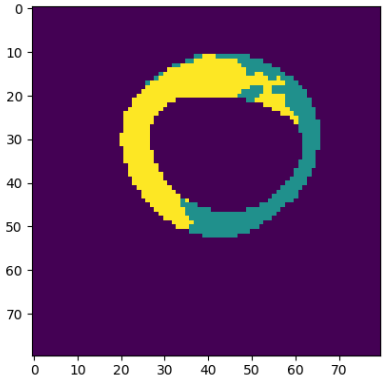
It can be noticed that there is a weight parameter T in equation (5.45). This parameter controls the strength of the neighbourhood information. In this study, three values, $T = 0.1, 1, 10$, are tested (larger values of T represent weaker spatial neighbourhood information). Figure 5.12 shows the classification maps generated by the GMM-MRF method using different weight parameters T. Larger values of T represent weaker spatial neighbourhood information. It can be seen that the single or small clusters are not completely removed from panel (c), Figure 5.12. However, the single or small clusters are completely eliminated from panels (a) and (b). Thus, when the weight of neighbourhood information is too small (or the value of T is too big), the

Algorithm 3: ICM algorithm for the Markov random field constrained Gaussian mixture model

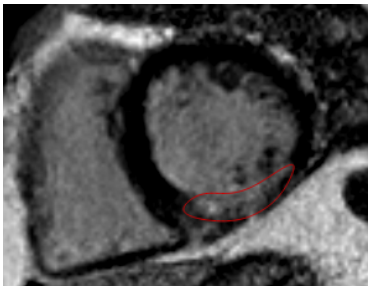
Data: $X = \{x_1, x_2, ..., x_N\}$
Output: $K = \{k_1, k_2, ..., k_N\}$
Generate initial values of K. The initial values can be fast obtained using simple methods,e.g. k-means [145] or GMM (see Sections 5.2). To avoid the local minima issue, multiple initial values can be generated;
Given K, estimate $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ using equations (5.51), (5.53), (5.52) and (5.54);
Given K, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$, calculate global negative logarithm posterior G^{old} using equation (5.50). Let $G^{\text{new}} = G^{\text{old}} - 0.001$ to enter the while loop;
while $G^{new} < G^{old}$ **do**
    Update G^{old} using G^{new};
    for $i \leftarrow 1$ **to** N **do**
        Given K^{old}, calculate $\boldsymbol{\mu}^{\text{old}}$ and $(\boldsymbol{\sigma}^2)^{\text{old}}$ using equations (5.51), (5.53), (5.52) and (5.54);
        Calculate old local negative logarithm posterior L_i^{old} for pixel i using equation (5.49);
        Give a new $k_i^{\text{new}} = |k_i^{\text{old}} - 1|$;
        Calculate new $\boldsymbol{\mu}^{\text{new}}$ and $(\boldsymbol{\sigma}^2)^{\text{new}}$ using equations (5.51), (5.53), (5.52) and (5.54);
        Given new k_i^{new}, $\boldsymbol{\mu}^{\text{new}}$ and $(\boldsymbol{\sigma}^2)^{\text{new}}$, calculate new local negative logarithm posterior L_i^{new} for pixel i using equation (5.49);
        If $L_i^{\text{new}} < L_i^{\text{old}}$, update old k_i^{old} using new k_i^{new}.
    end
    Given K, calculate $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ using equations (5.51), (5.53), (5.52) and (5.54);
    Given K, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$, calculate global negative logarithm posterior G^{new} using equation (5.50);
end
When the while loop stops, the values of K is the estimated MAP.



(a) ME



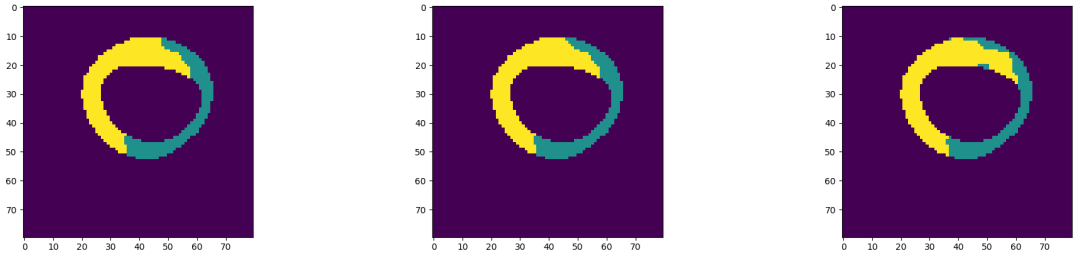
(b) GMM



(c) LGE

Figure 5.11: Panel (a) shows the ME image. Panel (b) shows the GMM classification map based on the ME image. The yellow pixels indicate the healthy tissues. The green pixels indicate the lesion tissues and the black pixels indicate the background (non-myocardium). Panel (c) shows the LGE image of this DCE-MRI dataset. the region inside the red curve shows the myocardial defect visually.

GMM-MRF method cannot remove all single or small clusters. The selection of the weight parameter is based on two criteria. Specifically, the first criterion is that the classified lesion is located in one connected space. This criterion is applied because of prior information that single and small isolated segments are physiologically unrealistic as described in Section 5.2. Furthermore, three main arteries supply blood to the myocardium and the myocardial defect is caused when one or more arteries are narrowed. Therefore, the myocardial defect is highly likely shown in a big continuous region because every two control regions of the main arteries are adjacent for the mid-cavity slice (see Figure 3.8). The limitation is that this method fails when there are two disconnected myocardial defects. This happens rarely but possibly. The second criterion is less prioritised than the first one. If there are multiple choices of the weight parameter that can generate the classification map that the classified lesion is located in one connected space, the parameter is selected by the negative conditional log likelihood. The negative conditional log likelihood can be calculated by equation (5.43). Therefore, the weight parameter is selected when the classification map has the classified lesion located in one connected space and has the



(a) $T = 0.1$ (b) $T = 1$ (c) $T = 10$

Figure 5.12: The GMM-MRF classifications to the original signals using different weight parameters. Panels (a-c) show the classification map using GMM-MRF method given $T = 0.1, 1, 10$ respectively. Larger values of T represent weaker spatial neighbourhood information. The yellow pixels indicate the healthy group. The green pixels indicate the lesion group and the black pixels indicate the background (non-myocardium).

minimum negative conditional log likelihood. Figure 5.13 shows the evolution of the negative conditional log likelihood given different weight parameters. When $T = 10$, the classified lesion regions are not located in a connected space. So this parameter is not selected. For the rest of parameter values, the case that $T = 1$ has the minimum negative conditional log likelihood, so it is considered to be the most appropriate one. It can also be noticed that though the negative conditional log likelihood is different for the cases $T = 0.1$ and $T = 1$, the classification differences are rather small, i.e. the labels for a few pixels in the boundary between healthy tissues and lesions are different. This shows the robustness of the GMM-MRF method.

From Figure 5.13, the GMM-MRF method reaches convergence very fast. Specifically, even for the most complex one $T = 0.1$, it only takes ten iterations to converge. This is a significant advantage of this algorithm. As a comparison, it takes about 500 iterations to reach the convergence for the SVFMM method illustrated in Section 5.3 when $\beta = 50$. Overall, it takes about 1 minute to run the GMM-MRF and more than 10 minutes to run the SVFMM.

The GMM-MRF method has also been applied to the MBF map in this study. Figure 5.14 shows the MBF estimation map and its corresponding GMM classification map. Many single or small clusters exist in panel (b), Figure 5.14. Similarly, the GMM-MRF method with three different weight parameters $T = 0.1, 1, 10$ are applied to panel (a) in Figure 5.14, and Figure 5.15 shows the corresponding classification results. When $T = 10$, the number of single or small clusters has been reduced. However, since the usage of the neighbourhood information is not sufficient, there are still some single or small clusters. On the other hand, when $T = 0.1$ and $T = 1$, the classification results are exactly the same (this can be validated by Figure 5.16). This shows the robustness of the GMM-MRF method. Similar to the original signals example, the negative conditional log likelihood can be used to select the weight parameter. Although the classification results are the same for these two weight parameters, the model selection phase is still necessary. For example, the convergence speed for different parameters might be differ-

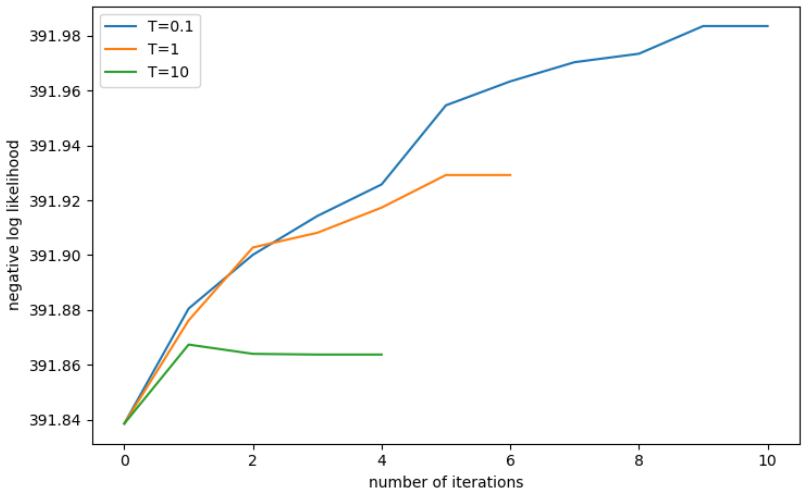


Figure 5.13: The negative conditional log likelihood evolution for different weight parameters based on original signals.

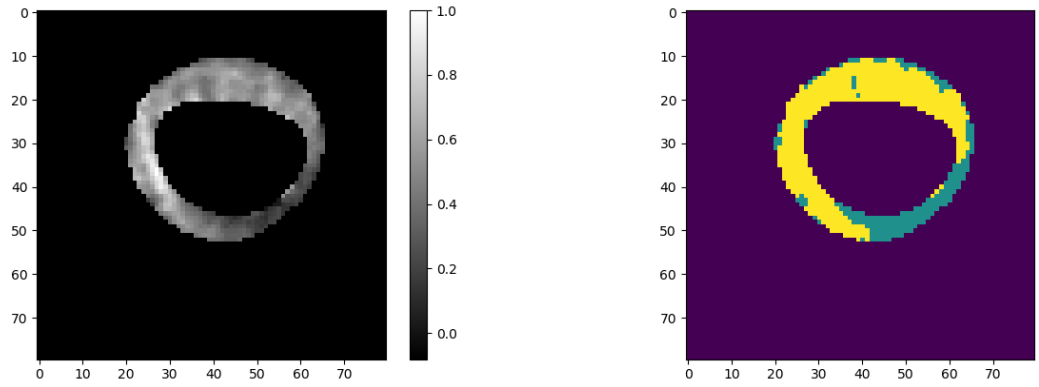
ent, and the fast one is preferred. Figure 5.16 shows the evolution of the negative conditional log likelihood for the MBF map. Surprisingly, the negative log likelihood decreases using the GMM-MRF method. This means that the initial value generated by the GMM using the EM algorithm gets stuck in a local minimum. Comparably, the convergence of the GMM-MRF for this case are also very fast. When $T = 0.1$ and $T = 1$, the algorithms stop at the same position, and the case, $T = 0.1$, is faster (2 iterations) and selected.

The GMM-MRF method converges very fast because the choice of the initial value is relatively close to the convergent state. Comparing panel (b) in Figure 5.14 and panel (a) in Figure 5.15, the main lesion cluster that is located at the right bottom of the myocardium in the image stays the same. All changes are to the single and small clusters.

Similar to the SVFMM method, the GMM-MRF method can also degenerate to the GMM. Specifically, when the weight parameter T approaches to infinity, the GMM-MRF changes to the GMM. In other words, the GMM method is a special case of the GMM-MRF method.

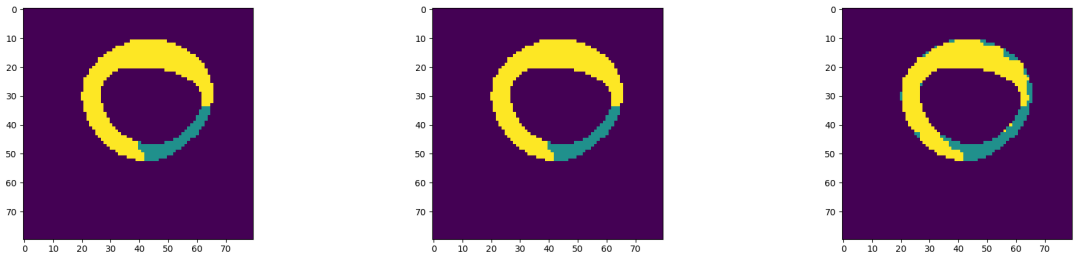
In conclusion, in this case, the GMM-MRF method can remove all single and small clusters given an appropriate weight parameter. This weight parameter can be selected based on two criteria. The first and the highest priority one is whether the classified lesion is located in one connected space. If multiple weight parameters can achieve this goal, they can be further selected by the negative conditional log likelihood and the convergence speed. Compared with the SVFMM method, the GMM-MRF method has a faster convergence speed. Moreover, the SVFMM cannot eliminate all single and small clusters, but the GMM-MRF, on the other hand, is able to remove all single and small clusters.

A disadvantage of the GMM-MRF method is that this method is highly dependent on the accuracy of the estimation method. In fact, all methods illustrated in this chapter have the same



(a) MBF (b) GMM

Figure 5.14: The MBF estimation map and its corresponding GMM classification map. In panel (b), the yellow pixels indicate the healthy tissues. The green pixels indicate the lesion tissues and the black pixels indicate the background (non-myocardium).



(a) $T = 0.1$ (b) $T = 1$ (c) $T = 10$

Figure 5.15: The GMM-MRF classifications to the MBF map using different weight parameters. Panels (a-c) show the classification map using GMM-MRF method given $T = 0.1, 1, 10$ respectively. The yellow pixels indicate the healthy group. The green pixels indicate the lesion group and the black pixels indicate the background (non-myocardium).

disadvantage. If the estimations of the features, e.g. the MBF and up-slope, are inaccurate, the methods introduced in this chapter might fail. Precisely, if the input parametric maps have the quality like panel (b) in Figure 4.6 or Figure 4.7, neither classification method can accurately predict the position of the hypo-perfusion region. The solution of this problem will be introduced in the conclusion and discussion section of this chapter, and other methods will be introduced in Chapter 6.

5.4.5 A temporal expansion of the GMM-MRF method

In the previous sections, the GMM-MRF has been introduced to make use of the spatial information to improve the classification results. In fact, this method can also introduce temporal information using the Markov random field priors. Especially, the original signals of the my-

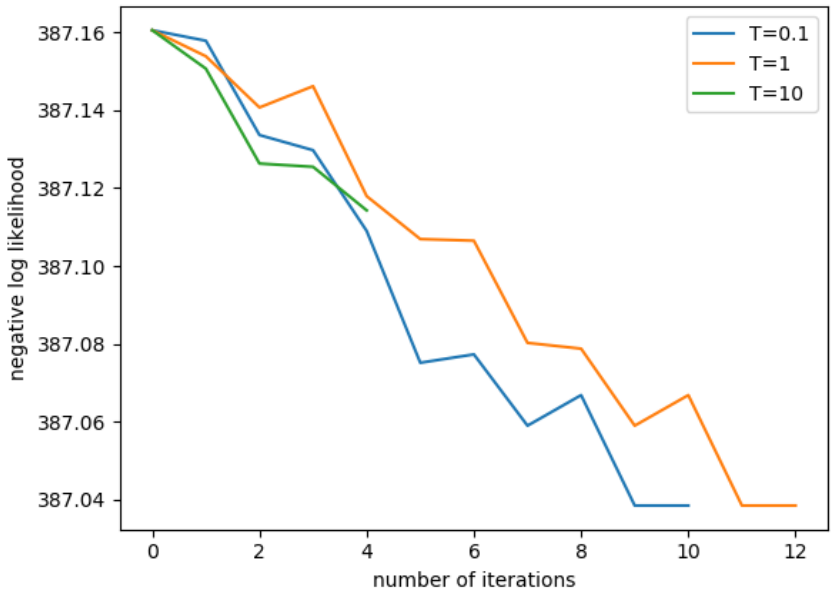


Figure 5.16: The negative conditional log likelihood evolution for different weight parameters based on the MBF map.

ocardial perfusion DCE-MRI is a series of images. Therefore, for any given pixel, its temporal neighbours, i.e. the same pixel at the adjacent time point, should also affect the label of the given pixel. Figure 5.17 shows the spatio-temporal neighbours of a given pixel.

The Markov random field priors for the spatio-temporal neighbours can be defined as an extension of the Markov random field priors for the spatial neighbours. Specifically, the expression of the Markov random field priors is

$$P(k_i|k_{-i}) = \frac{1}{Q}\exp(-U(k_i|k_{-i}) - V(k_i|k_{-i})) \tag{5.55}$$

where k_{-i} denotes the set of labels of spatio-temporal neighbours of pixel i. Q is a normalization constant of the prior distribution. The energy function $U(k_i|k_{-i})$ is defined the same as equation (5.45). The energy function $V(k_i|k_{-i})$ is defined as follows

$$V(k_i|k_{-i}) = \frac{1}{T_0}\sum_{i\sim l} v(k_i|k_l) \tag{5.56}$$

where $i \sim l$ represents the temporal neighbours l for pixel i and T_0 is the weight parameter. The

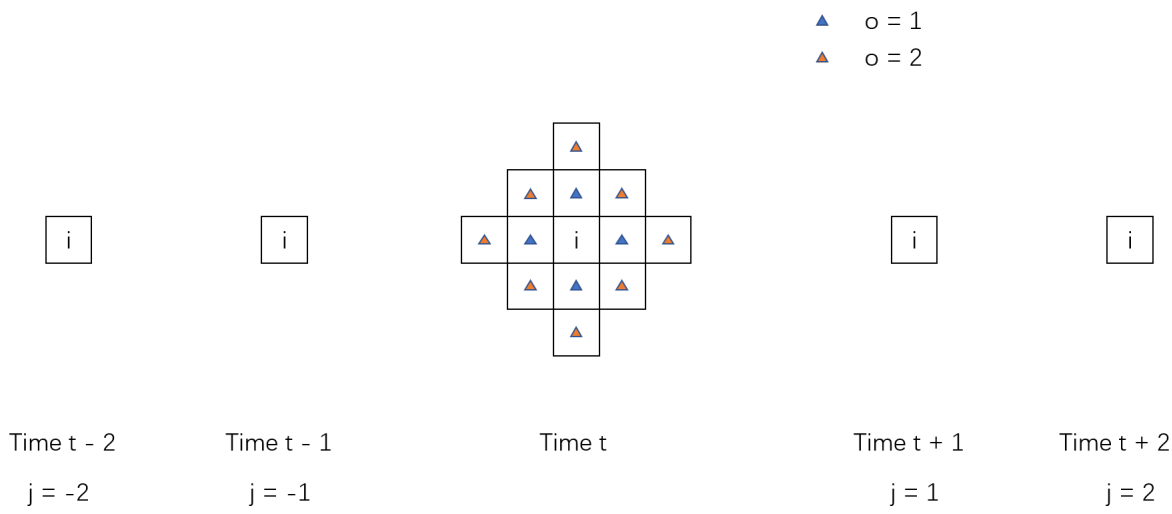


Figure 5.17: The spatio-temporal neighbours of a given pixel i.

definition of $v(k_i|k_l)$ is similar to $u(k_i|k_j)$. Specifically, we have

$$v(k_i|k_l) = \begin{cases} -\left(\frac{1}{2}\right)^{o-1} & k_i = k_l, \\ \left(\frac{1}{2}\right)^{o-1} & k_i \neq k_l. \end{cases} \tag{5.57}$$

Therefore, the local log conditional posterior turns to

$$-\log P(k_i|x_i,\boldsymbol{\mu},\boldsymbol{\sigma}^2,k_{-i}) \propto \begin{cases} \log\sigma_0 + \dfrac{(x_i-\mu_0)^2}{2\sigma_0^2} + \dfrac{1}{T}\sum\limits_{i\sim j} u(k_i|k_j) + \dfrac{1}{T_0}\sum\limits_{i\sim l} v(k_i|k_l) & k_i = 0, \\ \log\sigma_1 + \dfrac{(x_i-\mu_1)^2}{2\sigma_1^2} + \dfrac{1}{T}\sum\limits_{i\sim j} u(k_i|k_j) + \dfrac{1}{T_0}\sum\limits_{i\sim l} v(k_i|k_l) & k_i = 1 \end{cases} \tag{5.58}$$

and the global log conditional posterior for image t turns to

$$\begin{aligned} -\log P(K = K^{'}|X,\boldsymbol{\mu},\boldsymbol{\sigma}^2,K^{-}) \propto N_0\log\sigma_0 + N_1\log\sigma_1 + \sum_{m=1}^{N_0}\frac{(x_m-\mu_0)^2}{2\sigma_0^2} + \sum_{n=1}^{N_1}\frac{(x_n-\mu_1)^2}{2\sigma_1^2} \\ + \frac{1}{T}\sum_{i=1}^{N}\sum_{i\sim j} u(k_i|k_j) + \frac{1}{T_0}\sum_{i=1}^{N}\sum_{i\sim j} v(k_i|k_l). \end{aligned} \tag{5.59}$$

Since this expansion introduces a new dimension t, the global log conditional posterior for a serial images $t = 1,2,...,M$ can be defined as

$$-\sum_{t=1}^{M} P(K_t = K_t^{'}|X_t,\boldsymbol{\mu}_t,\boldsymbol{\sigma}_t^2,K_t^{-}) \tag{5.60}$$

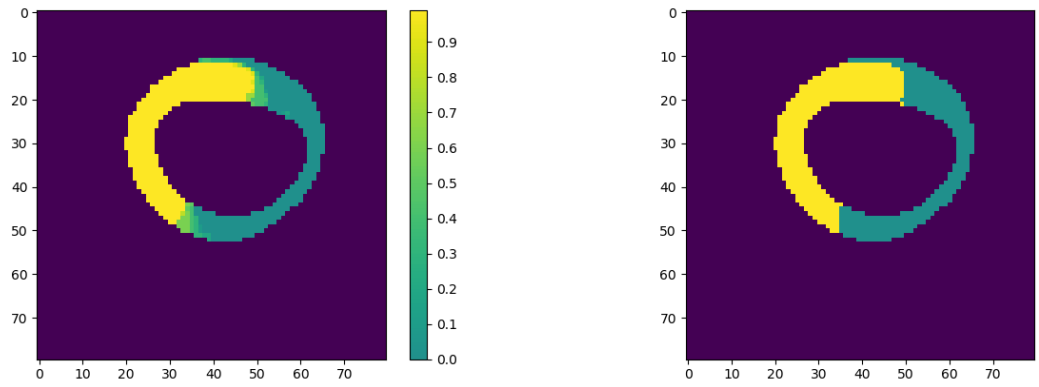
where the subscript t denotes the time index of the image. The ICM algorithm for the spatio-temporal GMM-MRF can be found in algorithm 4.

Algorithm 4: ICM algorithm for the spatio-temporal Markov random field constrained Gaussian mixture model

Data: $X = \{x_1, x_2, ..., x_N\}$
Output: $K = \{k_1, k_2, ..., k_N\}$
Generate initial values of $\{K_t\}$. The initial values can be fast obtained using simple methods,e.g. k-means [145] or GMM (see Sections 5.2). To avoid the local minima issue, multiple initial values can be generated ;
Given $\{K_t\}$, estimate $\{\boldsymbol{\mu}_t\}$ and $\{\boldsymbol{\sigma}_t^2\}$ using equations (5.51), (5.53), (5.52) and (5.54);
Given $\{K_t\}$, $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t^2$, calculate global negative logarithm posterior G^{old} using equation (5.60). Let $G^{\text{new}} = G^{\text{old}} - 0.001$ to enter the while loop;
while $G^{new} < G^{old}$ **do**
    Update G^{old} using G^{new};
    for $t \leftarrow 1$ **to** M **do**
        for $i \leftarrow 1$ **to** N **do**
            Given K_t^{old}, calculate $\boldsymbol{\mu}_t^{\text{old}}$ and $(\boldsymbol{\sigma}_t^2)^{\text{old}}$ using equations (5.51), (5.53), (5.52) and (5.54);
            Calculate old local negative logarithm posterior $L_{i,t}^{\text{old}}$ for pixel i image t using equation (5.58);
            Give a new $k_{i,t}^{\text{new}} = |k_{i,t}^{\text{old}} - 1|$;
            Calculate new $\boldsymbol{\mu}_t^{\text{new}}$ and $(\boldsymbol{\sigma}_t^2)^{\text{new}}$ using equations (5.51), (5.53), (5.52) and (5.54);
            Given new $k_{i,t}^{\text{new}}$, $\boldsymbol{\mu}_t^{\text{new}}$ and $(\boldsymbol{\sigma}_t^2)^{\text{new}}$, calculate new local negative logarithm posterior $L_{i,t}^{\text{new}}$ for pixel i image t using equation (5.58);
            If $L_{i,t}^{\text{new}} < L_{i,t}^{\text{old}}$, update old $k_{i,t}^{\text{old}}$ to new $k_{i,t}^{\text{new}}$.
        end
    end
    Given $\{K_t\}$, calculate $\{\boldsymbol{\mu}_t\}$ and $\{\boldsymbol{\sigma}_t^2\}$ using equations (5.51), (5.53), (5.52) and (5.54);
    Given $\{K_t\}$, $\{\boldsymbol{\mu}_t\}$ and $\{\boldsymbol{\sigma}_t^2\}$, calculate global negative logarithm posterior G^{new} using equation (5.60);
end
When the while loop stops, the values of $\{K_t\}$ is the estimated MAP for all images.

A problem of this spatio-temporal GMM-MRF method is that multiple classification maps are generated based on dynamic frames. However, the classification results should be consistent for all dynamic frames. Therefore, in this study, we use the average value of the classification maps to generate the final classification map, and use value 0.5 as the threshold of the hypo-perfusion and normal perfusion regions. Figure 5.18 shows the average values of the classification maps and the final classification map given $T = 1$ and $T_0 = 1$. The input images and their corresponding classification maps can be found in Figure 5.19 and 5.20.

Compared with the classification map in panel (b) Figure 5.12, the classification map in



(a) Average values

(b) Classification map

Figure 5.18: The average values of the classification maps and the final classification map. Panel (a) shows the average values of the multiple classification maps. Panel (b) shows the final classification map given threshold 0.5. The yellow pixels indicate the healthy tissues. The green pixels indicate the lesion tissues and the black pixels indicate the background (non-myocardium).

panel(b) Figure 5.18 does not change a lot. According to the visual inspection to panel (c), Figure 5.11, both spatial GMM-MRF and spatio-temporal GMM-MRF overestimate the hypo-perfusion region. On the other hand, the classification map based on the MBF map accurately predicts the position of the hypo-perfusion region. This is caused by the limitation of the original signals. For the spatial GMM-MRF method, the maximum enhancement image is used to classify the myocardial pixels. Similar to the CER map in Chapter 4, the maximum enhancement image cannot make use of the growth speed of the signals. It can only quantify the amplitude of the signal. Therefore, if the spatially correlated noise exists in the original signals, e.g. all pixels located on the right side have lower signals than the pixels located on the left side, the spatial GMM-MRF method will be invalid. Similarly, although the spatio-temporal GMM-MRF method uses temporal neighbourhood information, this method still cannot overcome the spatially correlated noise issue. Specifically, each classification map for the spatio-temporal GMM-MRF method is based on the image with spatially correlated noise, and thus all classification maps are affected by it. On the other hand, the MBF estimation methods standardise the image with the baseline and therefore reduces the spatially correlated noise.

5.5 Discussion and conclusion

In this chapter, three classification methods, i.e. the GMM, SVFMM and GMM-MRF, have been applied to the myocardial perfusion DCE-MRI original signals and parametric maps. These methods aim to detect the hypo-perfusion region using either original signals or the parametric maps developed using the methods illustrated in Chapter 4.

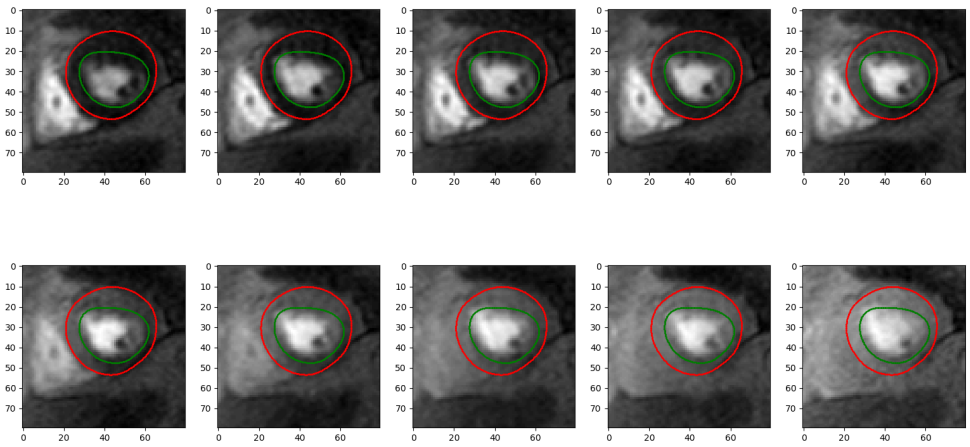


Figure 5.19: The input images for the spatio-temporal GMM-MRF methods.

The GMM classification method and its application to myocardial perfusion DCE-MRI are introduced in Section 5.2. The GMM method is fitted by the EM algorithm (see details in Section 2.4). Because the GMM classification method only takes the values of the signals or the parameters into account, some unrealistic single and small clusters appear in the classification map. This will happen more frequently when the SNR of the image is low (in practice, when the SNR is lower than 3, the SNR can be called low), or the parameter estimation method is inaccurate. Therefore, the neighbourhood information is introduced to reduce these unrealistic clusters using SVFMM and GMM-MRF methods.

In Section 5.3, the SVFMM classification method is introduced and applied to the original signals of the myocardial perfusion DCE-MRI. The SVFMM method is also fitted by the EM algorithm. It has been tested on both synthetic and clinical data. The number of single or small clusters pixels is used to quantify the performance of the SVFMM method for the clinical data. Moreover, the number of misclassified pixels is used to quantify the performance of the synthetic data. According to the quantification results, the SVFMM method dramatically decreases the number of single or small clusters. However, although large weight parameter for the neighbourhood information is given to the SVFMM, the single or small clusters still cannot be entirely removed. The reason is that the SVFMM indirectly introduces the neighbourhood information, i.e. to the probability of the label. Therefore, the GMM-MRF method is developed to introduce the neighbourhood information in a direct way, i.e. to the label.

In Section 5.4, the GMM-MRF method has been developed to detect the hypo-perfusion region using either original signals or the parametric map. The GMM-MRF method is fitted by the ICM algorithm, a greedy algorithm to search the MAP. Like the SVFMM method, the GMM-MRF method aims to eliminate the single and small clusters. For both cases, i.e. original

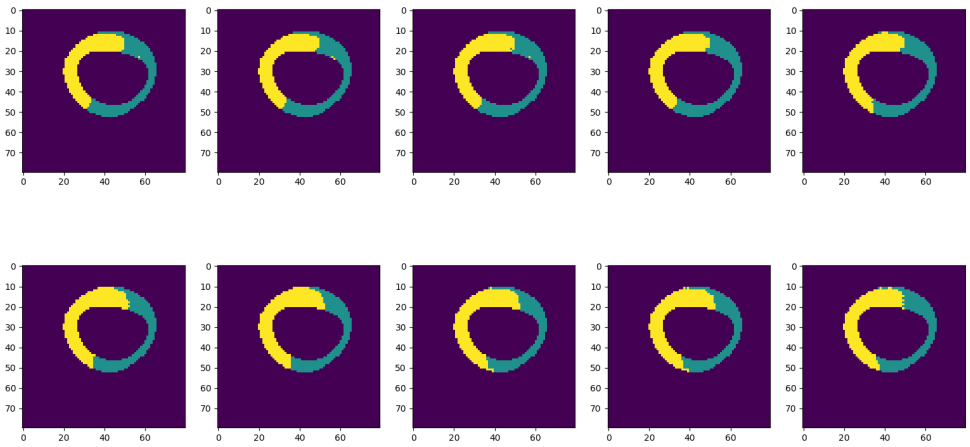


Figure 5.20: The corresponding classification maps for the input images in Figure 5.19.

signals and MBF map, the GMM-MRF method entirely removes all single and small clusters. Moreover, compared with the EM algorithm for the SVFMM, the ICM algorithm for the GMM-MRF is relatively faster. Generally, it takes about 1 minute to achieve convergence. On the other hand, it takes around 10 minutes to reach the convergence for the EM fitted SVFMM. However, for both spatial GMM-MRF and spatio-temporal GMM-MRF methods, their applications to the original signals overestimate the hypo-perfusion region. This is caused by the limitation of the original signals.

The limitation of the mixture models is that these methods are highly dependent on the quality of the data and the accuracy of the estimation methods. Precisely, if the input parametric maps have the quality like panel (b) in Figure 4.6 or Figure 4.7, neither classification method can accurately predict the position of the hypo-perfusion region. In other words, the detection of the hypo-perfusion region is separated into two phases. Firstly, the parametric map, e.g. the MBF map, is obtained using an independent method. Secondly, the classification map is obtained based on the MBF map. This is a single direct affection from the MBF map to the classification map. However, the classification map should also affect the MBF map because the healthy tissues are reasonable to have high MBF, and the lesion tissues reasonably have low MBF. Suppose a method can iteratively improve the MBF map using the classification map and improves the classification map using the MBF map, though the initial MBF map is not accurately estimated. In that case, it can be corrected finally because of the iterative improvements. This logic chain is also valid for the original signals because the classification map can reduce the noise of the original signal data. Therefore, in Chapter 6, a hierarchical Bayesian model is introduced to combine the MBF estimation method and the classification method. Moreover, in Chapter 7, a hierarchical Bayesian model is introduced to combine the

signal denoise (denoise from the original images) method and the classification method.

Chapter 6

Classification of Myocardial Blood Flow Based on DCE-MRI Using Hierarchical Bayesian Models

6.1 Introduction

In previous chapters, MBF estimations (see Chapter 4) and pixel classifications (see Chapter 5) have been introduced. Generally, the pixel classification approaches can be applied to the parametric maps, e.g. MBF maps, or the original images to generate classification maps. In other words, the classification maps are generated from the parametric maps, but the generation of parametric maps is free from the classification information. However, the effects of classification information to the parametric maps should not be neglected. Specifically, the lesion segments generally correspond to low MBF and the healthy segments correspond to high MBF. Therefore, a Fermi method (see Section 4.3.3) based hierarchical Bayesian model has been developed to integrate the process of parameter estimations and classifications in this chapter.

Hierarchical Bayesian modelling (HBM) has been extensively reviewed in Section 2.2.3. Hierarchical Bayesian models (HBMs) have been applied in a broad and diverse range of research and development areas, including healthcare medicine (see [146]), transportation networks (see [147]), clinical trials (see [148]), water resources (see [149]) and economics (see [150]). Moreover, HBMs have been applied widely in medical imaging, e.g. DCE-MRI. A HBM method has been developed to estimate kinetic parameters of DCE-MRI based on Gaussian Markov random fields priors in [103]. [151] introduced a HBM method to model the concentration time curve (CTC) based on a standard compartment model (see [152]) where a traditional nonlinear regression model was used for comparison. [153] illustrated an application of [151] to evaluate treatment response of squamous cell carcinoma. Of particular interest in the context of the present chapter are recent applications of HBMs to myocardial perfusion modelling. [97] modelled the non-linear attenuation of arterial input functions (AIF) and used HBMs to reconstruct AIFs

from magnetic resonance signal intensities in the myocardium. [98] extended the modelling in [97] by importing spatio-temporal constrains to the B-spline model using Gaussian Markov random fields. [105] illustrated an HBM method to analyze the myocardial perfusion DCE-MRI. In this method, a large-scale Bayesian spatio-temporal regression method was applied to model the data. Similarly, Gaussian Markov random fields priors were introduced to express the spatio-temporal constrains. [5] used HBMs to infer the parameters of tracer-kinetic models from contrast-enhanced magnetic resonance (MR) images. They further combined HBMs with Markov random fields to include prior knowledge on the physiological ranges of the model parameters and to integrate additional information on the physiological neighbourhood of a target voxel. The commonality of all these applications is that the models adopt hierarchical structures to incorporate potentially complex interactions and dependencies between their variables, and use state-of-the-art Markov chain Monte Carlo sampling techniques to address the analytical intractability of inference.

The work in this chapter aims to provide an automated identification of areas of hypo-perfused myocardium. To this end, we combine a myocardial perfusion model with a novel HBM and a Markov random fields (MRF) prior. The latter acts on discrete labels that we associate with the pixels of an MRI scan, to assign them to two distinct clusters: *lesion* versus *healthy* tissue. Loosely speaking, this clustering process can be viewed as a two-layered model, where the first (observed) layer represents the signal intensity (proportional to contrast agent concentration) and the second (hidden) layer contains the class labels (see [154]). A Markov random fields prior is used to provide context information, so that class assignment is not only based on a pixel's own signal intensity, but also takes spatial information from neighbouring pixels into account. As we will demonstrate, this is essential for noise reduction, good generalization and robust classification, i.e. obtaining classification patterns that are unaffected by random noise injections.

The content in this chapter has been published in Journal of the Royal Statistical Society: Series C at 12th May 2022 [155]:
***Yalei Yang**, Hao Gao, Colin Berry, David Carrick, Aleksandra Radjenovic and Dirk Husmeier. Classification of Myocardial Blood Flow Based on Dynamic Contrast Enhanced Magnetic Resonance Imaging Using Hierarchical Bayesian Models.*

6.2 Statistical method

6.2.1 Application of Markov random fields prior

The Markov random field priors has been applied in Chapter 5. Similarly, the Markov random field priors are also applied in this chapter to bring the neighbourhood information to decrease the unrealistic spurious singleton clusters.

Let $k_i \in \{0,1\}$ denote the label of pixel i. When $k_i = 0$, this pixel is marked as 'healthy tissue'. When $k_i = 1$, this pixel is marked as 'lesion'. The Markov random field of the label k_i is:

$$P(k_i|k_{-i}) = \frac{1}{Q}\exp(-U(k_i|k_{-i})) \tag{6.1}$$

where Q is a normalization constant to ensure that $P(k_i|k_{-i})$ is normalised and sums to 1, k_{-i} is the set of all other labels except k_i, $k_{-i} = \{k_j\}_{j\neq i}$, and

$$U(k_i|k_{-i}) = \frac{1}{T_{k_i}}\sum_{i\sim m} u(k_i|k_m), \tag{6.2}$$

in which T_{k_i} is a weight parameter, $i \sim m$ means pixel i and pixel m are neighbouring pixels and $u(k_i|k_m)$ is a local potential function, which we define to be

$$u(k_i|k_m) = \begin{cases} -\left(\frac{1}{2}\right)^{o-1} & k_i = k_m, \\ \left(\frac{1}{2}\right)^{o-1} & k_i \neq k_m \end{cases} \tag{6.3}$$

where 'o' indicates the neighbouring degree of order. In this work, we will use up to second order neighbourhoods, $o \in \{1,2\}$. The definition of spatial pixel cliques (pixel i and its neighbours) can be found in Figure 5.10. Increasing the neighbourhood of a pixel in the Markov random field to third order was found to lead to a substantial increase in the computational complexity, by about 100%. The restriction to second order is motivated by the fact that in nearly all our images, the thickness of the myocardium wall does not exceed 5 pixels. A third-order neighbourhood could potentially straddle the myocardium wall and combine regions on either side of it, which is an undesirable effect. The main justification for our restriction to second order, however, is our empirical finding that increasing the neighbourhood from second to third order was not found to make many difference in practice. Figure 6.1 shows a comparison between segmentations obtained with a 2nd-order and a 3rd-order neighbourhood, leading to very similar results visually.

For the Fermi parameters $\boldsymbol{\Theta}_i = \{A_i, \omega_i, \lambda_i\}$ (see Section 4.3.3), its Markov random fields prior $P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, k_i, k_{-i})$ is defined by

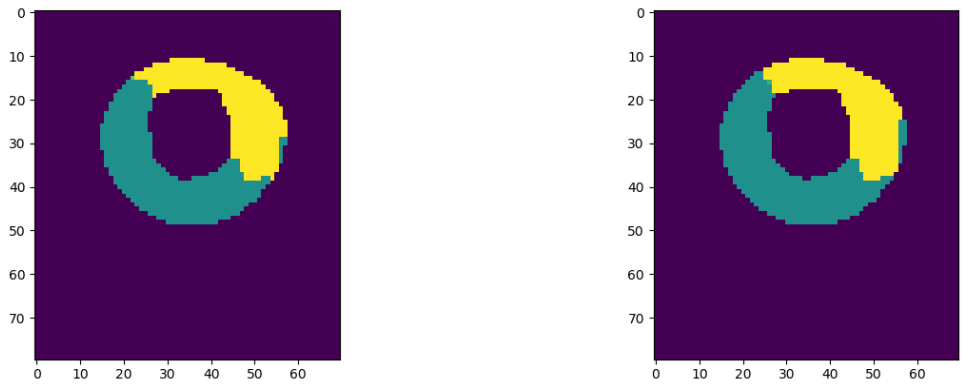
$$P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, k_i, k_{-i}) = \frac{1}{Q}\exp(-U(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, k_i, k_{-i})) \tag{6.4}$$

where

$$U(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, k_i, k_{-i}) = \frac{1}{T_{\text{Fermi}}}\sum_{i\sim m} v_{i,m}|F_b(\boldsymbol{\Theta}_i) - F_b(\boldsymbol{\Theta}_m)|. \tag{6.5}$$

and

$$F_b(\boldsymbol{\Theta}_i) = R_f(t=0) = \frac{A_i}{1+e^{-\omega_i/\lambda_i}}. \tag{6.6}$$



(a) 2 orders

(b) 3 orders

Figure 6.1: A comparison between the segmentations obtained with a 2nd-order (left panel) and a 3rd-order (right panel) neighbourhood for a pixel in the Markov random field prior.

Note that the dependence of the right-hand side of equation (6.5) on k_i and k_{-i} is via equation (6.7). In equation (6.5), T_{Fermi} is a weight parameter, and $i \sim m$ indicates that pixels i and m are neighbouring pixels, which are defined in Figure 5.10. The parameter family $\{v_{i,m}\}$ is used to guarantee edge-preserving properties, i.e. to prevent smoothing over edges that are indicative of tissue boundaries (e.g. delineating the area of a lesion); see [5] and [156] for details. When two adjacent pixels are in different categories, the parameters $\{v_{i,m}\}$ give the model the flexibility to switch off the interaction between the pixels and ensure they are not going to influence each other in the MBF. To be specific, in this work, $v_{i,m}$ is defined by

$$v_{i,m} = \begin{cases} 0 & k_i \neq k_m \\ 1 & k_i = k_m. \end{cases} \tag{6.7}$$

6.2.2 Hierarchical Bayesian model (HBM)

Figures 6.2 and 6.3 show the HBM model using DAGs. The descriptions of the parameters in Figures 6.2 and 6.3 can be found in Table 6.1. The structure of the HBM model reflects the three-layer causal relationships of the data-generating process. Specifically, the observed data is dependent on the contrast agent and its concentration, as described by the Fermi model (see Section 4.3.3), and the effect of the contrast agent is dependent on the status label. The rationality of the former dependence is from the central volume principle (see, e.g. [96]) and the Fermi model (see, e.g. [72]), and the rationality of the latter dependence follows from the physiological fact that the MBFs of healthy tissues and lesions are different (see, e.g. [157]). Moreover, both Fermi parameters and the status label are dependent on their neighbours using Markov random field priors (see Section 2.2.4) to introduce the spatial information. The details of the nodes in Figures 6.2 and 6.2 will be illustrated in Section 6.2.3.

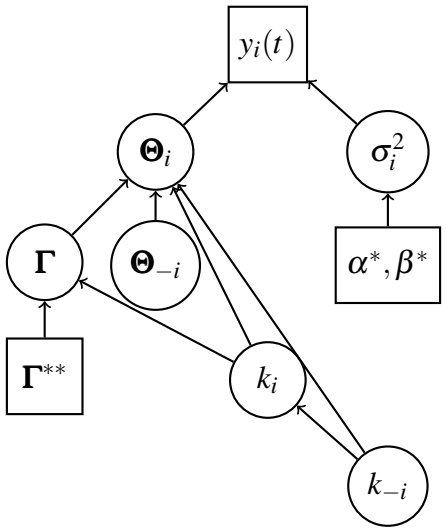


Figure 6.2: The structure of the hierarchical Bayesian model proposed in the work. We assume that the impulse response function is a Fermi function (see Section 4.3.3). The parameter groups are defined below equation (6.12). In this graph, the higher layers are conditionally dependent on the lower layers. The circle nodes denote variables and the rectangle nodes denote fixed values or observations. The explanations of parameters in this Figure can be found in Table 6.1.

Table 6.1: Overview of the HBM model parameters

Parameter	**Description**
$y_i(t)$	Observed contrast concentration for pixel i at time t
A_i, ω_i, λ_i	Fermi parameters for pixel i
σ_i^2	Variance of the iid Gaussian noise for pixel i
A_{-i}, ω_{-i}, λ_{-i}	Fermi parameters of the neighbours of pixel i
μ_{A_i}, μ_{ω_i}, μ_{λ_i}	Means of the prior distributions of the Fermi parameters for pixel i
$\sigma_{A_i}^2$, $\sigma_{\omega_i}^2$, $\sigma_{\lambda_i}^2$	Variances of the prior distributions of the Fermi parameters for pixel i
k_i	Label of pixel i
k_{-i}	Labels of the neighbours of pixel i
$*$ and $**$	The symbols with $*$ and $**$ are hyperparameters

6.2.3 Novel statistical model

Let $y_i(t)$ denote the observed contrast agent concentration of an image where i is a positive integer ($i = 1, 2, ..., N$) and t is a non-negative integer ($t = 0, 1, ..., M-1$). N is the total number of pixels in this image, M is the total number of time stamps of the data. For each pixel, k_i denotes the group of this pixel. In this case, we assume that there are two groups, healthy tissue and lesion. The proposed method can be easily to expand to more groups. However, two groups are more suitable for the data used in this study. Therefore, $k_i = 0$ when the i^{th} pixel is in the healthy group and $k_i = 1$ when the i^{th} pixel is in the lesion group. We assume the observed contrast concentration $y_i(t)$ is Gaussian distributed with mean $C(t|\boldsymbol{\Theta}_i)$ and variance

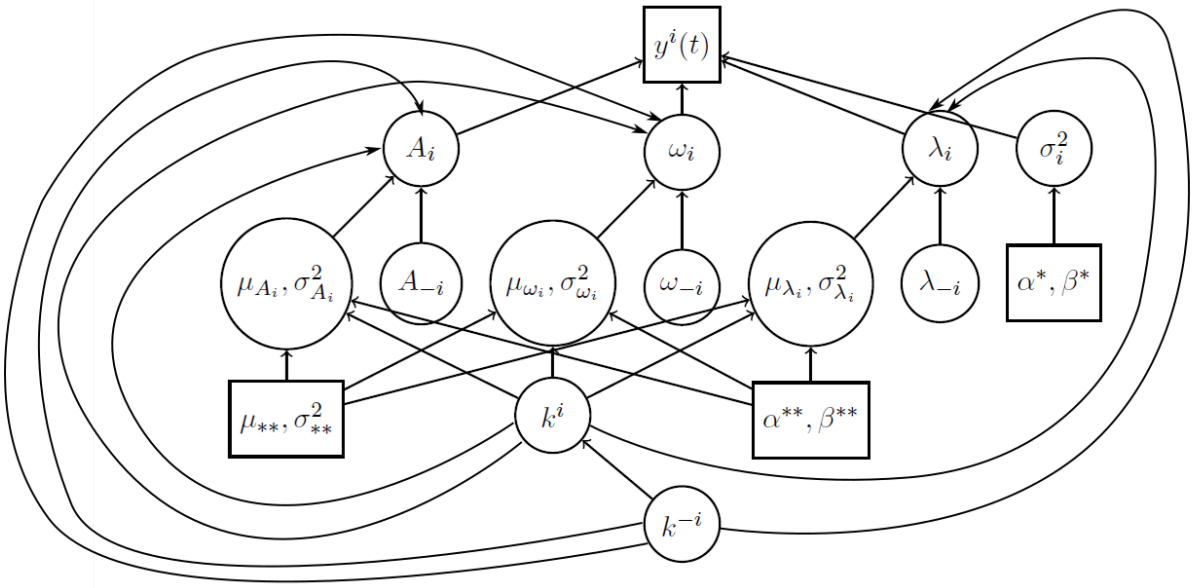


Figure 6.3: The hierarchical Bayesian model with all parameters shown. This figure is an extension of Figure 6.2 in which all parameters are shown individually. This inevitably leads to a more cluttered diagram, which has the advantage, though, that the conditional probabilities required for the Gibbs sampling scheme of Section 6.2.4 can directly be read off the graph based on its Markov blanket (see Section 2.2.3). The explanations of parameters in this Figure can be found in Table 6.1.

σ_i^2. Therefore, the expression of $y_i(t)$ is

$$y_i(t) = C(t|\mathbf{\Theta}_i) + \varepsilon_i \tag{6.8}$$

where $C(t|\mathbf{\Theta}_i)$ is the model calculated contrast agent concentration. Note that the notation $C(t|\mathbf{\Theta}_i)$ means the same as $C_{\text{myo}}(t)$ in equation (4.11), but we have made the dependence on the Fermi parameter vector $\mathbf{\Theta}_i$ explicit in the conditioning argument. $\varepsilon_i \sim N(0, \sigma_i^2)$ is assumed to be iid additive Gaussian noise (see [158] and [133]). We will test in our simulation studies how critical this distributional assumption is for the ultimate purpose of this study (automatic lesion detection). The parameters $\mathbf{\Theta}_i$ are the Fermi parameters defined in equation (4.15). According to the central volume principle (see [96]) in Section 4.3.3, the expression of contrast agent concentration $C(t|\mathbf{\Theta}_i)$ can be found in equation (4.11), which we rewrite to make the dependence on the Fermi parameters, defined in equation (4.15), explicit in our notation:

$$C(t|\mathbf{\Theta}_i) = R_f(t, \mathbf{\Theta}_i) * C_{\text{in}}(t) \tag{6.9}$$

where $R_f(t, \mathbf{\Theta}_i)$ can be found in equation (4.14). The new modified notation makes the dependence on the Fermi parameter vector $\mathbf{\Theta}_i$ explicit in the argument. The term $C_{\text{in}}(t)$ in equation (6.9) can be directly acquired from data. Specifically, the input contrast concentration $C_{\text{in}}(t)$ can

be extracted from the blood pool signal. Firstly, the blood pool in the images can be visually observed. Then, a 3×3 square is used to record the same blood pool area in all images. Finally, the average value of signal intensities in the square is calculated to be the blood pool signal and use equation 3.2 to transfer the signal values to the concentration values. We now make the dependence on the model parameters explicit in our notation. Since we assume the noise ε_i to be iid Gaussian distributed, we have

$$P(y_i(t)|\mathbf{\Theta}_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{-(y_i(t) - C(t|\mathbf{\Theta}_i))^2}{2\sigma_i^2}\right). \tag{6.10}$$

We assume that the parameters $\mathbf{\Theta}_i$ are independent of time t. The quantities $\mathbf{\Theta}_i$ are related to the patient's physiology and pathophysiology, which will change with time (e.g. as a consequence of treatment). However, these changes are slow, typically happening in the order of weeks or months. Our MRI time series, on the other hand, are acquired within a minute, that is, several orders of magnitude below. It is therefore justified to assume that for the duration of the measurement modelled in our study, the parameters $\mathbf{\Theta}_i$ can be taken as constant. For $\boldsymbol{y}_i = \{y_i(t=0), y_i(t=1), ..., y_i(t=M-1)\}$, the conditional density function $P(\boldsymbol{y}_i|\mathbf{\Theta}_i, \sigma_i^2)$ is

$$\begin{aligned} P(\mathbf{y}_i|\mathbf{\Theta}_i, \sigma_i^2) &= \prod_{t=0}^{M-1} P(y_i(t)|\mathbf{\Theta}_i, \sigma_i^2) \\ &\propto \exp\left(\sum_{t=0}^{M-1} \frac{-(y_i(t) - C(t|\mathbf{\Theta}_i))^2}{2\sigma_i^2}\right) \sigma_i^{-M}. \end{aligned} \tag{6.11}$$

As we have discussed in Section 2.2.3, according to the factorization rule, the joint distributions of a DAG is the product of conditional distributions determined by the respective parent nodes. Given the structure in Figure 6.2, the joint distribution is

$$\begin{aligned} P(\boldsymbol{y}_i, \mathbf{\Theta}_i, \sigma_i^2, \mathbf{\Gamma}, \mathbf{\Theta}_{-i}, \alpha^*, \beta^*, \mathbf{\Gamma}^{**}, k_i, k_{-i}) =& P(\boldsymbol{y}_i|\mathbf{\Theta}_i, \sigma_i^2) P(\mathbf{\Theta}_i|\mathbf{\Theta}_{-i}, \mathbf{\Gamma}, k_i, k_{-i}) \\ & P(\sigma_i^2|\alpha^*, \beta^*) P(\mathbf{\Gamma}|\mathbf{\Gamma}^{**}, k_i) \\ & P(k_i|k_{-i}) P(\mathbf{\Theta}_{-i}) P(k_{-i}) \end{aligned} \tag{6.12}$$

where $\mathbf{\Theta}_i = \{A_i, \omega_i, \lambda_i\}$, $\mathbf{\Theta}_{-i} = \{A_{-i}, \omega_{-i}, \lambda_{-i}\}$, $\mathbf{\Gamma} = \{\Gamma_A, \Gamma_\omega, \Gamma_\lambda\}$ and $\Gamma_A = \{\mu_{A,k_i}, \sigma^2_{A,k_i}\}$, $\Gamma_\omega = \{\mu_{\omega,k_i}, \sigma^2_{\omega,k_i}\}$, $\Gamma_\lambda = \{\mu_{\lambda,k_i}, \sigma^2_{\lambda,k_i}\}$, $\mathbf{\Gamma}^{**} = \{\mu_{**}, \sigma^2_{**}, \alpha^{**}, \beta^{**}\}$. The joint probability distribution in equation (6.12) has combined two classes of probabilistic graphical models: directed probabilistic graphical models, also called Bayesian networks (see Section 2.2.3), and undirected graphical models, also called Markov random fields (see Section 2.2.4). Both modelling paradigms can be integrated into the generalised concept of a partially directed acyclic graphical model, also called a chain event graph or just a chain graph; see e.g. [159] or the recent monograph by [160]. This combination of a Bayesian network with a Markov random field

is the natural approach for the present problem. The directed acyclic graph (DAG) structure of the Bayesian network represents the causal relationships of the data-generating process, as explained in the previous section. Spatial neighbourhood relations, on the other hand, lack a natural DAG structure, and an undirected graph, i.e. a Markov random field is a more natural representation. Our model can thus be interpreted as a particular instance of a chain graph, one in which the directed and undirected graph structures are clearly separated, and this separation simplifies the inference compared to the more general inference schemes discussed in [160].

We now explain each term on the right-hand side of equation (6.12). The likelihood $P(\boldsymbol{y}_i|\boldsymbol{\Theta}_i,\sigma_i^2)$ can be found in equation (6.11). $P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i},\boldsymbol{\Gamma},k_i,k_{-i})$ is the prior distribution of the Fermi parameter vector $\boldsymbol{\Theta}_i$, which combines explicit priors for the three parameter groups with the Markov random field prior of equation (6.4):

$$\begin{aligned}P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i},\boldsymbol{\Gamma},k_i,k_{-i}) \propto & N(\log(A_i)|\mu_{A,k_i},\sigma^2_{A,k_i})N(\log(\omega_i)|\mu_{\omega,k_i},\sigma^2_{\omega,k_i}) \\ & N(\log(\lambda_i)|\mu_{\lambda,k_i},\sigma^2_{\lambda,k_i})P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i},k_i,k_{-i}).\end{aligned} \tag{6.13}$$

$N(\log(A_i)|\mu_{A,k_i},\sigma^2_{A,k_i})$, $N(\log(\omega_i)|\mu_{\omega,k_i},\sigma^2_{\omega,k_i})$, $N(\log(\lambda_i)|\mu_{\lambda,k_i},\sigma^2_{\lambda,k_i})$ are Gaussian prior distributions. Pathologically, the values of MBF for healthy tissues and lesions are different. Since all Fermi parameters are non-negative, we assume the logarithms of the Fermi parameters are Gaussian distributed conditional on the labels k_i. We have also tried Gamma distributions as an alternative, and include the results in Appendix B.1. The difference in the results obtained with a log-Gaussian versus a Gamma prior was found to be minor, which suggests that the choice of functional family for the prior distributions on the Fermi parameters $\boldsymbol{\Theta}_i$ is not critical, as long as the distributions are consistent with the positivity constraint of the Fermi parameters, i.e. have positive support. For $N(\log(A_i)|\mu_{A,k_i},\sigma^2_{A,k_i})$, we have:

$$N(\log(A_i)|\mu_{A,k_i},\sigma^2_{A,k_i}) = \frac{1}{\sqrt{2\pi}\sigma_{A,k_i}}\exp\left(\frac{-(\log(A_i)-\mu_{A,k_i})^2}{2\sigma^2_{A,k_i}}\right). \tag{6.14}$$

with $k_i \in \{0,1\}$. For ω_i and λ_i, the definitions are similar. $P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i},k_i,k_{-i})$ is the Markov random field prior, defined in equations (6.4) and (6.5). We choose a vague inverse gamma distribution for the observational noise variance σ_i^2,

$$P(\sigma_i^2|\alpha^*,\beta^*) \;=\; \mathrm{IG}(\sigma_i^2|\alpha^*=0.1,\beta^*=0.1), \tag{6.15}$$

due to its conjugacy. $P(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{**},k_i)$ is the prior distribution for hyperparameter vector $\boldsymbol{\Gamma}$. It can be written as

$$P(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{**},k_i) \propto P(\mu_{A,k_i},\mu_{\omega,k_i},\mu_{\lambda,k_i}|\mu_{**},\sigma^2_{**})P(\sigma^2_{A,k_i},\sigma^2_{\omega,k_i},\sigma^2_{\lambda,k_i}|\alpha^{**},\beta^{**}) \tag{6.16}$$

where

$$P(\mu_{A,k_i=\psi},\mu_{\omega,k_i=\psi},\mu_{\lambda,k_i=\psi}|\mu_{**},\sigma^2_{**}) \propto N(\mu_{A,k_i=\psi}|\mu_{**},\sigma^2_{**})N(\mu_{\omega,k_i=\psi}|\mu_{**},\sigma^2_{**}) \\ N(\mu_{\lambda,k_i=\psi}|\mu_{**},\sigma^2_{**}) \tag{6.17}$$

and

$$P(\sigma^2_{A,k_i=\psi},\sigma^2_{\omega,k_i=\psi},\sigma^2_{\lambda,k_i=\psi}|\alpha^{**},\beta^{**}) \propto \mathrm{IG}(\sigma^2_{A,k_i=\psi}|\alpha^{**},\beta^{**})\mathrm{IG}(\sigma^2_{\omega,k_i=\psi}|\alpha^{**},\beta^{**}) \\ \mathrm{IG}(\sigma^2_{\lambda,k_i=\psi}|\alpha^{**},\beta^{**}) \tag{6.18}$$

with $\psi \in \{0,1\}$. $\{\mu_{**}=0,\sigma^2_{**}=10,\alpha^{**}=0.1,\beta^{**}=0.1\}$ are uninformative hyperparameters. $P(k_i|k_{-i})$, another term in equation (6.12), is the Markov random field prior of label k_i, which can be found in equations (6.1–6.3). $P(\mathbf{\Theta}_{-i})$ and $P(k_{-i})$ in equation (6.12) are the marginal distributions of $\mathbf{\Theta}_{-i}$ and k_{-i}. While these expressions are intractable, they do not enter the Gibbs sampling scheme for posterior inference, to be discussed in the Section 6.2.4.

6.2.4 Posterior inference

We now derive a Gibbs sampling scheme for posterior inference. In a DAG (the HBM model in Figure 6.3), the probability of a selected target parameter conditional on all other parameters is given by the probability of the target parameter conditional on its Markov blanket (see 2.2.3). Therefore, the relevant conditional probabilities can immediately be read off the graph in Figure 6.3 in Section 6.2.2.

For the parameter vector $\mathbf{\Theta}_i = \{A_i,\omega_i,\lambda_i\}$, according to Figure 6.2, its parents nodes are $\mathbf{\Theta}_{-i}$, $\mathbf{\Gamma}$, k_i and k_{-i}. Its child node is $y_i(t)$ and its co-parent node is σ_i^2. Therefore, its conditional posterior distribution is:

$$P(\mathbf{\Theta}_i|\mathbf{y}_i,\sigma_i^2,\mathbf{\Gamma},\mathbf{\Theta}_{-i},k_i,k_{-i}) \propto P(\mathbf{y}_i|\mathbf{\Theta}_i,\sigma_i^2)P(\mathbf{\Theta}_i|\mathbf{\Theta}_{-i},\mathbf{\Gamma},k_i,k_{-i}) \tag{6.19}$$

where $P(\mathbf{y}_i|\mathbf{\Theta}_i,\sigma_i^2)$ can be found in equation (6.11) and $P(\mathbf{\Theta}_i|\mathbf{\Theta}_{-i},\mathbf{\Gamma},k_i,k_{-i})$ can be found in equation (6.13). Note that $P(\mathbf{y}_i|\mathbf{\Theta}_i,\sigma_i^2)$ depends on $C(t|\mathbf{\Theta}_i)$; see equation (6.11). $C(t|\mathbf{\Theta}_i)$ is defined by a convolution integral – equation (6.9) – which is analytically intractable. Hence, the conditional posterior distribution of $\mathbf{\Theta}_i$ in equation (6.19) is analytically intractable.

According to Figure 6.2, the parent, child and co-parent nodes of σ_i^2 are α^*,β^*, $y_i(t)$ and $\mathbf{\Theta}_i$. The conditional distribution of the observational noise variance σ_i^2,

$$P(\sigma_i^2|\mathbf{y}_i,\mathbf{\Theta}_i,\alpha^*,\beta^*) \propto P(\mathbf{y}_i|\mathbf{\Theta}_i,\sigma_i^2)P(\sigma_i^2|\alpha^*,\beta^*) \tag{6.20}$$

is obtained by combining the likelihood for pixel i,

$$P(\boldsymbol{y}_i|\boldsymbol{\Theta}_i,\sigma_i^2) \propto \sigma_i^{-M} \exp\left(\frac{-\sum_{t=0}^{M-1}(y_i(t)-C(t|\boldsymbol{\Theta}_i))^2}{2\sigma_i^2}\right) \tag{6.21}$$

with the conjugate prior for $P(\sigma_i^2|\alpha^*,\beta^*)$ from equation (6.15) to give

$$P(\sigma_i^2|\boldsymbol{y}_i,\boldsymbol{\Theta}_i,\alpha^*,\beta^*) \;=\; \mathrm{IG}\left(\frac{M}{2}+\alpha^*, \frac{\sum_{t=0}^{M-1}(y_i(t)-C(t|\boldsymbol{\Theta}_i)^2)}{2}+\beta^*\right) \tag{6.22}$$

where $\alpha^* = \beta^* = 0.1$, from Section 6.2.3.

According to Figure 6.2, the parent, children and co-parent nodes of k_i are k_{-i}, $\boldsymbol{\Theta}_i$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}^{**}$ and $\boldsymbol{\Theta}_{-i}$. Its conditional posterior distribution is given by

$$P(k_i|\boldsymbol{\Theta}_i,\boldsymbol{\Theta}_{-i},\boldsymbol{\Gamma},\boldsymbol{\Gamma}^{**},k_{-i}) \propto P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i},\boldsymbol{\Gamma},k_i,k_{-i})P(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{**},k_i)P(k_i|k_{-i}). \tag{6.23}$$

$P(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i},\boldsymbol{\Gamma},k_i,k_{-i})$ can be found in equation (6.13); $P(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{**},k_i)$ can be found in equation (6.16), and $P(k_i|k_{-i})$ is the Markov random field prior for binary parameters, and its expression can be found in equations (6.1–6.3). The conditional posterior distribution of k_i is a Bernoulli distribution because $k_i \in \{0,1\}$. Specifically, when $k_i = 0$, we have

$$\begin{aligned}
q \propto & P(k_i=0|\boldsymbol{\Theta}_i,\boldsymbol{\Theta}_{-i},\boldsymbol{\Gamma},\boldsymbol{\Gamma}^{**},k_{-i}) \\
\propto & N(\log(A_i)|\mu_{A,k_i=0},\sigma^2_{A,k_i=0})N(\log(\omega_i)|\mu_{\omega,k_i=0},\sigma^2_{\omega,k_i=0})N(\log(\lambda_i)|\mu_{\lambda,k_i=0},\sigma^2_{\lambda,k_i=0}) \\
& \exp\left(-\frac{1}{T_{\mathrm{Fermi}}}\sum_{i\sim m}\nu_{i,m}|F_b(\boldsymbol{\Theta}_i)-F_b(\boldsymbol{\Theta}_m)|\right) N(\mu_{A,k_i=0}|\mu_{**},\sigma^2_{**}) \\
& N(\mu_{\omega,k_i=0}|\mu_{**},\sigma^2_{**})N(\mu_{\lambda,k_i=0}|\mu_{**},\sigma^2_{**})\mathrm{IG}(\sigma^2_{A,k_i=0}|\alpha^{**},\beta^{**})\mathrm{IG}(\sigma^2_{\omega,k_i=0}|\alpha^{**},\beta^{**}) \\
& \mathrm{IG}(\sigma^2_{\lambda,k_i=0}|\alpha^{**},\beta^{**})\exp\left(-\frac{1}{T_{k_i}}\sum_{i\sim m}u(k_i|k^m)\right).
\end{aligned} \tag{6.24}$$

Note that the term $\nu_{i,m}$ depends on the label k_i given equation (6.7). Similarly, when $k_i = 1$, we

have

$$
\begin{aligned}
p \propto & P(k_i = 1|\boldsymbol{\Theta}_i, \boldsymbol{\Theta}_{-i}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{**}, k_{-i}) \\
\propto & N(\log(A_i)|\mu_{A,k_i=1}, \sigma^2_{A,k_i=1}) N(\log(\omega_i)|\mu_{\omega,k_i=1}, \sigma^2_{\omega,k_i=1}) N(\log(\lambda_i)|\mu_{\lambda,k_i=1}, \sigma^2_{\lambda,k_i=1}) \\
& \exp\left(-\frac{1}{T_{\text{Fermi}}} \sum_{i\sim m} \nu_{i,m} |F_b(\boldsymbol{\Theta}_i) - F_b(\boldsymbol{\Theta}_m)|\right) N(\mu_{A,k_i=1}|\mu_{**}, \sigma^2_{**}) \\
& N(\mu_{\omega,k_i=1}|\mu_{**}, \sigma^2_{**}) N(\mu_{\lambda,k_i=1}|\mu_{**}, \sigma^2_{**}) \text{IG}(\sigma^2_{A,k_i=1}|\alpha^{**}, \beta^{**}) \text{IG}(\sigma^2_{\omega,k_i=1}|\alpha^{**}, \beta^{**}) \\
& \text{IG}(\sigma^2_{\lambda,k_i=1}|\alpha^{**}, \beta^{**}) \exp\left(-\frac{1}{T_{k_i}} \sum_{i\sim m} u(k_i|k^m)\right).
\end{aligned} \tag{6.25}
$$

In this way, the conditional posterior distribution for label k_i is

$$
P(k_i|\boldsymbol{\Theta}_i, \boldsymbol{\Theta}_{-i}, \boldsymbol{\Gamma}, \boldsymbol{\Gamma}^{**}, k_{-i}) = \left(\frac{p}{p+q}\right)^{k_i} \left(\frac{q}{p+q}\right)^{1-k_i}. \tag{6.26}
$$

According to Figure 6.3, the parent nodes, child node and co-parents nodes of μ_{A,k_i} and σ^2_{A,k_i} are $\{\mu_{**}, \sigma^2_{**}\}$, $\{\alpha^{**},\beta^{**}\}$, k_i, A_i, A_{-i} and k_{-i}. Therefore, the conditional posterior distributions are given by

$$
\begin{aligned}
& P(\mu_{A,k_i=\psi}|\{A_i, A_{-i}\}_{i|k_i=\psi}, \mu_{**}, \sigma^2_{**}, \sigma^2_{A,k_i=\psi}, k_{-i}) \\
\propto \; & N(\mu_{A,k_i=\psi}|\mu_{**}, \sigma^2_{**}) \prod_{i|k_i=\psi} N(\log(A_i)|\mu_{A,k_i=\psi}, \sigma^2_{A,k_i=\psi})
\end{aligned} \tag{6.27}
$$

and

$$
\begin{aligned}
& P(\sigma^2_{A,k_i=\psi}|\{A_i, A_{-i}\}_{i|k_i=\psi}, \alpha^{**}, \beta^{**}, \mu_{A,k_i=\psi}, k_{-i}) \\
\propto \; & \text{IG}(\sigma^2_{A,k_i=\psi}|\alpha^{**}, \beta^{**}) \prod_{i|k_i=\psi} N(\log(A_i)|\mu_{A,k_i=\psi}, \sigma^2_{A,k_i=\psi})
\end{aligned} \tag{6.28}
$$

where $\psi \in \{0, 1\}$ is the binary lesion indicator. As seen from equation (6.13), the term that includes the dependence on the neighbourhood, A_{-i} and k_{-i}, does not depend on μ_{A,k_i} and σ^2_{A,k_i}. Therefore the terms A_{-i} and k_{-i} drop out. By simplifying equation (6.27), the conditional posterior distribution for $\mu_{A,k_i=\psi}$ is

$$
P(\mu_{A,k_i=\psi}|\{A_i\}_{i|k_i=\psi}, \sigma^2_{A,k_i=\psi}, \mu_{**}, \sigma^2_{**}) = N\left(\frac{C}{B}, \frac{D}{B}\right). \tag{6.29}
$$

where $B = \sigma^2_{A,k_i=\psi} + \sum_{i|k_i=\psi} \sigma^2_{**}$, $C = \sigma^2_{A,k_i=\psi}\mu_{**} + \sigma^2_{**} \sum_{i|k_i=\psi} \log(A_i)$ and $D = \sigma^2_{**}\sigma^2_{A,k_i=\psi}$. By

simplifying equation (6.28), the conditional posterior distribution for $\sigma^2_{A,k_i=\psi}$ is

$$\begin{aligned} & P(\sigma^2_{A,k_i=\psi}|\{A_i\}_{i|k_i=\psi},\alpha^{**},\beta^{**},\mu_{A,k_i=\psi}) \\ \propto \quad & \text{IG}\left(\frac{\sum_{i|k_i=\psi}1}{2}+\alpha^{**},\frac{\sum_{i|i=\psi}(\log(A_i)-\mu_{A,k_i=\psi})^2}{2}+\beta^{**}\right). \end{aligned} \tag{6.30}$$

The conditional posterior distributions for $\{\mu_{\omega,k_i=\psi},\sigma^2_{\omega,k_i=\psi},\mu_{\lambda,k_i=\psi},\sigma^2_{\lambda,k_i=\psi}\}$ can be derived similarly.

In the Bayesian statistics framework, samples from the conditional posterior distributions of $\boldsymbol{\Theta}_i$, σ_i^2, k_i, and $\boldsymbol{\Gamma}$ can be approximately obtained with MCMC methods. In this work, the conditional posterior distribution of $\boldsymbol{\Theta}_i$ is analytically intractable because of the convolution form in equation (6.9). Therefore, we adopt a Metropolis-Hastings-within-Gibbs sampling scheme to sample the parameters. To be specific, $\boldsymbol{\Theta}_i$ is sampled with the Metropolis-Hastings algorithm based on adapted Gaussian proposals distributions. The acceptance rate is controlled within the range of $(0.23, 0.44)$, which has been suggested from [12]. All other parameters will be sampled directly from their conditional posterior distributions by Gibbs sampling. The details of our sampling method can be found in algorithm 5.

Algorithm 5: Metropolis-Hastings within Gibbs sampling for MBF classification

Input: Hyperparameters α^*, β^*, $\boldsymbol{\Gamma}^{**}$
Output: Fermi parameters $\boldsymbol{\Theta}_i$ and label k_i
Data: $\boldsymbol{y}_i = \{y_i(t)\}$ where $i = 1,2,...,N \quad t = 0,1,...,M-1$
Draw initial values of $\boldsymbol{\Gamma}$ from their prior distributions in equation (6.16);
for $i \leftarrow 1$ **to** N **do**
  Obtain initial values of $\boldsymbol{\Theta}_i$ using 'Least-Squares Fitting' based on equations (4.11), (4.14) and (4.16);
end
for $i \leftarrow 1$ **to** N **do**
  Obtain initial values of k_i, using EM algorithm (Gaussian mixture model) (see Section 5.2);
end
for $p \leftarrow 1$ **to** M_0 *(The number of MCMC samples)* **do**
  for $i \leftarrow 1$ **to** N **do**
    Given α^*, β^*, $\boldsymbol{\Theta}_i$, and $\boldsymbol{y}_i$ draw σ_i^2 from the inverse gamma distribution based on equation (6.22);
    Given σ_i^2, $\boldsymbol{\Gamma}$, k_i, k_{-i}, $\boldsymbol{\Theta}_{-i}$ and $\boldsymbol{y}_i(t)$, draw $\boldsymbol{\Theta}_i$ using random walk Metropolis-Hastings based on equation (6.19);
    Given $\boldsymbol{\Theta}_i$, $\boldsymbol{\Theta}_{-i}$, $\boldsymbol{\Gamma}^*$ and k_{-i}, draw k_i from Bernoulli distribution (the parameter k_i is either 0 or 1) based on equation (6.26);
  end
  Given $\{k_i\}$, $\{\boldsymbol{\Theta}_i\}$, $\boldsymbol{\Gamma}^{**}$, draw $\boldsymbol{\Gamma}$ based on equations (6.29) and (6.30);
end

6.3 Data and results

6.3.1 Clinical data

The myocardial perfusion DCE-MRI data (short-axis mid-cavity slice) from four patients (2 male, age range 56-60) were analyzed in this work. They are chosen randomly to show some example results for the proposed method. In the following analysis, we choose one patient to illustrate our method as an example. (56 year old male, scan acquired 2.7 days after percutaneous coronary intervention, with left anterior descending artery myocardial infarction). The results obtained in the remaining three datasets are presented in Appendix C.1. The performances of the results shown in the Appendix is quite similar to the results shown in the main pages. The conclusion obtained from the data shown in the main pages can be extended to the data shown in the Appendix. The general data information can be found in Section 3.3.

6.3.2 Synthetic data

Since it is challenging to obtain the ground truth of the size and location of the lesion in clinical data, we also use synthetic data to test our method. Two assumptions are made. Firstly, the noise for data generation is assumed to be either Gaussian or Rician noise. Due to the Central Limit Theorem, Gaussian noise is the most general and widely applied form of noise, while Rician noise is often assumed for MRI data due to the underlying physical characteristics (see e.g. [161]). Note that by generating data subject to Rician noise, we test the robustness of our method (which assumes Gaussian noise, as discussed in Section 6.2.3) with respect to model mis-specification. Secondly, the change of signal intensity for each pixel is a double exponential function, as explained below; see Equation (6.31) for the temporal variation of indicator concentration in the peripheral compartment (see [162]), which mimics the change of signal intensity of each pixel.

The synthetic data are designed to emulate the clinical data we have used in this work: 46 images (the same number as for the clinical data), i.e. the number of frames in one subject's mid-cavity slice, are simulated to mimic the acquired DCE-MRI data. The first 12 images are pre-contrast images, and the remaining 34 images are contrast enhanced images. We discard the first 12 images. The first image of the remaining 34 images corresponds to time point $t = 0$. We assume that there is one lesion, which is located in the anteroseptal region of the myocardium. The other pixels in the myocardium mimic healthy tissue with normal blood perfusion. For the pre-contrast image, the signal intensities for all pixels are drawn from a Gaussian distribution: $N(10, 1)$. These values have been chosen to ensure that, at this stage, all signal intensities are designed to be low. For the contrast enhanced images, the modelling of the signal intensities is more complex. We use a double exponential curve to mimic the change of signal intensity with

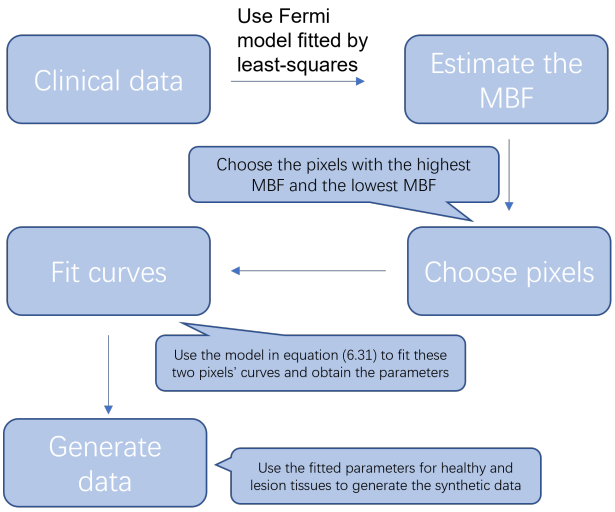


Figure 6.4: The process of generating realistic synthetic data based on real clinical data. See Section 6.3.2 for more detailed explanations.

time, which is given by the following equation:

$$s(t) = \frac{p_2 p_3}{(p_2 - p_1)} \times (e^{-p_1 t} - e^{-p_2 t}). \tag{6.31}$$

We use different parameters for healthy tissue and lesions. For lesion pixels, the parameters are set to $p_1 = 0.02$, $p_2 = 0.3$, $p_3 = 20$, while for healthy tissue pixels, the parameters are set to $p_1 = 0.01$, $p_2 = 0.4$, $p_3 = 25$. These values are chosen empirically to mimic our clinical data. The process to obtain these values can be found in Figure 6.4. For Gaussian noise, the signal intensity in the contrast enhancement stage is $s(t) + N(10, \sigma_*^2)$. The variance of the noise σ_*^2 is chosen to be 2 or 2.5, corresponding to SNRs of 4.1 and 3.72, respectively. For Rician noise, we use the Rayleigh distribution to generate the noise; see [161] for details. The value σ_*^2 is chosen as 1.8 with an SNR of 2.91. The choices of SNR here aim to emphasise the challenge of the work, and therefore high SNR examples are not simulated. Again, all these values have been selected so as to mimic our clinical data.

6.3.3 Alternative models for comparison

We have compared the performance of the HBM model proposed in this chapter with two alternative models for estimating the MBF. The first alternative model predicts signal intensities independently according to the Fermi model, as defined in equation (4.14), without taking any neighbourhood information of the image pixels into account. The parameters are estimated by minimizing the residual sum-of-squares score between measured and predicted signals, defined in equation (4.16). We refer to this model as the *Fermi model*. The second alternative model is a simplified HBM model where no explicit discrete labels for lesion indication are included. This model is similar to the model proposed in [5]. The only difference is that we have replaced

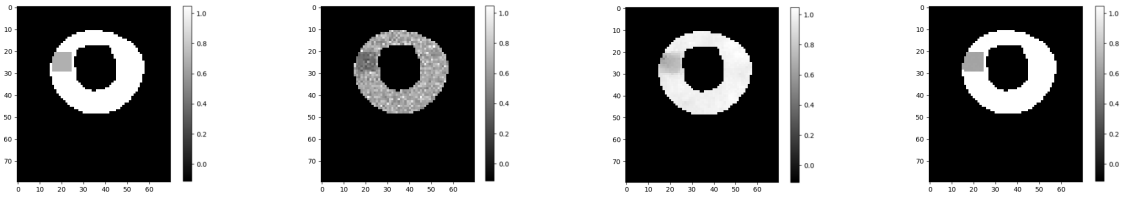
the authors' MBF model with the Fermi model of equation (4.14), for the reasons that our clinical datasets are limited to first-pass of the contrast agent through the vasculature, with very restricted coverage of the extravasation. This is why we chose to implement the Fermi model, as opposed to compartmental modelling approaches (see [80]). We emphasize, though, that the method we propose can work, in principle, with any model of $R_f(t)$, and the choice of the Fermi model is not critical to our approach. We refer to this model as the *HBM model without labels*. The model parameters are sampled from the posterior distribution with MCMC, using the same sampling scheme as described in [5].

We have also included a comparison with two alternative classification methods, for assigning pixels to the classes 'healthy tissue' versus 'lesion'. The first alternative scheme is a standard Gaussian mixture model, whose parameters are estimated in a maximum likelihood sense using the Expectation Maximization (EM) algorithm (see Section 5.2). We refer to the combined scheme as the *GMM based on Fermi* method. The second alternative classification method is an extension of the method just described with a morphological noise removal correction called the "opening and closing operation"; see ([163]) for details. We refer to this method as *GMM + O&C*.

6.3.4 Model selection and convergence diagnostics

The parameters of our model are sampled from the posterior distribution using the sampling algorithm described in Section 6.2.4. The exceptions are hyperparameters T_{k_i} and T_{Fermi} defined in equation (6.2) and (6.5). These hyperparameters define the strength of the neighbourhood influence in the Markov random field. They could in principle be sampled from the posterior distribution using MCMC methods, along with the other parameters. However, this would cause a considerable increase in the computational complexity, because convergence and mixing of lower-level hyperparameters tends to be happening at much lower rates than for higher-level parameters; see e.g. [164] for more details. For clinical decision making, excessive computational costs need to be avoided. We have therefore chosen 3 discrete values for these parameters, with equidistant intervals on a log_{10} scale. These values were obtained from pilot studies where the influence of the neighbourhood information in the Markov random field was explored. We then treat the selection of these parameters as a model selection problem, using the Watanabe information criterion (WAIC) introduced in [41] (see also in Section 2.6.3). This score, which can be directly computed from the MCMC trajectories, is asymptotically equivalent to Bayesian leave-one-out cross-validation, with lower values indicating models that are preferred by the data. The advantage is that we can now run MCMC simulations for different values of T_{k_i} and T_{Fermi} in parallel, and then select the parameters with the lowest WAIC scores. This leads to substantial computational savings compared to directly sampling these parameters from the posterior distribution with MCMC.

To assess the convergence of the MCMC simulations described in Section 6.2.4, we com-



(a) Ground truth (b) Fermi model (c) HBM without label (d) Proposed HBM

Figure 6.5: MBF estimations for synthetic data. These figures show four examples of the estimated MBF (rescaled within [0,1]) for the synthetic data with additive i.i.d. Gaussian noise of standard deviation 2.5 . The pixels inside and outside of the ring (myocardium) are background. Panel (a) shows the ground truth we designed. Panel (b) shows the MBF estimations using Fermi model fitted by least-squares. Panel (c) shows the MBF estimations using a method akin to [5]. Panel (d) shows the MBF estimations using the method proposed in the present work. For further details on the alternative methods used here, see Section 6.3.3.

puted the Geweke statistic (see e.g. [24]) directly from the MCMC trajectories, taking Z-scores in the range of $(-1,1)$ as absence of evidence for insufficient convergence.

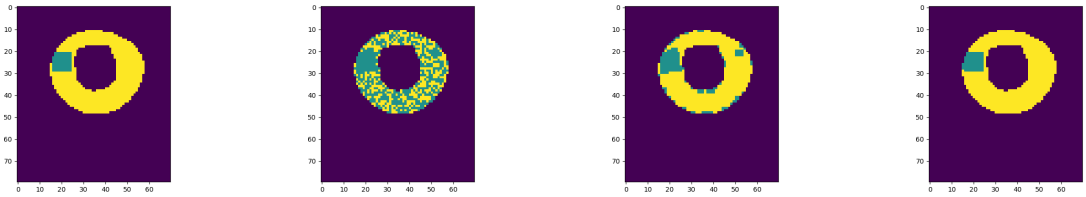
6.3.5 Results for synthetic data

We randomly generated 10 independent instantiations of the synthetic data, described in Section 6.3.2, to increase the reliability of our performance assessment. Figure 6.5 shows an example of the estimated MBF for the synthetic data. For each data set, we calculated the relative errors of the MBF estimations, E_{MBF}, as follows:

$$E_{\mathrm{MBF}} = \frac{1}{N_{\mathrm{total}}} \sum_{i=1}^{N_{\mathrm{total}}} \frac{|F^i_{\mathrm{est}} - F^i_{\mathrm{truth}}|}{F^i_{\mathrm{truth}}} \tag{6.32}$$

where N_{total} is the total number of pixels, F^i_{est} is the estimated MBF for the i^{th} pixel and F^i_{truth} is the ground truth of the MBF for the i^{th} pixel. Lower values of E_{MBF} are indicative of better performance. A comparison of the proposed HBM model with the two alternative methods for MBF estimation described in Section 6.3.3 is shown in Table 6.2. This table confirms quantitatively the qualitative impression already obtained from Figure 6.5: our method clearly outperforms the *Fermi model*, and performs slightly better than the *HBM without labels*. The reason for the performance improvement over the *Fermi model* is the inclusion of spatial context information via the Markov random field prior. The slight performance improvement over the *HBM without labels* demonstrates the edge-preserving properties afforded by the labels.

Figure 6.6 shows an example of automated tissue classification (lesion versus healthy tissue). We computed the marginal posterior probability of the pixel labels from the MCMC trajectories



(a) Ground truth (b) GMM based on Fermi (c) GMM + O&C (d) Proposed HBM

Figure 6.6: MBF classifications for synthetic data. These figures show four examples of the classification (posterior probability of the label) based on the MBF for the synthetic data with additive i.i.d. Gaussian noise of standard deviation 2.5. The pixels inside and outside of the ring (myocardium) are background. Panel (a) shows the ground truth we designed. Panel (b) shows the tissue classification using the Gaussian mixture model. Panel (c) shows the results based on GMM classification combined with the opening and closing method, which is a standard de-noising imaging processing technique. Panel (d) shows the MBF classification using the method proposed in the present work. For further details on the alternative methods used here, see Section 6.3.3.

as

$$\hat{P}(k_i = 1|D) = \frac{1}{M_0} \sum_{p=1}^{M_0} k_i^p \tag{6.33}$$

where M_0 is the length of the Markov chain, k_i^p is the p^{th} sample of label k_i and D denotes the set of all sampling values of k_i. Recall that $k_i = 1$ indicates that the ith pixel is in a lesion, whereas $k_i = 0$ indicates that the pixel lies in a healthy region. We also apply a decision threshold of 0.5 for automatic binary pixel classification. Compared with the posterior probability maps, the advantage of the binary pixel map is that it can provide a certain decision for the classification. Its disadvantage is that it does not contain the uncertainty quantification for the map. We compare the classification obtained with the new proposed HBM method with, firstly, the ground truth and, secondly, with the two alternative pixel classification schemes summarised in the final paragraph of Section 6.3.3. The figure suggests that the classification obtained with the proposed HBM is in very good agreement with the ground truth. The proposed HBM also clearly outperforms the *GMM based on Fermi* method. This is mainly due to the Markov random field prior, which achieves a substantial noise reduction. The *opening and closing* method, *GMM + O&C*, discussed in Section 6.3.3, also achieves a considerable noise reduction (see panel c in Figure 6.6). However, this noise reduction is not as reliable as the one achieved with our HBM model (panel d). In particular, the *opening and closing* method still leads to a spurious lesion, which is avoided by the proposed HBM.

To follow the visual inspection up with quantitative analysis, we did two things, both exploiting the fact that we have the ground truth labels. First, we used the estimated posterior probabilities of the pixel class labels to assign each pixel to one of the two classes (0 versus 1), based on a decision threshold of 0.5 (i.e. pixels with posterior probability above 0.5 are assigned

Table 6.2: The average relative errors of the MBF estimations

	SNR/sd*(Gaussian)		**SNR/sd(Rician)**
	4.1/2	3.27/2.5	2.91/1.8
Fermi model	0.09	0.12	0.24
HBM without label	0.04	0.04	0.19
Proposed HBM	0.03	0.04	0.18

*The abbreviation “sd" in the table stands for standard deviation of the Gaussian or Rician noise, and “SNR" means signal-to-noise ratio. The results are averages over 10 independent data instantiations.

Table 6.3: The average number of misclassified pixels

	SNR/sd*(Gaussian)		**SNR/sd(Rician)**
	4.1/2	3.27/2.5	2.91/1.8
GMM based on Fermi	64.0	234.4	183.4
GMM + O&C	17.2	42.0	36.6
Proposed HBM	2.9	4	3.5

*The abbreviation “sd" in the table stands for standard deviation of the Gaussian or Rician noise, and “SNR" means signal-to-noise ratio. The results are averages over 10 independent data instantiations.

to class 1, otherwise they’re assigned to class 0). This crisp assignment does not take posterior uncertainty into consideration, i.e. the posterior probabilities 0.51 and 0.99 are both mapped to the same class. To account for the different degrees of uncertainty, we therefore also computed the cross-entropy between the true labels and the predicted posterior probabilities. The results are shown in Tables 6.3 (misclassification) and 6.4 (cross-entropy), and clearly demonstrate that the proposed HBM outperforms the two alternative methods on both scores.

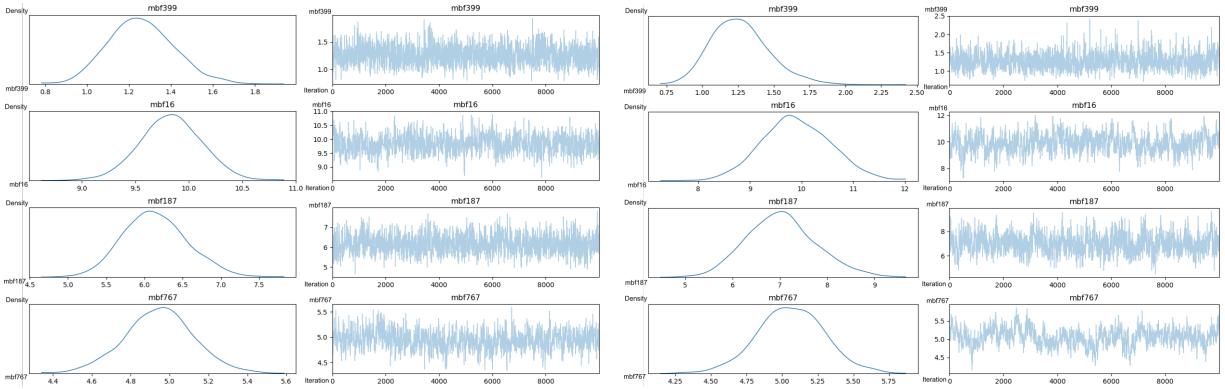
6.3.6 Results for clinical data

In this section, we discuss the application of the proposed method to the clinical data described in Section 6.3.1. We start with a discussion of our MCMC convergence diagnostics, and then continue with model selection, marginal posterior distribution analysis, MBF visualisation and

Table 6.4: Cross entropy between true labels and predicted posterior probabilities

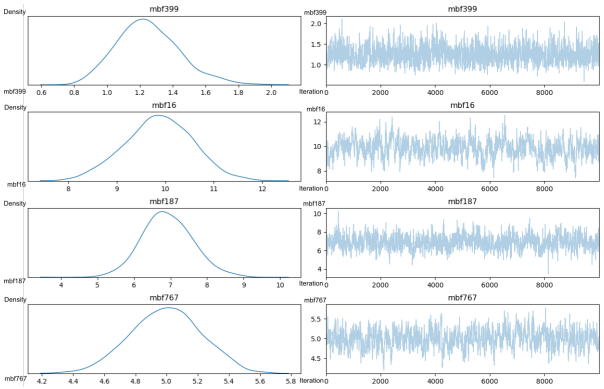
	SNR/sd*(Gaussian)		**SNR/sd(Rician)**
	4.1/2	3.27/2.5	2.91/1.8
GMM based on Fermi	2.35	8.6	6.73
GMM + O&C	0.5	1.42	1.23
Proposed HBM	0.11	0.15	0.13

*The abbreviation “sd" in the table stands for standard deviation of the Gaussian or Rician noise, and “SNR" means signal-to-noise ratio. The results are averages over 10 independent data instantiations.

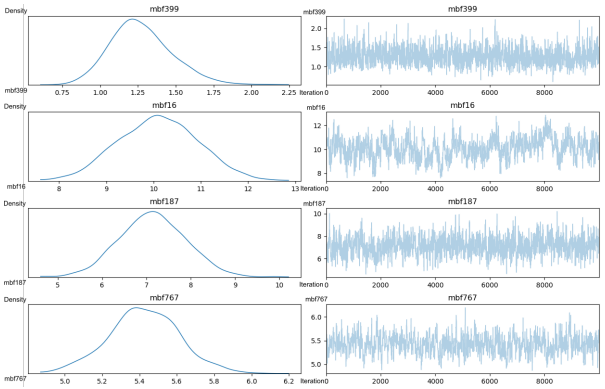


(a) $T_{\text{Fermi}} = 1, T_{k_i} = 0.1$

(b) $T_{\text{Fermi}} = 10, T_{k_i} = 0.1$



(c) $T_{\text{Fermi}} = 10, T_{k_i} = 0.01$



(d) $T_{\text{Fermi}} = 100, T_{k_i} = 0.01$

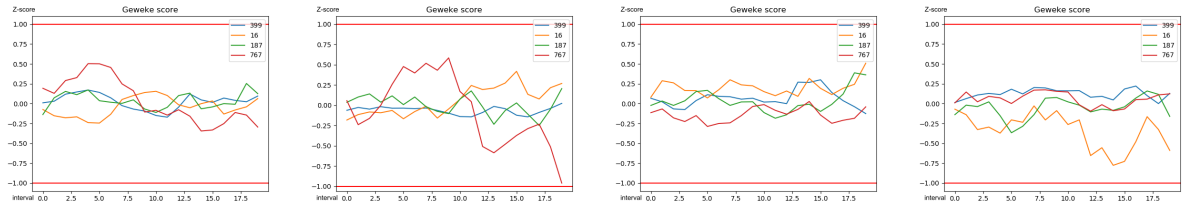
Figure 6.7: The traceplots and marginal posterior distributions estimated from the MCMC samples with a kernel density estimator of the MBF. This kernel density estimator used a Gaussian kernel and the bandwidth was selected according to Scott's rule (see details from [6]). Four typical pixels have been chosen (940 pixels overall). They are the lowest estimation (No. 399), the highest estimation (No. 16), the highest Sum Square Error (No. 187) and the lowest Sum Square Error (No. 767). The trace plots exclude the burn-in.

automated pixel classification.

MCMC convergence

Recall that our MCMC sampling algorithm is described in Algorithm 5. Figure 6.7 shows MBF traceplots for 4 typical pixels. The MCMC acceptance ratios were controlled within the range $(0.23, 0.44)$, which has been suggested to give the best mixing and convergence (see [12]). The traceplots show the exploration of the MBF samples for each pixel.

The traceplots for the MBF estimations of the four typical pixels look reasonable, with no trends to indicate insufficient convergence. As a more objective and quantitative convergence measure, we have computed the Geweke's statistic (see [24]). The results are shown in Figure 6.8, based on four independent MCMC simulations. Since the Geweke statistic stays consistently within the uncritical range $(-1, 1)$, there is no significant evidence for lack of convergence.



(a) $T_{\text{Fermi}} = 1, T_{k_i} = 0.1$ (b) $T_{\text{Fermi}} = 10, T_{k_i} = 0.1$ (c) $T_{\text{Fermi}} = 10, T_{k_i} = 0.01$ (d) $T_{\text{Fermi}} = 100, T_{k_i} = 0.01$

Figure 6.8: The Geweke test of the MBF samples. These figures show the Geweke test applied to the MBF samples for four specific pixels: the highest MBF (No. 16), the lowest MBF (No. 399), the highest SSE (No. 187) and the lowest SSE (No. 767). Two red horizontal lines indicate the upper and lower limits of the Z-score. If all values are within these limits, there is no significant evidence that the Markov chain is not converged.

Table 6.5: Watanabe-Akaike information criterion (WAIC). Lower values show better performance.

WAIC	T_{k_i}			
T_{Fermi}		0.01	0.1	1
	1	107506	**107437**	107441
	10	107566	107617	107602
	100	107715	107539	107884

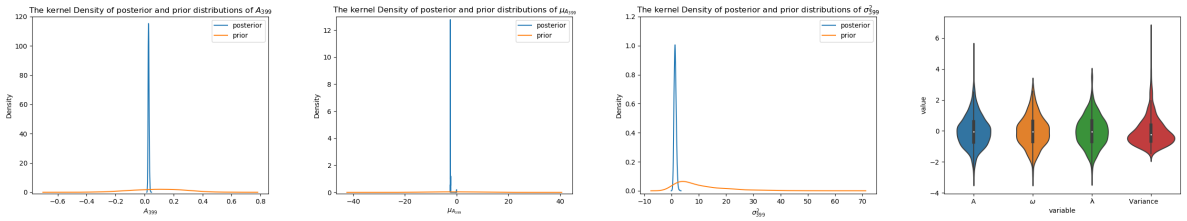
Figure 6.7 also shows the marginal posterior distribution of the MBF, which we obtained from the MCMC trajectories with a standard kernel density estimator. This provides a natural way to quantify the MBF estimation uncertainty inherent in the data.

Model selection

The Markov random field prior, introduced in Section 2.2.4, depends on two hyperparameters, T_{k_i} and T_{Fermi}; see equations (6.2) and (6.5). Rather than sampling them from the posterior distribution, we adopted a model selection approach: we ran the MCMC simulations for various fixed values of T_{k_i} and T_{Fermi}, and selected the best combination with WAIC; see Section 2.6.3 for details. The results are shown in Table 6.5, from which the best parameter combination can be obtained.

Marginal posterior distribution analysis

The posterior distributions are particularly interesting in comparison with the corresponding prior distributions. For illustration, we have randomly selected a pixel ($i = 399$) in the MR image, and estimated the posterior distributions for the associated parameters A_i, μ_{A_i} and σ_i^2. The reason we choose these three parameters is that they are representative of the other parameters. ω_i and λ_i are related to A_i, as these are all Fermi parameters. $\sigma_{A_i}^2$ is related to μ_{A_i}; both are hyperparameters of the prior distribution of A_i. Panels (a) - (c), Figure 6.9 show a comparison of the



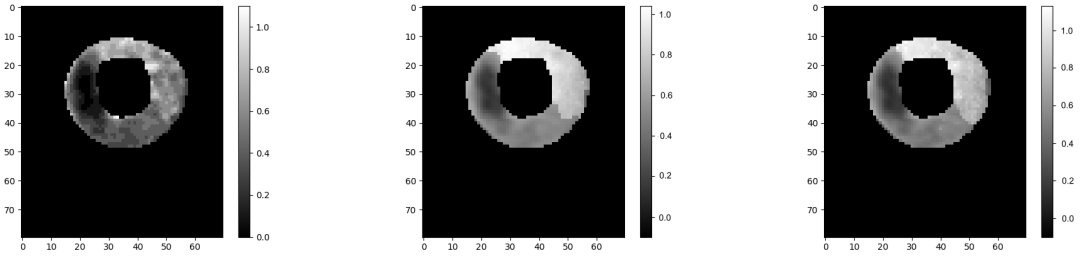
(a) A_{399} (b) $\mu_{A_{399}}$ (c) σ^2_{399} (d) Fermi parameters A_{399}, ω_{399}, λ_{399} and variance σ^2_{399}

Figure 6.9: Panels (a) - (c) show a comparison between prior and marginal posterior distributions for parameters A_i, μ_{A_i} and σ_i^2 associated with a randomly selected pixel $i = 399$. Panel (d) shows violin plots of the standardised marginal posterior distributions for Fermi parameters A_i, ω_i, λ_i and variance σ_i^2 given a randomly selected pixel $i = 399$.

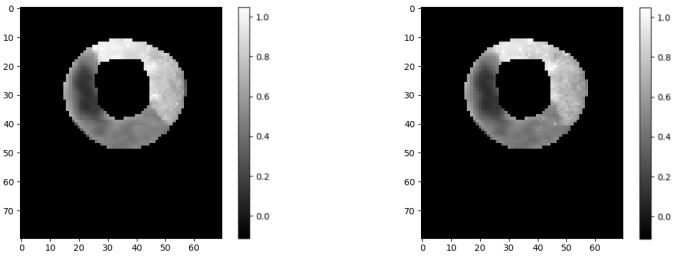
corresponding prior and posterior distributions. From the strong collapse of the posterior, whose effective support has shrunk considerably compared to that of the prior, we learn that the data are highly informative for estimating these parameters. Model developers interested in interpreting the model parameters can learn about their degree of practical identifiability from these plots. Panel (d), Figure 6.9 shows violin plots of the marginal posterior distributions of Fermi parameters A_i, ω_i, λ_i and variance σ_i^2 given a randomly selected pixel $i = 399$. These plots allow a quantification of the intrinsic posterior estimation uncertainty. In addition, they can be inspected for various patterns in the posterior distributions. For instance, the marginal posterior distributions of the parameters A_i, λ_i and ω_i are symmetric and bell-shaped, consistent with a Gaussian distribution. The marginal posterior distribution of σ_i^2, on the other hand, is skewed with a long upper tail. This is consistent with the prior distribution (log normal or gamma distribution) and suggests that despite the fact that the data are highly informative about this parameter, with a strong collapse of the posterior distribution, as discussed above, the functional form of the prior is maintained.

MBF visualisation

Figure 6.10 shows the estimation of the MBF with different alternative models: the traditional Fermi model fitted with least-squares (see Section 6.3.3), and the proposed HBM model with different values for the hyperparameters T_{k_i} and T_{Fermi}. The greyscale in the plots indicates the degree of hypo-perfusion, with darker greyscales indicating more severe hypo-perfusion. All predictions suggest that there is a severely hypo-perfused area on the left of each ring, and a slightly hypo-perfused area at the bottom. However, the HBM models predict a smoother contrast than the Fermi model, and the boundaries between hypo-perfused areas (indicative of lesions) and normally perfused areas (indicative of healthy tissue) are clearer. The best estimation of the MBF, based on our model selection scheme with WAIC, is shown in panel (b). However, the differences between the MBF estimations for different hyperparameters are fairly



(a) Fermi model (b) $T_{\text{Fermi}} = 1, T_{k_i} = 0.1$ (c) $T_{\text{Fermi}} = 10, T_{k_i} = 0.1$



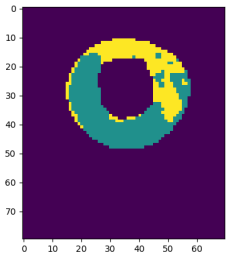
(d) $T_{\text{Fermi}} = 10, T_{k_i} = 0.01$ (e) $T_{\text{Fermi}} = 100, T_{k_i} = 0.01$

Figure 6.10: This figure shows, for different locations in the myocardium, the estimated MBF (rescaled within [0,1]), represented by different greyscales. Background pixels are shown in black. The different panels show the results for different models. Panel (a): traditional Fermi model, as described in Section 6.3.3. Panels (b-e): proposed HBM model with different hyperparameter values.

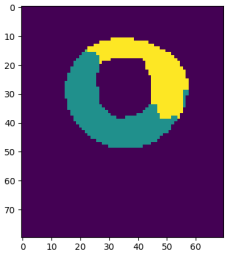
minor, indicating a certain degree of robustness of the proposed HBM model with respect to a variation in the hyperparameters

Automated pixel classification

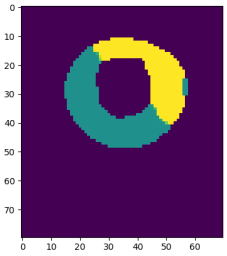
Figure 6.10 shows estimations of the MBF. This still requires manual inspection and interpretation. What we need for clinical decision support is automated classification of pixels and regions as healthy versus diseased, so as to automatically detect lesions. To this end, we compute the posterior probabilities of the labels from the MCMC trajectories using equation (6.33). Hence, a high posterior probability close to 1 is indicative of healthy tissue, whereas a value close to 0 indicates a lesion. Figure 6.11 shows the posterior probability of the class label for different models and different settings of the hyperparameters T_{k_i} and T_{Fermi}. The figure shows a crisp separation into regions with a high posterior probability of being a lesion (shown in dark green), and regions with a high posterior probability of being healthy (shown in yellow). There is only a small “uncertain" region with posterior probabilities in the order of 0.5, which we have enlarged in panel (f). For automatic pixel classification, we have to threshold these posterior probabilities: classify a pixel i as lesion if $\hat{P}(k_i = 0|D) \geq 1 - \tau$; classify a pixel i as healthy tissue if $\hat{P}(k_i = 0|D) \leq \tau$; and classify a pixel i as uncertain if $\tau < \hat{P}(k_i = 0|D) < 1 - \tau$. To



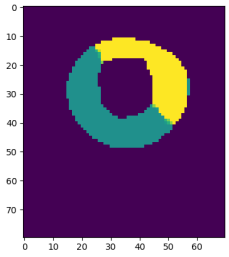
(a) GMM based on Fermi model



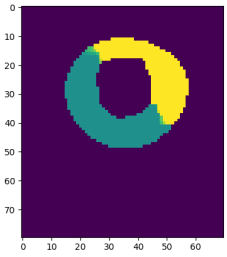
(b) $T_{\text{Fermi}} = 1, T_{k_i} = 0.1$



(c) $T_{\text{Fermi}} = 10, T_{k_i} = 0.1$



(d) $T_{\text{Fermi}} = 10, T_{k_i} = 0.01$



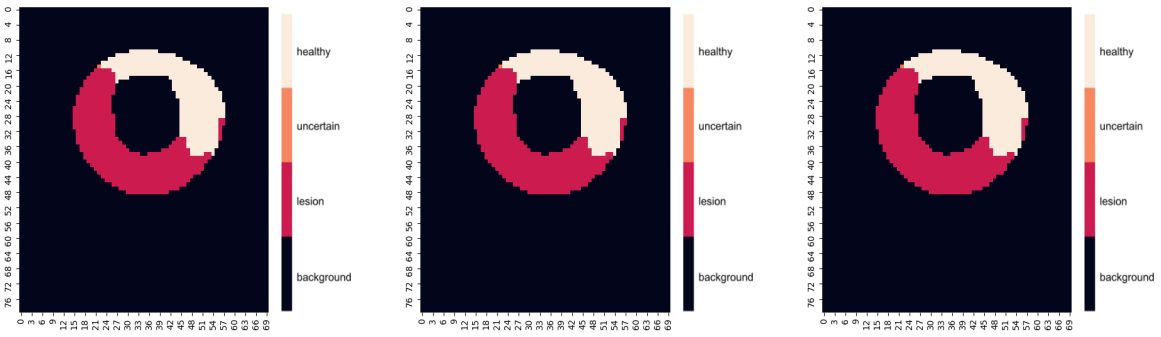
(e) $T_{\text{Fermi}} = 100, T_{k_i} = 0.01$

(f) Uncertain region in Panel (e)

Figure 6.11: Posterior probabilities for myocardial tissue classification. The figure shows the pixel-wise myocardial tissue posterior probability for *lesions* versus *healthy tissue*. *Dark green* indicates a high posterior probability that the corresponding pixel belongs to a lesion; *yellow* indicates a high posterior probability that the corresponding pixel belongs to healthy tissue. There is a small “uncertain" region with posterior probabilities in the order of 0.5, which we have enlarged in panel (f). The black colour marks the background of the myocardial ring. The different panels show the results for different models. *Panel (a):* traditional Fermi model combined with a Gaussian mixture model, as described in Section 6.3.3. *Panels (b-e):* proposed HBM model with different hyperparameters. Panel (f): the uncertain region in Panel (e), enlarged.

investigate the effect of this threshold, we have chosen the hyperparameters with the best WAIC score ($T_{k_i} = 0.1$ and $T_{\text{Fermi}} = 1$) and repeated the classification for three different settings of the threshold τ: $\tau \in \{0.025, 0.05, 0.1\}$. The results are shown in Figure 6.12, which demonstrates that the classification labels are not affected by the variation of the threshold. This is in line with Figure 6.11, which shows that most pixels have posterior probabilities close to 1 or 0 and are therefore not affected by the particular choice of threshold.

Figure 6.11 shows the posterior probabilities for both the GMM based on the traditional Fermi model (as described in Section 6.3.3), and the proposed HBM model with various settings of the hyperparameters T_{k_i} and T_{Fermi}. Note that the traditional method in Figure 6.11 panel (a) shows a fragmented structure, with various small and singleton lesions embedded in healthy tissue. This is physiologically unrealistic. Left ventricular myocardium is supplied by three major epicardial arteries, and any impairment in blood delivery is expected to affect consecutive regions of tissue (vascular territories), and not isolated small patches as shown in Figure 6.11(a). Interestingly, the physiologically unrealistic fragmentation and the prediction of

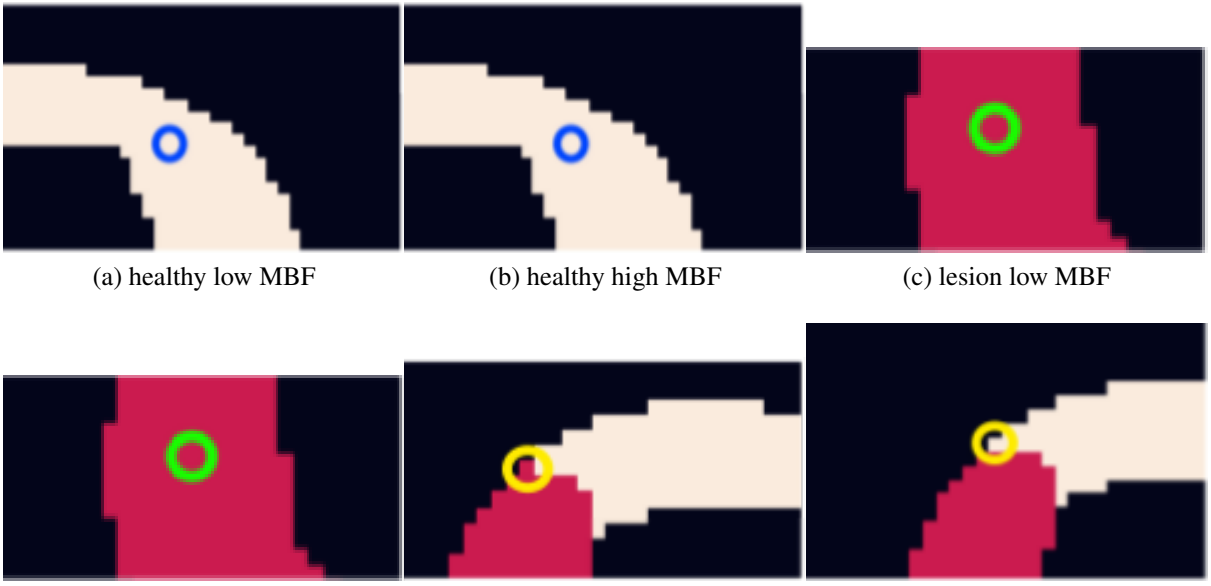


(a) $\tau = 0.025$ (b) $\tau = 0.05$ (c) $\tau = 0.1$

Figure 6.12: Automatic pixel classification. The figure shows the classification of pixels into the three classes *lesion*(red), *healthy tissue* (white) and *uncertain* (orange), by putting the posterior probabilities of Figure 6.11 (b) through the thresholding scheme described in the main text, for three different values of the threshold parameter τ.

spurious singleton clusters are suppressed with our HBM model, irrespective of the hyperparameter settings. This is shown in Figure 6.11 panels (b-e). This finding does not only demonstrate that the inclusion of spatial context information via the Markov random field prior achieves physiologically more realistic cluster assignments, but also demonstrates that the classification results show reasonable robustness with respect to a variation of the hyperparameters. Different values of the hyperparameters do not affect the location of the detected lesion much. Again, the best classification map can be found with our model selection scheme based on WAIC, shown in Table 6.5, which corresponds to panel Figure 6.11(b).

We have run an additional simulation study to focus on the interactions between the parameter and label uncertainties. We have chosen three pixels in distinct characteristic positions of the myocardium: the centre of the healthy tissue, the centre of the lesion, and the edge between healthy tissue and lesion. We set the Fermi parameters to their 10% and 90% posterior quantiles, as obtained from our previous MCMC simulations. All other parameters (including the Fermi parameters associated with the other pixels) were kept fixed. We ran 1 MCMC simulation with 1000 Gibbs sampling steps given different quantiles of Fermi parameters (specific pixel). We selected two combinations of Fermi parameters that represent the lowest MBF and the highest MBF respectively. The results can be found in Figure 6.13. It can be seen that the labels of the pixel inside the healthy tissue (marked by a blue circle) and the lesion (marked by a green circle) do not change and are invariant with respect to changing the quantiles of the Fermi parameters. However, the pixel near the boundary between the healthy tissue and the lesion (marked by a yellow circle), is affected by the quantiles of the Fermi parameters, with different quantiles leading to different labels. To explain this behaviour, note that the effect of the Markov random field prior on an individual target pixel is strong when the surrounding pixels have the same label, leading to low transition probabilities into different labels for the target pixel. A target pixel



(a) healthy low MBF (b) healthy high MBF (c) lesion low MBF

(d) lesion high MBF (e) boundary low MBF (f) boundary high MBF

Figure 6.13: The classification maps given fixed parameters. Panels (a) - (f) show the classification results given fixed parameters (10% and 90% quantiles of the Fermi parameters) in different locations (healthy region, lesion region, boundary region) of the myocardium successively. The white pixels indicate healthy tissue and the red pixels indicate lesions. The pixels inside the blue circles (healthy tissue), green circles (lesion) and yellow circles (boundary) have fixed Fermi parameters (10% and 90% quantiles).

near a boundary, on the other hand, is surrounded by pixels with different labels, whose influence effectively cancels out in the Markov random field. Consequently, the Fermi parameters now have a significant effect on the target pixel's label.

6.4 Discussion and conclusion

In this chapter, we have introduced a hierarchical Bayesian model with a Markov random field prior to classify myocardial tissues based on the MBF. This method systematically combines signal intensities from raw magnetic resonance images with spatial context information related to the individual image pixels to achieve physiologically more realistic myocardial tissue classification. Traditionally, least-squares fitting has been (and is still widely being) used to estimate MBF using the Fermi model, and we have used it as a benchmark for the MBF estimation in the present work. Our work is methodologically related to, and has been inspired by, the HBM model of myocardial perfusion MRI proposed in Scannell's work [5]. However, this study does not address the problem of automatic myocardial tissue classification, and we have found in general that this topic has only been addressed scarcely in the current cardio-physiological modelling literature. For that reason, we have used a standard and widely applied statistical model as a further benchmark for tissue classification from MBF data: the Gaussian mixture model,

with parameter estimation using the Expectation Maximisation (EM) algorithm. The proposed SVFMM and GMM-MRF methods are not compared here because a final comparisons for all methods will be present in Chapter 7.

A comparison with these two traditional methods has revealed three advantages of the proposed hierarchical Bayesian modelling framework. Firstly, it can generate a clearer, smoother and more realistic MBF map by taking physiological context information into account. Secondly, our inference method properly quantifies the uncertainty in both the MBF as well as the latent label assignment on which the myocardial tissue classification is based. Thirdly, the method exploits the relationship between the MBF estimation and the latent pixel labels for automatic myocardial tissue classification. This achieves automation (which is essential for clinical applications) at improved estimation and classification accuracy, paving the way to a future clinical decision support system.

The principal bottleneck of the proposed method for clinical applications is the high computational cost of the MCMC simulations. In our work, a single MCMC iteration took in the order of 2 seconds, leading to run times of about 5.5 hours for a typical MCMC chain length of 10,000 iterations (using the following hardware: Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz processor with 64GB memory). A potential solution is parallel processing to reduce these run times. This will exploit the fact that all relevant computations are invariant with respect to a permutation of the order in which the pixel is processed, and all pixel-related operations can be carried out in parallel rather than sequentially. A consequent improvement over CPU-based simulations would be the porting of our code to GPU platforms. In this way, the expected run time could be reduced to 1/10 according to [154]. While MCMC is intrinsically sequential, there are alternative sampling methods that can better exploit the scope for parallelisation, most notably sequential Monte Carlo (SMC) and adaptive multiple importance sampling (AMIS), which in combination with GPU architectures could significantly reduce the run times (see [165] and [166]). Another approach is to replace sampling based Bayesian inference with modal or distributional approximations, using the Laplace approximation, variational inference or expectation propagation; see details in [12] (Chapter 13).

A further limitation is related to the physical model used for quantifying the MBF. The MBF is approximated by the Fermi model, which depends on three shape parameters $\{A, \omega, \lambda\}$. For modelling the dependence on other kinetic parameters, like fractional plasma volume and fractional interstitial volume, more complex models need to be employed. However, as we have already mentioned before, the hierarchical Bayesian modelling framework we have proposed is agnostic to the specific form of the physical model employed. To paraphrase this: if a more complex physical model is employed, it can just take the place of the Fermi model in the proposed hierarchical Bayesian model, and all its parameters can be estimated methodologically the same way as proposed here for the Fermi model (exploiting the information-sharing capacity and uncertainty quantification of the hierarchical Bayesian model and its spatial Markov random field

prior). To be more specific, if a model satisfies the central volume principle in equation (4.9) and its structure of impulse response function $R_f(t)$ is given, this model can be put into the hierarchical structure presented in this work by replacing the Fermi parameters A, ω, λ in Figure 6.3 with the parameters in its impulse response function $R_f(t)$. In particular, the classification of the myocardial tissue for automatic detection of lesions will proceed in the same way as described here, while potentially benefitting from the richer structure of more advanced physical models.

Finally, there is scope for future improvement related to the choice of the prior distribution. In the present work, we have used vague prior distributions for the hyperparameters for Fermi parameters, i.e. the parameters $\mathbf{\Gamma}$ in Figure 6.2. This is a consequence of an intrinsic limitation of the Fermi model: while widely used in the literature, it depends on empirical shape parameters that lack any clear physical interpretation. However, for more advanced and realistic physical models, e.g. the tracer-kinetic models (see [77] and [78]), whose parameters have bio-physical and physiological interpretability, methods of cardio-physiological knowledge elicitation can be adopted to derive more informative prior distributions that have the potential to boost the predictive and diagnostic performance of the proposed methodology.

Chapter 7

Automatic Lesion Detection in Myocardial Perfusion DCE-MRI using Hierarchical Bayesian Models with Spatio-Temporal Markov Random Fields

7.1 Introduction

In Chapter 6, the hierarchical Bayesian modelling of myocardial perfusion DCE-MRI based on physiological models has been introduced. The performance of this method is highly dependent on the performance of the physiological model. Therefore, if the physiological model cannot correctly estimate the myocardial blood flow, the method illustrated in Chapter 6 is at risk of failure. To this end, a data-driven approach can avoid making use of physiological models and generate accurate classification maps.

In the chapter, we apply and evaluate a novel hierarchical Bayesian model (HBM) combined with spatio-temporal Markov random field (MRF) priors to the DCE-MRI myocardial perfusion data. We aim to classify the myocardial tissues into two categories, healthy tissue and lesion by using iteratively denoised images. Specifically, given the original images, the proposed method firstly generates an initial classification map. Then, based on the generated classification map, it generates denoised images. The proposed method repeats this procedure until it converges. Spatio-temporal Markov random field priors (see details in Section 2.2.4) are used to introduce the spatio-temporal information for each pixel. We have derived a posterior inference scheme for the parameters in the HBM model, using a MCMC (see details in Section 2.3) variant (Gibbs sampling) to approximately draw samples of the parameters from their posterior distribution. For an adequate trade-off between accuracy and computational efficiency, the hyperparameters are selected using model selection techniques based on the Watanabe Akaike information criterion (WAIC) (see details in Section 2.6.3). The proposed method is tested on both synthetic

data and clinical DCE-MRI scans, and the results are compared with two established methods: a Gaussian mixture model (GMM) classification method (see Section 5.2) and GMM based opening and closing operations [163, 167]. While the effectiveness of our classification method is demonstrated using raw DCE-MRI myocardial data, its formulation is applicable to any form of calibrated (transformed) datasets, where blood pool signal is used to derive absolute values of MBF [168].

7.2 Data

7.2.1 Clinical data

The myocardial perfusion DCE-MRI data (short-axis mid-cavity slice) from three patients (DCE-MRI datasets 1-3) were analyzed in this work. For one set of data (DCE-MRI dataset 1), we show all results (model selection, MCMC convergence test, classification maps and generated denoised MR images) in Section 7.5. For two extra sets of data, we only show classification maps and generated denoised MR images to emphasize the robustness of the proposed method. The general data information can be found in Section 3.3.

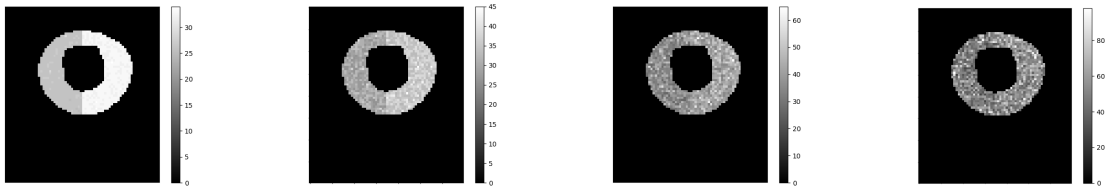
7.2.2 Synthetic data

The synthetic data was designed to mimic clinically observed lesion. A double exponential curve was used to model the signal intensity with time, e.g. see details in [162] (Chapter 2):

$$s(t) = \frac{p_2 p_3}{(p_2 - p_1)} \times (e^{-p_1 t} - e^{-p_2 t}). \tag{7.1}$$

Rician noise is usually assumed for MR images according to the work in [161]. In this way, the synthetic data was generated based on equation (7.1) by adding Rician noise with different scales. Specifically, we used different variances of Rayleigh distribution (a special case of Rician distribution) to design the degrees of noise. The variance values are 1^2 (very low noise), 2^2 (low noise), 3^2 (high noise) and 4^2 (very high noise). Figure 7.1 shows the maximum enhanced (ME) images with different noise variances. The ME images are the images at the time that the signals are enhanced to the maximum values.

In the proposed work, we have designed the synthetic images with different lesion sizes to test the proposed method and carried out a comparative evaluation with various alternative benchmark methods. Specifically, there were three sizes of lesions in the myocardium. A small lesion spanned 60°, a medium-size lesion spanned 120°, and a large lesion spanned 180° circumferentially. There were a total of 12 combinations based on different lesion sizes and values of noise variance. For each combination, 10 separate sets of synthetic images were designed to improve the robustness of our evaluation and reduce the effect of potential outliers. The



(a) Var 1^2 (b) Var 2^2 (c) Var 3^2 (d) Var 4^2

Figure 7.1: The noise added maximum enhanced images. Panels (a) - (d) show the noise added ME images with noise variance 1^2, 2^2, 3^2 and 4^2 respectively.

generations of these synthetic data can be found in Figure 7.2.

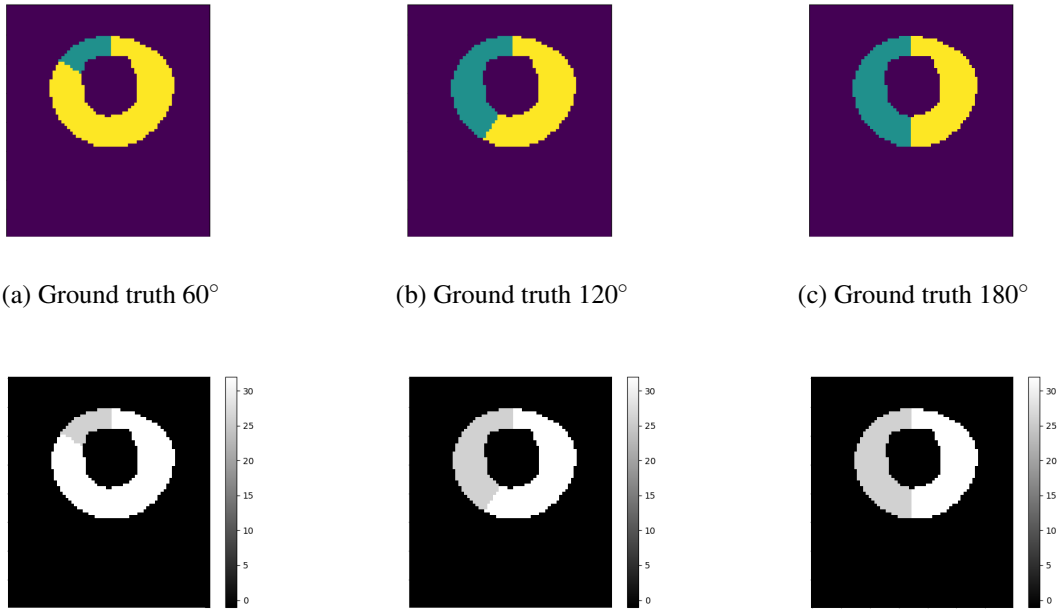
The parameters in equation (7.1) for healthy tissues and lesions are given based on empirical experiences. To be specific, we use equation (7.1) to fit a model to the clinical DCE-MRI data (serial 1) by least-squares estimation to obtain the values of the parameters. In detail, the values of parameters in equation (7.1) are set to $p_1 = 0.01$, $p_2 = 0.4$, $p_3 = 25$ for the healthy tissues. For the lesions, the parameters are set to $p_1 = 0.02$, $p_2 = 0.3$, $p_3 = 20$.

7.3 Method

7.3.1 Spatio-temporal information

Spatial methods have been widely applied for quantitative assessment of myocardial perfusion DCE-MRI [5, 105, 169, 170]. Moreover, the method introduced in Chapter 6 has also made use of spatial methods. We also introduce temporal information because the myocardial perfusion DCE-MRI data are time-series images. The signal intensity for a pixel at different times are related, which is illustrated in Figure 3.7. Specifically, there are three stages for the change of signals of myocardial pixels. In the beginning, the signal stays relatively low because the contrast agent has not flown into the myocardium. Then, it increases steadily because the contrast agent is flowing into the myocardium. Finally, it firstly decreases slightly and then keeps this value because the contrast agent is washed out from the myocardium. Therefore, we introduce spatio-temporal information to the DCE-MR SI images using Markov random field priors, which are explicitly described in Section 7.3.3.

Similarly, the proposed method illustrated in this chapter also introduces Markov random fields to the label. This aims to solve the issues of unclear boundary and single isolated segments. The details can be found in Sections 5.2 and 5.3.



(a) Ground truth 60° (b) Ground truth 120° (c) Ground truth 180°

(d) ME 60° (e) ME 120° (f) ME 180°

Figure 7.2: The simulations of synthetic data. Panels (a) - (c) show the design of lesions with 60°, 120° and 180° respectively. The dark blue region in these images denote the lesion. (d) - (f) show the maximum enhanced (ME) images for each size of lesion.

7.3.2 Hierarchical Bayesian model

Let $y^i(t)$ denote the logarithm of the signal intensity of an MR image pixel ($i = 1,2,...,N$) at time point t ($t = 1,2,...,M$), where N is the number of pixels and M is the number of time points. $k^i \in \{0,1\}$ is defined as the state (healthy and lesion) for the i^{th} pixel.

In this work, a hierarchical Bayesian model was used to classify the tissues based on myocardial perfusion DCE-MRI. This model can be found in Figure 7.3. The structure of the HBM reveals the three-layer causal relationships between parameters. Specifically, the observed signals are dependent on the true signals, and the true signals dependent on spatial-temporal neighbours and the labels. Table 7.1 shows the description of the parameters in Figure 7.3, and Figure 7.3 explicitly illustrates the relationships between the parameters in this HBM model. According to the factorization rule, which is specified in Section 2.2.3, the joint distribution for this hierarchical Bayesian model is the product of conditional distributions determined by their respective parent nodes. In this way, the joint distribution is:

$$
\begin{aligned}
& P\left(y^i(t), \psi^i(t), \sigma^2, \psi^i(t+j), \psi^{-i}(t), a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j), \mu_{k^i}(t), \sigma^2_{k^i}(t), \mathbf{\Gamma}, k^i, k^{-i}\right) \\
=& P\left(\psi^i(t) | \psi^i(t+j), \psi^{-i}(t), a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j), \mu_{k^i}(t), \sigma^2_{k^i}(t), k^i, k^{-i}\right) \\
& P\left(y^i(t) | \psi^i(t), \sigma^2\right) P\left(a_{k^i}(t+j) | \mu_{**}, \sigma^2_{**}\right) P\left(\mathbf{b}_{k^i}(t+j) | \mu_{**}, \sigma^2_{**}\right) P\left(\mu_{k^i}(t) | \mu_*, \sigma^2_*\right) \\
& P\left(\sigma^2_{k^i}(t) | \alpha_*, \beta_*\right) P\left(\sigma^2 | \alpha_{**}, \beta_{**}\right) P\left(k^i | k^{-i}\right) P\left(k^{-i}\right) P\left(\psi^i(t+j)\right) P\left(\psi^{-i}(t)\right)
\end{aligned} \tag{7.2}
$$

Table 7.1: Descriptions of the parameters for the HBM model

Parameter	Description
$y^i(t)$	the logarithm (log) of the observed signal for pixel i at time t
$\psi^i(t)$	the log of the true signal for pixel i at time t
σ^2	the variance of the iid Gaussian noise
k^i	the label for pixel i
k^{-i}	the labels for the neighbours of pixel i
$\psi^{-i}(t)$	the log of the true signals for the neighbours of pixel i at time t
$\psi^i(t+j)$	the log of the true signals for pixel i at time $t+j$
$a_{k^i}(t+j)$	autoregressive parameter (constant) for $\psi^i(t)$and $\psi^i(t+j)$
$\mathbf{b}_{k^i}(t+j)$	autoregressive parameter (coefficient) for $\psi^i(t)$and $\psi^i(t+j)$
$\mu_{k^i}(t)$	mean for the true signals conditional on the label at time t
$\sigma^2_{k^i}(t)$	variance for the true signals conditional on the label at time t
$*$ and $**$	the symbols with $*$ and $**$ are hyperparameters

where $\mathbf{\Gamma} = \{\mu_*, \sigma^2_*, \alpha_*, \beta_*, \mu_{**}, \sigma^2_{**}, \alpha_{**}, \beta_{**}\}$ is a vector containing all hyperparameters. All nodes in Figure 7.3 are illustrated in detail in this chapter. The logarithm of signal intensity $y^i(t)$ is assumed to be Gaussian distributed with mean $\psi^i(t)$ and variance σ^2, which can be written as:

$$P\left(y^i(t)|\psi^i(t),\sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\left(y^i(t)-\psi^i(t)\right)^2}{2\sigma^2}\right) \tag{7.3}$$

where $\psi^i(t)$ indicates the true signal intensity of pixel i at time t. σ^2 is assumed to be the variance of the iid Gaussian noise. The Rician noise will be further simulated in the synthetic data to test how critical the Gaussian assumption is. $\alpha_{**} = 0.1$ and $\beta_{**} = 0.1$ are fairly uninformative hyperparameters for variance σ^2. In this way, the prior distribution for σ^2, which is conjugate, is

$$P(\sigma^2|\alpha_{**},\beta_{**}) = \frac{\beta_{**}^{\alpha_{**}}}{\Gamma(\alpha_{**})}(\sigma^2)^{-\alpha_{**}-1}\exp\left(-\frac{\beta_{**}}{\sigma^2}\right). \tag{7.4}$$

The myocardial blood flows for healthy tissues and lesions are different. Moreover, the myocardial blood flows are positively correlated with the signal intensities for myocardial tissues. In this way, $\psi^i(t)$ is conditionally dependent on the label k^i. Specifically, the probability density of $\psi^i(t)$ conditional on the label k^i is

$$P\left(\psi^i(t)|\mu_\phi(t),\sigma^2_\phi(t)\right) = \frac{1}{\sqrt{2\pi}\sigma_\phi(t)}\exp\left(\frac{-\left(\psi^i(t)-\mu_\phi(t)\right)^2}{2\sigma^2_\phi(t)}\right) \tag{7.5}$$

with $\phi \in \{0,1\}$. $\mu_\phi(t)$ and $\sigma^2_\phi(t)$ are the simple forms of $\mu_{k^i=\phi}(t)$ and $\sigma^2_{k^i=\phi}(t)$. Let $\boldsymbol{\psi}_\phi(t) = \{\psi^i(t)\}_{i|k^i=\phi}$, we have:

$$P\left(\boldsymbol{\psi}_\phi(t)|\mu_\phi(t),\sigma^2_\phi(t)\right) = \left(2\pi\sigma^2_\phi(t)\right)^{-\frac{N_\phi}{2}}\exp\left(\frac{-\sum_{i|k^i=\phi}\left(\psi^i(t)-\mu_\phi(t)\right)^2}{2\sigma^2_\phi(t)}\right) \tag{7.6}$$

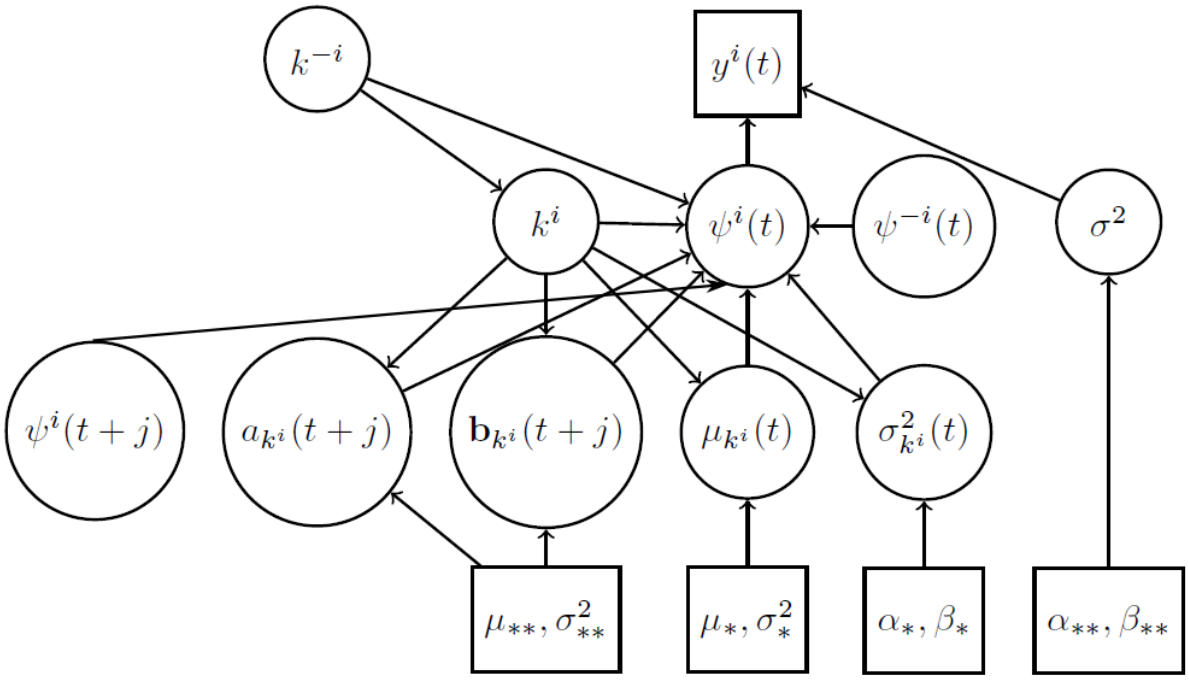


Figure 7.3: This figure shows the structure of the hierarchical Bayesian model proposed in this work. The circle nodes denote variables and the rectangle nodes denote fixed values. The descriptions of the parameters in this Figure can be found in Table 7.1.

with $\phi \in \{0,1\}$. N_ϕ is the number of pixels which satisfy label $k^i = \phi$. In a nutshell, all pixels in the myocardium are separated in two categories, healthy group and lesion group. The pixels in different groups have different means and standard deviations. $\mu_* = 0$, $\sigma_*^2 = 10$, $\alpha_* = 0.1$ and $\beta_* = 0.1$ are fairly uninformative hyperparameters for $\mu_{k^i}(t)$ and $\sigma^2_{k^i}(t)$. In this way, the prior distributions for $\mu_{k^i}(t)$ and $\sigma^2_{k^i}(t)$, which are conjugate, are

$$P\left(\mu_{k^i}(t)|\mu_*, \sigma_*^2\right) = \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left(\frac{-\left(\mu_{k^i}(t) - \mu_*\right)^2}{2\sigma_*^2}\right) \tag{7.7}$$

and

$$P\left(\sigma^2_{k^i}(t)|\alpha_*, \beta_*\right) = \frac{\beta_*^{\alpha_*}}{\Gamma(\alpha_*)} \left(\sigma^2_{k^i}(t)\right)^{-\alpha_*-1} \exp\left(-\frac{\beta_*}{\sigma^2_{k^i}(t)}\right). \tag{7.8}$$

For the true signal $\psi^i(t)$, we applied its spatial neighbours $\psi^{-i}(t)$ using Markov random fields. We also introduced the temporal neighbours of $\psi^i(t)$, $\psi^i(t+j)$, using Markov random fields. Similarly, the label k^i was also assumed to be conditional on its spatial neighbours k^{-i}. The definitions of $\psi^{-i}(t)$, $\psi^i(t+j)$ and k^{-i} are explicitly illustrated in Section 7.3.3.

7.3.3 Markov random fields

Markov random fields for labels

A Markov random field method was used to introduce the spatial neighbourhood information in this work. According to the Hammersley-Clifford theorem, a Markov random field can be written as a Gibbs distribution (see detials in Section 2.2.4),

$$P(k^i|k^{-i}) = \frac{1}{\mathrm{Q}} \exp\left(-U(k^i|k^{-i})\right) \tag{7.9}$$

where Q is a normalization constant to ensure the integral of $P(k^i|k^{-i})$ is 1. k^{-i} is the set of all other labels excluding k^i, $k^{-i} = \{k^m\}_{m\neq i}$. We defined $U(k^i|k^{-i})$ to be:

$$U(k^i|k^{-i}) = \frac{1}{\mathrm{T}} \sum_{i\sim m} u(k^i|k^m), \tag{7.10}$$

where T is a weight parameter. $i \sim m$ denotes the spatial neighbours of pixel i. The definition of $u(k^i|k^m)$ is

$$u(k^i|k^m) = \begin{cases} -\left(\frac{1}{2}\right)^{o-1} & k^i = k^m, \\ \left(\frac{1}{2}\right)^{o-1} & k^i \neq k^m \end{cases} \tag{7.11}$$

where o indicates the degree of neighbouring. Figure 5.17 shows the definitions of spatial and temporal neighbours in this work. We chose $o = 1$ in this work because it is computationally cheaper than higher degrees, e.g. $o = 2$. However, our method provides a straightforward mechanism for extension to higher degrees. In Chapter 6, $o = 2$ has been applied. The reason we applied $o = 1$ in this case is that $o = 1$ was tested and found to produce similar results to $o = 2$, and $o = 1$ is much computationally cheaper than $o = 2$.

Markov random fields for true signals

As mentioned in Section 7.3.1, we introduced spatial Markov random fields to the true signal $\psi^i(t)$. A Markov random field can be written as a Gibbs distribution, we therefore have:

$$P\left(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i}\right) = \frac{1}{Q_{\mathrm{ps}}} \exp\left(-U(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i})\right) \tag{7.12}$$

where $\mathrm{Q_{ps}}$ is a normalization constant. $\psi^{-i}(t)$ is the set of all other true signals excluding $\psi^i(t)$, $\psi^{-i}(t) = \{\psi^m(t)\}_{m\neq i}$. We defined $U\left(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i}\right)$ to be:

$$U\left(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i}\right) = \frac{1}{T_{\mathrm{ps}}} \sum_{i\sim m} \nu_{i,m} \left(\psi^i(t) - \psi^m(t)\right)^2 \tag{7.13}$$

where T_{ps} is a weight parameter and $\nu_{i,m}$ is the edge-preservation parameter defined by

$$\nu_{i,m} = \begin{cases} 0 & k^i \neq k^m(t) \\ 1 & k^i = k^m(t). \end{cases} \tag{7.14}$$

The edge-preservation parameter $\nu_{i,m}$ is used to prevent the smoothing effect of the neighbourhood information extend beyond the boundaries of different tissues, i.e. healthy tissues and lesions [5, 156].

In the proposed method, we used an auto-regressive model to describe the relationship for the true signals between different times. j is a non-zero integer to indicate the neighbouring time points. To be specific, the Markov random fields prior given temporal neighbour $\psi^i(t+j)$ is

$$\begin{aligned} &P\left(\psi^i(t)|\psi^i(t+j), a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j)\right) = \\ &\frac{1}{Q_{\text{pt}}} \exp\left(-U(\psi^i(t)|\psi^i(t+j), a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j))\right) \end{aligned} \tag{7.15}$$

where

$$\begin{aligned} &U\left(\psi^i(t)|\psi^i(t+j), a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j)\right) = \\ &\frac{1}{T_{\text{pt}}} \sum_{t\sim j} \left(\psi^i(t) - \sum_{p=0}^{q} b_{k^i}^p(t+j)\psi^i(t+j) - a_{k^i}(t+j))\right)^2 \end{aligned} \tag{7.16}$$

where Q_{pt} is a normalization constant. T_{pt} is a weight parameter and $\mathbf{b}_{k^i}(t+j) = \{b_{k^i}^0(t+j), ..., b_{k^i}^q(t+j)\}$. In our work, we assumed that the auto-regressive parameters, $\mathbf{b}_{k^i}(t+j)$ and $a_{k^i}(t+j)$ are dependent on the label k^i. This is physiologically realistic. For myocardial perfusion MRI data, the lesion tissues are hypoperfused and the healthy tissues are normally perfused. This means that not only are the signal intensities for the lesion pixels lower than for the healthy tissues, but also the growth rate of the signal intensities for the lesion pixels are lower than for the healthy tissues. In our work, we chose $j = -1$ and $q = 0$. Given $q = 0$ and $j = -1$, Gaussian prior distributions are assumed for both $b_{k^i}^0(t-1)$ and $a_{k^i}(t-1)$ with fairly uninformative hyperparameters $\mu_{**} = 0$ and $\sigma_{**}^2 = 10$. Specifically, the prior distributions for $b_{k^i}^0(t-1)$ and $a_{k^i}(t-1)$, which are conjugate, are

$$P\left(b_{k^i}^0(t-1)|\mu_{**}, \sigma_{**}^2\right) = \frac{1}{\sqrt{2\pi}\sigma_{**}} \exp\left(\frac{-(b_{k^i}^0(t-1) - \mu_{**})^2}{2\sigma_{**}^2}\right) \tag{7.17}$$

and

$$P\left(a_{k^i}(t-1)|\mu_{**}, \sigma_{**}^2\right) = \frac{1}{\sqrt{2\pi}\sigma_{**}} \exp\left(\frac{-(a_{k^i}(t-1) - \mu_{**})^2}{2\sigma_{**}^2}\right). \tag{7.18}$$

7.3.4 Posterior inference

In this section, the posterior inference for $\psi^i(t)$, $\mu_{k^i}(t)$, $\sigma^2_{k^i}(t)$, σ^2, $a_{k^i}(t+j)$, $\mathbf{b}_{k^i}(t+j)$ and k^i is illustrated. The hierarchical Bayesian model in Figure 7.3 is a directed acyclic graph (DAG). As illustrated in Section 2.2.3, in a DAG, the probability of a selected parameter conditional on all other parameters is given by the probability of this parameter conditional on its Markov blanket. Now, we will derive the conditional posterior distributions for all of these parameters, and these conditional posterior distributions can be used for Gibbs sampling.

The Markov blanket for $\psi^i(t)$ is $\{y^i(t), \psi^i(t+j), a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j), k^i, \psi^{-i}(t), k^{-i}, \mu_{k^i}(t), \sigma^2_{k^i}(t), \sigma^2\}$. In this way, its conditional posterior distribution is

$$\begin{aligned}
&P\left(\psi^i(t)|y^i(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j),k^i,\psi^{-i}(t),k^{-i},\mu_{k^i}(t),\sigma^2_{k^i}(t),\sigma^2\right)\\
\propto &P\left(y^i(t)|\psi^i(t),\sigma^2\right)P\left(\psi^i(t)|\mu_{k^i}(t),\sigma^2_{k^i}(t)\right)P\left(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i}\right)\\
&P\left(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\right).
\end{aligned} \tag{7.19}$$

where $P(y^i(t)|\psi^i(t),\sigma^2)$ can be found in equation (7.3). Given $k^i=\phi$ with $\phi\in\{0,1\}$, $P(\psi^i(t)|\mu_{k^i=\phi}(t),\sigma^2_{k^i=\phi}(t))$ can be found in equation (7.5). $P(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i})$ can be found in equations (7.12)-(7.14) and $P(\psi^i(t)|\psi^{-i}(t),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j))$ can be found in equations (7.15)-(7.16). By substituting these equations into equation (7.19), we have:

$$\begin{aligned}
&P\left(\psi^i(t)|y^i(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j),k^i,\psi^{-i}(t),k^{-i},\mu_{k^i}(t),\sigma^2_{k^i}(t),\sigma^2\right)\\
&\propto \exp\left(\frac{-A\times(\psi^i(t))^2+2B\times\psi^i(t)}{2C}\right)
\end{aligned} \tag{7.20}$$

where

$$A=\sigma^2_\phi(t)T_{\rm ps}T_{\rm pt}+\sigma^2T_{\rm ps}T_{\rm pt}+2\sigma^2\sigma^2_\phi(t)T_{\rm pt}\sum_{i\sim m}\nu_{i,m}+2\sigma^2\sigma^2_{phi}(t)T_{\rm ps}, \tag{7.21}$$

$$\begin{aligned}
B=&\sigma^2_\phi(t)T_{\rm ps}T_{\rm pt}y^i(t)+\sigma^2T_{\rm ps}T_{\rm pt}\mu_\phi(t)+2\sigma^2\sigma^2_\phi(t)T_{\rm pt}\sum_{i\sim m}\nu_{i,m}\psi^m(t)\\
&+2\sigma^2\sigma^2_\phi(t)T_{\rm ps}\left(b^0_\phi(t-1)\psi^i(t-1)+a_\phi(t-1)\right)
\end{aligned} \tag{7.22}$$

and

$$C=\sigma^2\sigma^2_\phi(t)T_{\rm ps}T_{\rm pt} \tag{7.23}$$

with $\phi\in\{0,1\}$, $q=0$ and $j=-1$. $a_\phi(t-1)$ and $b^0_\phi(t-1)$ are the simple forms of $a_{k^i=\phi}(t-1)$ and $b^0_{k^i=\phi}(t-1)$. Therefore, the conditional posterior distribution for the true signal $\psi^i(t)$ is

$$\begin{aligned}
&P\left(\psi^i(t)|y^i(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j),k^i,\psi^{-i}(t),k^{-i},\mu_{k^i}(t),\sigma^2_{k^i}(t),\sigma^2\right)\\
&=N\left(\frac{B}{A},\frac{C}{A}\right)
\end{aligned} \tag{7.24}$$

with $\phi \in \{0,1\}$, $q = 0$ and $j = -1$.

The Markov blanket for the parameter $\mu_{k^i}(t)$ is $\{\mu_*, \sigma_*^2, \psi^i(t), \sigma_{k^i}^2(t), k^i, \psi^{-i}(t), \psi^i(t+j), k^{-i}, a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j)\}$. According to equation (7.19), the parameter $\mu_{k^i}(t)$ is independent of $\psi^{-i}(t)$, $\psi^i(t+j)$, k^{-i}, $a_{k^i}(t+j)$ and $\mathbf{b}_{k^i}(t+j)$. Therefore, its conditional posterior distribution is

$$\begin{aligned} &P\left(\mu_\phi(t)|\{\psi^i(t)\}_{i|k^i=\phi}, \sigma_\phi^2(t), \mu_*, \sigma_*^2, \psi^{-i}(t), \psi^i(t+j), k^{-i}, a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j)\right) \\ &\propto P\left(\boldsymbol{\psi}_\phi(t)|\mu_\phi(t), \sigma_\phi^2(t)\right) P\left(\mu_{k^i}(t)|\mu_*, \sigma_*^2\right) \end{aligned} \tag{7.25}$$

with $\phi \in \{0,1\}$. $P(\boldsymbol{\psi}_\phi(t)|\mu_\phi(t), \sigma_\phi^2(t))$ can be found in equation (7.6). $P(\mu_{k^i}(t)|\mu_*, \sigma_*^2)$ can be found in equation (7.7). Since we choose a conjugate prior for $\mu_{k^i}(t)$, its conditional posterior distribution is

$$\begin{aligned} &P\left(\mu_\phi(t)|\{\psi^i(t)\}_{i|k^i=\phi}, \sigma_\phi^2(t), \mu_*, \sigma_*^2\right) = \\ &\mathrm{N}\left(\left(\frac{1}{\sigma_*^2} + \frac{N_\phi}{\sigma_\phi^2(t)}\right)^{-1}\left(\frac{\mu_*}{\sigma_*^2} + \frac{\sum_{i|k^i=\phi}\psi^i(t)}{\sigma_\phi^2(t)}\right), \left(\frac{1}{\sigma_*^2} + \frac{N_\phi}{\sigma_\phi^2(t)}\right)^{-1}\right) \end{aligned} \tag{7.26}$$

with $\phi \in \{0,1\}$.

The Markov blanket for the parameter $\sigma_{k^i}^2(t)$ is $\{\alpha_*, \beta_*, \psi^i(t), \mu_{k^i}(t), k^i, \psi^{-i}(t), \psi^i(t+j), k^{-i}, a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j)\}$. Similarly, according to equation (7.19), the parameter $\sigma_{k^i}^2(t)$ is independent of $\psi^{-i}(t)$, $\psi^i(t+j)$, k^{-i}, $a_{k^i}(t+j)$ and $\mathbf{b}_{k^i}(t+j)$. In this way, its conditional posterior distribution is

$$\begin{aligned} &P\left(\sigma_\phi^2(t)|\{\psi^i(t)\}_{i|k^i=\phi}, \mu_\phi(t), \alpha_*, \beta_*, \psi^{-i}(t), \psi^i(t+j), k^{-i}, a_{k^i}(t+j), \mathbf{b}_{k^i}(t+j)\right) \\ &\propto P\left(\boldsymbol{\psi}_\phi(t)|\mu_\phi(t), \sigma_\phi^2(t)\right) P\left(\sigma_{k^i}^2(t)|\alpha_*, \beta_*\right) \end{aligned} \tag{7.27}$$

with $\phi \in \{0,1\}$. $P\left(\boldsymbol{\psi}_\phi(t)|\mu_\phi(t), \sigma_\phi^2(t)\right)$ can be found in equation (7.6). Meanwhile, $P\left(\sigma_{k^i}^2(t)|\alpha_*, \beta_*\right)$ can be found in equation (7.8). Similarly, we choose a conjugate prior for $\sigma_{k^i}^2(t)$, i.e. an inverse-gamma (IG) distribution $\mathrm{IG}(\alpha_*, \beta_*)$, so that its conditional posterior distribution is

$$P\left(\sigma_{k^i}^2(t)|\{\psi^i(t)\}_{i|k^i=\phi}, \mu_\phi(t), \alpha_*, \beta_*\right) = \mathrm{IG}\left(\alpha_* + \frac{N_\phi}{2}, \beta_* + \frac{\sum_{i|k^i=\phi}(\psi^i(t) - \mu_\phi(t))^2}{2}\right) \tag{7.28}$$

with $\phi \in \{0,1\}$.

The Markov blanket for the parameter σ^2 is $\{y^i(t), \psi^i(t), \alpha_{**}, \beta_{**}\}$. Its conditional posterior distribution is

$$P\left(\sigma^2|\{y^i(t)\}_{i,t}, \{\psi^i(t)\}_{i,t}, \alpha_{**}, \beta_{**}\right) \propto \prod_{i=1}^{N}\prod_{t=1}^{M} P\left(y^i(t)|\psi^i(t), \sigma^2\right) P\left(\sigma^2|\alpha_{**}, \beta_{**}\right) \tag{7.29}$$

where $P(y^i(t)|\psi^i(t),\sigma^2)$ can be found in equation (7.3) and $P(\sigma^2|\alpha_{**},\beta_{**})$ can be found in equation (7.4). Since we choose a conjugate prior for σ^2, its conditional posterior distribution is

$$\begin{aligned} &P\left(\sigma^2|\{y^i(t)\}_{i,t},\{\psi^i(t)\}_{i,t},\alpha_{**},\beta_{**}\right) \\ =&\mathrm{IG}\left(\alpha_{**}+\frac{N}{2},\beta_{**}+\frac{\sum_{i=1}^{N}\sum_{t=1}^{M}\left(y^i(t)-\psi^i(t)\right)^2}{2}\right). \end{aligned} \tag{7.30}$$

The Markov blanket for the parameter $a_{k^i}(t+j)$ is $\{\psi^i(t),\ \psi^i(t+j),\ \mathbf{b}_{k^i}(t+j),\ \mu_{**},\ \sigma^2_{**},\ k^i,\ \psi^{-i}(t),\ k^{-i},\ \mu_{k^i}(t),\ \sigma^2_{k^i}(t)\}$. According to equation (7.19), $a_{k^i}(t+j)$ is independent of $\psi^{-i}(t)$, k^{-i}, $\mu_{k^i}(t)$ and $\sigma^2_{k^i}(t)$. Therefore, its conditional posterior distribution is

$$\begin{aligned} &P\left(a_\phi(t+j)|\{\psi^i(t),\psi^i(t+j)\}_{i|k^i=\phi},\mathbf{b}_\phi(t+j),\mu_{**},\sigma^2_{**},\psi^{-i}(t),k^{-i},\mu_{k^i}(t),\sigma^2_{k^i}(t)\right) \\ &\propto \prod_{i|k^i=\phi} P(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j))P(a_{k^i}(t+j)|\mu_{**},\sigma^2_{**}) \end{aligned} \tag{7.31}$$

where $P\left(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\right)$ can be found in equations (7.15) and (7.16). $P\left(a_{k^i}(t+j)|\mu_{**},\sigma^2_{**}\right)$ can be found in equation (7.18). Given $j=-1$ and $q=0$, the conditional posterior distribution for $a_\phi(t-1)$ is

$$P\left(a_\phi(t-1)|\{\psi^i(t)\}_{i|k^i=\phi},\{\psi^i(t-1)\}_{i|k^i=\phi},b^0_\phi(t-1),\mu_{**},\sigma^2_{**}\right)=\mathrm{N}(D,E) \tag{7.32}$$

where

$$D=\left(\frac{1}{\sigma^2_{**}}+\frac{2N_\phi}{T_{\mathrm{pt}}}\right)^{-1}\left(\frac{\mu_{**}}{\sigma^2_{**}}+\frac{2\sum_{i|k^i=\phi}\left(\psi^i(t)-b^0_{k^i=\phi}(t-1)\psi^i(t-1)\right)}{T_{\mathrm{pt}}}\right) \tag{7.33}$$

and

$$E=\left(\frac{1}{\sigma^2_{**}}+\frac{2N_\phi}{T_{\mathrm{pt}}(t-1)}\right)^{-1} \tag{7.34}$$

with $\phi\in\{0,1\}$.

The Markov blanket for $\mathbf{b}_{k^i}(t+j)$ is $\{\psi^i(t),\ \psi^i(t+j),\ a_{k^i}(t+j),\ \mu_{**},\ \sigma^2_{**},\ k^i,\ \psi^{-i}(t),\ k^{-i},\ \mu_{k^i}(t),\ \sigma^2_{k^i}(t)\}$. According to equation (7.19), $\mathbf{b}_{k^i}(t+j)$ is independent of $\psi^{-i}(t)$, k^{-i}, $\mu_{k^i}(t)$ and $\sigma^2_{k^i}(t)$. Therefore, its conditional posterior distribution is

$$\begin{aligned} &P\left(\mathbf{b}_\phi(t+j)|\{\psi^i(t),\psi^i(t+j)\}_{i|k^i=\phi},a_\phi(t+j),\mu_{**},\sigma^2_{**},\psi^{-i}(t),k^{-i},\mu_{k^i}(t),\sigma^2_{k^i}(t)\right) \\ &\propto \prod_{i|k^i=\phi} P\left(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\right)P\left(\mathbf{b}_{k^i}(t+j)|\mu_{**},\sigma^2_{**}\right) \end{aligned} \tag{7.35}$$

where $P\left(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\right)$ can be found in equations (7.15) and (7.16). Given $j=-1$ and $q=0$, $P\left(b^0_{k^i}(t+j)|\mu_{**},\sigma^2_{**}\right)$ can be found in equation (7.17). Therefore,

the explicit form of the conditional posterior distribution for $b^0_{k^i=\phi}(t-1)$ is

$$P\left(b^0_\phi(t-1)|\{\psi^i(t)\}_{i|k^i=\phi},\{\psi^i(t-1)\}_{i|k^i=\phi},a_\phi(t-1),\mu_{**},\sigma^2_{**}\right)=\mathrm{N}\left(\frac{B^{'}}{A^{'}},\frac{C^{'}}{A^{'}}\right) \tag{7.36}$$

where

$$A^{'}=T_{\mathrm{pt}}+2\sigma^2_{**}\sum_{i|k^i=\phi}\left(\psi^i(t-1)\right)^2 \tag{7.37}$$

$$B^{'}=\mu_{**}T_{\mathrm{pt}}+2\sigma^2_{**}\sum_{i|k^i=\phi}\psi^i\left(\psi^i(t)-a_{k^i}(t-1)\right) \tag{7.38}$$

and

$$C^{'}=\sigma^2_{**}T_{\mathrm{pt}} \tag{7.39}$$

The Markov blanket for k^i is $\{\psi^i(t),\ \psi^i(t+j),\ a_{k^i}(t+j),\ \mathbf{b}_{k^i}(t+j),\ \mu_{k^i}(t),\ \sigma^2_{k^i}(t),\ k^{-i},\ \psi^{-i}(t),\ \mu_*,\ \sigma^2_*,\ \mu_{**},\ \sigma^2_{**},\ \alpha_*,\ \beta_*\}$. According to equations (7.7), (7.8), (7.17) and (7.18), k^i is independent of parameters $\mu_*,\ \sigma^2_*,\ \mu_{**},\ \sigma^2_{**},\ \alpha_*,\ \beta_*$. In this way, its conditional posterior distribution is

$$\begin{aligned}
&P\left(k^i=\phi|k^{-i},\{\psi^i(t),\mu_\phi(t),\sigma^2_\phi(t),\psi^{-i}(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\}_t\right)\\
\propto&P(k^i|k^{-i})\prod_{t=1}^{M}P\left(\psi^i(t)|\mu_\phi(t),\sigma^2_\phi(t)\right)P\left(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i}\right)\\
&\prod_{t=1}^{M}P\left(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\right)
\end{aligned} \tag{7.40}$$

where $P(k^i|k^{-i})$ can be found in equations (7.9)-(7.11). Meanwhile, $P\left(\psi^i(t)|\mu_\phi(t),\sigma^2_\phi(t)\right)$ can be found in equation (7.5). $P\left(\psi^i(t)|\psi^{-i}(t),k^i,k^{-i}\right)$ can be found in equations (7.12)-(7.14). $P\left(\psi^i(t)|\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\right)$ can be found in equations (7.15) and (7.16). Since k^i can only be either 0 or 1, we set it to be a Bernoulli distribution conditional on its Markov blanket. Given $j=-1$ and $q=0$, when $k^i=0$, we have

$$\begin{aligned}
q\propto&P\left(k^i=0|k^{-i},\{\psi^i(t),\mu_{k^i}(t),\sigma^2_{k^i}(t),\psi^{-i}(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\}_t\right)\\
\propto&\exp\left(-\frac{1}{T}\sum_{i\sim m}u\left(k^i|k^m\right)\right)\exp\left(-\sum_{t=1}^{M}\frac{\left(\psi^i(t)-\mu_{k^i}(t)\right)^2}{2\sigma^2_{k^i}(t)}\right)\\
&\exp\left(-\sum_{t=1}^{M}\frac{1}{T_{\mathrm{ps}}}\sum_{i\sim m}v_{i,m}\left(\psi^i(t)-\psi^m(t)\right)^2\right)\\
&\exp\left(-\sum_{t=1}^{M}\frac{1}{T_{\mathrm{pt}}}\left(\psi^i(t)-b^0_{k^i}(t-1)\psi^i(t-1)-a_{k^i}(t-1)\right)^2\right).
\end{aligned} \tag{7.41}$$

Similarly, given $j=-1$ and $q=0$, when $k^i=1$, we have

$$\begin{aligned} p \propto & P\left(k^i=1|k^{-i},\{\psi^i(t),\mu_{k^i}(t),\sigma^2_{k^i}(t),\psi^{-i}(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\}_t\right) \\ \propto & \exp\left(-\frac{1}{T}\sum_{i\sim m}u\left(k^i|k^m\right)\right)\exp\left(-\sum_{t=1}^{M}\frac{\left(\psi^i(t)-\mu_{k^i}(t)\right)^2}{2\sigma^2_{k^i}(t)}\right) \\ & \exp\left(-\sum_{t=1}^{M}\frac{1}{T_{\mathrm{ps}}}\sum_{i\sim m}v_{i,m}\left(\psi^i(t)-\psi^m(t)\right)^2\right) \\ & \exp\left(-\sum_{t=1}^{M}\frac{1}{T_{\mathrm{pt}}}\left(\psi^i(t)-b^0_{k^i}(t-1)\psi^i(t-1)-a_{k^i}(t-1)\right)^2\right). \end{aligned} \tag{7.42}$$

Therefore, the explicit conditional posterior distribution for k^i is

$$\begin{aligned} & P(k^i|k^{-i},\{\psi^i(t),\mu_{k^i}(t),\sigma^2_{k^i}(t),\psi^{-i}(t),\psi^i(t+j),a_{k^i}(t+j),\mathbf{b}_{k^i}(t+j)\}_t) = \\ & \left(\frac{p}{p+q}\right)^{k^i}\left(\frac{q}{p+q}\right)^{1-k^i}. \end{aligned} \tag{7.43}$$

The conditional posterior distributions of $\psi^i(t)$, $\mu_{k^i}(t)$, $\sigma^2_{k^i}(t)$, σ^2, $a_{k^i}(t+j)$, $\mathbf{b}_{k^i}(t+j)$ and k^i has been derived in this section, and these conditional posterior distributions will be used for MCMC simulations.

7.3.5 MCMC simulations

In the present work, we applied the proposed HBM method to both clinical and synthetic data. In practice, the samples of the HBM parameters can be approximately drawn using an MCMC method, the Gibbs sampling method, from their posterior distributions. Specifically, we have already derived the conditional posterior distributions of all HBM parameters in Section 7.3.4, which is used in the Gibbs sampling scheme described in detail in Algorithm 6.

7.4 Benchmark methods

Similar to Chapter 6, we used two alternative methods as the benchmark methods in this work. The first benchmark method is the Gaussian mixture model classification method (see Section 5.2). As mentioned in Section 5.2, this classification method is susceptible to spurious singleton clusters. Therefore, we further used an image processing approach called the "opening and closing operations" [163], as another benchmark method. The opening and closing operations aim to improve the GMM classification results by reducing spurious singleton clusters. This approach applies mathematical morphology (erosion and dilation) to reduce the occurrence of spurious clusters. Both closing and opening operations are derived from the fundamental operations called "erosion" and "dilation". In general, the opening operation can remove small

Algorithm 6: Gibbs sampling for myocardial perfusion DCE-MRI classification

Input: Hyperparameters α_*, β_*, α_{**}, β_{**}, μ_*, σ_*^2, μ_{**}, σ_{**}^2, T, T_{ps}, T_{pt}
Output: $\{\psi^i(t)\}_{i,t}$, $\{\mu_{k^i}(t)\}_t$, $\{\sigma_{k^i}^2(t)\}_t$, σ^2, $\{a_{k^i}(t+j)\}_t$, $\{\mathbf{b}_{k^i}(t+j)\}_t$ and $\{k^i\}_i$
Data: $\{y^i(t)\}_{i,t}$ where $i = 1,2,...,N$ and $t = 1,2,...,M$
Set the initial values of $\psi^i(t)$ to be $y^i(t)$; set the initial value of σ^2 to be 1;
for $i \leftarrow 1$ **to** N **do**
  Based on image at t_{m} (the time that the signals achieve to peak values), obtain initial values of k^i, using the EM algorithm (Gaussian mixture model) (see details from section 9 in [18]);
end
for $t \leftarrow 1$ **to** M **do**
  Given k^i and $\psi^i(t)$, obtain the initial values of $\mu_{k^i}(t)$ and $\sigma_{k^i}^2(t)$ based on maximum likelihood estimation (MLE);
  Given k^i and $\psi^i(t)$, obtain the initial values of $a_{k^i}(t+j)$ and $\mathbf{b}_{k^i}(t+j)$ using least-squares fitting;
end
for $p \leftarrow 1$ **to** M_0 *(The number of MCMC samples)* **do**
  for $t \leftarrow 1$ **to** M **do**
    for $i \leftarrow 1$ **to** N **do**
      Given $y^i(t)$, $\psi^i(t+j)$, $a_{k^i}(t+j)$, $\mathbf{b}_{k^i}(t+j)$, $\mu_{k^i}(t)$, $\sigma_{k^i}^2(t)$, k^i, k^{-i}, $\psi^{-i}(t)$, T_{ps}, T_{pt} and σ^2, draw $\psi^i(t)$ based on equation (7.24);
    end
    Given $\{\psi^i(t)\}_i$, $\sigma_{k^i}^2(t)$, $\{k^i\}_i$, μ_* and σ_*^2, draw $\mu_{k^i}(t)$ based on equation (7.26);
    Given $\{\psi^i(t)\}_i$, $\mu_{k^i}(t)$, $\{k^i\}_i$, α_* and β_*, draw $\sigma_{k^i}^2(t)$ based on equation (7.28);
    Given $\{\psi^i(t)\}_i$, $\{\psi^i(t+j)\}_i$, $\mathbf{b}_{k^i}(t+j)$, T_{pt}, k^i, μ_{**} and σ_{**}^2, draw $a_{k^i}(t+j)$ based on equation (7.32);
    Given $\{\psi^i(t)\}_i$, $\{\psi^i(t+j)\}_i$, $a_{k^i}(t+j)$, T_{pt}, k^i, μ_{**} and σ_{**}^2, draw $\mathbf{b}_{k^i}(t+j)$ based on equation (7.36);
  end
  for $i \leftarrow 1$ **to** N **do**
    Given $\{\psi^i(t)\}_t$, $\{\mu_{k^i}(t)\}_t$, $\{\sigma_{k^i}^2(t)\}_t$, k^{-i}, $\{\psi^{-i}(t)\}_t$, $\{\psi^i(t+j)\}_t$, $\{a_{k^i}(t+j)\}_t$, $\{\mathbf{b}_{k^i}(t+j)\}_t$, T, T_{ps} and T_{pt}, draw k^i based on equation (7.43);
  end
  Given $\{y^i(t)\}_{i,t}$, $\{\psi^i(t)\}_{i,t}$, α_{**} and β_{**}, draw σ^2 based on equation (7.30);
end

Table 7.2: Descriptions of the benchmark methods

Abbreviation of methods	Description
GMM	Gaussian mixture model classification
GMM_{C}	GMM based closing operation with kernel size 2×2
$\text{GMM}_{\text{C\&O}}$	GMM based closing and opening operation with kernel size 2×2
$\text{GMM}^{\text{max}}_{\text{C\&O}}$	the result of $\text{GMM}_{\text{C\&O}}$ which keeps the largest lesion

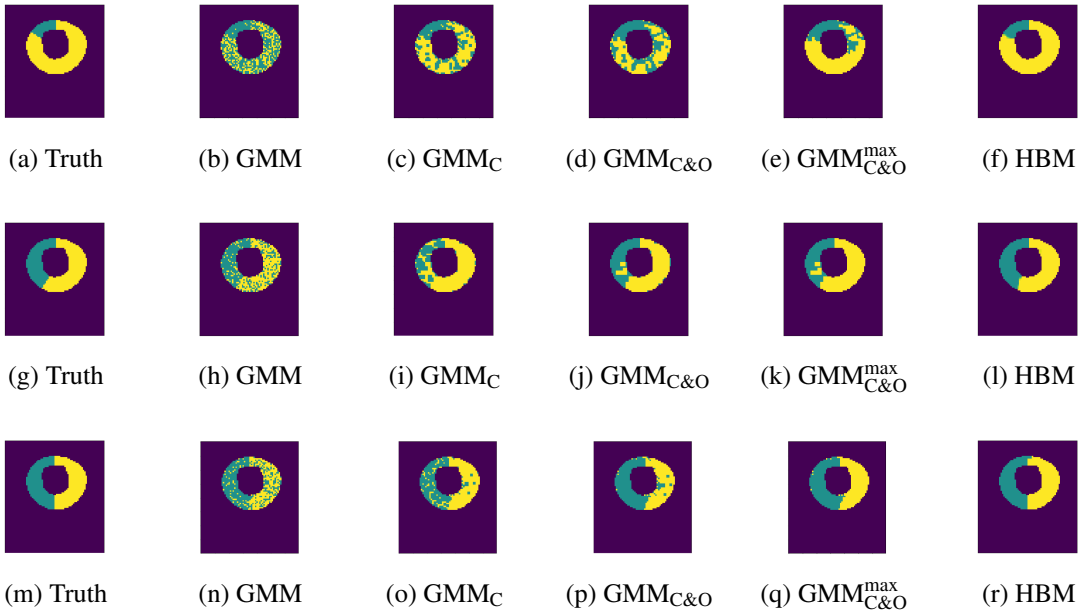
lesion clusters inside healthy tissues and the closing operation can remove small healthy clusters inside lesions. Since the GMM with opening and closing operations ($\text{GMM}_{\text{C\&O}}$) usually cannot remove spurious medium-sized clusters, we also applied an approach where we only kept the largest lesion region as in [171], named as "$\text{GMM}^{\text{max}}_{\text{C\&O}}$" (see Table 7.2). We found that when noise is strong, $\text{GMM}^{\text{max}}_{\text{C\&O}}$ is more successful than $\text{GMM}_{\text{C\&O}}$ for removing spurious clusters. The description of these benchmark methods can be found in Table 7.2. The kernel size for the opening and closing operators was chosen as 2×2 in the present work for the reason that it can produce the closest classification result to the ground truth. In Figure 7.6, it can be observed that the performances with kernel size 2×2 is better than the performances with kernel size 3×3 (less misclassified pixels).. Thus, we have used the best (the closest to the ground truth) alternative approach to compare with the proposed method, thereby giving the alternative schemes a competitive advantage over our method.

7.5 Results

7.5.1 Results for synthetic data

Visual inspection

Figure 7.4 shows the classification results for the synthetic examples based on different methods. The descriptions of the benchmark methods can be found in Table 7.2. We show the case that the noise is very high (noise variance 4^2). According to Figure 7.4, none of the methods except for the method proposed in the present work manage to reproduce the ground truth. The GMM classification method predicts many spurious small clusters. "GMM_{C}", "$\text{GMM}_{\text{C\&O}}$", "$\text{GMM}^{\text{max}}_{\text{C\&O}}$" (see Table 7.2) reduce the noise to some degree. However, all these methods either misclassify some lesion pixels in the healthy region, or misclassify some healthy pixels in the lesion region. Generally, in all benchmark methods, the method "$\text{GMM}^{\text{max}}_{\text{C\&O}}$" achieves the closest results to the ground truth. However, it still misclassifies some healthy pixels in the lesion region in Figure 7.4(k) and misclassifies some lesion pixels in the healthy region in Figure 7.4(e). Moreover, in Figure 7.4(q), the boundaries between the healthy region and the lesion region are misclassi-



(a) Truth (b) GMM (c) GMM_{C} (d) $\text{GMM}_{\text{C\&O}}$ (e) $\text{GMM}^{\text{max}}_{\text{C\&O}}$ (f) HBM

(g) Truth (h) GMM (i) GMM_{C} (j) $\text{GMM}_{\text{C\&O}}$ (k) $\text{GMM}^{\text{max}}_{\text{C\&O}}$ (l) HBM

(m) Truth (n) GMM (o) GMM_{C} (p) $\text{GMM}_{\text{C\&O}}$ (q) $\text{GMM}^{\text{max}}_{\text{C\&O}}$ (r) HBM

Figure 7.4: Lesion delineation for the synthetic data. These figures show lesion delineations for the synthetic examples with additive i.i.d. Rician noise with signal-to-noise (SNR) 1.4 (noise variance 4^2). The pixels inside and outside of the ring (myocardium) are background. The three rows denote the classification results for 60°, 120° and 180° respectively. Panels (a), (g) and (m) show the ground truth. Panels (b) - (e), (h) - (k) and (n) - (q) show the lesion delineation using alternative methods, which are explicitly explained in Section 7.4 and Table 7.2. Panels (f), (l) and (r) show the lesion delineation using the proposed HBM.

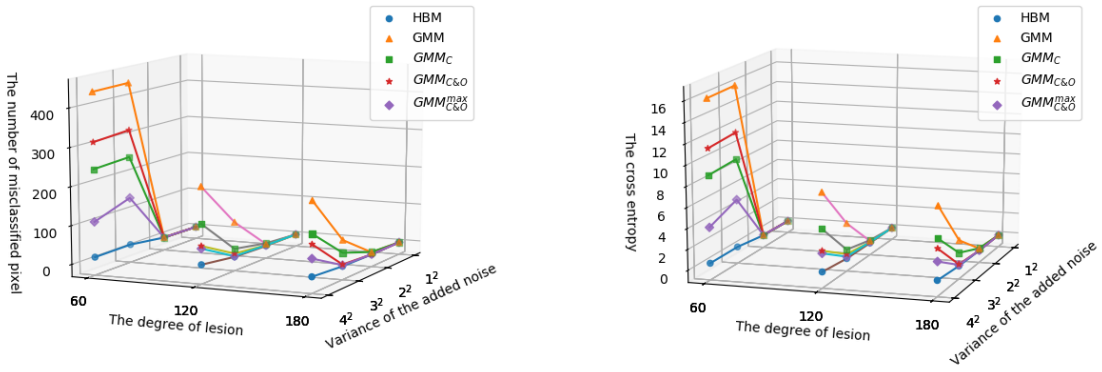
fied. On the other hand, the HBM method proposed in the present work successfully reproduces the ground truth with only very few misclassified pixels located in the boundary between healthy tissues and the lesion for all three lesion size cases. Therefore, a visual inspection clearly reveals that the results generated with the proposed HBM method are the closest to the ground truth and outperform all alternative benchmark methods included in our comparative evaluation study.

Prediction accuracy

To quantify the classification accuracy, we first counted the number of misclassified pixels, and then calculated the cross-entropy between the ground truth labels and the predicted posterior probabilities (see equation (7.44)).

$$H(p_i, q_i) = -\sum_i [p_i \log(q_i) + (1 - p_i) \log(1 - q_i)]. \tag{7.44}$$

In equation (7.44), the subscript i indicates the pixel index, $p_i \in \{0, 1\}$ indicates the true label given the ground truth and q_i indicates the predicted posterior probabilities. Similar to the work in Chapter 6, the reason we used the cross-entropy as an alternative quantitative criterion, in addition to the misclassification rate, is to assess if the methods can properly deal with uncer-



(a) Misclassified numbers

(b) Cross-entropy

Figure 7.5: 3D plots for the numbers of misclassified pixels and cross-entropy. These figures show the 3D plots for the numbers of misclassified pixels (in (a)) and cross-entropy (in (b)). They are the visualizations of Tables 7.3 and 7.4. Lines are used to link the points with different noise variances to show the trends of their changes.

tainty quantification, i.e. correctly predict the uncertainty of their predictions. This shifts the focus from the classification label to the posterior probability of finding a pixel in a given class. For example, predicting the posterior probability of a pixel to be lesion to be 0.51 indicates a considerably lower level of confidence than predicting a posterior probability of, say, 0.99. This difference is duly captured by the cross-entropy, but not the misclassification rate, which would not distinguish between the above two scenarios.

Tables 7.3, 7.4 and Figure 7.5 show the average misclassified pixels and cross-entropy based on different methods for the synthetic examples. When the noise is very low (noise variance 1^2), all methods reproduce the ground truth. For other degrees of noise, the performance for the proposed HBM method is consistently the best and outperforms all competing methods. A standard t-test confirms that this difference, and hence the performance improvement achieved with the proposed HBM, is significant. It is worth mentioning that the results for 120° and for 180° are similar. For these two sizes, when the noise is low (noise variance 2^2), the methods "$\text{GMM}^{\text{max}}_{\text{C\&O}}$" and "$\text{GMM}_{\text{C\&O}}$", (see Table 7.2), have slightly lower prediction accuracies than the proposed HBM method. However, when the noise is getting higher, their prediction accuracies are getting lower. This means that the methods "$\text{GMM}^{\text{max}}_{\text{C\&O}}$" and "$\text{GMM}_{\text{C\&O}}$" are very sensitive to the degree of the noise. For the 60°, all methods perform similarly when the noise is low (noise variance 2^2). However, the prediction accuracies for all competing methods decrease dramatically when the noise is high (noise variance 3^2) and very high (noise variance 4^2). The prediction accuracy for the proposed HBM method is much higher than for all competing methods.

Note that all methods included in our comparative evaluation depend on various regularisation parameters. As we will demonstrate below, one of the advantages of the proposed HBM

Table 7.3: The average number of misclassified pixels for the synthetic data

60°	VAR/SNR (Rician noise)			
	1^2/22.9	2^2/5.7	3^2/2.5	4^2/1.4
GMM	0.0	2.1	440.0	438.1
GMM_{C}	0.0	2.0	246.9	241.6
$\text{GMM}_{\text{C\&O}}$	0.0	2.3	317.4	310.6
$\text{GMM}^{\text{max}}_{\text{C\&O}}$	0.0	2.3	139.2	106.3
HBM	0.0	2.0	15.0	14.7
120°	VAR/SNR (Rician noise)			
	1^2/22.9	2^2/5.7	3^2/2.5	4^2/1.4
GMM	0.0	5.9	95.4	216.3
GMM_{C}	0.0	7.9	25.6	122.0
$\text{GMM}_{\text{C\&O}}$	0.0	2.6	13.3	65.7
$\text{GMM}^{\text{max}}_{\text{C\&O}}$	0.0	2.6	6.5	58.9
HBM	0.0	0.7	4.6	17.8
180°	VAR/SNR (Rician noise)			
	1^2/22.9	2^2/5.7	3^2/2.5	4^2/1.4
GMM	0.0	5.8	70.3	200.7
GMM_{C}	0.0	6.6	37.6	118.0
$\text{GMM}_{\text{C\&O}}$	0.0	0.3	7.4	92.0
$\text{GMM}^{\text{max}}_{\text{C\&O}}$	0.0	0.3	5.8	56.1
HBM	0.0	0.0	2.8	10.4

method is that these regularisation parameters can be set objectively, based on the available data and established statistical information criteria. This is a substantial advantage over the alternative benchmark methods included in our comparative evaluation, where these regularisation parameters are “user-defined," meaning that they have to be set by the user based on intuition, experience, and less objective criteria. Specifically, the order and kernel size for the opening and closing operations affect the classification results significantly. We test the performances of the opening and closing operations with different orders and kernel sizes (see Tables 7.5) using the synthetic data at different noise scales. Figure 7.6 and Table 7.5 show the classification results for different orders and kernel sizes of the closing and opening operations. The classification results for them are quite different. Without the input of extra information (the ground truth), the order and kernel size for the opening and closing operations are difficult to choose, which would be methodologically inconsistent. The values of the hyperparameters also affect the classification results for the proposed HBM method. However, as mentioned above, these values can be selected based on established information criteria, e.g. Watanabe Akaike information criterion (WAIC). This selection process is thus less subjective and more reproducible than for the alternative methods, and less reliant on the user's experience and intuition.

Table 7.4: Cross-entropy between true labels and predicted posterior probabilities for the synthetic data (smaller values mean better results)

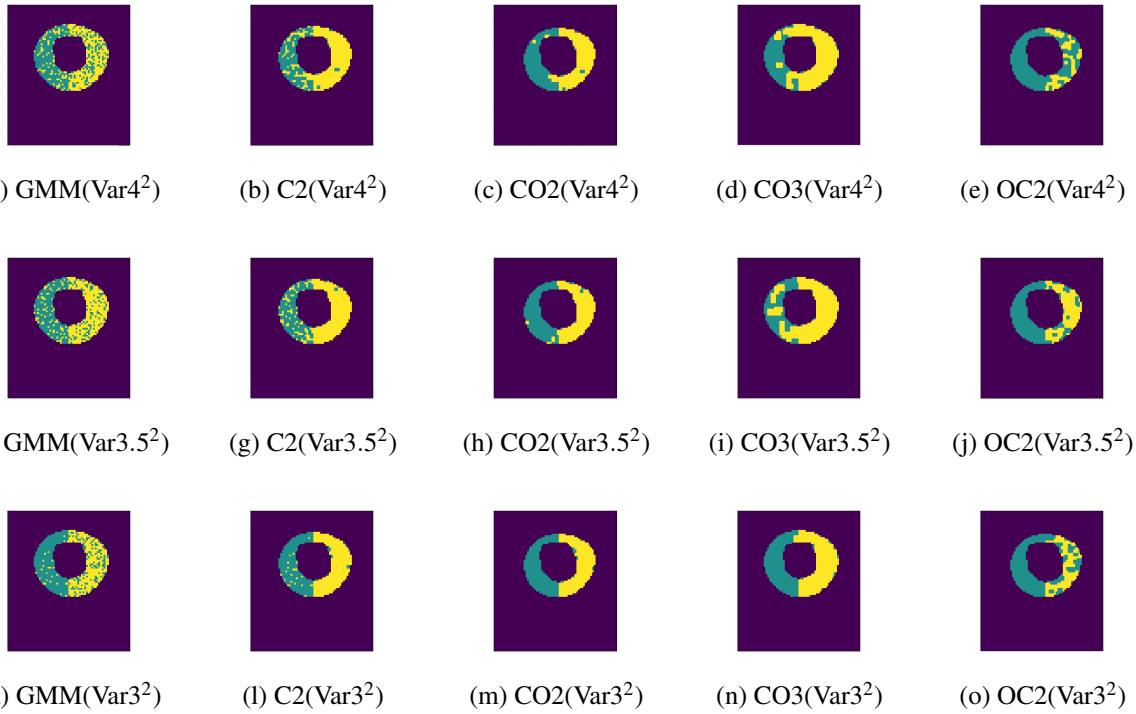
60°	VAR/SNR (Rician noise)			
	1^2/22.9	2^2/5.7	3^2/2.5	4^2/1.4
GMM	0.00	0.07	16.15	16.08
GMM_{C}	0.00	0.05	9.06	8.87
$\text{GMM}_{\text{C\&O}}$	0.00	0.08	11.65	11.40
$\text{GMM}_{\text{C\&O}}^{\max}$	0.00	0.08	5.11	3.90
HBM	0.00	0.05	0.44	0.44
120°	VAR/SNR (Rician noise)			
	1^2/22.9	2^2/5.7	3^2/2.5	4^2/1.4
GMM	0.00	0.22	3.50	7.94
GMM_{C}	0.00	0.29	0.93	4.48
$\text{GMM}_{\text{C\&O}}$	0.00	0.10	0.49	2.41
$\text{GMM}_{\text{C\&O}}^{\max}$	0.00	0.10	0.24	2.16
HBM	0.00	0.00	0.05	0.39
180°	VAR/SNR (Rician noise)			
	1^2/22.9	2^2/5.7	3^2/2.5	4^2/1.4
GMM	0.00	0.21	2.58	7.37
GMM_{C}	0.00	0.24	1.38	4.33
$\text{GMM}_{\text{C\&O}}$	0.00	0.01	0.27	3.38
$\text{GMM}_{\text{C\&O}}^{\max}$	0.00	0.01	0.21	2.06
HBM	0.00	0.00	0.10	0.38

Table 7.5: The descriptions of opening and closing operations with different orders and kernel sizes in Figure 7.6

Abbreviation in Figure 7.6	Description
C2	only closing with kernel 2×2
CO2	closing with kernel 2×2 and opening with kernel 2×2
CO3	closing with kernel 3×3 and opening with kernel 3×3
OC2	opening with kernel 2×2 and closing with kernel 2×2

Denoised image generations

We finally focus on the challenges of image noise reduction and filtering. An additional benefit of the proposed HBM method is that it can, to some extent, filter out the noise and generate denoised images from their original noisy counterparts. Figure 7.7 shows the ME SI images for the synthetic data. Since the added noise to the ground truth is very high, the lesion pattern is very hard to distinguish based on the noise-corrupted images. After applying the proposed HBM method in the present work, the noise is visually reduced. For the generated images, the lesion pattern is distinguishable and very close to the ground truth.



(a) GMM(Var4^2) (b) C2(Var4^2) (c) CO2(Var4^2) (d) CO3(Var4^2) (e) OC2(Var4^2)

(f) GMM(Var3.5^2) (g) C2(Var3.5^2) (h) CO2(Var3.5^2) (i) CO3(Var3.5^2) (j) OC2(Var3.5^2)

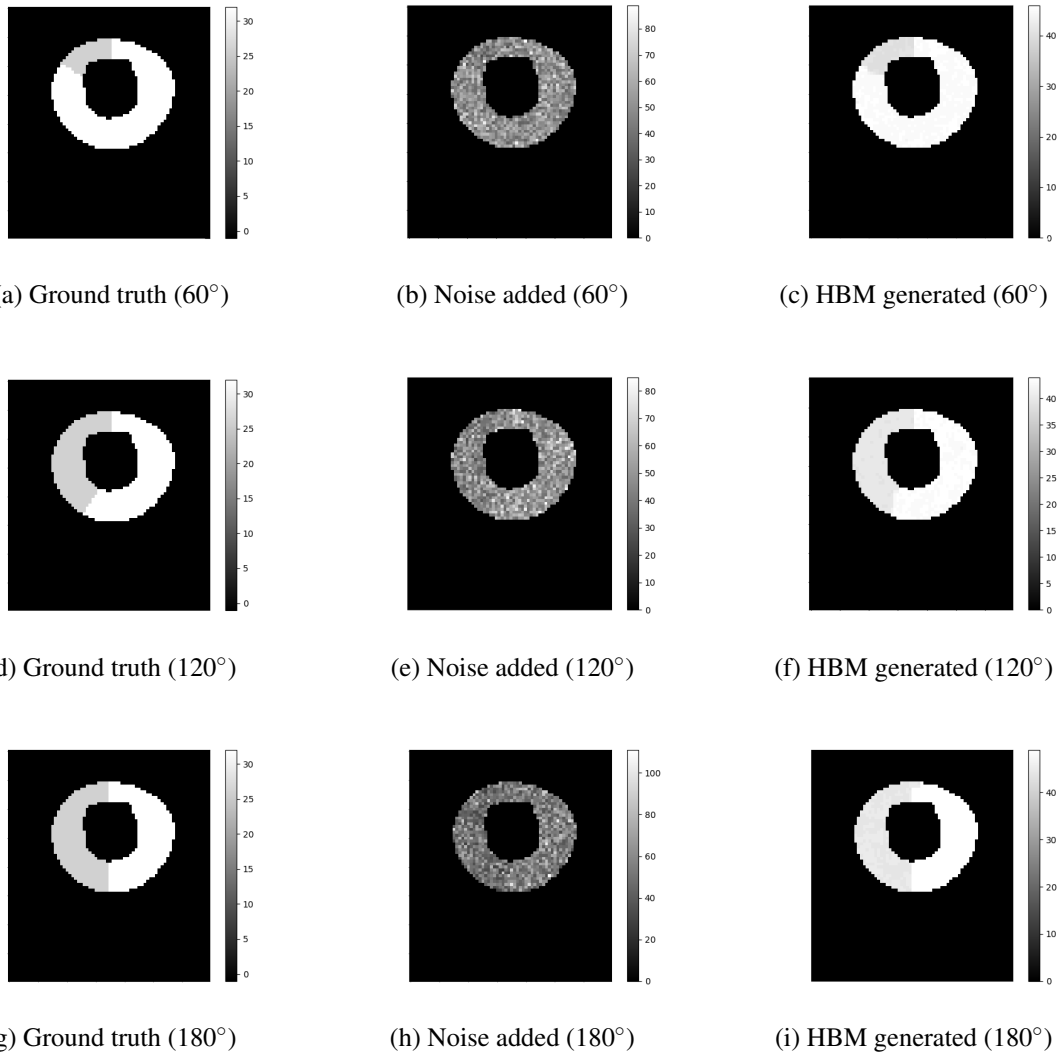
(k) GMM(Var3^2) (l) C2(Var3^2) (m) CO2(Var3^2) (n) CO3(Var3^2) (o) OC2(Var3^2)

Figure 7.6: Closing and opening operations segmentation results for different noise scales. These figures show the classification results by GMM based closing and opening operations for different noise scales. The first, second and third row show the results corresponding to noise variance 4^2, 3.5^2 and 3^2. Panels (a), (f) and (k) show the GMM classification results. The descriptions for the other panel names can be found in Table 7.5. These results show the inconsistency of the closing and opening operations.

7.5.2 Results for clinical data

Model selection

In the proposed HBM method, there are three user-defined hyperparameters T, T_{ps} and T_{pt}, which can be found in equations (7.10), (7.13) and (7.16). Firstly, we used WAIC values to explore the best combination of the hyperparameters. In principle, they could be sampled from the posterior distribution with MCMC techniques, along with the other parameters. However, this would lead to a substantial increase in the computational complexity, given that convergence and mixing of hyperparameters tends to be happening at much lower rates than for parameters; see e.g. [164]. For clinical decision making, excessive computational costs need to be avoided, and we also want to better exploit parallel computing resources for computational cost reduction. This can be achieved by computing advanced information criteria, like WAIC, for a set of candidate values in parallel, using high-performance computer clusters, and then selecting the results corresponding to those hyperparameters that have obtained the lowest WAIC score. In our work, we have processed Algorithm 6 separately for different combinations of the hyperparameters on different processors. Table 7.6 shows the WAIC values for different combinations of the hyperparameters.



(a) Ground truth (60°) (b) Noise added (60°) (c) HBM generated (60°)

(d) Ground truth (120°) (e) Noise added (120°) (f) HBM generated (120°)

(g) Ground truth (180°) (h) Noise added (180°) (i) HBM generated (180°)

Figure 7.7: Maximum enhanced images. These figures show the SI images at the ME time for the synthetic data. Three rows show the images for 60°, 120° and 180° respectively. Panels (a), (d) and (g) show the ground truth ME images. Panels (b), (e) and (h) show the noisy images after adding Rician noise. Panels (c), (f) and (i) show the reconstructed noise-reduced images using the proposed HBM method.

Table 7.6: Watanabe-Akaike information criterion (WAIC)

T = 0.01	T_{pt}			
T_{ps}		1	10	100
	0.1	79113	12665	10009
	1	83785	45537	38527
	10	91290	98814	97253
T = 0.1	T_{pt}			
T_{ps}		1	10	100
	0.1	78886	12859	**9675**
	1	84172	45588	38348
	10	91637	98895	97444
T = 1	T_{pt}			
T_{ps}		1	10	100
	0.1	79277	13615	9990
	1	84004	45778	38321
	10	91879	98732	97750

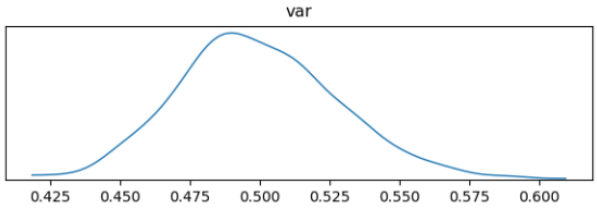
We find that the combination $T = 0.1$, $T_{ps} = 0.1$ and $T_{pt} = 100$ achieves the lowest (i.e. best) WAIC value, and these hyperparameter settings were subsequently used for all our simulations (including the synthetic data).

Convergence test

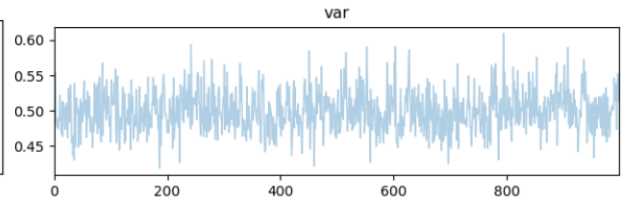
All parameters were sampled from their posterior distributions using Gibbs sampling according to Algorithm 6. Figure 7.8 inspects the convergence of the parameter σ^2. The reason we chose this parameter to assess the convergence of the Gibbs sampling is that it can reveal the uncertainty of differences between observed and generated signals. The traceplot and kernel density plot for the MCMC samples look reasonable, without any signs of insufficient convergence. We also use Geweke's test [24] to test for convergence. If the Geweke scores consistently stay within the range (-1,1), there is no significant evidence for insufficient convergence. Therefore, given Figure 7.8, there is no significant evidence that the parameter σ^2 has not converged. We also repeat the convergence explorations for other parameters ($\psi^i(t)$, $\mu_{k^i}(t)$, ...). There is no significant evidence for lack of convergence here either.

Classification results and denoised image generations

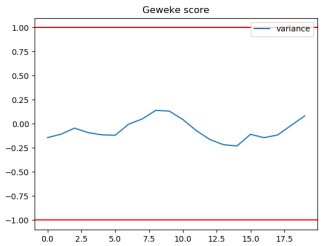
Figure 7.9 shows the classification results for the clinical data based on three methods, GMM, "$\text{GMM}_{\text{C\&O}}$" (see Table 7.2) and the proposed HBM. There are some small and singleton healthy clusters located inside of the myocardium lesion for the GMM result. As mentioned in Section 5.2, this phenomenon is physiologically unrealistic. Small and singleton healthy clusters are reduced but not removed by the method "$\text{GMM}_{\text{C\&O}}$" (see Table 7.2). However, these small



(a) Kernal density of the posterior distribution



(b) Traceplot of the posterior distribution



(c) Geweke test

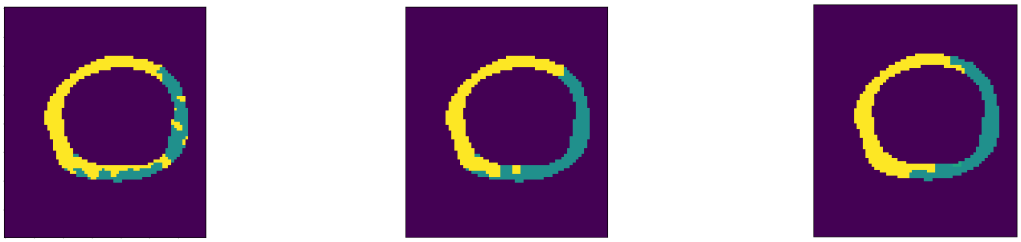
Figure 7.8: MCMC convergence. These figures inspect the convergence of the MCMC simulation for varaince σ^2. (a) shows the kernel density plot of the MCMC samples for parameter σ^2 after burn-in. (b) shows the traceplot of the MCMC samples for parameter σ^2 after burn-in. (c) shows Geweke scores based on the MCMC samples of σ^2. The red horizontal lines show the value 1 and -1. If the Geweke scores are within the range $(-1, 1)$, there is no significant evidence that this parameter has not converged.

and singleton lesion clusters are successfully removed by the proposed HBM method after introducing the spatio-temporal context information (details in Figure 7.9(c)). This means that the proposed HBM method improves the classification results.

Figure 7.10 shows the comparison between original and generated images. The noise of the original image is reduced significantly after applying the method proposed in this work. Moreover, the boundary between the hypo-perfused area and the normal-perfused area is clearer as a result of the edge-preservation property mentioned in Section 7.3.3. These results indicate that the proposed HBM method successfully reduces noise from the original images and hence, generates the noise reduced images.

Figure 7.11 shows the results for two extra sets of data. As mentioned in Section 7.2.1, only classification maps and generated denoised images are shown for these two sets of data. These results are used to emphasize the validation of the proposed HBM method. The unrealistic small and singleton clusters cannot be avoided by the benchmark methods, GMM and "$\text{GMM}_{\text{C\&O}}$" (see Table 7.2). However, the unrealistic small and singleton clusters are removed by the proposed HBM method. Thus, the proposed HBM method shows robustness based on the results using different sets of data.

The comparisons among all methods in the thesis have been carried out. However, since different chapters and sections are submitted as scientific papers, the comparisons are not been presented in each chapter/section. Therefore, the overall comparisons are presented at the end



(a) GMM (b) $GMM_{C\&O}$ (c) HBM

Figure 7.9: Classifications of clinical data. These figures show the classifications of the clinical data by three methods. The black areas inside and outside of the myocardium ring indicate the background. The yellow pixels indicate the healthy tissue and the dark green pixels indicate the lesions. Panel (a) shows the classification given by GMM. Panel (b) shows the classification given by "$GMM_{C\&O}$" (see Table 7.2). Panel (c) shows the classification given by the proposed HBM method.

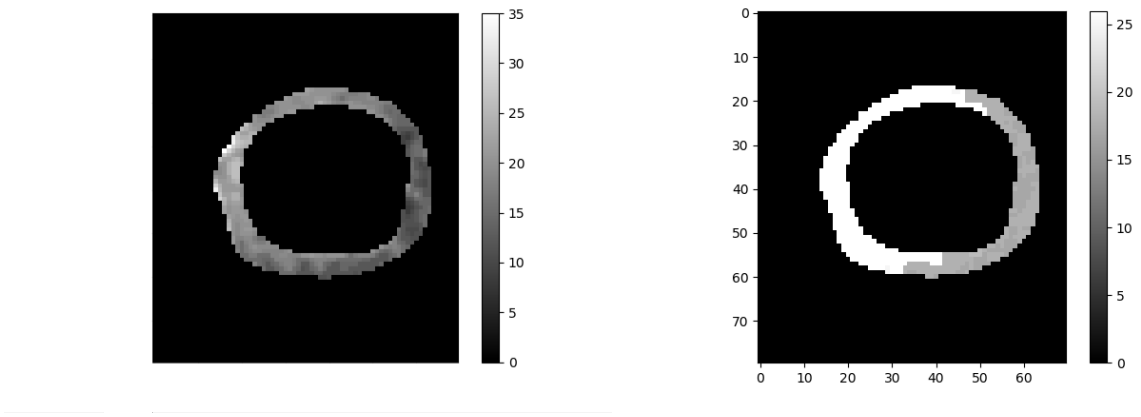
Table 7.7: The comparisons of different methods

	GMM	SVFMM	GMM-MRF	Fermi HBM	Data HBM
misclassified rate at SNR = 4	11%	6%	5%	1%	0%
misclassified rate at SNR = 3	32%	24%	13%	2%	1%
Computation time	1 second	10 minutes	1 minute	2 hours	1 hour

of this chapter to give the reader an overview of these methods. The specific comparisons can be found in Table 7.7. The GMM is very fast because it is applied by build-in package and other methods are applied by the codes writen by the author. In general, the methods with higher accuracy are computationally slow, and the methods with lower accuracy are computationally faster. Different methods provide the information of the balance between computational efficiency and accuracy.

7.6 Discussion and conclusion

In this chapter, we have proposed a classification approach based on a hierarchical Bayesian model (HBM) with a spatio-temporal Markov random field (MRF) prior for automatic lesion detection in myocardial perfusion DCE-MRI scans. This method is a fully data driven approach, meaning that all parameters and hyperparameters can be consistently inferred from the data, without need for any heuristic user-defined tuning parameters. Two alternative established methods, the Gaussian mixture model (GMM) and the Gaussian mixture model with opening and closing operations, have been applied for comparison, using both synthetic and clinical data. For the former, the ground truth is known. This allowed us to compute two objective quantitative criteria, the number of misclassified pixels and the cross-entropy between true and predicted labels. Both quantitative criteria demonstrate that the proposed HBM clearly outper-



(a) Original image

(b) Generated image

Figure 7.10: Signal comparison between original image and generated image. These figures show the comparison between original image and generated image. Panel (a) shows the original signal intensity map of the maximum enhanced image. Panel (b) shows the generated signal intensity map of the maximum enhanced image.

forms the established alternative methods used for comparison, which we further corroborated with a standard statistical hypothesis test.

For the clinical data, we evaluated the performance of the methods based on the physiological insight that the blockage of a blood vessel affects extended connected regions of the myocardial tissue, and that large numbers of tiny unconnected lesions are unrealistic. We therefore used the number of spurious singleton clusters as a performance criterion, with larger numbers being in increasing disagreement with our physiological insight. This criterion can be visually observed from the classification results and we found that the proposed HBM clearly outperformed the alternative established methods according to this performance measure, systematically avoiding singleton clusters and leading to predictions that are physiologically more plausible.

The proposed HBM was also found to succeed in generating denoised magnetic resonance (MR) images. For the synthetic data, where the ground truth is known, the generated denoised MR images tend to be much closer to the ground truth than the original noisy MR images. While the lesion patterns for the original noisy MR images are hard to distinguish visually, the lesion patterns for the generated MR images are clearly distinguishable and the boundaries between lesions and healthy tissues are easy to discern. For the clinical data, too, the generated MR images show clearer evidence of noise reduction and crisper boundaries between lesions and healthy tissues.

For inference, we have sampled the parameters from the posterior distribution with Gibbs sampling, while the lower level hyperparameters were selected based on an advanced information criterion (WAIC), which can be directly computed from the Gibbs samples. The motivation for this combined approach over a full sampling scheme is faster mixing and convergence and the facilitation of the exploitation of parallel computer clusters (see further below).

(a) Original image (b) Generated image (c) GMM (d) $\text{GMM}_{\text{C\&O}}$ (e) Proposed HBM

(f) Original image (g) Generated image (h) GMM (i) $\text{GMM}_{\text{C\&O}}$ (j) Proposed HBM

Figure 7.11: Classification results and generated denoised images for two more sets of data. These figures show classification and generation results for two different data sets. Panels (a) - (e) show the maximum enhanced original image, generated maximum enhanced image, GMM classification, "$\text{GMM}_{\text{C\&O}}$" (see Table 7.2) and proposed HBM classification for case 1 respectively. Panels (f) - (j) show the maximum enhanced original image, generated maximum enhanced image, GMM classification, "$\text{GMM}_{\text{C\&O}}$" (see Table 7.2) and proposed HBM classification for case 2 respectively.

In summary, there are four main advantages of the proposed HBM over the alternative established methods. Firstly, the prediction accuracy of the proposed HBM is significantly better than for the alternative established methods, based on the two performance criteria of classification error and cross-entropy. Secondly, as opposed to the alternative established methods – GMM and GMM with closing and opening operations – the proposed HBM succeeds in generating denoised images. Thirdly, the proposed HBM is more robust and consistent than the competing alternative methods, in that it does not depend on any subjective user-defined tuning parameters; rather, all parameters and hyperparameters are inferred from the data. Fourthly, our Bayesian inference scheme for the proposed HBM automatically provides uncertainty quantification for both generated signals and class labels. This is a substantial advantage over the two alternative established classification methods (GMM and GMM with closing and opening operations), which are optimisation-based methods without any uncertainty quantification.

The main disadvantage of the proposed HBM is its higher computational complexity. The two methods used for comparison are computationally cheap: it typically takes less than a minute to obtain the results. The Gibbs sampling simulations for our HBM, on the other hand, are computationally expensive. We typically required about 3000 Gibbs sampling steps to reach an acceptable level of convergence, based on the Geweke convergence test, which was equivalent to a computational costs of about one hour on the hardware we were using (Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz processor with 64GB memory).

To reduce the computational complexity, we can draw on parallelisation: both the pixel-level computations on which the Gibbs sampling scheme is based and the WAIC computations for

hyperparameter selection can easily be run in parallel. While for the present proof-of-concept study this parallelisation has not been implemented yet, it is conceptually straightforward and can be expected to lead to a substantial reduction in the computational costs.

The proposed method is aimed at addressing one of the two essential components of myocardial perfusion assessment, namely identification of the lesion's spatial extent (i.e. its presence, location and size). To address the remaining component (lesion severity), absolute MBF values also need to be extracted from raw DCE-MRI datasets. Without any loss of generality, the form of our method could be applied to calibrated DCE-MRI data, and thus enable comprehensive and robust assessment of myocardial perfusion in the future.

Chapter 8

Polar Projection of Myocardial Perfusion DCE-MRI

8.1 Introduction

The longitudinal analysis of myocardial perfusion derived from the DCE-MRI is important because the nature of changes of myocardial microcirculation after infarction is of crucial importance clinically, as this has direct impact on morbidity and mortality. The AHA mapping (see details in Section 3.6) provides a way to apply regional longitudinal analysis because myocardial wall can be segmented into 17 regions, or 6 regions within a single mid-ventricular slice. The myocardial perfusion estimates of the six regions can be quantified using average values. Many MBF estimation methods have been based on the AHA mapping approach [169,172–174]. However, the definition of discrete myocardial segments is subject to operator bias, and there is an inevitable loss of granularity caused by signal averaging.

An alternative approach to the longitudinal analysis is pixelwise analysis. Specifically, longitudinal comparisons between different DCE-MRI datasets can be carried out pixel-by-pixel, which would solve the operator bias and signal averaging issues. However, the shape and size of the myocardial tissues from different DCE-MRI datasets (see details in Chapter 3) are different. Therefore, a registration procedure will be needed to establish a pixelwise correspondence to compare different DCE-MRI datasets, for example, by applying some specifically standardised process to the dataset.

In this chapter, a standardisation approach based on a polar projection has been introduced to project contoured myocardium to a standardised annulus. This projection method is based on linear interpolation to deform the myocardium in the image, and it will be applied to different parametric maps illustrated in Chapters 4 and 5. The proposed method can be used in a variety of settings: to overlay different sequences obtained within the same MRI exam, to overlay different parameter maps acquired within different MRI exams, and also to combine images or maps from different subjects into standardised atlases.

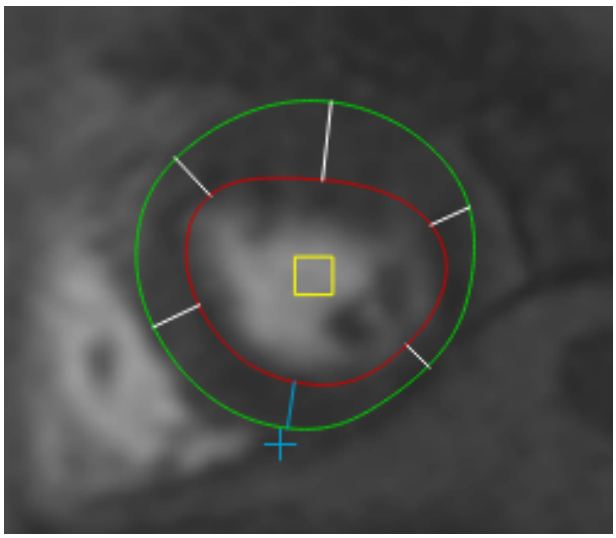


Figure 8.1: The manual contouring of the myocardium. The green curve denotes the epicardium. The red curve denotes the endocardium. The tissues inside the yellow square denote the LV blood pool samples, and the blue cross denotes the reference point. Five white lines and one blue line that link the endocardium and epicardium separate the myocardium to six AHA segments [4].

8.2 Method

8.2.1 Overview

The segmented myocardium used in this chapter are obtained manually from the DCE-MRI, see Section 3.3. The critical elements for delineating the myocardium include "endocardial contour", "epicardial contour" and "reference point". Figure 8.1 shows the manual contour of the myocardium. From this figure, all critical elements related to the delineation of the myocardium can be observed. The green curve denotes the epicardium. The red curve denotes the endocardium. The tissues inside the yellow square denote the LV blood pool, and the blue cross denotes the reference point (the inferior insertion point located at the junction of the right ventricle and the left ventricle). This point is used to locate the relative position of the myocardium. Five white lines and one blue line separate the myocardium into six segments based on the AHA mapping (see Section 3.6). The area located between the green and red curves are the myocardium. The projection approach proposed in this chapter aims to project any contoured myocardium to a standardised annulus.

This projection process can be roughly divided into two steps:

1. Project the myocardium to a rectangle. It is a transformation from the Cartesian coordinate system to a polar coordinate system. Specifically, the center of mass of the blood pool in the original image is the pole in the projected rectangle. The row of the rectangle represents the polar axis in the polar coordinate system and the column of the rectangle represents the angle in the polar coordinate system. The correspondence of the original

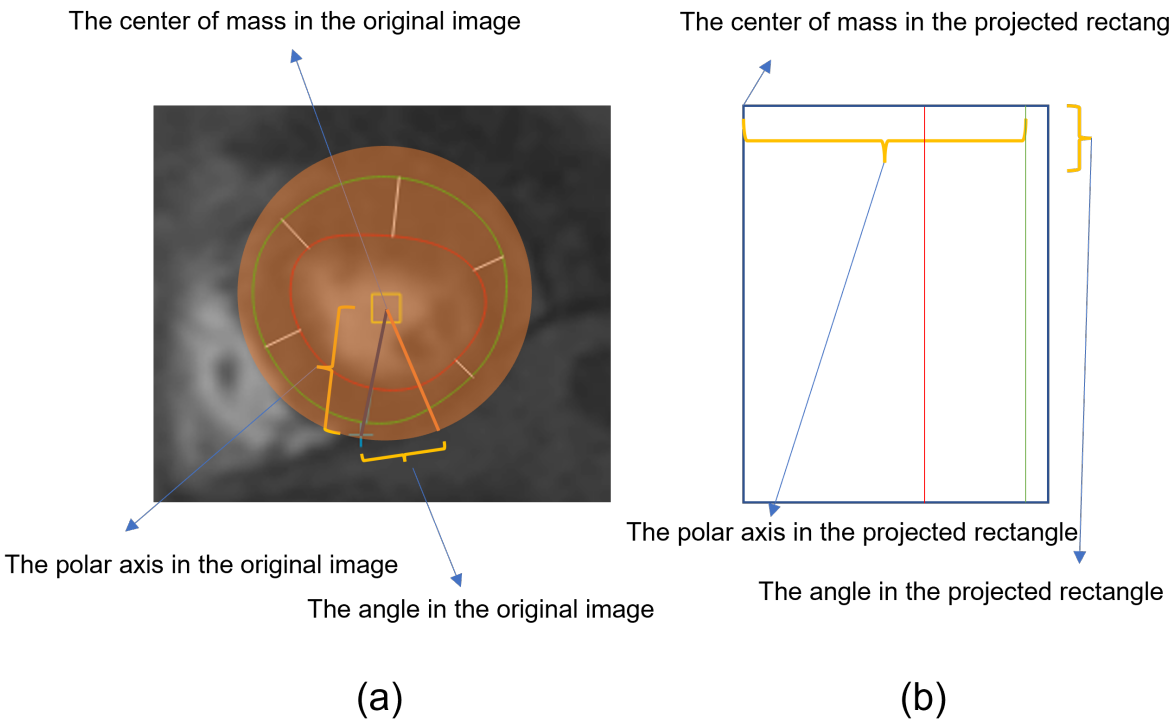


Figure 8.2: The correspondence between the original image and the projected rectangle. The green and red lines in the projected rectangle represent the epicardium and the endocardium in the original image. The signals inside the shaded orange circle will be projected to the rectangle.

image and the projected rectangle can be found in Figure 8.2. The signals inside the shaded orange circle will be projected to the rectangle.

2. Project the rectangle to an annulus. It is a transformation from the polar coordinate system to the Cartesian coordinate system. The correspondence of the projected rectangle and the projected annulus can be found in Figure 8.3.

8.2.2 Step 1: the myocardium to the rectangle projection

The first step of the projection process is to project the myocardium from the original image to a rectangle. Specifically, the proposed projection method projects all signals inside a circle with centre O (the center of mass of blood pool) and radius r (the line segment between O and the reference point) to a rectangle (see Figure 8.2). Linear interpolations and trigonometric functions are used in this step. In specific:

1. Link the centre of mass O of the blood pool in the original image and the reference point. This line segment is defined as the 0-degree. This line segment is shown as a purple line in panel (a), Figure 8.2.

2. Define the move step d, i.e. the degree of clockwise move for the line segment from 0-degree. The move step controls one dimension of the standard rectangle. Specifically, if

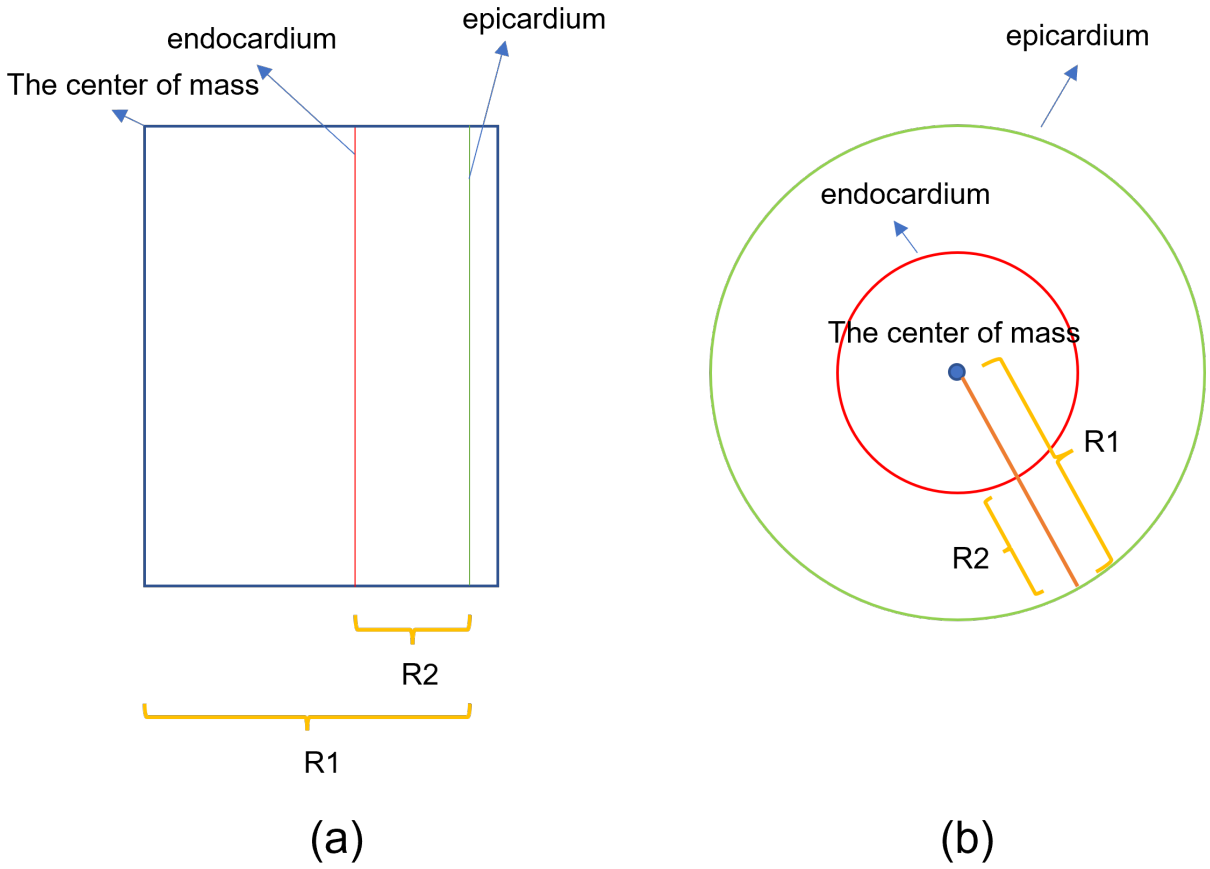


Figure 8.3: The correspondence between the projected rectangle and projected annulus. The green and red lines in panels (a) and (b) represent the epicardium and the endocardium in the original image. R1 and R2 are the projected radiuses of the epicardium and endocardium respectively.

the move step is defined as $d = 4$ degrees, then there will be $360/4 = 90$ line segments every 4 degrees clockwise, and accordingly 90 rows for the rectangle.

3. Calculate the coordinates of the end points for each line segment. The length of the line segment linking the centre of mass and the reference point is defined as r, and therefore the whole myocardial wall is included in the defined circle with radius r. By clockwise moving the 0-degrees line segment according to the move step, calculate the coordinates of the end point of each line segment. Particularly, the end point for the 0-degree line segment is the reference point.

4. Define the number of interpolation points for the line segment to be l. For each line segment, calculate the coordinates of these points by:

$$x_i = x_0 + i * \frac{x_{l-1} - x_0}{l-1} \tag{8.1}$$

and

$$y_i = y_0 + i * \frac{y_{l-1} - y_0}{l-1} \tag{8.2}$$

where $i \in \{0,1,2,...,l-1\}$ is the index of the points. The coordinate of the center of mass is (x_0, y_0) and the coordinate of the end point is (x_{l-1}, y_{l-1}). Establish two coordinate matrices (x-axis and y-axis). For the x-axis matrix, its first row contains the x-axis coordinates of points in 0 degrees radius. Its second row contains the x-axis coordinates of points in d degree radius and so on. The design of the y-axis matrix is similar. Both coordinate matrices have size $\frac{360}{d} \times l$. These two coordinate matrices represent the location correspondence between the pixel in the original image and the projected rectangle.

5. Calculate the projected values of the rectangle in panel (b), Figure 8.2 based on the coordinate matrices. Specifically, the value of the pixel in the projected rectangle is the same as its corresponding pixel in the original image, and the correspondence between two images is defined in the previous step. For any pixel in the projected rectangle, the x-axis and y-axis coordinates of its corresponding pixel in the original image (panel (a), Figure 8.2) can be divided to integer part x_{int}, y_{int} and decimal part x_{dec}, y_{dec}. In other words, its coordinate is $(x_{\text{int}} + x_{\text{dec}}, y_{\text{int}} + y_{\text{dec}})$ where $0 \leq x_{\text{dec}} \leq 1$ and $0 \leq y_{\text{dec}} \leq 1$. Therefore, The value of any pixel in the projected rectangle can be calculated by the values of the vertices of a square showing in Figure 8.4. The four vertices of the square have coordinates $(x_{\text{int}}, y_{\text{int}})$, $(x_{\text{int}} + 1, y_{\text{int}})$, $(x_{\text{int}}, y_{\text{int}} + 1)$, and $(x_{\text{int}} + 1, y_{\text{int}} + 1)$ respectively. Let s_1, s_2, s_3 and s_4 be the values of the four vertices respectively. The value for the pixel in the projected rectangle s_p can be calculated as follows:

$$\begin{aligned} s_p =& s_1 \times (1 - x_{\text{dec}}) \times (1 - y_{\text{dec}}) + s_2 \times x_{\text{dec}} \times (1 - y_{\text{dec}}) \\ &+ s_3 \times (1 - x_{\text{dec}}) \times y_{\text{dec}} + s_4 \times x_{\text{dec}} \times y_{\text{dec}}. \end{aligned} \tag{8.3}$$

In this way, the shaded orange region in panel (a), Figure 8.2 can be projected to a rectangle. However, the endocardium and epicardium are not straight line in the rectangle (see Figure 8.5). Two further steps need to be implemented to modify the curves of the endocardium and epicardium to straight lines.

6. Find the interpolation points that are closest to the endocardium and epicardium. For each line segment, there are $360/d$ points, and there will be one point that is closest to the endocardium (if coincidentally two points have the same distance, pick the point closer to the center of mass of the blood pool). Similarly, there will be one point that is closest to the epicardium (if coincidentally two points have the same distance, pick the point further to the center of mass of the blood pool). These two points are shown in Figure 8.6. The

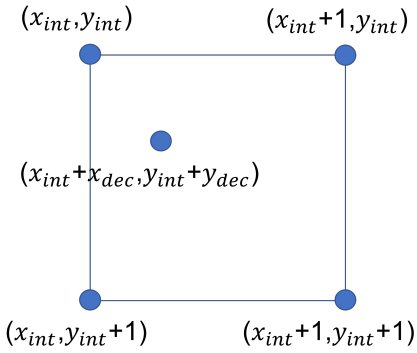


Figure 8.4: The four blue circles located at the vertices of the square denote four adjacent pixels in the original image. The blue circle inside the square denotes the corresponding pixel in the original image. The values in the brackets denote the coordinates of the pixels.

distance between two points is defined by two-dimensional Euclidean distance:

$$d(p,q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}. \tag{8.4}$$

Once the number of interpolation points l is big enough, the distances between the endocardium or epicardium points and the interpolation points will be small enough. The value of l will be tested in the subsequent section.

7. For each radius, use the two chosen points (point A and B in Figure 8.6) as the new start and end points. Define a new interpolation number m, and use linear interpolation to regenerate the row of the value matrix. In fact. it is the projection from Figure 8.5 to panel (b), Figure 8.2. The newly generated matrix with size $\frac{360}{d} \times m$ represents the rectangle with red and green sides in panel (b), Figure 8.2.

The processes shown above introduced the first step of the projection process, i.e. project the myocardium to a rectangle. This method makes use of linear interpolation to transform a circle (shaded orange region in panel (a), Figure 8.2) into a rectangle. Moreover, the myocardium region in the circle is standardised to a rectangle (the region between red and green lines in panel (b), Figure 8.2).

8.2.3 Step 2: the rectangle to annulus projection

The second step of the projection process is to project the rectangle to an annulus. This step has been divided as follows:

1. Define the 0 degrees radius in the annulus. Define the radius of the larger circle to be R_1 and the smaller circle to be R_2. Define the center of mass of the annulus to be $(x_0^{'}, y_0^{'})$. Specifically, in this study, $x_0^{'} = y_0^{'} = \frac{R_1}{2}$. The superscript $'$ denotes the coordinates in the annulus. The

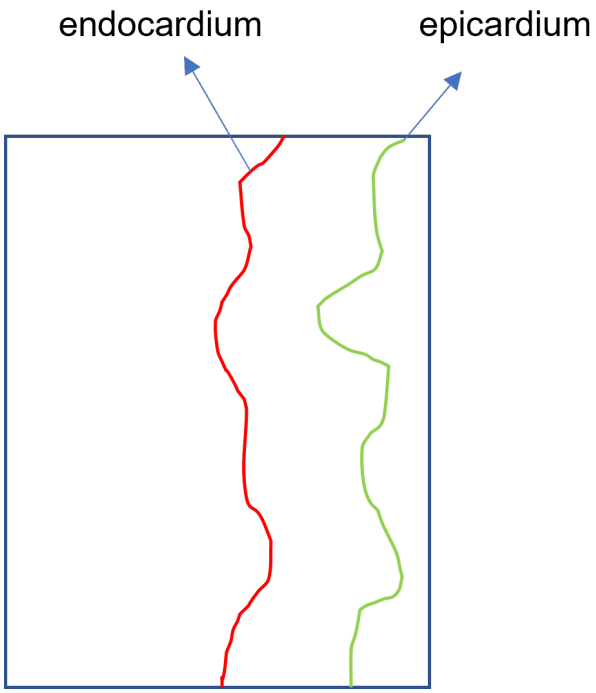


Figure 8.5: The contours of endocardium and epicardium in the rectangle.

2. For each pixel i in the annulus, its location coordinate can be described by the polar coordinate system. For example, the i^{th} pixel is located at $(x_i^{'}, y_i^{'})$. The first dimension of the polar coordinate system is the distance between this pixel and the centre of mass. This distance can be calculated as:

$$d(0,i) = \sqrt{(x_0^{'} - x_i^{'})^2 + (y_0^{'} - y_i^{'})^2}. \tag{8.5}$$

   The second dimension of the polar coordinate system is the angle $a(0,i)$ between the 0 degrees radius and the line segment that is linking the center of mass and the given pixel. $a(0,i)$ can be calculated from inverse trigonometric functions. If $d(0,i) > R_1$ or $d(0,i) < R_2$, the value of pixel i is set as the background value. If $R_2 \leq d(0,i) \leq R_1$, the value for pixel i under the polar coordinate system with coordinate $(d(0,i), a(0,i))$ in the annulus is exactly the same as the value for pixel $i^{''}$ under the Cartesian coordinate system with coordinate $(d(0,i), a(0,i))$ in the rectangle. For both $d(0,i)$ and $a(0,i)$, their values might not correspond to the specific pixels in the rectangle. However, they can be calculated using the interpolation method described in Figure 8.4 and equation (8.3).

Now, the complete process for the myocardium projection has been introduced. Since the type of the image is not restricted, theoretically, all types of the images, e.g. the original image, the up-slope map, the MBF map, etc., can be applied by this approach once the three elements (endocardium contour, epicardium contour and reference point) are given.

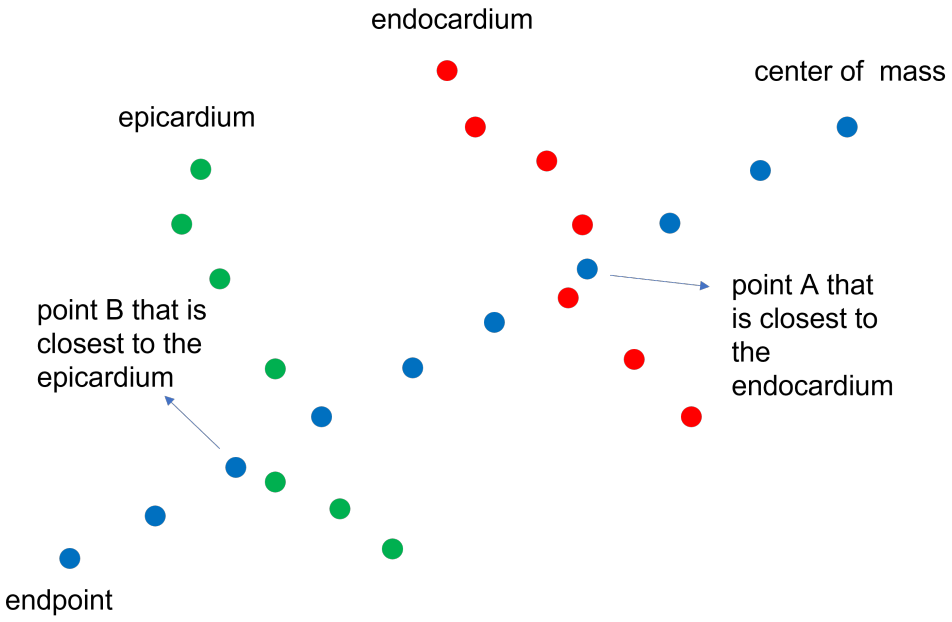


Figure 8.6: The blue circles indicate the interpolation points in the line segment. The red circles indicate the endocardium contour, and the green circles indicate the epicardium contour. Point A is closest to the endocardium and point B is closest to the epicardium.

8.3 Application of myocardium projection to DCE-MRI

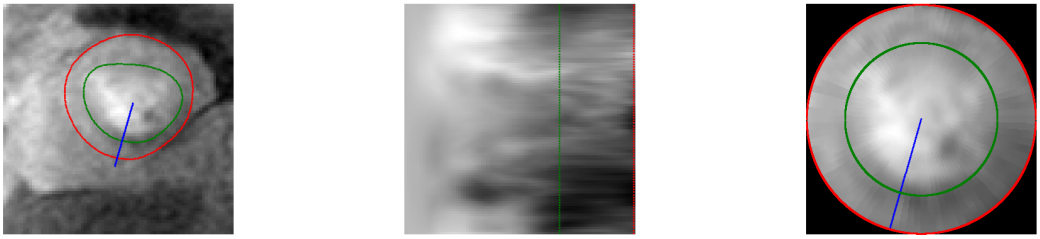
In this section, the myocardium projection approach has been applied to the original image and three parametric maps, CER, up-slope, MBF and their corresponding GMM classification maps (see details in Section 5.2.3). Moreover, since the purpose of the proposed projection method is to compare longitudinal data, an additional DCE-MRI dataset obtained from the same patient will also be used in this section.

Firstly, the values of the parameter l have been tested. Given different values of l, the distances between the chosen points in the interpolation line segment and their closest contour points are used to quantify the information lose. In other words, the shorter the distance between point A and its closest endocardium point is, the less the information lose. This principle is the same for point B and its closest epicardium point. Since the density of the endocardium and epicardium points will also affect the distances, the numbers of endocardium and epicardium points will be extra variables to quantify the distances. The distance between two adjacent pixels in the image is 1, and when $l = 30$, the distance between two interpolation points is around 1. Table 8.1 shows the average distances between the chosen points in the interpolation line segment and their closest contour points given different values of l or different numbers of endocardium and epicardium points. Therefore, even for the most crude case, $l = 30$, the distance is rather small. Moreover, when l is larger than 360, the improvement is negligible. However, when the number of interpolation points is too big, the computational complexity will also be high especially when the models are also complex, e.g. the models illustrated in

Table 8.1: The average distances between the chosen points and contour points

	number of contour points: 199/271*	number of contour points: 1990/2710**
$l = 30$	0.307	0.262
$l = 36$	0.257	0.213
$l = 60$	0.192	0.132
$l = 90$	0.158	0.086
$l = 360$	0.127	0.026
$l = 900$	0.124	0.016

* 199/271 means that the endocardium contains 199 points and the epicardium contains 271 points.
** 1990/2710 means that the endocardium contains 1990 points and the epicardium contains 2710 points.



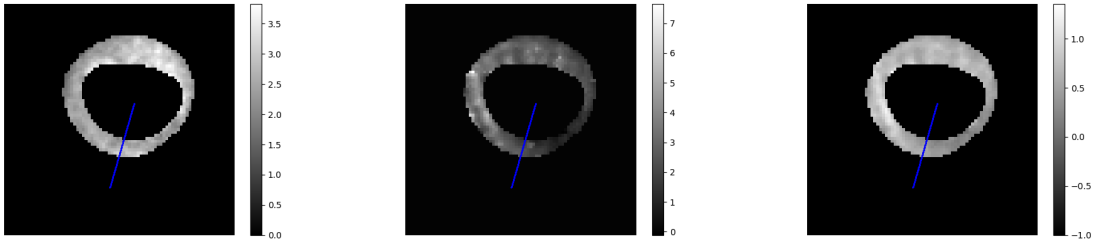
(a) Origin (b) Rectangle (c) Annulus

Figure 8.7: Polar projection of the maximum enhancement image. Panels (a) - (c) show the maximum enhancement image, its rectangle projection and its annulus projection respectively. The red and green lines represent the endocardium and epicardium in different images respectively. The blue lines in panels (a) and (c) indicate the line segment linked by the center of mass and the reference point.

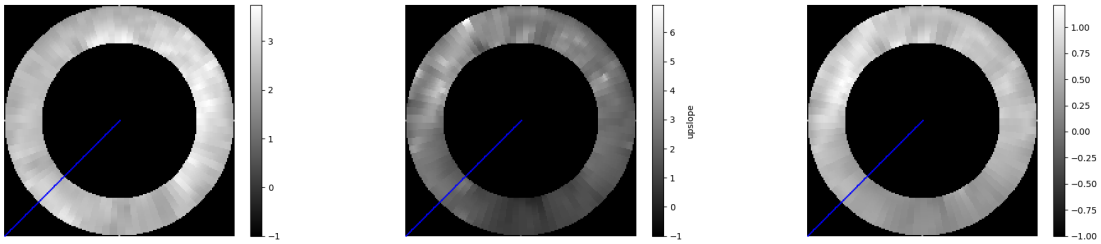
Chapters 6 and 7. Therefore, two suggested parameters setting are $l = 90$, $d = 4$ and $m = 30$; $l = 360$, $d = 1$ and $m = 120$ for the reason that the projected rectangle is a square and the rectangle of the myocardium occupies 1/3 of the square in area.

Figure 8.7 shows the maximum enhancement image and its corresponding rectangle and annulus projections. The projection parameters are set to: $d = 4$, $l = 90$, $m = 30$, $R_1 = 90$ and $R_2 = 60$. All pixels inside the red curve in panel (a), Figure 8.7 have been projected to the rectangle in panel (b) and the annulus in panel (c) using the method illustrated in Section 8.2. Firstly, all patterns shown in panel (a) can be observed in the same relative position in panel (c). For example, the hypo-perfusion region (the low SI region which is dark in the image) is located in the right bottom of the myocardium in the image in panel (a), and this can also be observed in panel (c) in the same relative position.

A further application of the polar projection method is to project the parametric maps of the myocardium to annuluses. Figures 8.8 and 8.9 show the comparisons for different parametric maps between original and projected maps. The values of projection parameters introduced in



(a) Original CER (b) Original Up-slope (c) Original MBF

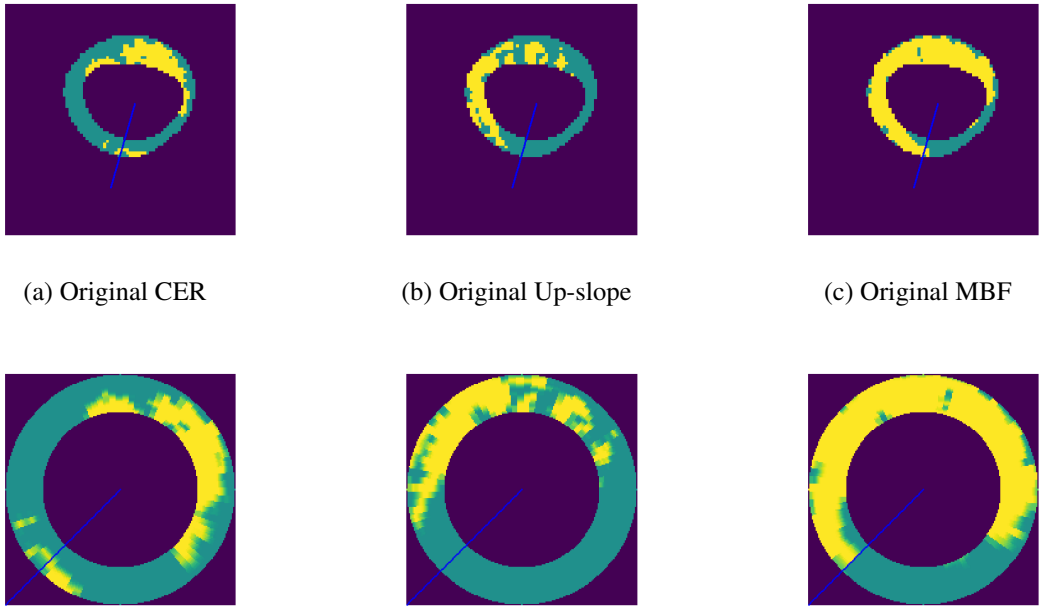


(d) Projected PSI (e) Projected Up-slope (f) Projected MBF

Figure 8.8: Comparisons of different parametric maps between original and projected maps. Panels (a) - (c) show the original parametric maps of the CER, up-slope and MBF respectively. Panels (d) - (f) show the projected annulus for these three parameters. The blue lines indicate the line segment linked by the center of mass and the reference point. The values of background pixels are set to -1, which can avoid any potential conflicts since these parameters are all non-negative.

Section 8.2 are set to: $d = 4$, $l = 90$, $m = 30$, $R_1 = 90$ and $R_2 = 60$.

The first advantage of the myocardium projection is that the quantification analysis will be easier in a standard shape than in an irregular shape. Since three main blood vessels provide the blood to the myocardium, 120° of the myocardium corresponds to one main vessel. Therefore, it is more accurate to quantify the hypo-perfused region based on the angle than other elements, e.g. the area. Specifically, the standard segmentation in [4] also makes use of the angle to quantify the proportion of the hypo-perfused region. However, it is complex to use the angle to quantify it in an irregular shape. On the other hand, it is easy to make use of the area to quantify the hypo-perfused region, but the quantification is inaccurate. The reason is that the myocardial wall in an irregular shape has different thicknesses in different positions. For example, Figure 8.9, panel (c) shows the classification map using the GMM classification method (see Section 5.2) based on the MBF maps. Suppose we use the proportion of hypo-perfused area to estimate the proportion of the hypo-perfusion region. In that case, the result will be underestimated because the myocardial wall in the hypo-perfusion region is thinner than the myocardial wall in the normal-perfused region. However, for the projected GMM map in panel (f), Figure 8.9, the proportion of the hypo-perfused area can accurately estimate the proportion of the hypo-perfused region because the thickness of the myocardial wall has been standardised



(a) Original CER (b) Original Up-slope (c) Original MBF

(d) Projected PSI (e) Projected Up-slope (f) Projected MBF

Figure 8.9: Comparisons of classification maps of different parameters between original and projected maps. Panels (a) - (c) show the original classification maps of the CER, up-slope and MBF respectively. Panels (d) - (f) show the projected annulus for these classification maps. The blue lines indicate the line segment linked by the center of mass and the reference point. The yellow segments indicate healthy tissues. The dark green segments indicate lesion tissues and the black pixels indicate the background.

in different positions. Moreover, given the coordinates of any two points, the angle is much easier to be obtained in a regular annulus than in an irregular shape.

The second advantage of the myocardium projection is that different DCE-MRI datasets can be compared pixelwise, which can be further used in longitudinal time series analysis, regression analysis and statistical predictions, and these statistical analyses will improve disease diagnosis. Figure 8.10 show longitudinal comparisons of MBF classification maps from one patient. According to panels (a) and (b) in Figure 8.10, only qualitative comparisons, e.g. the hypo-perfused area (marked as dark green) is enlarged, can be applied. However, since the sizes of the myocardium are different, the relative position of the newly added lesion area is difficult to quantify. Panels (c) and (d) in Figure 8.10 show the projected GMM classification maps based on MBF values. Since the shape and size of the projected maps are always consistent, quantitative analysis, e.g. the relative position of the newly added lesion area, can be easily obtained.

It should be noticed that the longitudinal comparisons for MBF maps are not included in this section. The reason is that the MBF estimations from different DCE-MRI datasets are not calibrated. It is illustrated that the estimated values of the MBF in this study are systematically overestimated (see Section 3.5) though the relative distribution is not affected. Therefore, we

(a) Original GMM 1 (b) Original GMM 2 (c) Projected GMM 1 (d) Projected GMM 2

Figure 8.10: Longitudinal comparisons for MBF classification maps. Panels (a) and (b) show the original classification maps for DCE-MRI datasets 1 and 2. Panels (c) and (d) show the projected classification maps for DCE-MRI datasets 1 and 2.

only compare the classification maps longitudinally. To further compare the estimated MBF maps, the calibration of the MBF values should be applied.

8.4 Discussion and conclusion

The proposed projection method aims to project any contoured myocardial map to a standard annulus. It is emphasised that this method can not only be applied to myocardial perfusion images, but also be applied to any type of myocardial images such as LGE images. The projected map is a visualisation of the AHA mapping. In clinical practice, is is easier for the clinicians to find the specific segments of the myocardium, e.g. anterior, anteroseptal, inferoseptal, inferior, inferolateral and anterolateral, and further to find the corresponding main branches of the narrowed coronary artery. For example, given panel (f), Figure 8.9, the clinician can easily conclude that the hypo-perfusion region of the specific heart is in inferior and its corresponding narrowed coronary artery branch is RCA.

In the near future, the proposed projection method will be applied to longitudinal datasets. It provides an easier comparison between different sets of data for the same patient. Theoretically, given the projected classification maps for different sets of the data, the trend of the changes of the classification maps can be inferred, and therefore the trend of the hypo-perfusion region of the patient's heart can be inferred. In clinical practice, the clinician can therefore justify the condition of the patient's heart longitudinally, and make clinical decisions based on the longitudinal classification maps.

In conclusion, the proposed projection approach provides an opportunity to apply pixelwise longitudinal analysis. It also makes the justification based on the area and angle consistent. Therefore the straightforward approach (the area based) can be used to replace the complex approach (the angle based). However, there are also some limitations to this approach. The application of linear interpolation generates some intermediate values. For example, the GMM classification maps only contain two different values, i.e. 0 for lesion and 1 for healthy tissue (see Figure 8.9, panel (a)). However, after applying the projection approach, some values between 0

and 1 are interpolated. Moreover, the application of linear interpolation loses information. For example, the thickness of the myocardial wall will also provide some useful information, but this information is removed during the standardization process. Although the primary information, e.g. the spatial distributions of values, is preserved during the projection, the projection still cannot fully preserve all information. Some information is lost because of the nature of the interpolation. This problem can be alleviated by using large values of interpolation numbers l (see Table 8.1).

Chapter 9

Future Work, Discussion and Conclusion

9.1 Future work

9.1.1 Longitudinal analysis

Most studies proposed in this thesis aim to apply analysis to a single frame or a series of dynamic frames of the myocardial perfusion DCE-MRI data. However, the most valuable part of the data used in this study is its longitudinal structure (see details in Chapter 3.3). Scaling up the approaches in this thesis to capture longitudinal aspects of data could be the focus of future work. The author has carried out some preparatory work specified in this section.

In Chapter 8, a polar projection method that can project any myocardium to an annulus has been introduced. This is the first step for the pixelwise longitudinal analysis. The next pre-process work is to calibrate different DCE-MRI datasets. Firstly, the non-linearity between the SI and the contrast concentration should be reviewed. In Chapter 3.5, it is mentioned that the high dose of contrast agent will cause the non-linearity between the SI and the contrast concentration. Moreover, the blood pool SI will be saturated. The blood pool saturation can be observed from Figure 9.1. There is a plateau of the SI from time point 11 to time point 17. This phenomenon is called blood pool saturation, and it will cause a dramatic overestimation of the MBF. If the analysis is proceeded based on a single DCE-MRI dataset, the relative MBF estimates for different pixels will not be severely affected because the estimates are obtained using the same blood pool SI. However, for different DCE-MRI datasets, the degrees of the blood pool saturation will be different, and moreover, the underestimations of the MBF will be different. Therefore, the MBF estimates for different DCE-MRI datasets cannot be compared directly without calibrations.

There are many either simple or complex methods to calibrate the MBF estimates. The simplest method standardises the MBF estimates to "reasonable" values. The word "reasonable" means the value is physiologically normal. For example, the MBF for a healthy rest heart is commonly around 1 mL/min/mL [85]. Therefore, the maximal MBF estimates can be stan-

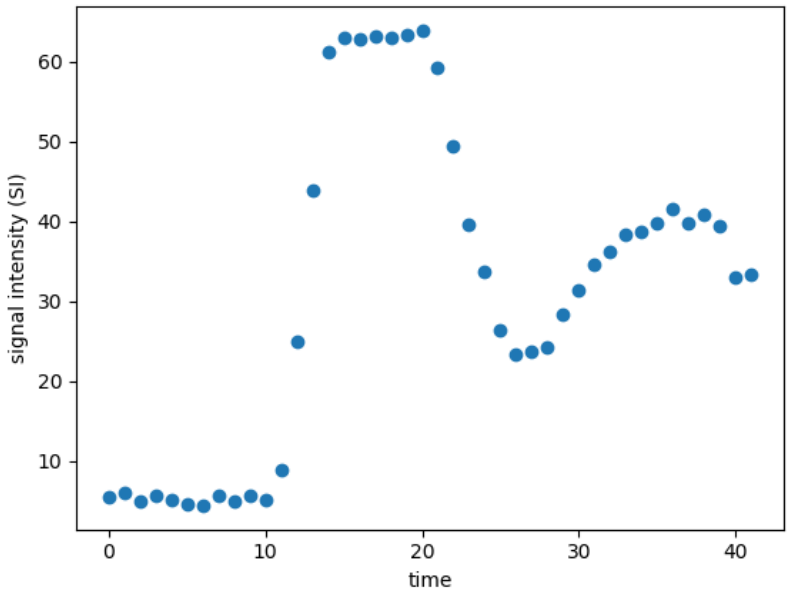


Figure 9.1: Blood pool SI.

dardised to 1 mL/min/mL for different DCE-MRI datasets. However, this method can only be applied to the patient whose heart still has normal function parts. If a patient's heart does not have normal function parts, e.g. the maximal MBF for this heart is only 0.8 mL/min/mL, this calibration method will cause systematical bias. There are also some complex methods to fix the blood pool saturation issue. One method is to eliminate the plateau of the SI using up-slope replacement. Specifically, the up-slope of the curve can reflect the growth speed of the SI before the SI reaches the saturated SI value, e.g. 60 in Figure 9.1. Therefore, a TIC without the saturation plateau can be generated by replacing the up-slope in the saturated region (time point 14 - 20 in Figure 9.1) by the up-slope in increasing (time point 10 - 14 in Figure 9.1) and decreasing (time point 20 - 25 in Figure 9.1) regions. Figure 9.2 shows a diagram of the up-slope replacement. A vital part of this method is the uncertainty of the generation. Specifically, the up-slope values are within a range in the increasing and decreasing regions. Different choices of the up-slope values will generate different results. The uncertainty of the up-slope replacement will cause uncertainty in the MBF estimation and it is worth applying the sensitivity analysis of the up-slope replacement.

Both the polar projection and blood pool SI calibration lead to the longitudinal analysis of the myocardial perfusion DCE-MRI. The longitudinal analysis aims to predict the heart conditions based on the longitudinal data. This analysis can be justified by two types of approaches, i.e. classification and estimation.

The classification analysis aims to determine the area of the hypo-perfusion region for longitudinal data. This can be incorporated by the results from the previous work. In Chapters 6 and 7, model-driven and data-driven methods are introduced to classify the pixels in the myocardium

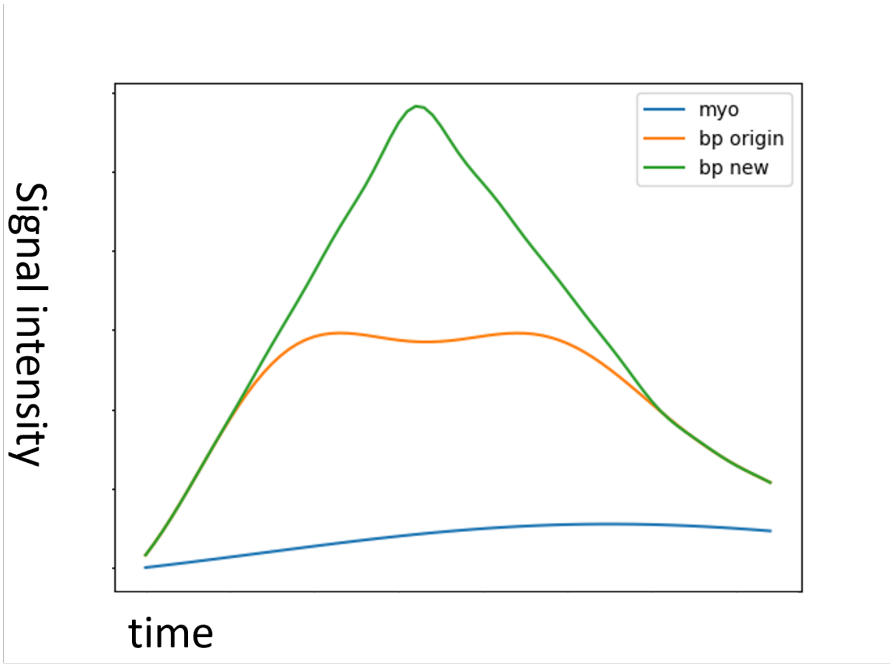


Figure 9.2: Generation of Blood pool TIC without saturation plateau. The blue line shows the TIC of a pixel from myocardium. The orange line shows the TIC of blood pool that has saturation plateau. The green line shows the TIC of blood pool without saturation plateau using up-slope replacement.

into two different groups, i.e. healthy group and the lesion group. The classification results can be used for the classification analysis. A straightforward method can be applied by comparing the proportion of the lesion pixels between different serials of the same patient. For instance, there are 30% pixels in the lesion group for patient A, dataset 1 (day 1), and 35% pixels in the lesion group for patient A, dataset 2 (day 2). It is believed that the heart condition of patient A is getting worse. This example is used to describe the classification analysis. This procedure can be carried out by more complex, but more accurate methods. In [175], a support vector machine (SVM) method is applied to classify the brain MRI for Alzheimer patients to implement the longitudinal analysis. Hierarchical classification is applied in [176] to process longitudinal analysis of Alzheimer patients. For multiple sclerosis patients, a Bayesian classification method is applied to the longitudinal data in [177].

The estimation analysis aims to quantify the degree of hypo-perfusion of the myocardium using MBF estimates for longitudinal data. The methods illustrated in Chapters 4 and 6 can be used to estimate the MBF firstly. Then, the MBF estimates of pixels from different serials can be the input data of some time-series methods. The final target of the estimation analysis is to predict the MBF values given the longitudinal data. Autoregressive–moving-average (ARMA) model [178] is a general method to analyze time series data. A Gaussian process can also be applied to proceed with the longitudinal time series analysis [179].

Generally speaking, an advantage of the classification analysis is that the blood pool SI

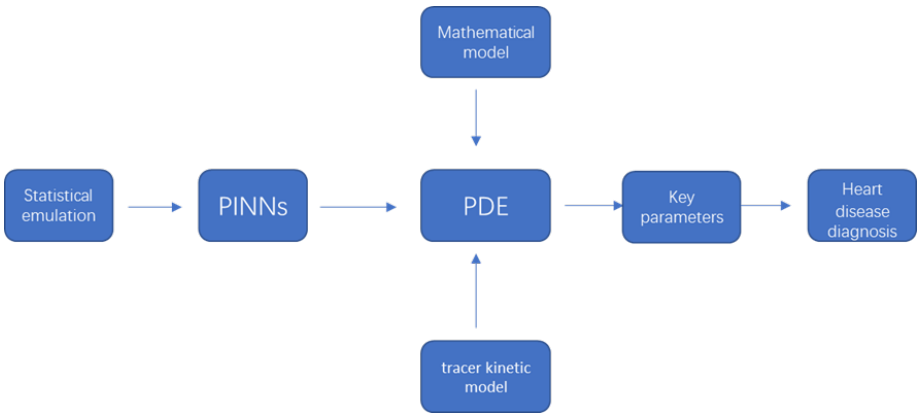


Figure 9.3: A flow chart for the statistical emulation for PINNs.

calibration is not strictly necessary for it. The reason is that the classification map is based on the relative MBF values, which are generated by the same saturated blood pool SI. In other words, although the blood pool saturation will overestimate the MBF values, the degrees of the overestimation are the same for all pixels. On the other hand, the classification analysis cannot quantify the degree of hypo-perfusion. For example, for the same patient, if the proportions of the lesion are the same for DCE-MRI dataset 1 and DCE-MRI dataset 2, the classification analysis will justify that the patient is neither getting better nor getting worse. However, the estimation analysis can detect the values of the MBF for different DCE-MRI datasets. If the estimated MBF values are getting higher from dataset 1 to dataset 2, the estimation analysis will justify that the patient is getting better. Moreover, the absolute values will help for longitudinal comparisons because the relative values for different sets of data cannot be compared directly.

9.1.2 Statistical emulation for PINNs using myocardial perfusion DCE-MRI

In Chapter 4, a traditional MBF quantification method, the Fermi method, has been reviewed. In recent years, deep learning methods have been introduced to quantify the MBF [180, 181]. Potential future work is to make use of the most recent method to quantify the key perfusion parameter, the MBF. Moreover, some mathematical models are applied to derive the MBF [182]. The combination of DCE-MRI and mathematical modelling may improve the diagnosis of a wide range of cardiovascular pathologies. Specifically, a fast data-driven method for quantifying key physiological parameters, e.g., MBF, using myocardial perfusion DCE-MRI combined with mathematical modelling of the human cardiovascular system can be developed. Generally, the key physiological parameters are formed as partial differential equations (PDE) in both mathematical models [182] and DCE-MRI tracer-kinetic models [183]. A data-driven method named physics-informed neural networks (PINNs) [184] has been developed to solve PDEs, and it has been applied to myocardial perfusion DCE-MRI [181, 185]. However, the training

of data-driven methods are usually computationally complex, and time-consuming [181]. Statistical emulation for PINNs is a potential method to tackle the computationally complex and time-consuming issue [186]. Therefore, the potential future work aims to develop statistical emulation of PINNs that will realise the combination of mathematical modelling and DCE-MRI tracer-kinetic modelling of key physiological parameters using myocardial perfusion DCE-MRI. A flow chart shown in Figure 9.3 describes the process of this method.

9.2 Discussion and conclusion

The main target of the work in this thesis is to detect the hypo-perfusion region using classification methods from myocardial perfusion DCE-MRI data. Traditionally, qualitative, semi-quantitative and quantitative methods are used to diagnose stenosis (which is shown as hypo-perfusion in the myocardial perfusion DCE-MRI). The main contribution of the work in this thesis is to design classification methods to the semi-quantitative and quantitative derived parameters to obtain the classification map of the myocardial tissues in the image. The most beneficial aspect of the work is that it can clearly delineate the boundary between the healthy tissues and the lesions without any need for operator input. Given an MBF map of the myocardium image, different operators may delineate different boundaries of the healthy tissues and the lesions because manual image contouring is highly user-dependent. However, the classification methods can distinguish different groups based on the mean value and the covariance matrix of the input data. Moreover, customised prior information can be incorporated by introducing prior distributions under the Bayesian statistical framework. Specifically, in this study, spatial and temporal information is usually introduced for physiological reasons. The study in this thesis also initiates the longitudinal analysis of the myocardial perfusion DCE-MRI. Specifically, the classification maps of the longitudinal data can be used as the input of some longitudinal models.

Starting from the background information in Chapters 2 and 3, the statistical tools and basic concepts of the heart physiology and myocardial perfusion DCE-MRI have been reviewed. In Chapter 4, semi-quantitative and quantitative analysis of the myocardial perfusion parameters, e.g. the up-slope, the MBF, are first applied. These parameters are used to reflect the degree of ischaemia. Moreover, once the estimates of these parameters are obtained, they can also be employed as the input of the classification models. In the same chapter, the regressions of TIC using different methods are also introduced and compared. The smooth TIC generated by the regression methods improves both semi-quantitative and quantitative estimation of parameters. In Chapter 5, three classification methods, i.e. GMM, SVFMM and GMM-MRF, are applied to either original images or parametric maps generated in Chapter 4. The GMM method is a general application of a Gaussian mixture model. There are unrealistic single and small clusters shown in the classification maps. To fix this issue, the SVFMM has been applied to introduce spatial information. However, because the spatial information is introduced by an indirect way, not all

single and small clusters can be removed. The GMM-MRF method, on the other hand, makes use of the Markov random field prior to introduce the spatial (or spatio-temporal) information directly to the label of the pixel, and it is tested to remove all unrealistic single and small clusters. By comparing all three methods, the GMM-MRF method performs the best (removing all single and small clusters). The limitation of the classification methods is that it is highly dependent on the quantified parametric maps generated in Chapter 4. The classification methods may fail (when the SNR is lower than 3) if the parametric maps are too noisy or incorrect. The core issue is that the estimation phase in Chapter 4 and the classification phase in Chapter 5 are causal related. Specifically, the classification maps are dependent on the parametric maps, but the parametric maps are independent of the classification map. However, they should interact with each other. To this end, in Chapters 6 and 7, hierarchical Bayesian models are used to combine the estimation phase and the classification phase.

In Chapter 6, a hierarchical Bayesian model with a Markov random field prior has been introduced. The Fermi-deconvolution method has been incorporated in the hierarchical model to estimate the MBF. Overall, this method aims to estimate the MBF while classifying the MBF. Since the Fermi parameters and the label parameter are connected by directed links, their posterior distributions interact with each other. Moreover, an MCMC sampling method has been applied to draw parameters from their corresponding target posterior distributions. Model selections based on WAIC values are carried out to compare different hyperparameters. The comparisons of the proposed method and the GMM classification method have been implemented based on both clinical and synthetic data, and the proposed hierarchical Bayesian model is tested to outperform the benchmark GMM method. The main limitation of the hierarchical Bayesian model is that it is computationally expensive. Further modification can be focused on the parallelisation of the MCMC samplings.

In Chapter 7, a hierarchical Bayesian model with spatio-temporal Markov random field priors has been introduced. Different from the method illustrated in Chapter 6, the proposed method in this chapter is an entirely data-driven method. Expressly, the method illustrated in Chapter 6 is incorporated with the Fermi model that is used to estimate the MBF. However, the proposed method in this chapter only makes use of the original data to generate the classification map. Moreover, spatio-temporal Markov random field priors have been applied to generate denoised images. An MCMC sampling method is used to draw samples of label parameters and true SI from their posterior distributions. Model selections have been applied to compare different hyperparameters using the WAIC values. Two benchmark methods, i.e. GMM classification method, closing and opening operation, have been used to compare with the proposed hierarchical Bayesian model using both clinical and synthetic data. The clinical and synthetic data results indicate that the proposed hierarchical Bayesian model has the best performance with the least classification errors. The comparisons of all proposed methods are applied. In general, the methods with higher accuracy are computationally slow, and the methods with lower accuracy

are computationally faster. Therefore, different methods can be chosen based on the aims of the analysis.

In Chapter 8, a polar projection method has been introduced to project any myocardium to an annulus. This method aims to provide a way to implement pixelwise comparisons between different myocardium images. Specifically, given the myocardium images from different DCE-MRI datasets, the shapes and sizes of the myocardium for different serials are different, and therefore pixelwise analysis cannot be applied without first perfoming non-rigid transformation. The polar projection method can project different myocardium images to annuluses with the same shape and size. The work in this chapter is the preparatory work of the longitudinal analysis that is the most critical future work stemming from this thesis.

Appendix A

Derivation of the equation

A.1 The derivation of equation 2.55

According to equation 2.54, we have

$$\ln P(X,Z|\theta) = \ln P(Z|X,\theta) + \ln P(X|\theta). \tag{A.1}$$

We move the term $\ln P(Z|X,\theta)$ to the left hand side of the equation, and we have

$$\ln P(X|\theta) = \ln P(X,Z|\theta) - \ln P(Z|X,\theta). \tag{A.2}$$

Since $q(Z)$ is a density function, we have

$$\sum_Z q(Z) = 1. \tag{A.3}$$

We can multiply by $q(Z)$ and then sum over Z to both sides of equation A.2, and we have

$$\ln P(X|\theta) = \sum_Z q(Z) \ln P(X,Z|\theta) - \sum_Z q(Z) \ln P(Z|X,\theta). \tag{A.4}$$

Then we add and subtract $\sum_Z q(Z) \ln q(Z)$ to the right hand side of equation A.4, it turns to be

$$\ln P(X|\theta) = \sum_Z q(Z) \ln P(X,Z|\theta) - \sum_Z q(Z) \ln q(Z) - \left(\sum_Z q(Z) \ln P(Z|X,\theta) - \sum_Z q(Z) \ln q(Z) \right). \tag{A.5}$$

Finally, we have got

$$\ln P(X|\theta) = \sum_Z q(Z) \ln \left\{ \frac{P(X,Z|\theta)}{q(Z)} \right\} - \sum_Z q(Z) \ln \left\{ \frac{P(Z|X,\theta)}{q(Z)} \right\}. \tag{A.6}$$

A.2 The derivation of equation 4.11

According to equation 4.9, we have

$$C_{\text{myo}}(t) = F_{\text{b}} \int_0^t [C_{\text{in}}(s) - C_{\text{out}}(s)]ds. \tag{A.7}$$

After substituting equation 4.10 into equation A.7, we have

$$C_{\text{myo}}(t) = F_{\text{b}} \int_0^t [C_{\text{in}}(s) - C_{\text{in}}(s) * h(t)]ds \tag{A.8}$$

$$C_{\text{myo}}(t) = F_{\text{b}} \int_0^t C_{\text{in}}(s) * [\delta(s) - h(s)]ds \tag{A.9}$$

$$C_{\text{myo}}(t) = F_{\text{b}} \left(C_{\text{in}}(t) * \left[\int_0^t \delta(s)ds - \int_0^t h(s)ds \right] \right) \tag{A.10}$$

$$C_{\text{myo}}(t) = F_{\text{b}} \left(C_{\text{in}}(t) * \left[1 - \int_0^t h(s)ds \right] \right). \tag{A.11}$$

We define the impulse response function $R_f(t)$ as:

$$R_f(t) = F_{\text{b}} \left[1 - \int_0^t h(s)ds \right]. \tag{A.12}$$

Therefore, equation 4.9 becomes

$$C_{\text{myo}}(t) = R_f(t) * C_{\text{in}}(t) \tag{A.13}$$

The derivation from equation A.8 to equation A.9 is based on

$$f(s) * \delta(s) = f(s), \tag{A.14}$$

and the derivation from equation A.9 to equation A.10 is based on

$$\int_0^t f(s) * g(s)ds = f(t) * G(t). \tag{A.15}$$

where G is the primitive function of g.

A.3 The derivation of equation 5.8

By substituting equations 5.5 and 5.7 into equation 5.8, equation 5.8 becomes:

$$P(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \prod_{j=1}^{K} \pi_j^{k_j} \prod_{j=1}^{K} N(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)^{k_j} \tag{A.16}$$

$$= \sum_{\boldsymbol{k}} \prod_{j=1}^{K} \left(\pi_j N\left(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)\right)^{k_j} \tag{A.17}$$

Since $\boldsymbol{k}$ have a 1-of-K representation, given the value j, $k_j = 1$ and $k_{p\neq j} = 0$.

$$P(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \left(\pi_1 N(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)^{k_1} \cdot \ldots \cdot \pi_K N(\boldsymbol{x}|\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)^{k_K} \right) \tag{A.18}$$

$$= (\pi_1 N(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \cdot 1 \cdot \ldots \cdot 1) + \ldots + (1 \cdot \ldots \cdot 1 \cdot \pi_K N(\boldsymbol{x}|\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K)) \tag{A.19}$$

$$= \sum_{j=1}^{K} \pi_j N(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \tag{A.20}$$

which is exactly the same as equation 5.1.

A.4 The derivation of equation 5.18

Set the derivatives of the quantity in equation 5.17 with respect π_j to be 0. The equation turns to be:

$$0 = \sum_{n=1}^{N} \frac{N(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \Sigma_j)}{\sum_p \pi_p N(\boldsymbol{x}_n|\boldsymbol{\mu}_p, \Sigma_p)} + \lambda. \tag{A.21}$$

Since $\sum_{j=1}^{K} \pi_j = 1$, both sides of equation A.21 multiply π_j and sum over j would not change the equation:

$$0 = \sum_{j=1}^{K} \sum_{n=1}^{N} \frac{\pi_j N(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \Sigma_j)}{\sum_p \pi_p N(\boldsymbol{x}_n|\boldsymbol{\mu}_p, \Sigma_p)} + \sum_{j=1}^{K} \pi_j \lambda \tag{A.22}$$

$$0 = N + \lambda \tag{A.23}$$

$$\lambda = -N \tag{A.24}$$

Substitute $\lambda = -N$ into equation A.21 and multiply π_j on both sides:

$$0 = \sum_{n=1}^{N} \frac{\pi_j N(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \Sigma_j)}{\sum_p \pi_p N(\boldsymbol{x}_n|\boldsymbol{\mu}_p, \Sigma_p)} - N\pi_j \tag{A.25}$$

$$\pi_j = \frac{N_j}{N} \tag{A.26}$$

Appendix B

Alternative solutions

B.1 Alternative Gamma priors for Fermi parameters

In Section 6.2.3, log-Gaussian priors are used for the logarithm of parameters $\mathbf{\Theta}_i$. There is an alternative selection. We can use Gamma distributions as their prior distributions. To be specific, for the term $P(\mathbf{\Theta}_i|\mathbf{\Theta}_{-i},\mathbf{\Gamma},k^i,k^{-i})$, we have:

$$\begin{aligned} P(\mathbf{\Theta}_i|\mathbf{\Theta}_{-i},\mathbf{\Gamma},k^i,k^{-i}) \propto & \Gamma(A_i|\alpha_{A,k^i},\beta_{A,k^i})\Gamma(\omega_i|\alpha_{\omega,k^i},\beta_{\omega,k^i}) \\ & \Gamma(\lambda_i|\alpha_{\lambda,k^i},\beta_{\lambda,k^i})P(\mathbf{\Theta}_i|\mathbf{\Theta}_{-i},k^i,k^{-i}). \end{aligned} \tag{B.1}$$

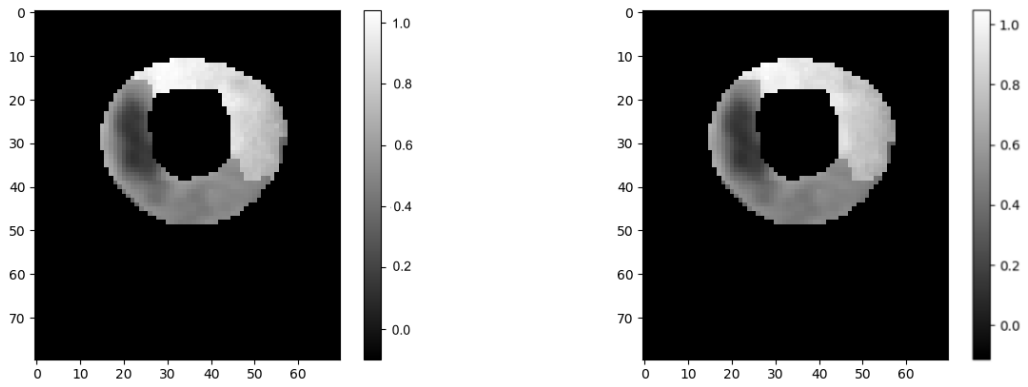
Now, we derive the conditional posterior distributions for the hyperparameters $\mathbf{\Gamma} = \{\alpha_{A,k^i},\beta_{A,k^i},\alpha_{\omega,k^i},\beta_{\omega,k^i},\alpha_{\lambda,k^i},\beta_{\lambda,k^i}\}$. $P(\mathbf{\Gamma}|\alpha^{**},\beta^{**})$ is the prior distribution for hyperparameters $\mathbf{\Gamma}$ where we let $(\alpha^{**}=0.01,\beta^{**}=0.01)$. In this way, $P(\alpha_{A,k^i},\beta_{A,k^i}|\alpha^{**},\beta^{**})$ is:

$$P(\alpha_{A,k^i},\beta_{A,k^i}|\alpha^{**},\beta^{**}) = \Gamma(\alpha_{A,k^i}|\alpha^{**},\beta^{**})\Gamma(\beta_{A,k^i}|\alpha^{**},\beta^{**}). \tag{B.2}$$

The prior distributions for other hyperparameters can be set similarly. $P(\mathbf{\Gamma}|\alpha^{**},\beta^{**})$ is the product of all of them because all these hyperparameters are independent. The conditional distribution of $\alpha_{A,k^i},\beta_{A,k^i}$ is given by

$$\begin{aligned} & P(\alpha_{A,k^i=\psi},\beta_{A,k^i=\psi}|\{A_i,A_{-i}\}_{i|k^i=\psi},k^i=\psi,\alpha^{**},\beta^{**}) \\ = \; & P\Big(\alpha_{A,k^i=\psi},\beta_{A,k^i=\psi}|\alpha^{**},\beta^{**}\Big)\prod_{i|k^i=\psi}\Gamma\Big(A_i|\alpha_{A,k^i=\psi},\beta_{A,k^i=\psi}\Big) \end{aligned} \tag{B.3}$$

where $\psi\in\{0,1\}$ is the binary lesion indicator, and $P(\alpha_{A,k^i=\psi},\beta_{A,k^i=\psi}|\alpha^{**},\beta^{**})$ can be found in equation (B.2). The conditional posterior distributions for $\{\alpha_{\omega,k^i=\psi},\beta_{\omega,k^i=\psi},\alpha_{\lambda,k^i=\psi},\beta_{\lambda,k^i=\psi}\}$



(a) log-Gaussian prior

(b) Gamma prior

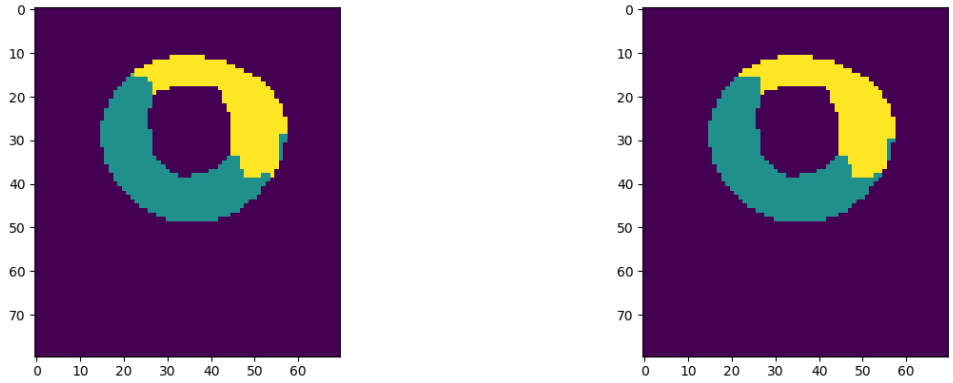
Figure B.1: MBF Comparisons between different priors. These figures show the estimations of the MBF (rescaled within [0,1]) between different prior distributions for Fermi parameters. The greyscale denotes the value of MBF. The pixels inside and outside of the ring (myocardium) are background. Panel (a) shows the estimations using log-Gaussian priors based HBM. Panel (b) shows the estimations using Gamma priors based HBM.

can be derived similarly. For $\beta_{A,k^i=\psi}$, its conditional posterior distribution is:

$$\begin{aligned} & P(\beta_{A,k^i=\psi}|\{A_i\}_{i|k^i=\psi},\alpha_{A,k^i=\psi},\alpha^{**},\beta^{**},k^i=\psi) \\ = \;& \Gamma\left(\alpha^{**}+\sum_{i|k^i=\psi}\alpha_{A,k^i=\psi},\beta^{**}+\sum_{i|k^i=\psi}A_i\right). \end{aligned} \tag{B.4}$$

For $\alpha_{A,k^i=\psi}$, the samples from its conjugate prior cannot be sampled directly. In this way, a Gamma prior is used for it and Metropolis-Hastings algorithm will be used in the sampling scheme.

Figures (B.1) and (B.2) show the comparisons using the Gamma prior and Gaussian prior given $T_{k^i}=0.1$ and $T_{\text{Fermi}}=1$. The difference in the results obtained with a log-Gaussian versus a Gamma prior was found to be minor, which suggests that the choice of functional family for the prior distributions on the Fermi parameters $\boldsymbol{\Theta}_i$ is not critical, as long as the distributions are consistent with the positivity constraint of the Fermi parameters, i.e. have positive support.



(a) log-Gaussian prior

(b) Gamma prior

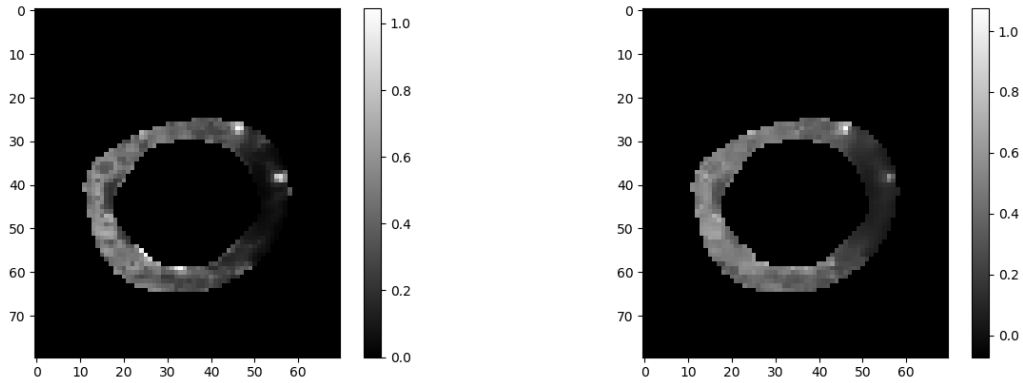
Figure B.2: Classification comparisons between different priors. The figure shows the pixel-wise myocardial tissue classification between different prior distributions for Fermi parameters. The yellow region indicates healthy tissues and dark green region indicates lesions. The blue colour marks the background of the myocardial ring. The class assignments are based on the estimated posterior class probabilities, as explained in the main text. Panel (a) shows the classification using log-Gaussian priors based HBM. Panel (b) shows the classification using Gamma priors based HBM.

Appendix C

Extra results

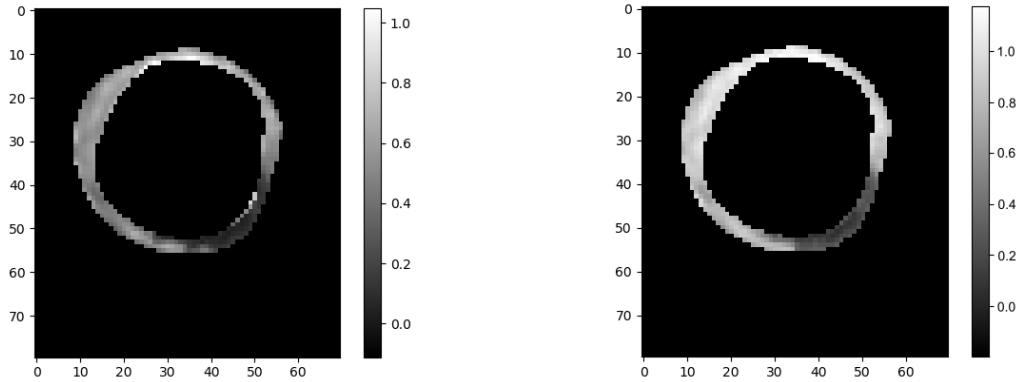
C.1 Extra results for Fermi based HBM

In this section, we show the comparisons of our Fermi based HBM method and alternative models based on 3 sets of DCE-MRI myocardial perfusion data. According to the model selection result in the main page, the values of hyperparameters are chosen to be $T_{k^i} = 0.1$ and $T_{\text{Fermi}} = 1$. Figures (C.1), (C.2) and (C.3) show the comparisons of the MBF estimations between the Fermi model and the HBM method proposed in this work. The performances of the comparisons are consistent with the results in Chapter 6. Figures (C.4), (C.5) and (C.6) show the comparisons of the classifications bewtween the GMM based on Fermi and the HBM method proposed in Chapter 6. The performances of the comparisons are consistent with the results in Chapter 6.



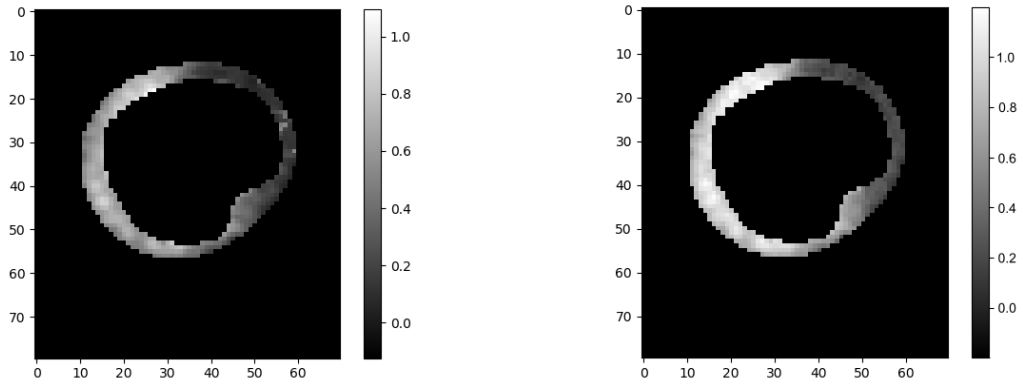
(a) Fermi model (b) proposed HBM

Figure C.1: MBF estimations for DCE-MRI dataset 25. These figures show the estimations of the MBF (rescaled within [0,1]). The greyscale denotes the value of MBF. The pixels inside and outside of the ring (myocardium) are background. Panel (a) shows the Fermi model fitted by least-squares. Panel (b) shows the MBF estimations using the HBM illustrated in this thesis.



(a) Fermi model (b) proposed HBM

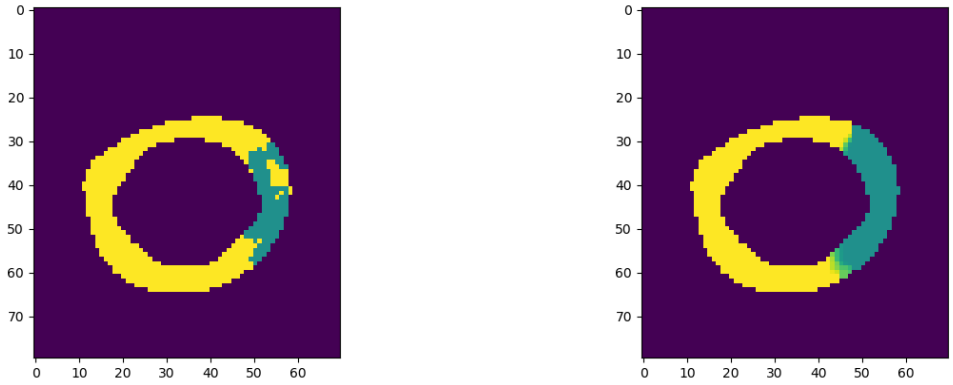
Figure C.2: MBF estimations for DCE-MRI dataset 26. These figures show the estimations of the MBF (rescaled within [0,1]). The greyscale denotes the value of MBF. The pixels inside and outside of the ring (myocardium) are background. Panel (a) shows the Fermi model fitted by least-squares. Panel (b) shows the MBF estimations using the HBM illustrated in this thesis.



(a) Fermi model

(b) proposed HBM

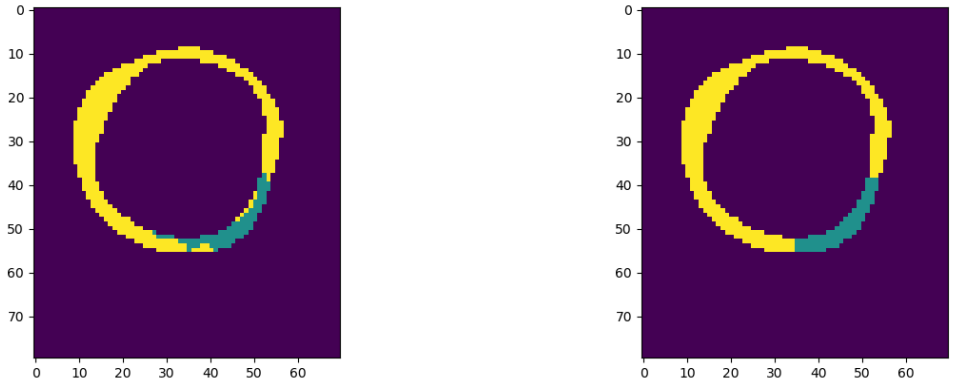
Figure C.3: MBF estimations for DCE-MRI dataset 27. These figures show the estimations of the MBF (rescaled within [0,1]). The greyscale denotes the value of MBF. The pixels inside and outside of the ring (myocardium) are background. Panel (a) shows the Fermi model fitted by least-squares. Panel (b) shows the MBF estimations using the HBM illustrated in this thesis.



(a) GMM based on Fermi model

(b) proposed HBM

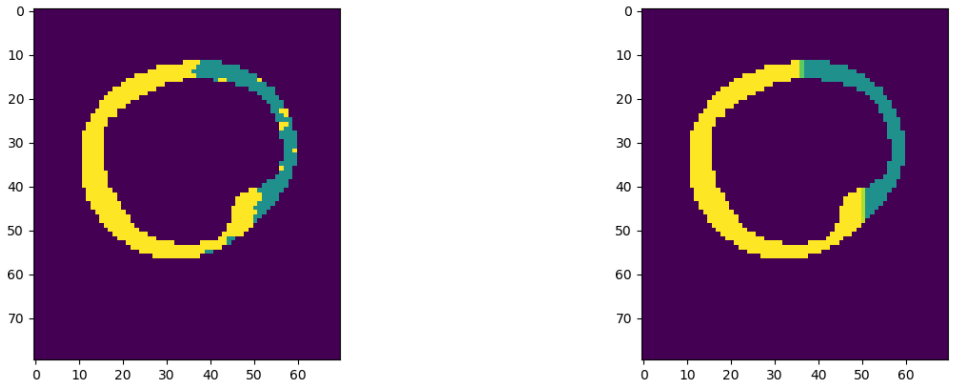
Figure C.4: The classifications for DCE-MRI dataset 25. The figure shows the pixel-wise myocardial tissue classification into the three classes *lesion* (dark green), *healthy* (yellow) and *uncertain* (light green). The blue colour marks the background of the myocardial ring. The class assignments are based on the estimated posterior class probabilities, as explained in the main text. Panel (a) shows the classification based on Fermi estimated MBF using Gaussian Mixture Model. Panel (b) shows the classification result based on the HBM illustrated in this work.



(a) GMM based on Fermi model

(b) proposed HBM

Figure C.5: The classifications for DCE-MRI dataset 26. The figure shows the pixel-wise myocardial tissue classification into the three classes *lesion* (dark green), *healthy* (yellow) and *uncertain* (light green). The blue colour marks the background of the myocardial ring. The class assignments are based on the estimated posterior class probabilities, as explained in the main text. Panel (a) shows the classification based on Fermi estimated MBF using Gaussian Mixture Model. Panel (b) shows the classification result based on the HBM illustrated in this work.



(a) GMM based on Fermi model

(b) proposed HBM

Figure C.6: The classifications for DCE-MRI dataset 27. The figure shows the pixel-wise myocardial tissue classification into the three classes *lesion* (dark green), *healthy* (yellow) and *uncertain* (light green). The blue colour marks the background of the myocardial ring. The class assignments are based on the estimated posterior class probabilities, as explained in the main text. Panel (a) shows the classification based on Fermi estimated MBF using Gaussian Mixture Model. Panel (b) shows the classification result based on the HBM illustrated in this work.

Appendix D

Extra tables

Table D.1: Commonly used kernel functions reviewed in Chapter 4, [2]

kernel function	expression
Constant	σ_0^2
Linear	$\sum_{d=1}^{D} \sigma_d^2 x_d x_d^{'}$
polynomial	$(\boldsymbol{x} \cdot \boldsymbol{x}^{'} + \sigma_0^2)^p$
squared exponential	$\exp(-\frac{r^2}{2l^2})$
Matern	$\frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{l} r\right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}}{l} r\right)$
exponential	$\exp(-\frac{r}{l})$
γ-exponential	$\exp\left(-\left(\frac{r}{l}\right)^{\gamma}\right)$
rational quadratic	$(1+\frac{r^2}{2\alpha l^2})^{-\alpha}$
neural network	$\sin^{-1}(\frac{2\tilde{x}^{\intercal}\Sigma\tilde{x}^{'}}{\sqrt{(1+2\tilde{x}^{\intercal}\Sigma\tilde{x})(1+2\tilde{x}^{'\intercal}\Sigma\tilde{x}^{'})}})$

. The r in the table denotes $|x - x^{'}|$. All other symbols represent hyperparameters of the kernel function.

Table D.2: Likelihoods that has common probability distributions and their conjugate prior distributions [3]

likelihood	model parameters	conjugate prior	prior hyperparameters	posterior hyperparameters
Binomial	p	Beta	α,β	$\alpha+\sum_{i=1}^{n}x_i$, $\beta+\sum_{i=1}^{n}N_i-\sum_{i=1}^{n}x_i$
Normal (σ^2 known)	μ	Normal	μ_0, σ_0^2	$\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n}x_i}{\sigma^2}\right)$, $\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}$
Normal (μ known)	σ^2	Inverse Gamma	α, β	$\alpha+\frac{n}{2}$, $\beta+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}$
Exponential	λ	Gamma	α, β	$\alpha+n$, $\beta+\sum_{i=1}^{n}x_i$
Gamma (α known)	β	Gamma	α_0, β_0	$\alpha_0+n\alpha$, $\beta_0+\sum_{i=1}^{n}x_i$
Gamma (β known)	α	$\propto\frac{a^{\alpha-1}\beta^{\alpha c}}{\Gamma(\alpha)^b}$	a, b, c	$a+\prod_{i=1}^{n}x_i$, $b+n$, $c+n$

Bibliography

[1] K. R. Nandalur, B. A. Dwamena, A. F. Choudhri, M. R. Nandalur, R. C. Carlos, Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis, Journal of the American College of Cardiology 50 (14) (2007) 1343–1353.

[2] C. E. Rasmussen, K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, MA., 2006.

[3] D. Fink, A compendium of conjugate priors (1997).

[4] A. H. A. W. G. on Myocardial Segmentation, R. for Cardiac Imaging:, M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, et al., Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association, Circulation 105 (4) (2002) 539–542.

[5] C. M. Scannell, A. Chiribiri, A. D. Villa, M. Breeuwer, J. Lee, Hierarchical Bayesian myocardial perfusion quantification, Medical image analysis 60 (2020) 101611.

[6] D. W. Scott, Multivariate density estimation: theory, practice, and visualization, John Wiley & Sons, 2015.

[7] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, A. Alonso, A. Z. Beaton, M. S. Bittencourt, A. K. Boehme, A. E. Buxton, A. P. Carson, Y. Commodore-Mensah, et al., Heart disease and stroke statistics—2022 update: a report from the american heart association, Circulation 145 (8) (2022) e153–e639.

[8] M. Jerosch-Herold, Quantification of myocardial perfusion by cardiovascular magnetic resonance, Journal of Cardiovascular Magnetic Resonance 12 (1) (2010) 1–16.

[9] D. A. Reynolds, Gaussian mixture models., Encyclopedia of biometrics 741 (659-663) (2009).

[10] S. Sanjay-Gopal, T. J. Hebert, Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm, IEEE Transactions on Image Processing 7 (7) (1998) 1014–1028.

[11] T. M. Nguyen, Q. J. Wu, Fast and robust spatially constrained Gaussian mixture model for image segmentation, IEEE transactions on circuits and systems for video technology 23 (4) (2012) 621–635.

[12] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis, Chapman and Hall/CRC, 1995.

[13] E. T. Jaynes, Prior probabilities, IEEE Transactions on systems science and cybernetics 4 (3) (1968) 227–241.

[14] E. Süli, D. F. Mayers, An introduction to numerical analysis, Cambridge university press, 2003.

[15] J. Nocedal, S. Wright, Numerical optimization, Springer Science & Business Media, 2006.

[16] C. Andrieu, A. Doucet, Simulated annealing for maximum a posteriori parameter estimation of hidden Markov models, IEEE Transactions on Information Theory 46 (3) (2000) 994–1004.

[17] J. Besag, On the statistical analysis of dirty pictures, Journal of the Royal Statistical Society: Series B (Methodological) 48 (3) (1986) 259–279.

[18] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.

[19] A. P. Dawid, Conditional independence for statistical operations, The Annals of Statistics 8 (3) (1980) 598–617.

[20] P. Clifford, Markov random fields in statistics, Disorder in physical systems: A volume in honour of John M. Hammersley (1990) 19–32.

[21] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics 21 (6) (1953) 1087–1092.

[22] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications (1970).

[23] C. P. Robert, G. Casella, G. Casella, Introducing Monte Carlo methods with r, Vol. 18, Springer, 2010.

[24] J. Geweke, et al., Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, Vol. 196, Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, 1991.

[25] A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences, Statistical science 7 (4) (1992) 457–472.

[26] M. K. Cowles, B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: a comparative review, Journal of the American Statistical Association 91 (434) (1996) 883–904.

[27] C. J. Geyer, Practical Markov chain Monte Carlo, Statistical science (1992) 473–483.

[28] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39 (1) (1977) 1–22.

[29] G. J. McLachlan, T. Krishnan, The EM algorithm and extensions, Vol. 382, John Wiley & Sons, 2007.

[30] S. Kullback, R. A. Leibler, On information and sufficiency, The annals of mathematical statistics 22 (1) (1951) 79–86.

[31] M. Jamshidian, R. I. Jennrich, Conjugate gradient acceleration of the EM algorithm, Journal of the American Statistical Association 88 (421) (1993) 221–228.

[32] X.-L. Meng, D. B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework, Biometrika 80 (2) (1993) 267–278.

[33] R. M. Neal, G. E. Hinton, A view of the EM algorithm that justifies incremental, sparse, and other variants, in: Learning in graphical models, Springer, 1998, pp. 355–368.

[34] R. B. Gramacy, Surrogates: Gaussian process modeling, design, and optimization for the applied sciences, Chapman and Hall/CRC, 2020.

[35] T. Gneiting, A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, Journal of the American statistical Association 102 (477) (2007) 359–378.

[36] I. J. Good, Rational decisions, in: Breakthroughs in statistics, Springer, 1992, pp. 365–377.

[37] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Selected papers of hirotugu akaike, Springer, 1998, pp. 199–213.

[38] J. Schwitter, D. Nanz, S. Kneifel, K. Bertschinger, M. Buchi, P. Knusel, B. Marincek, T. Luscher, G. K. von Schulthess, Assessment of myocardial perfusion in coronary artery disease by magnetic resonance: a comparison with positron emission tomography and coronary angiography, Circulation 103 (18) (2001) 2230–2235.

[39] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit, Journal of the royal statistical society: Series b (statistical methodology) 64 (4) (2002) 583–639.

[40] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models, Statistics and computing 24 (6) (2014) 997–1016.

[41] S. Watanabe, A widely applicable Bayesian information criterion, Journal of Machine Learning Research 14 (2013) 867–893.

[42] M. Stone, Cross-validatory choice and assessment of statistical predictions, Journal of the royal statistical society: Series B (Methodological) 36 (2) (1974) 111–133.

[43] J. E. Hall, M. E. Hall, Guyton and Hall textbook of medical physiology e-Book, Elsevier Health Sciences, 2020.

[44] T. Vos, C. Allen, M. Arora, R. M. Barber, Z. A. Bhutta, A. Brown, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015, The lancet 388 (10053) (2016) 1545–1602.

[45] I. Euroaspire, Lifestyle and risk-factor management and use of drug therapies in coronary patients from 15 countries, European heart journal 22 (7) (2001) 554–72.

[46] D. McCormick, J. H. Gurwitz, D. Lessard, J. Yarzebski, J. M. Gore, R. J. Goldberg, Use of aspirin, β-blockers, and lipid-lowering medications before recurrent acute myocardial infarction: missed opportunities for prevention?, Archives of internal medicine 159 (6) (1999) 561–567.

[47] S. Yusuf, S. Macmahon, R. Collins, R. Peto, Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials, The Lancet 331 (8594) (1988) 1088–1092.

[48] R. J. Kim, E. Wu, A. Rafael, E.-L. Chen, M. A. Parker, O. Simonetti, F. J. Klocke, R. O. Bonow, R. M. Judd, The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction, New England Journal of Medicine 343 (20) (2000) 1445–1453.

[49] V. Gupta, H. A. Kirişli, E. A. Hendriks, R. J. van der Geest, M. van de Giessen, W. Niessen, J. H. Reiber, B. P. Lelieveldt, Cardiac MR perfusion image processing techniques: a survey, Medical image analysis 16 (4) (2012) 767–785.

[50] L. Axel, J. Costantini, J. Listerud, Intensity correction in surface-coil MR imaging, American Journal of Roentgenology 148 (2) (1987) 418–420.

[51] A. E. Arai, Magnetic resonance first-pass myocardial perfusion imaging, Topics in magnetic resonance imaging 11 (6) (2000) 383–398.

[52] W. G. Schreiber, M. Schmitt, P. Kalden, O. K. Mohrs, K.-F. Kreitner, M. Thelen, Dynamic contrast-enhanced myocardial perfusion imaging using saturation-prepared TrueFISP, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 16 (6) (2002) 641–652.

[53] E. Di Bella, D. Parker, A. Sinusas, On the dark rim artifact in dynamic contrast-enhanced mri myocardial perfusion studies, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 54 (5) (2005) 1295–1299.

[54] P. Kellman, A. E. Arai, Imaging sequences for first pass perfusion—a review, Journal of Cardiovascular Magnetic Resonance 9 (3) (2007) 525–537.

[55] H. Köstler, C. Ritter, M. Reiss-Zimmermann, M. Beer, D. Hahn, J. Sandstede, Correction for partial volume errors in MR heart perfusion imaging, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 51 (4) (2004) 848–852.

[56] P. Storey, Q. Chen, W. Li, R. R. Edelman, P. V. Prasad, Band artifacts due to bulk motion, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 48 (6) (2002) 1028–1036.

[57] M. Fenchel, U. Helber, O. P. Simonetti, N. I. Stauder, U. Kramer, C.-N. Nguyen, J. P. Finn, C. D. Claussen, S. Miller, Multislice first-pass myocardial perfusion imaging: Comparison of saturation recovery (SR)-TrueFISP-two-dimensional (2D) and SR-TurboFLASH-2D pulse sequences, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 19 (5) (2004) 555–563.

[58] C. L. Schultz, R. Alfidi, A. D. Nelson, S. Y. Kopiwoda, M. E. Clampitt, The effect of motion on two-dimensional Fourier transformation magnetic resonance images., Radiology 152 (1) (1984) 117–121.

[59] A. M. Weissler, W. S. Harris, C. D. Schoenfeld, Systolic time intervals in heart failure in man, Circulation 37 (2) (1968) 149–159.

[60] A. E. Holland, J. W. Goldfarb, R. R. Edelman, Diaphragmatic and cardiac motion during suspended breathing: preliminary experience and implications for breath-hold MR imaging., Radiology 209 (2) (1998) 483–489.

[61] A. D. Scott, J. Keegan, D. N. Firmin, Motion in cardiovascular MR imaging, Radiology 250 (2) (2009) 331–351.

[62] K. McLeish, D. L. Hill, D. Atkinson, J. M. Blackall, R. Razavi, A study of the motion and deformation of the heart due to respiration, IEEE transactions on medical imaging 21 (9) (2002) 1142–1150.

[63] M. A. Costa, S. Shoemaker, H. Futamatsu, C. Klassen, D. J. Angiolillo, M. Nguyen, A. Siuciak, P. Gilmore, M. M. Zenni, L. Guzman, et al., Quantitative magnetic resonance perfusion imaging detects anatomic and physiologic coronary artery disease as measured by coronary angiography and fractional flow reserve, Journal of the American College of Cardiology 50 (6) (2007) 514–522.

[64] A. Huber, S. Sourbron, V. Klauss, J. Schaefer, K. U. Bauner, M. Schweyer, M. Reiser, E. Rummeny, J. Rieber, Magnetic resonance perfusion of the myocardium: semiquantitative and quantitative evaluation in comparison with coronary angiography and fractional flow reserve, Investigative radiology 47 (6) (2012) 332–338.

[65] E. Nagel, C. Klein, I. Paetsch, S. Hettwer, B. Schnackenburg, K. Wegscheider, E. Fleck, Magnetic resonance perfusion measurements for the noninvasive detection of coronary artery disease, Circulation 108 (4) (2003) 432–437.

[66] N. Al-Saadi, E. Nagel, M. Gross, A. Bornstedt, B. Schnackenburg, C. Klein, W. Klimek, H. Oswald, E. Fleck, Noninvasive detection of myocardial ischemia from perfusion reserve based on cardiovascular magnetic resonance, Circulation 101 (12) (2000) 1379–1383.

[67] T. Giang, D. Nanz, R. Coulden, M. Friedrich, M. Graves, N. Al-Saadi, T. Lüscher, G. Von Schulthess, J. Schwitter, Detection of coronary artery disease by magnetic resonance myocardial perfusion imaging with various contrast medium doses: first European multi-centre experience, European heart journal 25 (18) (2004) 1657–1665.

[68] J. Rieber, A. Huber, I. Erhard, S. Mueller, M. Schweyer, A. Koenig, T. M. Schiele, K. Theisen, U. Siebert, S. O. Schoenberg, et al., Cardiac magnetic resonance perfusion imaging for the functional assessment of coronary artery disease: a comparison with coronary angiography and fractional flow reserve, European heart journal 27 (12) (2006) 1465–1471.

[69] S. Plein, A. Radjenovic, J. P. Ridgway, D. Barmby, J. P. Greenwood, S. G. Ball, M. U. Sivananthan, Coronary artery disease: myocardial perfusion MR imaging with sensitivity encoding versus conventional angiography, Radiology 235 (2) (2005) 423–430.

[70] A. Larghat, J. Biglands, N. Maredia, J. P. Greenwood, S. G. Ball, M. Jerosch-Herold, A. Radjenovic, S. Plein, Endocardial and epicardial myocardial perfusion determined by semi-quantitative and quantitative myocardial perfusion magnetic resonance, The international journal of cardiovascular imaging 28 (6) (2012) 1499–1511.

[71] L. Axel, Tissue mean transit time from dynamic computed tomography by a simple deconvolution technique., Investigative radiology 18 (1) (1983) 94–99.

[72] M. Jerosch-Herold, N. Wilke, A. E. Stillman, R. F. Wilson, Magnetic resonance quantification of the myocardial perfusion reserve with a Fermi function model for constrained deconvolution, Medical physics 25 (1) (1998) 73–84.

[73] D. A. Broadbent, J. D. Biglands, A. Larghat, S. P. Sourbron, A. Radjenovic, J. P. Greenwood, S. Plein, D. L. Buckley, Myocardial blood flow at rest and stress measured with dynamic contrast-enhanced MRI: Comparison of a distributed parameter model with a fermi function model, Magnetic resonance in medicine 70 (6) (2013) 1591–1597.

[74] M. Jerosch-Herold, C. Swingen, R. T. Seethamraju, Myocardial blood flow quantification with MRI by model-independent deconvolution, Medical physics 29 (5) (2002) 886–897.

[75] S. L. Keeling, T. Kogler, R. Stollberger, Deconvolution for DCE-MRI using an exponential approximation basis, Medical image analysis 13 (1) (2009) 80–90.

[76] P. Batchelor, A. Chiribiri, N. Z. Nooralipour, Z. Cvetković, ARMA regularization of cardiac perfusion modeling, in: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2010, pp. 642–645.

[77] S. Sourbron, D. L. Buckley, Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability, Physics in Medicine & Biology 57 (2) (2011) R1.

[78] S. P. Sourbron, D. L. Buckley, On the scope and interpretation of the Tofts models for DCE-MRI, Magnetic resonance in medicine 66 (3) (2011) 735–745.

[79] J.-P. M. Vallée, H. D. Sostman, J. R. Macfall, T. R. Degrado, J. Zhang, L. Sebbag, F. R. Cobb, T. Wheeler, L. W. Hedlund, T. G. Turkington, et al., Quantification of myocardial perhsion by MRI after coronary occlusion, Magnetic resonance in medicine 40 (2) (1998) 287–297.

[80] J. D. Biglands, D. R. Magee, S. P. Sourbron, S. Plein, J. P. Greenwood, A. Radjenovic, Comparison of the diagnostic performance of four quantitative myocardial perfusion estimation methods used in cardiac MR imaging: CE-MARC substudy, Radiology 275 (2) (2015) 393–402.

[81] T. Ichihara, M. Ishida, K. Kitagawa, Y. Ichikawa, T. Natsume, N. Yamaki, H. Maeda, K. Takeda, H. Sakuma, Quantitative analysis of first-pass contrast-enhanced myocardial perfusion MRI using a patlak plot method and blood saturation correction, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 62 (2) (2009) 373–383.

[82] N. A. Pack, E. V. DiBella, Comparison of myocardial perfusion estimates from dynamic contrast-enhanced magnetic resonance imaging with four quantitative analysis methods, Magnetic Resonance in Medicine 64 (1) (2010) 125–137.

[83] A. Cernicanu, L. Axel, Theory-based signal calibration with single-point T1 measurements for first-pass quantitative perfusion MRI studies, Academic radiology 13 (6) (2006) 686–693.

[84] W. Utz, T. Niendorf, R. Wassmuth, D. Messroghli, R. Dietz, J. Schulz-Menger, Contrast–dose relation in first-pass myocardial MR perfusion imaging, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 25 (6) (2007) 1131–1135.

[85] J. D. Biglands, Quanitifying myocardial blood flow using dynamic contrast enhanced cardiac magnetic resonance imaging, University of Leeds, 2012.

[86] P. S. Tofts, 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) (1991) 357–367.

[87] H. K. Thompson Jr, C. F. Starmer, R. E. Whalen, H. D. McIntosh, Indicator transit time considered as a gamma variate, Circulation research 14 (6) (1964) 502–515.

[88] M. Jerosch-Herold, R. T. Seethamraju, C. M. Swingen, N. M. Wilke, A. E. Stillman, Analysis of myocardial perfusion MRI, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 19 (6) (2004) 758–770.

[89] N. Wilke, M. Jerosch-Herold, Assessing myocardial perfusion in coronary artery disease with magnetic resonance first-pass imaging, Cardiology clinics 16 (2) (1998) 227–246.

[90] J. Barkhausen, P. Hunold, M. Jochims, J. F. Debatin, Imaging of myocardial perfusion with magnetic resonance, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 19 (6) (2004) 750–757.

[91] L.-Y. Hsu, K. L. Rhoads, J. E. Holly, P. Kellman, A. H. Aletras, A. E. Arai, Quantitative myocardial perfusion analysis with a dual-bolus contrast-enhanced first-pass mri technique in humans, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 23 (3) (2006) 315–322.

[92] M. Jerosch-Herold, X. Hu, N. S. Murthy, C. Rickers, A. E. Stillman, Magnetic resonance imaging of myocardial contrast enhancement with MS-325 and its relation to myocardial blood flow and the perfusion reserve, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 18 (5) (2003) 544–554.

[93] J. T. Keijer, A. C. van Rossum, M. J. van Eenige, A. J. Karreman, M. B. Hofman, J. Valk, C. A. Visser, Semiquantitation of regional myocardial blood flow in normal human subjects by first-pass magnetic resonance imaging, American heart journal 130 (4) (1995) 893–901.

[94] N. Al-Saadi, E. Nagel, M. Gross, B. Schnackenburg, I. Paetsch, C. Klein, E. Fleck, Improvement of myocardial perfusion reserve early after coronary intervention: assessment with cardiac magnetic resonance imaging, Journal of the American College of Cardiology 36 (5) (2000) 1557–1564.

[95] P. Meier, K. L. Zierler, On the theory of the indicator-dilution method for measurement of blood flow and volume, Journal of applied physiology 6 (12) (1954) 731–744.

[96] K. L. Zierler, Equations for measuring blood flow by external monitoring of radioisotopes, Circulation research 16 (4) (1965) 309–321.

[97] V. J. Schmid, P. D. Gatehouse, G.-Z. Yang, Attenuation resilient AIF estimation based on hierarchical Bayesian modelling for first pass myocardial perfusion MRI, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2007, pp. 393–400.

[98] V. J. Schmid, Voxel-based adaptive spatio-temporal modelling of perfusion cardiovascular MRI, IEEE transactions on medical imaging 30 (7) (2011) 1305–1313.

[99] B. Neyran, M. F. Janier, C. Casali, D. Revel, E. P. Canet Soulas, Mapping myocardial perfusion with an intravascular MR contrast agent: robustness of deconvolution methods at various blood flows, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 48 (1) (2002) 166–179.

[100] H. Larsson, M. Stubgaard, J. Frederiksen, M. Jensen, O. Henriksen, O. Paulson, Quantitation of blood-brain barrier defect by magnetic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumors, Magnetic Resonance in Medicine 16 (1) (1990) 117–131.

[101] G. Brix, W. Semmler, R. Port, L. R. Schad, G. Layer, W. J. Lorenz, Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging., Journal of computer assisted tomography 15 (4) (1991) 621–628.

[102] P. S. Tofts, G. Brix, D. L. Buckley, J. L. Evelhoch, E. Henderson, M. V. Knopp, H. B. Larsson, T.-Y. Lee, N. A. Mayr, G. J. Parker, et al., Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 10 (3) (1999) 223–232.

[103] V. J. Schmid, B. Whitcher, A. R. Padhani, N. J. Taylor, G.-Z. Yang, Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging, IEEE transactions on medical imaging 25 (12) (2006) 1627–1636.

[104] N. Dikaios, D. Atkinson, C. Tudisca, P. Purpura, M. Forster, H. Ahmed, T. Beale, M. Emberton, S. Punwani, A comparison of bayesian and non-linear regression methods for robust estimation of pharmacokinetics in DCE-MRI and how it affects cancer diagnosis, Computerized Medical Imaging and Graphics 56 (2017) 1–10.

[105] J. Lehnert, C. Kolbitsch, G. Wubbeler, A. Chiribiri, T. Schaeffter, C. Elster, Large-scale Bayesian spatial-temporal regression with application to cardiac MR-perfusion imaging, SIAM Journal on Imaging Sciences 12 (4) (2019) 2035–2062.

[106] J. D. Biglands, M. Ibraheem, D. R. Magee, A. Radjenovic, S. Plein, J. P. Greenwood, Quantitative myocardial perfusion imaging versus visual analysis in diagnosing myocardial ischemia: a CE-MARC substudy, JACC: Cardiovascular Imaging 11 (5) (2018) 711–718.

[107] R. C. Cury, C. A. Cattani, L. A. Gabure, D. J. Racy, J. M. de Gois, U. Siebert, S. S. Lima, T. J. Brady, Diagnostic performance of stress perfusion and delayed-enhancement MR imaging in patients with coronary artery disease, Radiology 240 (1) (2006) 39–45.

[108] J. P. Greenwood, N. Maredia, J. F. Younger, J. M. Brown, J. Nixon, C. C. Everett, P. Bijsterveld, J. P. Ridgway, A. Radjenovic, C. J. Dickinson, et al., Cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary heart disease (CE-MARC): a prospective trial, The Lancet 379 (9814) (2012) 453–460.

[109] N. Ishida, H. Sakuma, M. Motoyasu, T. Okinaka, N. Isaka, T. Nakano, K. Takeda, Noninfarcted myocardium: correlation between dynamic first-pass contrast-enhanced myocardial MR imaging and quantitative coronary angiography, Radiology 229 (1) (2003) 209–216.

[110] K. Yoshio, N. Masaki, H. Katsunori, O. Kazue, Y. Junichi, Assessment of coronary artery disease with nicorandil stress magnetic resonance imaging, Osaka city medical journal 50 (2) (2004) 87–94.

[111] I. Klem, J. F. Heitner, D. J. Shah, M. H. Sketch, V. Behar, J. Weinsaft, P. Cawley, M. Parker, M. Elliott, R. M. Judd, et al., Improved detection of coronary artery disease by stress perfusion cardiovascular magnetic resonance with the use of delayed enhancement infarction imaging, Journal of the American College of Cardiology 47 (8) (2006) 1630–1638.

[112] I. Klem, S. Greulich, J. F. Heitner, H. Kim, H. Vogelsberg, E.-M. Kispert, S. R. Ambati, C. Bruch, M. Parker, R. M. Judd, et al., Value of cardiovascular magnetic resonance stress perfusion testing for the detection of coronary artery disease in women, JACC: Cardiovascular Imaging 1 (4) (2008) 436–445.

[113] I. Paetsch, C. Jahnke, A. Wahl, R. Gebker, M. Neuss, E. Fleck, E. Nagel, Comparison of dobutamine stress magnetic resonance, adenosine stress magnetic resonance, and adenosine stress magnetic resonance perfusion, Circulation 110 (7) (2004) 835–842.

[114] A. R. Patel, P. F. Antkowiak, K. R. Nandalur, A. M. West, M. Salerno, V. Arora, J. Christopher, F. H. Epstein, C. M. Kramer, Assessment of advanced coronary artery disease: advantages of quantitative cardiac magnetic resonance perfusion analysis, Journal of the American College of Cardiology 56 (7) (2010) 561–569.

[115] G. Pilz, P. Bernhardt, M. Klos, E. Ali, M. Wild, B. Höfling, Clinical implication of adenosine-stress cardiac magnetic resonance imaging as potential gatekeeper prior to invasive examination in patients with AHA/ACC class II indication for coronary angiography, Clinical research in cardiology 95 (10) (2006) 531–538.

[116] S. Plein, J. P. Greenwood, J. P. Ridgway, G. Cranny, S. G. Ball, M. U. Sivananthan, Assessment of non–ST-segment elevation acute coronary syndromes with cardiac magnetic resonance imaging, Journal of the American College of Cardiology 44 (11) (2004) 2173–2181.

[117] H. Sakuma, N. Suzawa, Y. Ichikawa, K. Makino, T. Hirano, K. Kitagawa, K. Takeda, Diagnostic accuracy of stress first-pass contrast-enhanced myocardial perfusion MRI compared with stress myocardial perfusion scintigraphy, American Journal of Roentgenology 185 (1) (2005) 95–102.

[118] J. Schwitter, C. M. Wacker, A. C. van Rossum, M. Lombardi, N. Al-Saadi, H. Ahlstrom, T. Dill, H. B. Larsson, S. D. Flamm, M. Marquardt, et al., MR-IMPACT: comparison of perfusion-cardiac magnetic resonance with single-photon emission computed tomography for the detection of coronary artery disease in a multicentre, multivendor, randomized trial, European heart journal 29 (4) (2008) 480–489.

[119] J. Schwitter, C. M. Wacker, N. Wilke, N. Al-Saadi, E. Sauer, K. Huettle, S. O. Schönberg, A. Luchner, O. Strohm, H. Ahlstrom, et al., MR-IMPACT II: Magnetic resonance imaging for myocardial perfusion assessment in coronary artery disease trial: perfusion-cardiac magnetic resonance vs. single-photon emission computed tomography for the detection of coronary artery disease: a comparative multicentre, multivendor trial, European heart journal 34 (10) (2013) 775–781.

[120] B. Takase, M. Nagata, T. Kihara, A. Kameyawa, K. Noya, T. Matsui, F. Ohsuzu, M. Ishihara, A. Kurita, Whole-heart dipyridamole stress first-pass myocardial perfusion MRI for the detection of coronary artery disease, Japanese heart journal 45 (3) (2004) 475–486.

[121] S. Watkins, R. McGeoch, J. Lyne, T. Steedman, R. Good, M.-J. McLaughlin, T. Cunningham, V. Bezlyak, I. Ford, H. J. Dargie, et al., Validation of magnetic resonance myocardial perfusion imaging with fractional flow reserve for the detection of significant coronary heart disease, Circulation 120 (22) (2009) 2207–2213.

[122] S. Wolff, J. Schwitter, R. Coulden, M. Friedrich, D. Bluemke, R. Biederman, E. Martin, A. Lansky, F. Kashanian, T. Foo, et al., Myocardial first-pass perfusion magnetic resonance imaging: a multicenter dose-ranging study, Circulation 110 (6) (2004) 732–737.

[123] M. Doyle, A. Fuisz, E. Kortright, R. W. Biederman, E. G. Walsh, E. T. Martin, L. Tauxe, W. J. Rogers, C. N. Bairey Merz, C. Pepine, et al., The impact of myocardial flow reserve on the detection of coronary artery disease by perfusion imaging methods: an NHLBI wise study: Perfusion imaging, Journal of Cardiovascular Magnetic Resonance 5 (3) (2003) 475–485.

[124] K. L. Zierler, Theoretical basis of indicator-dilution methods for measuring flow and volume, Circulation Research 10 (3) (1962) 393–407.

[125] F. G. Zöllner, S. Li, J. Roervik, A. Lundervold, L. R. Schad, Segmentation of renal compartments in DCE-MRI of human kidney, in: 2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, 2011, pp. 744–748.

[126] S. H. Han, E. Ackerstaff, R. Stoyanova, S. Carlin, W. Huang, J. Koutcher, J. Kim, G. Cho, G. Jang, H. Cho, Gaussian mixture model-based classification of dynamic contrast enhanced MRI data for identifying diverse tumor microenvironments: preliminary results, NMR in Biomedicine 26 (5) (2013) 519–532.

[127] M. Aristophanous, B. C. Penney, M. K. Martel, C. A. Pelizzari, A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography, Medical physics 34 (11) (2007) 4223–4235.

[128] M. R. Divine, P. Katiyar, U. Kohlhofer, L. Quintanilla-Martinez, B. J. Pichler, J. A. Disselhorst, A population-based Gaussian mixture model incorporating 18F-FDG PET and diffusion-weighted MRI quantifies tumor tissue classes, Journal of Nuclear Medicine 57 (3) (2016) 473–479.

[129] L. Tesar, D. Smutek, Ultrasonography diagnostics using Gaussian mixture model, in: 6th International PhD Workshop on Systems and Control. Izola, Slovenia: Jozef Stefan Institute, 2005.

[130] D. Fabre, T. Hueber, L. Girin, X. Alameda-Pineda, P. Badin, Automatic animation of an articulatory tongue model from ultrasound images of the vocal tract, Speech Communication 93 (2017) 63–75.

[131] Y. Yang, H. Gao, C. Berry, A. Radjenovic, D. Husmeier, Quantification of myocardial perfusion lesions using spatially variant finite mixture modelling of DCE-MRI, Proceedings of the International Conference on Statistics: Theory and Applications (ICSTA) (2019).

[132] T. Layer, M. Blaickner, B. Knäusl, D. Georg, J. Neuwirth, R. P. Baum, C. Schuchardt, S. Wiessalla, G. Matz, Pet image segmentation using a Gaussian mixture model and Markov random fields, EJNMMI physics 2 (1) (2015) 1–15.

[133] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximization algorithm, IEEE transactions on medical imaging 20 (1) (2001) 45–57.

[134] A. Banerjee, P. Maji, Rough sets and stomped normal distribution for simultaneous segmentation and bias field correction in brain MR images, IEEE Transactions on Image Processing 24 (12) (2015) 5764–5776.

[135] Z. Ji, Y. Xia, Y. Zheng, Robust generative asymmetric GMM for brain MR image segmentation, Computer methods and programs in biomedicine 151 (2017) 123–138.

[136] F. Dong, J. Peng, Brain MR image segmentation based on local Gaussian mixture model and nonlocal spatial regularization, Journal of Visual Communication and Image Representation 25 (5) (2014) 827–839.

[137] G. J. McLachlan, K. E. Basford, Mixture models: Inference and applications to clustering, Vol. 38, M. Dekker New York, 1988.

[138] G. J. McLachlan, S. X. Lee, S. I. Rathnayake, Finite mixture models, Annual review of statistics and its application 6 (2019) 355–378.

[139] K. Krishna, M. N. Murty, Genetic K-means algorithm, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29 (3) (1999) 433–439.

[140] P. J. Green, Bayesian reconstructions from emission tomography data using a modified EM algorithm, IEEE transactions on medical imaging 9 (1) (1990) 84–93.

[141] C. Nikou, N. P. Galatsanos, A. C. Likas, A class-adaptive spatially variant mixture model for image segmentation, IEEE Transactions on Image Processing 16 (4) (2007) 1121–1130.

[142] W. Li, S. Prasad, J. E. Fowler, Hyperspectral image classification using Gaussian mixture models and Markov random fields, IEEE Geoscience and Remote Sensing Letters 11 (1) (2013) 153–157.

[143] M. Welling, K. Kurihara, Bayesian K-means as a "maximization-expectation" algorithm, in: Proceedings of the 2006 SIAM international conference on data mining, SIAM, 2006, pp. 474–478.

[144] T. Hebert, R. Leahy, A generalized EM algorithm for 3-D Bayesian reconstruction from poisson data using Gibbs priors, IEEE transactions on medical imaging 8 (2) (1989) 194–202.

[145] A. Likas, N. Vlassis, J. J. Verbeek, The global k-means clustering algorithm, Pattern recognition 36 (2) (2003) 451–461.

[146] V. Kantorová, M. C. Wheldon, P. Ueffing, A. N. Dasgupta, Estimating progress towards meeting women's contraceptive needs in 185 countries: A Bayesian hierarchical modelling study, PLoS medicine 17 (2) (2020) e1003026.

[147] A. R. Pitombeira-Neto, C. F. G. Loureiro, L. E. Carvalho, A dynamic hierarchical Bayesian model for the estimation of day-to-day origin-destination flows in transportation networks, Networks and Spatial Economics 20 (2) (2020) 499–527.

[148] Y. Chu, Y. Yuan, A bayesian basket trial design using a calibrated bayesian hierarchical model, Clinical Trials 15 (2) (2018) 149–158.

[149] C. Bracken, K. Holman, B. Rajagopalan, H. Moradkhani, A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis, Water Resources Research 54 (1) (2018) 243–255.

[150] R. Meager, Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments, American Economic Journal: Applied Economics 11 (1) (2019) 57–91.

[151] V. J. Schmid, B. Whitcher, A. R. Padhani, N. J. Taylor, G.-Z. Yang, A Bayesian hierarchical model for the analysis of a longitudinal dynamic contrast-enhanced MRI oncology study, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 61 (1) (2009) 163–174.

[152] G. J. Parker, et al., Measuring contrast agent concentration in T 1-weighted dynamic contrast-enhanced MRI, in: Dynamic contrast-enhanced magnetic resonance imaging in oncology, Springer, 2005, pp. 69–79.

[153] B. Whitcher, V. J. Schmid, D. J. Collins, M. R. Orton, D.-M. Koh, I. D. de Corcuera, M. Parera, J. M. del Campo, N. M. DeSouza, M. O. Leach, et al., A Bayesian hierarchical model for DCE-MRI to evaluate treatment response in a phase ii study in advanced squamous cell carcinoma of the head and neck, Magnetic Resonance Materials in Physics, Biology and Medicine 24 (2) (2011) 85–96.

[154] H. Sui, F. Peng, C. Xu, K. Sun, J. Gong, GPU-accelerated MRF segmentation algorithm for SAR images, Computers & Geosciences 43 (2012) 159–166.

[155] Y. Yang, H. Gao, C. Berry, D. Carrick, A. Radjenovic, D. Husmeier, Classification of myocardial blood flow based on dynamic contrast enhanced magnetic resonance imaging using hierarchical bayesian models, Journal of the Royal Statistical Society: Series C (Applied Statistics) (2022).

[156] J. M. Bardsley, Laplace-distributed increments, the Laplace prior, and edge-preserving regularization, J. Inverse Ill-Posed Probl 20 (3) (2012) 271–285.

[157] P. Knaapen, O. Bondarenko, A. M. Beek, M. J. Götte, R. Boellaard, A. P. van der Weerdt, C. A. Visser, A. C. van Rossum, A. A. Lammertsma, F. C. Visser, Impact of scar on water-perfusable tissue index in chronic ischemic heart disease, Molecular imaging and biology 8 (4) (2006) 245–251.

[158] S. P. Chatzis, T. A. Varvarigou, A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation, IEEE Transactions on Fuzzy Systems 16 (5) (2008) 1351–1361.

[159] D. Koller, N. Friedman, Probabilistic graphical models: principles and techniques, MIT press, 2009.

[160] R. A. Collazo, C. Görgen, J. Q. Smith, Chain Event Graphs, CRC Press, 2018.

[161] H. Gudbjartsson, S. Patz, The Rician distribution of noisy MRI data, Magnetic resonance in medicine 34 (6) (1995) 910–914.

[162] A. Radjenovic, Measurement of physiological variables by dynamic Gd-DTPA enhanced MRI, Ph.D. thesis, University of Leeds (2003).

[163] S. Chen, R. M. Haralick, Recursive erosion, dilation, opening, and closing transforms, IEEE Transactions on image processing 4 (3) (1995) 335–345.

[164] R. Neal, Bayesian Learning for Neural Networks, Springer, 1996.

[165] J.-M. Cornuet, J.-M. Marin, A. Mira, C. P. Robert, Adaptive multiple importance sampling, Scandinavian Journal of Statistics 39 (4) (2012) 798–812.

[166] A. Doucet, N. De Freitas, N. Gordon, Sequential Monte Carlo methods in practice, Springer, 2001.

[167] K. A. M. Said, A. B. Jambek, N. Sulaiman, A study of image processing using morphological opening and closing processes, International Journal of Control Theory and Applications 9 (31) (2016) 15–21.

[168] P. Kellman, M. S. Hansen, S. Nielles-Vallespin, J. Nickander, R. Themudo, M. Ugander, H. Xue, Myocardial perfusion cardiovascular magnetic resonance: optimized dual sequence and reconstruction for quantification, Journal of Cardiovascular Magnetic Resonance 19 (1) (2017) 1–14.

[169] V. J. Schmid, G.-Z. Yang, Spatio-temporal modelling of first-pass perfusion cardiovascular MRI, in: World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany, Springer, 2009, pp. 45–48.

[170] L. Cordero-Grande, G. Vegas-Sánchez-Ferrero, P. Casaseca-de-la Higuera, J. A. San-Román-Calvar, A. Revilla-Orodea, M. Martín-Fernández, C. Alberola-López, Unsupervised 4D myocardium segmentation with a Markov random field based deformable model, Medical image analysis 15 (3) (2011) 283–301.

[171] H. Gao, K. Kadir, A. R. Payne, J. Soraghan, C. Berry, Highly automatic quantification of myocardial oedema in patients with acute myocardial infarction using bright blood T2-weighted CMR, Journal of Cardiovascular Magnetic Resonance 15 (1) (2013) 1–12.

[172] K. P. Kunze, S. G. Nekolla, C. Rischpler, S. H. Zhang, C. Hayes, N. Langwieser, T. Ibrahim, K.-L. Laugwitz, M. Schwaiger, Myocardial perfusion quantification using simultaneously acquired 13nh3-ammonia PET and dynamic contrast-enhanced MRI in patients at rest and stress, Magnetic resonance in medicine 80 (6) (2018) 2641–2654.

[173] C. Zakkaroff, A. Radjenovic, J. D. Biglands, S. Plein, J. P. Greenwood, D. R. Magee, Registration of coronary MRA to DCE-MRI myocardial perfusion series improves diagnostic accuracy through the computation of patient-specific coronary supply territories: a CE-MARC sub-study, Journal of Cardiovascular Magnetic Resonance 16 (1) (2014) 1–3.

[174] J. Biglands, D. Magee, R. Boyle, A. Larghat, S. Plein, A. Radjenović, Evaluation of the effect of myocardial segmentation errors on myocardial blood flow estimates from DCE-MRI, Physics in Medicine & Biology 56 (8) (2011) 2423.

[175] J. Zhang, M. Liu, L. An, Y. Gao, D. Shen, Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images, IEEE journal of biomedical and health informatics 21 (6) (2017) 1607–1616.

[176] M. Huang, W. Yang, Q. Feng, W. Chen, Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease, Scientific reports 7 (1) (2017) 1–13.

[177] C. Elliott, S. J. Francis, D. L. Arnold, D. L. Collins, T. Arbel, Bayesian classification of multiple sclerosis lesions in longitudinal MRI using subtraction images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2010, pp. 290–297.

[178] E. J. Hannan, Multiple time series, John Wiley & Sons, 2009.

[179] M. Lorenzi, G. Ziegler, D. C. Alexander, S. Ourselin, Efficient Gaussian process-based modelling and prediction of image time series, in: International Conference on Information Processing in Medical Imaging, Springer, 2015, pp. 626–637.

[180] C. M. Scannell, P. v. d. Bosch, A. Chiribiri, J. Lee, M. Breeuwer, M. Veta, Deep learning-based prediction of kinetic parameters from myocardial perfusion MRI, arXiv preprint arXiv:1907.11899 (2019).

[181] R. L. van Herten, A. Chiribiri, M. Breeuwer, M. Veta, C. M. Scannell, Physics-informed neural networks for myocardial perfusion MRI quantification, arXiv preprint arXiv:2011.12844 (2020).

[182] L. Fan, R. Namani, J. S. Choy, G. S. Kassab, L. C. Lee, Effects of mechanical dyssynchrony on coronary flow: insights from a computational model of coupled coronary perfusion with systemic circulation, Frontiers in physiology (2020) 915.

[183] M. Ingrisch, S. Sourbron, Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer, Journal of pharmacokinetics and pharmacodynamics 40 (3) (2013) 281–300.

[184] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational physics 378 (2019) 686–707.

[185] E. Martín-González, E. Alskaf, A. Chiribiri, P. Casaseca-de-la Higuera, C. Alberola-López, R. G. Nunes, T. Correia, Physics-informed self-supervised deep learning reconstruction for accelerated first-pass perfusion cardiac MRI, in: International Workshop on Machine Learning for Medical Image Reconstruction, Springer, 2021, pp. 86–95.

[186] U. Noè, A. Lazarus, H. Gao, V. Davies, B. Macdonald, K. Mangion, C. Berry, X. Luo, D. Husmeier, Gaussian process emulation to accelerate parameter estimation in a mechanical model of the left ventricle: a critical step towards clinical end-user relevance, Journal of the Royal Society Interface 16 (156) (2019) 20190114.