Regularized estimation and model selection in compartment models

Julia C. Sommer

München 2013

Regularized estimation and model selection in compartment models

Julia C. Sommer

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Julia C. Sommer

München, den 26. Februar 2013

Erstgutachter: Prof. Dr. Volker Schmid
Zweitgutachter: Brandon Whitcher, PhD
Tag der mündlichen Prüfung: 26. April 2013

Danksagung

Ich möchte mich ganz herzlich bei allen bedanken, die mich beim Erstellen dieser Arbeit unterstützt haben.

Besonders danke ich Volker Schmid dafür, dass er mir interessante Einblicke in die Welt der medizinische Bildgebung und Bayes-Inferenz ermöglicht hat. Ich bin sehr dankbar dafür, dass er sich viel Zeit für fachliche Fragen und Diskussionen genommen hat. Außerdem danke ich Brandon Whitcher für die Bereitschaft, meine Arbeit zu begutachten.

Ich bedanke mich ganz herzlich bei Jan Gertheiss für die angenehme Zusammenarbeit und seinen kompetenten fachlichen Rat. Kapitel 6 ist im Rahmen dieser Zusammenarbeit entstanden. Danke auch an Sonja Greven für fachliche Anregung und Diskussion.

Danke an die DFG für die Finanzierung meiner Arbeit und an Anwar Padhani und Lothar Schermelleh für das Bereitstellen von Daten.

Ein herzliches Dankeschön an Martina Feilke und Nora Fenske für Korrekturvorschläge und Anmerkungen zu dieser Arbeit und an Franziska Kärcher für die Hilfe bei der graphischen Illustration (Abbildungen 3.2 und 3.5).

Bei allen Kolleginnen und Kollegen am Institut bedanke ich mich ganz herzlich für das nette Arbeitsklima.

Meinen Eltern und Geschwistern bin ich sehr dankbar für ihren Rückhalt.

Danke an Clemens für alles.

München, den 7. Mai 2013 Julia C. Sommer

Zusammenfassung

Zeitreihen von dynamischen Bildern aus Medizin und Biologie werden oft mit Hilfe von Kompartimentmodellen untersucht. Kompartimentmodelle liefern eine parametrische, nichtlineare Funktion von interpretierbaren Parametern, die beschreibt, wie sich eine beobachtbare Konzentration im Zeitverlauf entwickelt. Will man die kinetischen Parameter schätzen, so führt dies zu einem nichtlinearen Regressionsproblem. In vielen Anwendungen ist die Anzahl der im Modell benötigten Kompartimente aus biologischen Gesichtspunkten unbekannt und sollte vielmehr gemeinsam mit den kinetischen Parametern aus den Daten geschätzt werden. Da die Daten in medizinischen und biologischen Experimenten oft als Bilder vorliegen, muss dabei die räumliche Struktur der Bilder mit berücksichtigt werden.

Diese Arbeit befasst sich mit der Aufgabe der Parameterschätzung und Modellselektion in Kompartimentmodellen. Neben einem penalisierten likelihoodbasierten Ansatz werden zur Lösung dieser Aufgabe mehrere Bayesianische Ansätze vorgeschlagen und bewertet, darunter ein hierarchisches Modell mit Gauss-Markov-Zufallsfeld-Prioris und ein Ansatz mit flexibler Modelldimension. Bestehende Methoden werden für die Parameterschätzung und Modellselektion in komplexeren Kompartimentmodellen erweitert. Im nichtlinearen Regressionsmodell und besonders in komplexeren Kompartimentmodellen können jedoch Redundanzprobleme auftreten. Diese Arbeit untersucht die Schwierigkeiten, die durch Parameterredundanzen auftreten und zeigt mehrere Lösungsansätze auf, die Probleme durch Regularisierung des Parameterraumes zu mildern.

Wir bewerten das Potential der vorgeschlagenen Schätz- und Modellselektionsverfahren anhand von Simulationsstudien sowie für zwei *in vivo* Bildgebungsanwendungen: eine Brustkrebsstudie mit Bildgebung durch dynamische kontrastmittelverstärkte Magnetresonanztomographie und eine Studie zum Bindungsverhalten von Molekülen in lebenden Zellkernen, welches in einem Photobleichungsexperiment mit Hilfe der Fluoreszenzmikroskopie beobachtet wird.

Abstract

Dynamic imaging series acquired in medical and biological research are often analyzed with the help of compartment models. Compartment models provide a parametric, nonlinear function of interpretable, kinetic parameters describing how some concentration of interest evolves over time. Aiming to estimate the kinetic parameters, this leads to a nonlinear regression problem. In many applications, the number of compartments needed in the model is not known from biological considerations but should be inferred from the data along with the kinetic parameters. As data from medical and biological experiments are often available in the form of images, the spatial data structure of the images has to be taken into account.

This thesis addresses the problem of parameter estimation and model selection in compartment models. Besides a penalized maximum likelihood based approach, several Bayesian approaches—including a hierarchical model with Gaussian Markov random field priors and a model state approach with flexible model dimension—are proposed and evaluated to accomplish this task. Existing methods are extended for parameter estimation and model selection in more complex compartment models. However, in nonlinear regression and, in particular, for more complex compartment models, redundancy issues may arise. This thesis analyzes difficulties arising due to redundancy issues and proposes several approaches to alleviate those redundancy issues by regularizing the parameter space.

The potential of the proposed estimation and model selection approaches is evaluated in simulation studies as well as for two *in vivo* imaging applications: a dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) study on breast cancer and a study on the binding behavior of molecules in living cell nuclei observed in a fluorescence recovery after photobleaching (FRAP) experiment.

Contents

List of abbreviations

ACF	autocorrelation function
AIC	Akaike information criterion
AIF	arterial input function
BIC	Bayesian information criterion
CNR	contrast to noise ratio
CT	X-ray computed tomography
CTC	concentration time curve
DCE-CT	dynamic contrast enhanced computed tomography
DCE-MRI	dynamic contrast enhanced magnetic resonance imaging
DIC	deviance information criterion
Dnmt1	DNA methyltransferase 1
DSC-MRI	dynamic susceptibility contrast magnetic resonance imaging
EES	extravascular extracellular space
e.g.	*exempli gratia*
fMRI	functional magnetic resonance imaging
FRAP	fluorescence recovery after photobleaching
Gd-DTPA	gadolinium diethylenetriaminepentaacetic acid
GFP	green fluorescent protein
i.e.	*id est*
MCMC	Markov chain Monte Carlo
MDCT	multidetector row computed tomography
ML	maximum likelihood
MRI	magnetic resonance imaging
PDF	probability density function
PET	positron emission tomography
RJMCMC	reversible jump MCMC
ROI	region of interest
SNR	signal to noise ratio
SPECT	single-photon emission computed tomography
SSE	sum of squared errors
1Comp	one tissue compartment model
2Comp	two tissue compartment model
2CXM	two compartment exchange model
qComp	multi tissue compartment model

Chapter 1

Introduction

Compartment models form a powerful model class to describe dynamical systems. With the help of such models the exchange of material or energy between different pools, called compartments, can be described analytically. Like this, the evolvement of material over time can be parametrically described with the help of meaningful kinetic parameters describing size and exchange properties of the compartments. Enjoying the advantage of providing interpretable model parameters, compartment models are at present widely used in medical (Slifstein and Laruelle, 2001; Parker and Buckley, 2005) and biological research (Sprague and McNally, 2005), but they have formerly also been used in other fields, for example, in social sciences (Herbst, 1963) or in ecology (Eriksson, 1971).

The function describing the evolution of the compartmental system is—given as solution of differential equations—always nonlinear and comprises exponentials of kinetic parameters and time. Quantitative analysis with compartment models uses noisy observations of the state of material over time, and, from this time series, one aims to infer on the underlying kinetics of the system. In the nonlinear regression problem arising, time is the only measurable explanatory variable of the system, and the parameters describing size and kinetic properties of the compartments have to be estimated. Though the number of model parameters is rather small, parameter redundancy is frequently an issue due to the nonlinear form of the regression model (Seber and Wild, 1989). In this thesis, we propose different estimation approaches for compartment models. Those approaches address redundancy issues by using different ways of regularizing the parameter space.

For a given compartment model, the form of the nonlinear function can analytically be derived. However, the choice of a compartment model with suitable architecture and adequate model complexity is an unsolved question in many applications. Especially, the number of compartments needed to adequately describe a dynamic system may not be known *a priori*, for instance, from biological or physical considerations. Often, the number of compartments itself is characteristic for the behavior of the system and should, hence, be inferred from the observed behavior. For example, the number of compartments may indicate how heterogeneous cancerous tissue is in the analysis of dynamic contrast enhanced magnetic resonance imaging (DCE-MRI). In biochemical experiments, the number of binding partners may be interesting to draw conclusions about the diffusion and binding

pattern of some protein.

Motivated by those questions arising in medical and biological applications, this thesis addresses the problem of parameter estimation and model selection in compartment models. The focus is on developing and extending statistical tools to approach this task. The proposed approaches are thoroughly evaluated and compared to existing approaches. Furthermore, finding a suitable model goes hand in hand with understanding possible underlying biological processes and taking this knowledge into account.

Prior knowledge—for example about biological processes—is most intuitively accounted for in a Bayesian framework. Basing inference on the posterior distribution, the Bayesian framework considers both observed data and prior knowledge. Most of the estimation approaches for compartment models presented in this thesis are based on Bayesian hierarchical models. Those models allow to impose prior distributions on model parameters, *e.g.*, reflecting typical expected ranges for kinetic parameters. Furthermore, in this thesis, spatial smoothing of parameters is incorporated in a Bayesian hierarchical model using Gaussian Markov random field priors. Finally, the number of model parameters itself is considered a random variable and, hence, treated as part of the parameter space, transforming the Bayesian hierarchical model into a model state approach.

In this thesis, inference is mainly done for the nonlinear regression problem as is without linearization. When studying whether an estimation approach is suitable for nonlinear regression models, an important point is to analyze its ability to cope with redundancy issues. For comparison, we linearize the nonlinear regression model and evaluate estimation and model selection in the linearized model based on a newly developed spatially penalized maximum likelihood approach.

Software and applications

In Bayesian nonlinear regression the posterior distribution is typically analytically not approachable and has to be calculated with the help of Markov chain Monte Carlo (MCMC) simulations. As part of this thesis, the corresponding algorithms for all Bayesian estimation approaches considered were implemented in C and interfaced in R. The Bayesian approaches for DCE-MRI were implemented extending the R-package `dcemriS4` (Whitcher and Schmid, 2009, 2011). As computationally competitive alternative, we present a penalized maximum likelihood based inference approach for compartment models. The corresponding algorithm was implemented in R.

All inference approaches considered are evaluated with simulation studies as well as for real data from biological and medical imaging applications. Most estimation and model selection approaches are designed and evaluated for the quantitative analysis of DCE-MRI. This noninvasive *in vivo* imaging technique makes the perfusion of blood observable with the aid of a tracer. Typically, one acquires two-dimensional or three-dimensional images over time. We analyze DCE-MRI data from a breast cancer study using the time series in each voxel, since the resolution of the data allows to do so. Like this, parameter estimates are obtained per voxel and local heterogeneities of the tissue can be modeled. This is important in the analysis of tumorous tissue as tumors are believed to be heterogeneous

and information on tissue heterogeneity is diagnostically informative.

The second data set used in this thesis is from a fluorescence recovery after photobleaching (FRAP) experiment. This imaging technique produces a series of two-dimensional images over time. FRAP is used to study biochemical experiments, and, hence, for this imaging application completely different research questions arise. Nevertheless, the data structure and the compartment models used for the analysis of DCE-MRI and FRAP data share some similarities. Both data sets consist of time series of images acquired over time. As the resolution in the FRAP data set is considerably lower, it is analyzed on a region of interest level here, that is, working with concentrations averaged over all voxels of interest.

Outline

This thesis is organized as follows. Chapter 2 introduces compartment models and its potential applications, it describes the underlying assumptions of those models and establishes the differential equations arising from compartment models. Based on this, inference approaches are introduced and parameter redundancy issues arising in corresponding nonlinear regression models are characterized.

Chapter 3 raises research questions analyzed with the help of DCE-MRI. An overview of existing compartment models for the quantitative analysis of DCE-MRI is given and—in order to account for tumor heterogeneity—more complex models are proposed to describe DCE-MRI data on a voxel level. At the end of this chapter, we introduce a DCE-MRI *in vivo* breast cancer study serving as real data example in several chapters.

The first model selection approach, described in Chapter 4, is to fit models with one or two tissue compartments and to decide which model is more suitable based on the deviance information criterion (DIC). We base estimation on a hierarchical Bayesian framework incorporating prior information on the parameters and estimate parameters for each voxel independently. The analysis of parameter redundancy issues arising in a two tissue compartment model is an important, new contribution. We analyze the effects of parameter redundancy and the results of model selection for simulated data as well as for the DCE-MRI breast cancer study. Chapter 4 is considered preliminary work for Chapter 5.

In Chapter 5, we propose and analyze a spatial Bayesian model for a two tissue compartment model. With this approach, spatial information is modeled assuming Gaussian Markov random field priors on the kinetic parameters. Like this, parameter maps are spatially smoothed and the parameter space is regularized. We compare the estimation results of the newly developed spatial two tissue compartment model with an existing spatial one tissue compartment model (Schmid et al., 2006; Kelm et al., 2009) and with the corresponding models without spatial smoothing. This is done for simulated data and for the DCE-MRI breast cancer study.

In Chapter 6, the nonlinear regression problem is transformed into a linear regression problem. Due to construction, explanatory variables in this linear regression problem are highly correlated and pure maximum likelihood estimates are unstable. Hence, we propose spatially penalized maximum likelihood estimation with an elastic net for the analysis of DCE-MRI. With this approach, a flexible number of compartments is chosen per voxel and

parameters are estimated simultaneously. We evaluate how the newly developed spatial version of the elastic net performs compared to the ordinary elastic net for simulated data and for the DCE-MRI breast cancer study.

Finally, in Chapter 7, we propose a Bayesian model space approach, allowing to simultaneously estimate model parameters as well as the model complexity in a compartment model with the help of a reversible jump Markov chain Monte Carlo (RJMCMC) procedure. To this end, birth and death steps are derived for the estimation in a nonlinear regression problem. The RJMCMC procedure is evaluated for simulated data and for data from a FRAP experiment.

In Chapter 8, we conclude and discuss.

Several chapters of this thesis are to large parts based on the following publications and working papers. Some of them are currently under review. The manuscripts were adapted for this thesis such that the notation is consistent and redundancy is avoided.

Chapter 4 is based on

> Kärcher, J.C., Schmid, V.J.: *Two tissue compartment model in DCE-MRI: A Bayesian Approach*. IEEE International Symposium on Biomedical Imaging. From Nano to Macro. 724-727. Peer-reviewed proceedings. (Kärcher and Schmid, 2010),

Chapter 5 and parts of Chapters 3 and 4 are based on

> Sommer, J.C., Schmid, V.J.: *Spatial two tissue compartment model for DCE-MRI*. arXiv:1209.0901. Under review. (Sommer and Schmid, 2012),

Chapter 6 is based on

> Sommer, J.C., Gertheiss, J., Schmid, V.J.: *Spatially regularized estimation for the analysis of DCE-MRI data*. Department of Statistics: Technical Reports, Nr. 132. Under review. (Sommer et al., 2012),

and preliminary work for Chapter 6 can be found in

> Gertheiss, J., Kärcher, J.C., Schmid, V.J.: *Analysis of DCE-MRI Data using a Nonnegative Elastic Net*. Department of Statistics: Technical Reports, Nr. 90. (Gertheiss et al., 2010).

Chapter 2

Compartment models

In this chapter, we give a general introduction to compartment models starting with some applications of such models in Section 2.1. We present a general compartment model with one basic compartment and K non-nested compartments which includes the applications presented in this thesis as special cases. We introduce the system of differential equations which generally describes a compartment model and describe the form of solution in such systems (Section 2.2). Then, we present possible inference approaches for the corresponding nonlinear regression problems arising from compartmental systems (Section 2.3). Special focus is on parameter redundancy—a well known problem in nonlinear regression.

2.1 Applications

We start with giving some examples of different fields that benefit from the analysis with compartment models. In social sciences, human behavior is studied with the help of compartmental models, *e.g.*, the number of entrants in an organization is modeled as time dependent quantity depending on the current number of members in the organization (Herbst, 1963). In physiology, organs are modeled as compartments exchanging blood and, hence, oxygen (Jacquez, 1972). In ecological systems, the exchange of energy is modeled with the help of compartments (Eriksson, 1971). In this thesis, we focus on estimation and model selection in compartment models used in medical and biological imaging applications.

For example, compartment models are frequently used in medical tracer experiments. In such experiments one is interested in the uptake of a tracer to certain sections of the body. The tracer's uptake over time can be observed using *in vivo* imaging techniques like magnetic resonance imaging (MRI), positron-emission tomography (PET), or single-photon emission computed tomography (SPECT). The quantitative analysis of DCE-MRI is based on compartment models to describe the exchange of blood (and hence tracer) between different, well-mixed compartments. Typically, contributing compartments are the arterial plasma, other plasma compartments and interstitial space compartments (see Chapter 3). Compartment models allow to link physiological properties of interest, like capillary permeability and perfusion, to the observed uptake of tracer. Chapters 3–6 build

on compartmental analysis for a DCE-MRI breast cancer study.

In quantitative models for DCE-MRI, the compartments can be thought of being physically separated, *e.g.*, by a membrane. However, this does not need to be the case as compartments can also represent a concept or a state such as, for example, the state of molecules in a chemical process. This is the case in the analysis of PET, SPECT and FRAP data as described in the following.

Compartment models—in this context also referred to as kinetic models or pharmacokinetic models—are very commonly used in PET and SPECT neuroreceptor imaging studies. The aim of such studies is to quantitatively describe how a target receptor is distributed throughout the brain and to estimate receptor parameters, *i.e.* the binding rates of a tracer. A radiotracer is injected and its uptake in the brain is recorded. The arterial plasma concentration of tracer forms the basic compartment (and is often calculated from blood samples during the PET scan). The activity of radiotracer in brain tissue can be decomposed in three contributing compartments: unbound, non-specifically bound, and specifically bound radiotracer. Often, one combines unbound and non-specifically bound tracer into one compartment, assuming that those compartments come to equilibrium very fast compared to the kinetics of specific binding. Slifstein and Laruelle (2001) give an overview of how PET and SPECT studies are analyzed with the help of compartment models.

Besides those medical imaging applications, compartmental modeling is also used in biochemical experiments that give insight to the binding behavior of molecules in living cells (Sprague and McNally, 2005). In such experiments, each molecule can adopt different states: it can either be free (unbound) or it can be bound to one of several binding partners. The class of unbound molecules forms one compartment, the class of molecules bound to a binding partner forms an additional compartment. Such models are used in the analysis of fluorescence recovery after photobleaching (FRAP) experiments. In Chapter 7, an approach to estimate the number of compartments—corresponding to classes of binding partners with different binding properties—is proposed and evaluated for data from a FRAP experiment.

2.2 Compartment models: assumptions and model equations

As the concept of compartmental systems is broad and used in various fields, a definition of such a system sounds rather abstract. A compartmental system consists of a certain number of pools—or compartments—that exchange some material with each other. Additionally, there may be inflows to and outflows from the system (open system with sources and sinks). If there are no in- and outflows in the system, the system is called closed. Compartmental systems are often depicted in connectivity diagrams as, for example, shown in Figure 2.1. The system depicted consists of two compartments A_1 and A_2, exchanging some material, as well as an external source with time dependent inflow $b_1(t)$ to compartment

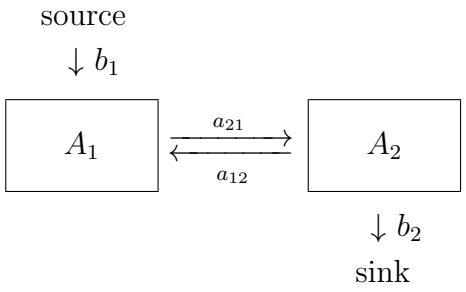

Figure 2.1: Connectivity diagram for a model with two compartments, an inflow and an outflow

A_1, and time dependent outflow $b_2(t)$ from compartment A_2 to an external sink.

When modeling with compartmental systems, one usually makes several assumptions: Within one compartment, the material is assumed to be well-mixed and homogeneous. Furthermore, the exchange between compartments is assumed to follow certain rules. The amount of outflow from compartment A_1 to compartment A_2 is assumed to be proportional to the current amount in compartment A_1 and *vice versa*. The proportionality factors, a_{12} and a_{21}, however, are constant with respect to time and are called rate constants. Like this, the amount of material in both compartments, denoted with $C_1(t)$, $C_2(t)$, varies with time. The underlying parameters describing the exchange, a_{12} and a_{21}, however, are constant.

Systems of ordinary differential equations

Generally, compartmental systems which obey the above assumptions can be described by a system of linear ordinary differential equations, more precisely, a non-homogeneous linear system of differential equations with constant coefficients (Seber and Wild, 1989):

$$\begin{aligned}
\frac{d}{dt}C_1 &= a_{11}C_1(t) + a_{12}C_2(t) + \ldots + a_{1K}C_K(t) + b_1(t) \\
\frac{d}{dt}C_2 &= a_{21}C_1(t) + a_{22}C_2(t) + \ldots + a_{2K}C_K(t) + b_2(t) \\
&\vdots \\
\frac{d}{dt}C_K &= a_{K1}C_1(t) + a_{K2}C_2(t) + \ldots + a_{KK}C_K(t) + b_K(t)
\end{aligned} \tag{2.1}$$

with starting conditions $C_1(0) = c_{10}, \ldots, C_K(0) = c_{K0}$. Here, $C_k(t)$ denotes the concentration or fraction of material in the kth compartment at time t. The rate constant a_{ki} determines the rate of flow from compartment i to compartment k, and b_k is the input/output to/from compartment k from/to an external source/sink. In Matrix-Notation this system of differential equations can be summarized as

$$\frac{d}{dt}\mathbf{C}(t) = \mathbf{A}\mathbf{C}(t) + \mathbf{b}(t), \mathbf{C}(0) = \mathbf{c}_0. \tag{2.2}$$

The system has a unique solution (see for example Heuser, 1995, Chapter VII) of the form

$$\mathbf{C}(t) = \exp(\mathbf{A}t)\mathbf{c}_0 + \int_0^t \exp(\mathbf{A}(t-\tau))\mathbf{b}(\tau)d\tau. \tag{2.3}$$

Hence, the evolution of concentration over time can be described with a known parametric function that depends on the architecture of the model and, of course, on its parameters. Typically, one aims to determine the model parameters given the model architecture and after observing the concentration over time in some compartments. Often, not even the model architecture is known *a priori* and in this case, both the estimation of model parameters and the selection of the model need to be based on the observed concentration over time. Different approaches for this model selection problem are proposed in this thesis.

For the computation of the matrix valued exponentials in equation (2.3), the spectral decomposition $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$ has to be computed. Here, $\mathbf{S}$ consists of eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_K)$ is a diagonal matrix with the corresponding eigenvalues at the diagonal. The integral in equation (2.3) is derived to be (Seber and Wild, 1989, p. 381)

$$\int_0^t \exp(\mathbf{A}(t-\tau))\mathbf{b}(\tau)d\tau = \mathbf{S}\int_0^t \exp(\mathbf{\Lambda}(t-\tau))\mathbf{S}^{-1}\mathbf{b}(\tau)d\tau \tag{2.4}$$

where

$$\left[\int_0^t \exp(\mathbf{\Lambda}(t-\tau))\mathbf{S}^{-1}\mathbf{b}(\tau)d\tau\right]_j = \sum_{k=1}^{K} s^{jk} \underbrace{\int_0^t \exp(\lambda_j(t-\tau))b_k(\tau)d\tau}_{=b_k * \exp(\lambda_j t)} \tag{2.5}$$

and s^{jk} are the elements of $\mathbf{S}^{-1}$. Here, $*$ denotes the convolution operator, *i.e.* $(f*g)(t) = \int_0^t f(\tau)g(t-\tau)d\tau$. This means that given the initial condition $\mathbf{c}_0 = \mathbf{0}$, the concentration in each compartment can be written as a sum of up to K convolved exponentials. The exponential rates are the eigenvalues of the connectivity matrix $\mathbf{A}$.

Compartment models considered

In this thesis, we will consider compartment models with a special model architecture depicted in Figure 2.2. We assume a basic compartment with known concentration $C_B(t)$, which may be constant or varying over time. This basic compartment exchanges with K other compartments with concentrations $C_1(t), \ldots, C_K(t)$. As the compartments $1, \ldots, K$ do not exchange material among each other, we call this system non-nested. The exchange into and out of the kth compartment happens at constant rates b_k^{in} and b_k^{out} and depends only on the actual concentrations in the basic and the kth compartment:

$$\frac{d}{dt}C_k(t) = -b_k^{out}C_k(t) + b_k^{in}C_B(t), \tag{2.6}$$

with boundary condition $C_k(0) = c_{k0}$. This compartmental system is a special case of the system given by equation (2.2) with connectivity matrix $\mathbf{A} = diag(-b_1^{out}, \ldots, -b_K^{out})$. Here, the basic compartment takes the role of an external source determining the input $\mathbf{b}$ to all

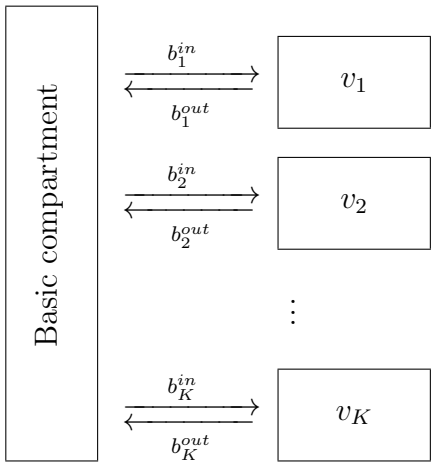

Figure 2.2: Connectivity diagram of a non-nested compartment model with K compartments and a basic compartment

other compartments: $\mathbf{b}(t) = (b_1^{in} C_B(t), \ldots, b_k^{in} C_B(t))^T$. This non-nested compartmental system is general enough to comprise the biomedical applications discussed in this thesis. In this system, the concentration C_k in the kth compartment has a simple functional form given by the solution of (2.6):

$$C_k(t) = c_{k0} \exp(-b_k^{out} t) + b_k^{in} C_B * \exp(-b_k^{out} t). \tag{2.7}$$

However, in the imaging applications considered in this thesis, the concentrations $C_k(t)$ in each compartment cannot be observed separately, but only the total concentration is observed and modeled. The total concentration of all compartments, $C(t)$, is given as sum of the individual concentrations

$$C(t) = v_B C_B(t) + \sum_{k=1}^{K} v_k C_k(t) \tag{2.8}$$

where v_B denotes the volume fraction of the basic compartment, and v_k the volume fraction of the k-th compartment. For the boundary condition $c_{k0} = 0$ and with C_k given by (2.7) for $k = 1, \ldots, K$ the total concentration is then

$$C(t) = v_B C_B(t) + C_B * \sum_{k=1}^{K} v_k b_k^{in} \exp(-b_k^{out} t). \tag{2.9}$$

For the FRAP application considered in Chapter 7, the concentration of the basic compartment, $C_B(t)$, is assumed to be constant. For $C_B(t) \equiv 1$, the convolution with an exponential simplifies to $C_B * \exp(-bt) = \frac{1}{b} - \frac{1}{b} \exp(-bt)$ (see Appendix A.1.2) and the

total concentration is described by

$$\begin{aligned} C(t) &= v_B + \sum_{k=1}^{K} v_k b_k^{in} \left(\frac{1}{b_k^{out}} - \frac{1}{b_k^{out}} \exp(-b_k^{out} t) \right) \\ &= v_B + \sum_{k=1}^{K} \frac{v_k b_k^{in}}{b_k^{out}} - \sum_{k=1}^{K} \frac{v_k b_k^{in}}{b_k^{out}} \exp(-b_k^{out} t). \end{aligned} \tag{2.10}$$

So, for a constant concentration in the basic compartment the convolution is simple, and the concentration can directly be described by a sum of K exponentials.

In the model for DCE-MRI considered in Chapters 3–6, however, $C_B(t)$ is a non-constant function and, hence, one deals with sums of convolved exponentials. In DCE-MRI, one assumes the influx and outflux rates to be the same: $b_k = b_k^{in} = b_k^{out}$. Then, the volumes of the compartments can be determined from the parameters a_k and b_k: $v_k = \frac{a_k}{b_k}$.

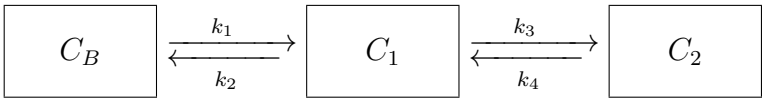

Figure 2.3: Connectivity diagram for model with two nested compartments

Nested versus non-nested model architecture

In DCE-MRI and PET applications some authors have proposed models of different architecture than the one depicted in Figure 2.2. For example, in Gunn et al. (2002) a two tissue compartment model for PET experiments is described with two nested tissue compartments only one of which exchanges with the basic compartment, see Figure 2.3. For this model, the corresponding model equations are

$$\begin{aligned} \frac{d}{dt} C_1 &= -(k_2 + k_3) C_1(t) + k_4 C_2(t) + k_1 C_B(t) \\ \frac{d}{dt} C_2 &= k_3 C_1(t) - k_4 C_2(t). \end{aligned} \tag{2.11}$$

Here, the connectivity matrix $\mathbf{A} = \begin{pmatrix} -(k_2 + k_3) & k_4 \\ k_3 & -k_4 \end{pmatrix}$ is not diagonal and, therefore, the concentration cannot be described as directly with exponentials of the rate constants as in equation (2.9). However, the solution of (2.11) is still a sum of two convolved exponentials (Gunn et al., 2002)

$$C(t) = C_1(t) + C_2(t) = C_B * \sum \psi_k \exp(\lambda_k t). \tag{2.12}$$

In this case, the coefficients ψ_k and the exponential rates λ_k are computed from the eigen decomposition of $\mathbf{A}$ and, hence, functions of (k_1, k_2, k_3, k_4). Therefore, one would have to

estimate the parameters ψ_k and λ_k in such a model and then translate those to physiological interpretable model parameters k_1, k_2, k_3, and k_4.

Similarly, models with nested tissue compartments have been proposed for the analysis of DCE-MRI (Sourbron and Buckley, 2011). The functional form of the total concentration and, hence, the nonlinear regression problem arising in models with two nested compartments (2.12), is mathematically the same as in the models with two non-nested compartments (2.9). This is why one can always translate a model with nested compartments to an equivalent model with non-nested compartments. In this thesis, we prefer to assume model architectures with non-nested compartments as given in Figure 2.2, as for those models the estimated parameters directly correspond to physiologically interpretable model parameters. A large part of our methods and conclusions, however, can be transfered to nested model architectures as well.

2.3 On inference and parameter identifiability in compartment models

In Section 2.2 we have shown how—with the help of compartment models—some concentration can analytically be described as a function of time and physiologically interpretable model parameters. For example, for the model with K non-nested compartments described by equation (2.9) the model parameters are the volume fractions and rate constants, summarized in the parameter vector $\boldsymbol{\theta} = (v_B, v_1, \ldots, v_K, b_1, \ldots, b_K)$. Assuming that we observe a noisy version y_j of the concentration at time point t_j described by the parametric function $C(\boldsymbol{\theta}; t_j)$, this leads to

$$y_j = C(\boldsymbol{\theta}; t_j) + \epsilon_j, j = 1, \ldots, n \tag{2.13}$$

where ϵ_j is a stochastic noise term. We will assume additive Gaussian noise terms throughout this thesis. Equation (2.13) describes a nonlinear regression problem when $C(\boldsymbol{\theta}; t_j)$ is nonlinear in $\boldsymbol{\theta}$. This is the case when $C(\boldsymbol{\theta}; t_j)$ is the concentration described by a compartment model. Here, the time points $t_1, \ldots, t_n$ are the explanatory variables and $\boldsymbol{\theta}$ are the unknown model parameters; we denote more generally $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.

Then, we aim to find an estimate for $\boldsymbol{\theta}$ such that this estimator is optimal in some sense. Best known inference approaches are least squares, maximum likelihood or Bayesian inference. Theoretical properties of parameter estimation in nonlinear regression models are thoroughly discussed in Seber and Wild (1989), Bates and Watts (1988) and Ratkowsky (1990). Estimation in nonlinear regression models differs from estimation in linear regression models concerning several aspects. In Seber and Wild (1989, p. 91 ff) three typical, interconnected problem areas are described: First, iterative methods searching for extrema may fail to converge. Second, estimates in a nonlinear regression problem may not have desired properties (like unbiased, asymptotically normally distributed and minimum variance estimation). Third, one often deals with ill-conditioning in parameter estimation. Here, we give a short introduction to inference in nonlinear regression with focus on compartment models.

Inference in nonlinear regression

The *least-squares* estimator $\hat{\boldsymbol{\theta}}$, *e.g.*, is found by minimizing the sum of squares

$$S(\boldsymbol{\theta}) = \|\boldsymbol{y} - C(\boldsymbol{\theta})\|^2 = \sum_{j=1}^{n} \left(y_j - C(t_j, \boldsymbol{\theta})\right)^2 . \tag{2.14}$$

When $C(t_j, \boldsymbol{\theta})$ is nonlinear in $\boldsymbol{\theta}$, $S(\boldsymbol{\theta})$ may have several relative minima. If $C(t_j, \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$, then the argmin $\hat{\boldsymbol{\theta}}$ of $S(\boldsymbol{\theta})$ will satisfy the necessary condition

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \theta_r} = \left(y_j - C(t_j, \boldsymbol{\theta})\right) \frac{\partial C(t_j, \boldsymbol{\theta})}{\partial \theta_r}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \tag{2.15}$$

for $j = 1, ..., n$ and $r = 1, ..., p$ written as

$$C_{\cdot}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - C(\hat{\boldsymbol{\theta}})) = 0. \tag{2.16}$$

In contrast to the linear regression model, here, the normal equation (2.16) cannot be solved analytically, and iterative procedures or search algorithms have to be applied. Those algorithms start with an initial guess $\boldsymbol{\theta}_0$ for the parameter vector and then iteratively improve the current estimate until convergence. The algorithms can find local extrema of a nonlinear function, however, convergence to an absolute extremum is not guaranteed. Common algorithms searching for local extrema of a nonlinear function are the Gauss-Newton method—using a linear approximation of C—and the Levenberg-Marquardt method (Ahearn et al., 2005)—which interpolates between Gauss-Newton increments and increments in the direction of gradient descent (ascent). The linear approximation corresponds to the first order Taylor series

$$C(\boldsymbol{\theta}) \cong C(\boldsymbol{\theta}^*) + C_{\cdot}(\boldsymbol{\theta}^*)(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \tag{2.17}$$

and therefore a lot depends on the behavior of $C_{\cdot}(\boldsymbol{\theta}^*)$ and on how good this linear approximation is. Other iterative methods use quadratic approximations. Several of those iterative procedures are available in R, see *e.g.* the functions *nls*, *nls.lm* and *optim*. If the error terms, ϵ_j, are independently identically normally distributed, the *maximum likelihood* estimator corresponds to the least squares estimator (see Seber and Wild, 1989, p. 32ff), and so the same methods apply.

In linear regression models with normally distributed additive errors, least squares estimators and maximum likelihood estimators, respectively, do have nice properties: They are unbiased, asymptotically normal and minimum variance estimators. In contrast, in nonlinear regression models, least squares estimators achieve those desired properties only asymptotically, that is, for infinite sample size, given some regularity assumptions (see Seber and Wild (1989), Chapter 12). Ratkowsky (1990) calls nonlinear regression models '*close-to-linear*' if the estimators are close to being unbiased, normally distributed, minimum variance estimators for relatively small sample sizes. Otherwise, models are classified to be '*far-from-linear*'.

Inference based on least squares or maximum likelihood assumes that there is one true underlying parameter vector $\boldsymbol{\theta}$, and the aim is to find an optimal point estimate for $\boldsymbol{\theta}$. In

contrast, in the Bayesian approach $\boldsymbol{\theta}$ itself is considered to be a random variable with some distribution, and the aim is to describe this distribution based both on prior knowledge and the observed data (Robert, 2001; Gelman et al., 2004).

More precisely, *Bayesian inference* on a parameter vector $\boldsymbol{\theta}$ given the observed data $\boldsymbol{y}$ is based on the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ which by the Bayesian theorem can be expressed by

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{2.18}$$

Probabilistic models need to be specified both for the data, specified by the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$, and for the parameters, specified by the prior $p(\boldsymbol{\theta})$. Ignoring the normalizing constant in the denominator, the posterior distribution is proportional to the product of the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ and the prior distribution $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.19}$$

In nonlinear regression models, the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$, on which Bayesian estimation relies, is usually not analytically approachable. Hence, MCMC algorithms are needed to generate samples from the posterior distribution. For an introduction to MCMC see Gilks et al. (1996) and Green (2000). Theoretically, after a sufficiently large burnin phase, MCMC samples can be considered as a Markov chain having the posterior as its stationary distribution. In order to achieve stationarity in finite sampling time, however, it is crucial to have well mixing MCMC chains.

In the Bayesian approach, one is not as much interested in optimal properties of an estimator, because inference is not mainly based on point estimators, but on the whole posterior distribution. Like this, uncertainty in estimation becomes visible in the posterior densities of the parameters. We claim that 'far-from-linear' models will be recognized by flat, plateau-shaped or multimodal posterior densities. Thus, in a Bayesian framework one becomes well aware of possible difficulties such as estimation uncertainty or correlation of parameters. Furthermore, using an appropriate prior distribution may help to regularize the parameter space. Different ways to do so are proposed in this thesis.

For the inference in compartment models one approach considered in this thesis is to transform the nonlinear regression problem into a linear regression problem, using adequate nonlinear basis functions as predictors. A penalized maximum likelihood approach for such a linearized regression problem is used in Chapter 6. In Chapters 4, 5 and 7 we base inference and model selection on hierarchical Bayesian approaches working with the original nonlinear regression model.

Ill-conditioning, parameter identifiability, parameter redundancy

In nonlinear regression, parameter estimation may often be of poor precision. It is important to note that—while some identifiability issues are similar to problems known from estimation in linear regression—the problem of parameter redundancy is different and intrinsic to some nonlinear models.

The problem of *ill-conditioning* as well occurs in the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

when the columns of the regressor $\boldsymbol{X}$ are highly correlated. This problem is known as multi-collinearity. In general, the sum of squares surface has elliptical contours in the linear model, and least squares estimators of $\boldsymbol{\beta}$ can be calculated analytically. Confidence intervals of $\hat{\boldsymbol{\beta}}$ are ellipsoids. In the case of multi-collinearity, $\boldsymbol{X'X}$ is nearly singular and has small eigenvalues. The confidence ellipsoid expands in the corresponding direction, and estimation precision is poor. The same problems may arise in the nonlinear model, in case $C.(\boldsymbol{\theta})'C.(\boldsymbol{\theta})$ is ill-conditioned.

Additionally, in nonlinear regression, complications arise when the contours of the sum of squares surface $S(\boldsymbol{\theta})$ are curved. In curved contours one finds long curves along which $S(\boldsymbol{\theta})$ remains constant. In this case, any iterative optimization algorithm will fail to converge to global extrema of $S(\boldsymbol{\theta})$. In a Bayesian framework, MCMC simulations will show bad mixing behavior and posteriors will be flat or multimodal (see Chapter 4). Note that the posterior distribution will partly depend on the likelihood and, hence, on the sum of squares $S(\boldsymbol{\theta})$. Sometimes, the problem of ill-conditioning can be eased by reparametrizing $\boldsymbol{\theta}$ such that the contours of $S(\boldsymbol{\theta})$ become less curved. Or careful choice of the time points t_j may help.

Besides ill-conditioning problems, which might be solved by reparametrization, one cannot overcome *parameter redundancy* with reparametrization, optimal design nor reduced observation error. Parameter redundancy is also termed *approximate lack of identifiability* (Seber and Wild, 1989), meaning that (at least) two parameter vectors exist that—even without observation error—describe approximately the same curve

$$\exists \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \text{ with } \boldsymbol{\theta}_1 \ncong \boldsymbol{\theta}_2 \text{ and such that } C(t, \boldsymbol{\theta}_1) \cong C(t, \boldsymbol{\theta}_2) \text{ for all } t. \tag{2.20}$$

Those parameter redundancy problems are different from identifiability problems encountered in linear regression as they are intrinsic to the model and may exist independently from the data or the noise level. There is a measure of parameter redundancy that—prior to data collection—aims to measure how redundant the parameters of a model are as described in the following.

Redundancy measure

A desirable property of a model is sensitivity to changes in its parameters. If the output of a model changes a lot when varying its parameters, it is easier to estimate those parameters given a measured output. In contrast, if a model lacks sensitivity, one deals with parameter redundancy. As redundancy is high when sensitivity is low, Reich (1981) proposes a measure for parameter redundancy based on the following sensitivity matrix:

$$M_{\boldsymbol{\theta}} = \frac{1}{t_{max}} \int_0^{t_{max}} \frac{\delta C(t, \boldsymbol{\theta})}{\delta \boldsymbol{\theta}} \frac{\delta C(t, \boldsymbol{\theta})}{\delta \boldsymbol{\theta}'} dt. \tag{2.21}$$

The elements m_{ii} of the matrix $M_{\boldsymbol{\theta}}$ measure sensitivity with respect to one parameter θ_i, the elements m_{ij} sensitivity with respect to two parameters θ_i, θ_j varying together (cross-sensitivity). The sensitivity matrix is scaled

$$R_{\boldsymbol{\theta}} = [r_{i,j}] = D^{-1} M_{\boldsymbol{\theta}} D^{-1} \tag{2.22}$$

with $r_{ij} = \frac{m_{ij}}{\sqrt{m_{ii} m_{jj}}}$. Here, $D = diag(\sqrt{m_{11}}, \sqrt{m_{22}}, \ldots)$ and d_{ii} measures sensitivity. Then, the inverse of its determinant, $|R_{\boldsymbol{\theta}}|^{-1}$, is used as measure for parameter redundancy. Reich (1981) proposes to use a threshold value for $|R_{\boldsymbol{\theta}}|^{-1}$.

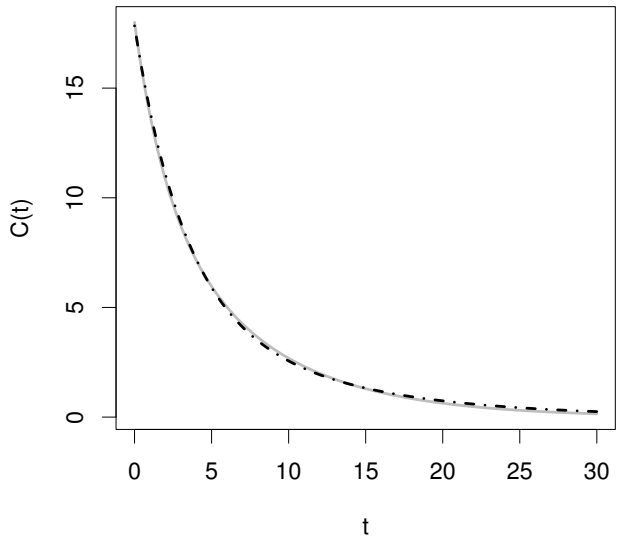

Figure 2.4: Concentration $C(t, \boldsymbol{\theta})$ from a sum of two exponentials for $\boldsymbol{\theta_1} = (\alpha_1, \alpha_2, \beta_1, \beta_2) = (7, 11, 1/2, 1/7)$ (gray curve) and $\boldsymbol{\theta_2} = (11.78, 6.06, 1/3.1, 1/9.4)$ (black dashed curve); modified from Seber and Wild (1989)

Redundancy in compartment models

As described in Section 2.2, the concentration in a compartment model can be described by a sum of exponentials or a sum of convolved exponentials. The nonlinear function

$$C(t, \boldsymbol{\theta}) = \alpha_1 \exp(-\beta_1 t) + \alpha_2 \exp(-\beta_2 t), \tag{2.23}$$

a sum of two exponentials, has been thoroughly discussed in nonlinear regression literature: In Ratkowsky (1990, p. 112 f.) this function is classified as often 'far-from-linear'. Furthermore, in Seber and Wild (1989, p. 118 ff) sum of exponential models are considered as a typical example for parameter redundancy. In Figure 2.4 we show $C(t, \boldsymbol{\theta})$ for two parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Though $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are quite different, the two curves $C(t, \boldsymbol{\theta}_1)$ and $C(t, \boldsymbol{\theta}_2)$ almost coincide. For the model described by a sum of two exponentials (2.23),

redundancy has been investigated in more detail by Reich (1981) who used $|R_{\boldsymbol{\theta}}|^{-1}$ in order to specify how redundant it is for different parameters $\boldsymbol{\theta}$.

The scaled sensitivity matrix $R_{\boldsymbol{\theta}}$ depends on the decay rates β_1 and β_2 only and is independent of the coefficients α_1 and α_2. With $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2)$ the partial derivatives are simple exponentials and the calculation of the integrals m_{ij} is straightforward (Reich, 1981). Due to the special simple model structure, the determinant of $|R_{\boldsymbol{\theta}}|$ and hence the redundancy measure depends only on the ratio of the decay rates β_1 and β_2. One can conclude that parameter redundancy is high, if the decay rates are too similar. Using a threshold of 100 for $|R_{\boldsymbol{\theta}}|^{-1}$ the parameters are judged to be redundant if the decay rates differ by less than a factor of five (Seber and Wild, 1989, p.124, example 3.6). This result inspired us to use regularizing priors on the decay rates, *a priori* assuming that decay rates of different compartments in expectation differ by a factor of five and, like this, avoiding parameter redundancy, see Chapter 7. The calculation of the redundancy measure $R_{\boldsymbol{\theta}}$ for a sum of K exponentials is given in Appendix A.2.

For the model with two convolved exponentials,

$$C(t, \boldsymbol{\theta}) = C_B(t) * (\alpha_1 \exp(-\beta_1 t) + \alpha_2 \exp(-\beta_2 t)) \,, \tag{2.24}$$

the analysis of parameter redundancy is not as easy. This type of model is relevant in the analysis of DCE-MRI with the help of a two tissue compartment model as proposed in Chapters 3–5. For this model the partial derivatives are much more complicated, and the integrals for the calculation of $R_{\boldsymbol{\theta}}$ become difficult (see Appendix A.3). We do not expect that for this model the redundancy measure can analytically be shown to depend on the ratio of the decay rates only. Yet, it is evident that even when convolved with the concentration in the basic compartment, parameters will be redundant if the decay rates differ too little. This is also what we experienced in the simulation studies described in Chapter 4.

Regularized estimation as solution to redundancy issues

Due to redundancy issues, for the different estimation and model selection approaches for compartment models that are proposed in this thesis, the crucial point is to regularize the parameter space taking into account additional, prior information about the parameters and their relationship. We investigate different methods of regularizing the parameter space of compartment models. For the analysis of a FRAP experiment in Chapter 7, we use a regularizing prior on the binding rates inspired by the quantification of redundancy in a sum of two exponentials model developed by Reich (1981). For the analysis of DCE-MR images parameters are estimated per voxel, and hence spatial smoothing is an evident option. Therefore, we propose two different ways of spatially regularizing the parameter space, see Chapters 5 and 6. Like this, voxelwise estimates borrow strength from neighboring voxels and, hence, spatial regularization takes into account the intrinsic information contained in an image.

Regularized estimation is a common technique in generalized linear regression models, especially in non-parametric and semi-parametric regression. Often, in such regression

models there is a large number of predictors, and pure maximum likelihood estimators may be unstable or not unique. Therefore, the parameter space is regularized by imposing a penalty term on the parameters or—in a Bayesian framework—by assuming priors on the parameters. Penalized maximum likelihood based approaches include ridge regression, the lasso, and the elastic net (Hastie et al., 2009). Those shrinkage methods differ in the penalty term that is imposed on the parameters and are also known as continuous alternatives to subset selection procedures. Bayesian automated variable selection approaches include Gibbs variable selection, stochastic search variable selection, and adaptive shrinkage (O'Hara and Sillanpää, 2009). With hierarchical mixture prior formulations promising subsets of predictors may be found in terms of large posterior probability.

However, those methods of regularized estimation are designed for generalized linear regression models only, and hence, in order to use those methods for compartment models, the corresponding nonlinear problem has to be linearized first. A nonlinear model can be linearized by using a bundle of nonlinear basis functions as potential predictors (see Chapter 6). Like this, the nonlinear model is approximated by a linear model on a grid.

Working with the nonlinear regression problem as is (not linearized) is conceptually different. In nonlinear regression, redundancy issues may occur even for a small number of parameters, and hence different methods are suitable. Bayesian methods in image analysis using spatial priors have been proposed for a nonlinear model arising in the analysis of DCE-MRI (Schmid et al., 2006; Kelm et al., 2009) and for linear models arising in the analysis of functional MRI (fMRI) (Gössl et al., 2001; Brezger et al., 2007). Spatial smoothing is an efficient way of regularizing the parameter space in image analysis (see Chapter 5). However, to the best of our knowledge there is no literature analyzing how regularization helps to alleviate parameter redundancy issues in more complex—and hence potentially parameter redundant—nonlinear regression models.

Chapter 3

Introduction to the analysis of DCE-MRI

Large parts of this thesis deal with the regularized estimation and model selection in compartmental models used for the analysis of DCE-MRI. DCE-MRI is an imaging technique which allows to image the blood supply to some tissue of interest *in vivo*. After injection of a contrast medium, a series of images is obtained. This series shows the uptake dynamics of the contrast medium into the tissue over time.

In this chapter, we describe what is measured with DCE-MRI and motivate research questions that can potentially be analyzed with the help of DCE-MRI. Then, we introduce different compartment models which are typically used to quantitatively describe DCE-MRI data. As observed concentration time curves (CTC) are often not adequately described by the commonly used Tofts model (see Section 3.2.1), we propose a two tissue compartment model (2Comp) and as further extension a multi tissue compartment model (qComp) which account for heterogeneity in tumorous tissue. This is important as tumor heterogeneity is diagnostically informative. The statistical performance of those models—applied on a voxel level—is analyzed and compared to the Tofts model in Chapters 4–6. Finally, we present data from a DCE-MRI breast cancer study for which the methods used in Chapters 4–6 are evaluated.

Parts of Section 3.2 are based on Sommer and Schmid (2012).

3.1 DCE-MRI: physical and physiological background

What is measured with DCE-MRI?

DCE-MRI is based on a dynamic series of conventional MRI acquired using a contrast agent enhancing specific magnetic properties. Generally, conventional MRI uses the property of atomic nuclei to align in a magnetic field (nuclear magnetic resonance property). The MRI scanner produces a varying electromagnetic field. Each time this field is turned off, the spins of the atomic nuclei return to their normal state (they relax), energy is set free, and this energy is recorded. One MR image records the T_1 relaxation time for every

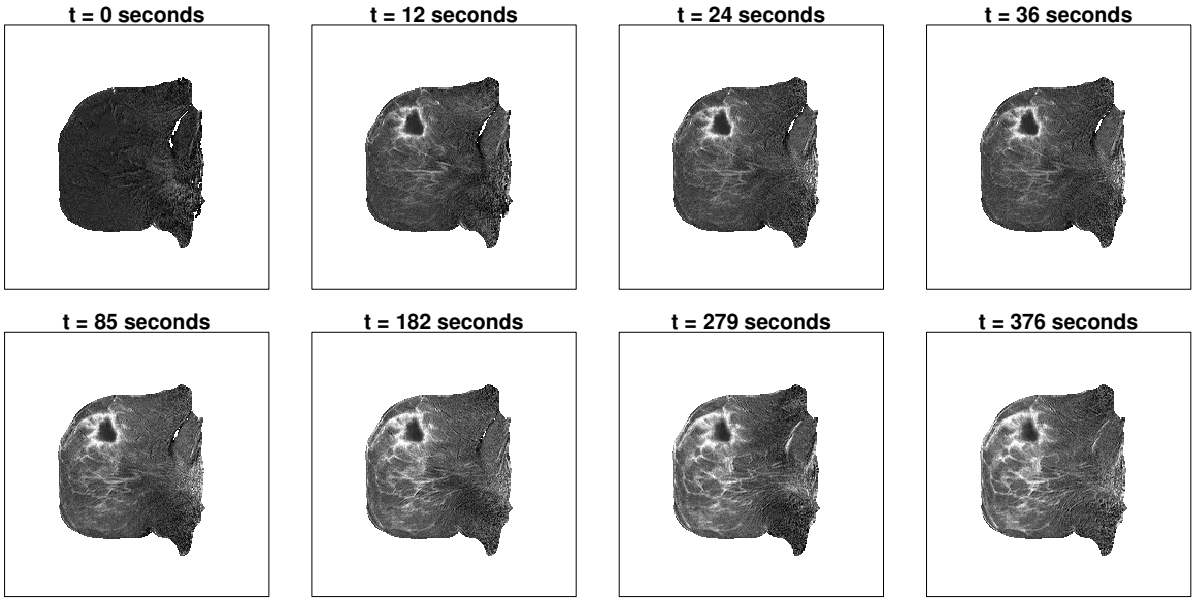

Figure 3.1: DCE-MRI scan of a breast cancer patient's breast showing the contrast agent concentration over time (patient 6, central slice)

voxel after the MRI-device transmitted a magnetic disturbing pulse. For every voxel, the average relaxation time T_1 is measured. For a detailed introduction to the physics of MRI see Edelmann and Warach (1993) and Dössel (2000). In a gradient procedure, several disturbing pulses with different direction are used. Therefore, the average relaxation time can be calculated more precisely and two-dimensional or three-dimensional images can be obtained (Parker and Padhani, 2003).

With DCE-MRI a whole series of MR images is recorded. At the beginning of the image acquisition, a low-molecular-weight contrast medium—typically a Gadolinium complex, *e.g.* gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA) (Mitchell, 1997)—is injected to the patient. The contrast medium acts as a tracer, its distribution within the tissue of interest can be assessed over time. The contrast medium shortens the relaxation times and hence induces increased contrast (Caravan, 2006). From this change in relaxation time one infers on the concentration of the contrast medium.

Commonly, in the analysis of DCE-MRI one assumes that the contrast agent concentration C_t at time t is related to the change in relaxation rate in the following way

$$C_t(t) = \frac{1}{r_1} * \left[\frac{1}{T_1(t)} - \frac{1}{T_{10}} \right]. \tag{3.1}$$

This means that the change in the T_1 signal intensity (due to injection of a contrast medium) is assumed to be proportional to the concentration of contrast medium (Tofts, 1997; Parker and Padhani, 2003). Here, T_{10} is the T_1 of the tissue in absence of contrast agent and r_1 is the longitudinal relaxivity of the contrast agent which is influenced by

the strength of the magnetic field as well as by properties of the tissue and the contrast agent (Parker and Padhani, 2003). Buckley and Parker (2005) give an overview of how contrast agent concentration is calculated from the T_1 signal intensity changes. The imaging technique is called T_1-weighted DCE-MRI to distinguish it from T_2- and T_2^*- weighted DCE-MRI which has a higher temporal and lower spatial resolution. Here, we will generally write DCE-MRI when we actually mean T_1-weighted DCE-MRI.

Summarizing, a dynamic series of contrast enhanced MR images enables to noninvasively observe the uptake of contrast medium by the tissue and its washout from the tissue *in vivo*. Assuming that the contrast agent is transported via the blood plasma, one observes blood perfusion. Figure 3.1 depicts parts of a patient's DCE-MRI series from a study on breast cancer described in Section 3.3 which will be used later on. Even at first glance, one can identify regions with increased and accelerated blood supply or regions with low blood supply.

Why is the tracer concentration of any interest?

The analysis of dynamic contrast medium uptake is especially important in the analysis of tumorous tissue. In tumors, one observes increased blood flow due to abnormal growth of blood vessels (angiogenesis) and due to increased permeability of the capillary wall (capillary leakage) (Parker and Buckley, 2005). Thus, in order to understand the disease process and to analyze the disease progression, one needs to characterize perfusion and permeability of the tissue of interest. Another property observed in tumors is heterogeneity of the tissue (Schmid, 2010). Commonly, compartmental kinetic models are used in order to obtain quantifiable parameters characterizing perfusion and permeability. In this thesis, we aim to additionally characterize and account for tissue heterogeneity by allowing for increased model complexity on a voxel level.

Quantitative analysis of DCE-MRI is successfully used to early detect and diagnose cancer and allows to specify malignancy, type, and grading of tumors. As tumor growth largely depends on the tumor blood supply, many therapies target on destroying tumor vasculature. Other therapies aim to stop the process of angiogenesis. Thus, DCE-MRI helps to check on changes in tumor angiogenesis, to evaluate the response to therapy and, hence, to plan further treatment (Parker and Buckley, 2005). Furthermore, DCE-MRI is useful in the clinical evaluation and development of antiangiogenic and vascular disrupting compounds (Leach et al., 2005; O'Connor et al., 2007; Steingoetter et al., 2011).

DCE-MRI is most frequently used for the assessment of breast cancer (Padhani et al., 2005) but has also shown good results in cancer diagnosis of many other organs, *e.g.*, the brain (Parker and Padhani, 2003; Sourbron et al., 2009), the prostate (McMahon et al., 2009), and the cervix (Donaldson et al., 2010). However, in organs that move with the breathing of the patient, image preprocessing becomes more important, image registration is quite challenging, and signal to noise ratios are comparable low.

Contrast-enhanced perfusion MRI is not only useful in the analysis of cancer but also in quantifying the blood flow in myocardial tissue, *e.g.*, for the evaluation of ischemic heart disease (Schmitt et al., 2002; Schmid, 2011). Furthermore, characteristics like perfusion

and permeability have been shown to be of interest for the diagnosis of retinal disease and multiple sclerosis, and hence DCE-MRI has shown to be useful for the analysis of those diseases as well (Parker and Buckley, 2005). Finally, DCE-MRI has been used in rheumatology for many years, *e.g.*, to assess disease progression and treatment response in rheumatoid arthritis (Hodgson et al., 2008).

Comparison to other imaging modalities

DCE-MRI provides functional information about the tissue of interest combined with morphological information. With DCE-MRI the patient is not exposed to ionizing radiation. This is the main advantage compared to alternative imaging modalities like single photon emission computed tomography (SPECT), positron emission computed tomography (PET), and perfusion computed tomography—also known as dynamic contrast enhanced computed tomography (DCE-CT) or multidetector row computed tomography (MDCT). DCE-MRI offers a better spatial resolution compared to PET and SPECT, but a lower spatio-temporal resolution compared to DCE-CT (Schmid, 2010).

DCE-CT uses contrast agents based on iodine and can potentially be used for similar examinations in oncology as DCE-MRI. A main advantage over DCE-MRI is that in DCE-CT the relation between contrast concentration enhancement and iodine concentration is linear, facilitating the absolute quantification of perfusion. Goh and Padhani (2006) give a detailed overview of the similarities and differences between DCE-MRI and DCE-CT, and for some anatomic regions (upper abdomen, mediastinum, and pulmonary hila) explicitly advise the use of DCE-CT. For serial examinations and brain imaging, however, the use of DCE-MRI is recommended as the radiation burden of DCE-CT becomes unacceptable (Goh and Padhani, 2006).

Another related imaging technique is dynamic susceptibility contrast MRI (DSC-MRI). Similar to DCE-MRI, magnetic properties of a contrast agent are used to describe perfusion in tissue. However, DSC-MRI contrast agents are designed to alter the tissue's magnetic susceptibility resulting in changes of T_2 and T_2^* (Rosen et al., 1990). DSC-MRI is very common in brain perfusion imaging and is, for instance, helpful in the analysis of acute and chronic cerebrovascular disorders (Wintermark et al., 2005).

One has to trade off temporal and spatial resolution when defining an acquisition protocol for DCE-MRI. A high spatial resolution is important when morphological features are of interest. This is, for example, the case when one tries to detect primary tumors or to assess the extent of a known tumor. A high temporal resolution allows a better description of dynamic features needed, for instance, to determine if a previously identified region is malignant (Padhani et al., 2005). Sometimes, diagnosis may benefit from combining information from several imaging modalities. For example, functional imaging techniques like DCE-MRI, PET and SPECT can be combined with morphological information obtained with X-ray computed tomography (CT).

In the screening of breast cancer, other radiologic imaging techniques are successfully used, such as X-ray mammography and ultrasound. However, X-ray mammography shows a reduced sensitivity in tumor detection for dense breasts, it exposes the patient to ra-

diation, and it has only limited use in postsurgical examinations. Therefore, the use of DCE-MRI is recommended for those cases (Padhani et al., 2005).

3.2 Compartment models for DCE-MRI

Model-free analysis—also referred to as semi-quantitative analysis—of contrast medium uptake quantifies simple characteristics like the peak concentration, the time to maximal enhancement, and the initial area under the curve. Those characteristics are easy to derive, but they provide quantities that are difficult to interpret, often largely effected by noise, and hence not reproducible (Padhani et al., 2005). In contrast, compartment models allow to describe the observed uptake behavior with the help of physiologically meaningful parameters linked to properties like blood flow and capillary permeability. Those tissue properties are helpful when diagnosing tumor type and malignancy and when quantifying treatment success. Also, quantitative analysis with compartmental models makes DCE-MRI studies less dependent from scanners, sites, injection protocols, etc. and, hence, more objective, better reproducible, and easier to compare (Padhani et al., 2005). Compartment models assuming various tissue architectures of different complexities have been proposed for the quantitative analysis of DCE-MRI data as described in the following.

3.2.1 One tissue compartment model (1Comp): Tofts and extended Tofts model

Around 1990, three groups independently proposed very similar compartment models to describe signal enhancement of measured T_1 series as a function of time (Brix et al., 1991; Larsson et al., 1990; Tofts and Kermode, 1991). Tofts (1997) gives an overview of similarities and differences of those models, and in Tofts et al. (1999), a standardized notation is proposed to make the results of different groups comparable. There are two versions of the compartment model assuming negligible or non-negligible amount of intravascular tracer. In the context of DCE-MRI analysis the first model is most often referred to as the (standard) "Tofts model", the second one as "extended Tofts model".

Tofts model

Both Tofts models describe the diffusion of some contrast medium with the help of the arterial plasma compartment and one interstitial space compartment (Tofts, 1997; Tofts et al., 1999). The contrast medium is assumed to enter only the interstitial space and not the cells. This assumption is typically fulfilled for low-molecular-weight tracers such as Gd-DTPA. To emphasize that blood plasma and cells are not included in the interstitial space, it is often referred to as extravascular extracellular space (EES). Figure 3.2 depicts the spaces considered in the Tofts model.

In the standard Tofts model, also referred to as 1Comp model here, the observed CTC

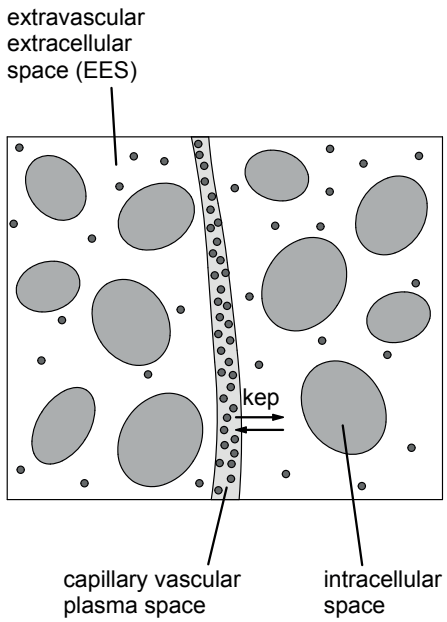

Figure 3.2: Physiological model for distribution of contrast medium (small circles) between plasma and EES modified from Parker and Padhani (2003)

$C_t(t)$ is described by

$$\frac{d}{dt}C_t(t) = K^{\text{trans}}C_p(t) - k_{\text{ep}}C_t(t). \tag{3.2}$$

Given the initial condition $C_t(0) = 0$, the concentration in the observed tissue can be described by the solution of this differential equation:

$$C_t(t) = K^{\text{trans}}C_p(t) * \exp(-k_{\text{ep}}t). \tag{3.3}$$

Here, $C_p(t)$ describes the concentration of contrast agent in the blood plasma over time. The standard Tofts model assumes negligible blood plasma volumes in the observed tissue region such that the observed concentration is due to the concentration in the EES exclusively, see Figure 3.3. The concentration is hence described by two parameters, *i.e.* the transfer rate from blood plasma to EES, K^{trans}, and the rate constant for transfer between EES and plasma, k_{ep}. Alternatively, the model can be described using the EES fractional volume v_e, which relates rate and transfer constants by $k_{\text{ep}} = \frac{K^{\text{trans}}}{v_e}$ (Tofts et al., 1999). As any two of the three parameters k_{ep}, K^{trans}, and v_e suffice to describe the concentration, one can use different parametrizations.

Extended Tofts model

In the extended Tofts model, one assumes that the volume fraction of plasma within the region of interest, denoted as v_p, cannot be neglected. Then, the observed concentration of

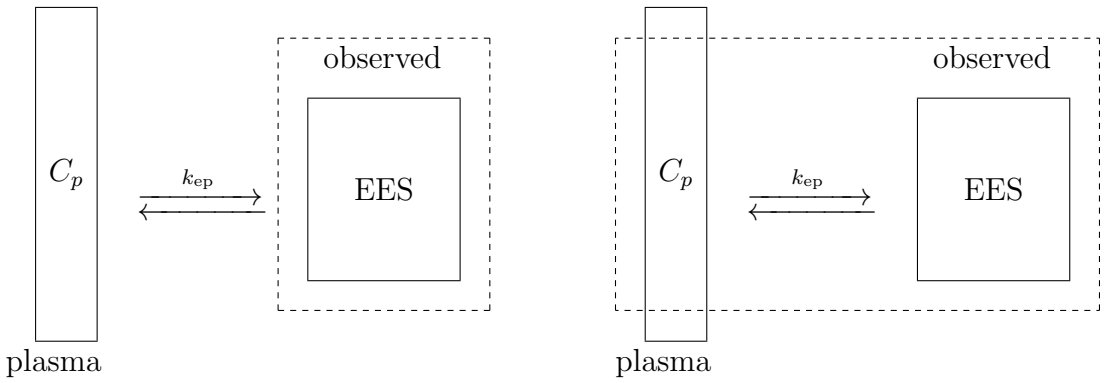

Figure 3.3: Connectivity diagram for the standard Tofts model (left) and extended Tofts model (right)

contrast medium is given as sum of the concentration in plasma and in the EES weighted with the corresponding fractional volumes, $C_t = v_p C_p + v_e C_e$, resulting in

$$C_t(t) = v_p C_p(t) + K^{\text{trans}} C_p(t) * \exp(-k_{\text{ep}} t). \tag{3.4}$$

Thus, the exchange mechanism is the same as in the simple Tofts model, but in the extended Tofts model the plasma compartment is additionally included in the observed space, see Figure 3.3. The inclusion of blood plasma volume in the observed space is important for tissue with high blood flow; for tissue with slow blood flow it can be neglected and the standard Tofts model applies.

Physiological interpretation of parameters

Generally, the rate constant k_{ep} describes the shape of the CTC, whereas the transfer constant K^{trans} and the fractional volume v_e relate to absolute values of the CTC. There are several physiological interpretations of the transfer constant. In a situation where tracer flux is flow limited (capillary permeability is high), K^{trans} is equal to the blood plasma flow per unit volume of tissue. In a situation where tracer flux is permeability limited (low permeability, high plasma flow), K^{trans} is equal to the permeability surface area product between blood plasma and the EES (Tofts et al., 1999).

In Buckley (2002) the uncertainty in estimation of the kinetic parameters has been investigated with the help of a simulation study. Results of this study indicate that K^{trans} tends to be overestimated and v_p tends to be underestimated. In Sourbron and Buckley (2011) conditions are identified under which the Tofts and extended Tofts models can be applied. The authors find that the model parameters can be accurately interpreted only if the tissue is weakly vascularized or highly perfused. In tissues for which those conditions do not hold, the fit to the data may be bad or, despite a good fit, the physiological interpretation of the parameters may be unclear.

Historical development and relation to other compartment models

The Tofts model is a special case of the general compartment model given by equation (2.6), where the arterial plasma compartment forms the basic compartment, and there is only one additional compartment, namely the EES compartment. Therefore, we refer to this model as 1Comp model, also. For simplicity, we will sometimes refer to the compartments as plasma and tissue compartments.

Note that similar compartment models have been proposed much earlier, but using different notation and assumptions on model parameters (Kety, 1960; Renkin, 1959; Crone, 1963). The Kety model was developed to model the distribution of an inert gas into the whole tissue including the intracellular spaces and assuming high permeability (Kety, 1960). Renkin (1959) defined an extraction ratio, assuming that capillary blood flow is reduced when it passes through tissue. The Renkin-Crone model relates the entities flow, permeability, and extraction (Renkin, 1959; Crone, 1963). The relation of those models to the Tofts model are described in Tofts et al. (1999).

A refinement of the Tofts model has been suggested by St. Lawrence and Lee (1998). With a tissue homogeneity model (Johnson and Wilson, 1966) differences of plasma concentration along the capillaries are accounted for, and the tracer concentration in plasma is modeled as a function of space and time. Using an adiabatic approximation, that is, assuming that concentration of the contrast medium in the interstitial space changes slowly compared to changes in the capillary plasma space, a closed form solution can be derived which is related to the Tofts model and makes the tissue homogeneity model practically applicable (St. Lawrence and Lee, 1998). Other extensions of the Tofts model result in models with more compartments and are described in the next section.

Arterial input function

The concentration in plasma, C_p, is also called arterial input function (AIF) as it describes the input of contrast agent through the blood stream. Ideally, one observes the concentration of tracer in the plasma compartment and the concentration in the EES compartment separately. This is, for instance, the case when one samples blood from the patient every some seconds and in parallel a DCE-MRI series is recorded. However, blood sampling is inconvenient for the patient, and the calculation of the AIF is technically demanding. Therefore, one usually prefers a noninvasive examination of the patient and records DCE-MR images only. In breast DCE-MRI, however, the input function cannot be measured from an artery due to the lack of big vessels in the field of view. Hence, using a population based AIF from literature is the standard procedure, and an ideal AIF is assumed which is derived from typical blood sampling curves observed for a group of healthy volunteers.

Frequently, a population based AIF is used which was proposed by Tofts and Kermode (1991). Tofts and Kermode (1991) use a bi-exponential function of the form

$$C_p(t) = D(a_1 \exp(-m_1 t) + a_2 \exp(-m_2 t)) \tag{3.5}$$

with dose D depending on the experimental conditions and fixed values $a_1 = 3.99\ kg/l$, $a_2 = 4.78\ kg/l$, $m_1 = 0.144\ min^{-1}$, $m_2 = 0.0111\ min^{-1}$. Precise parameters of this

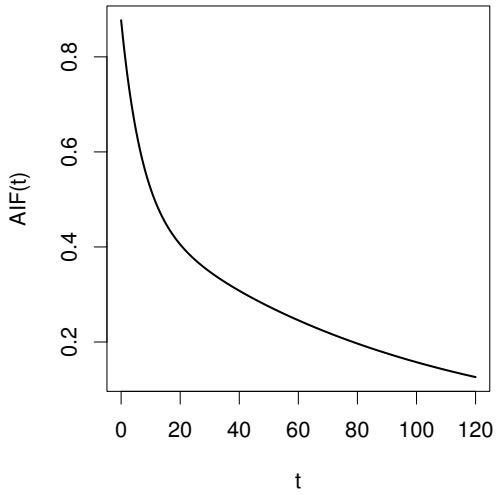

Figure 3.4: Form of population based AIF proposed by Tofts and Kermode (1991) (assuming dose $D = 0.1$, time t in minutes)

AIF were derived by fitting plasma concentrations which were measured in normal control subjects by Weinmann et al. (1984) after a bolus injection of Gd-DTPA. The form of this AIF is depicted in Figure 3.4. The concentration of contrast medium in plasma is highest directly after its injection; then, there is a fast washout phase followed by a slow washout phase. Throughout this thesis, we will assume this population based AIF. With this AIF, the convolved exponential (3.3) describing the observed concentration is of the form

$$\begin{aligned} C_t(t) &= K^{\text{trans}} C_p(t) * \exp(-k_{\text{ep}} t) \\ &= DK^{\text{trans}} \left(a_1 \frac{\exp(-k_{\text{ep}} t) - \exp(-m_1 t)}{m_1 - k_{\text{ep}}} + a_2 \frac{\exp(-k_{\text{ep}} t) - \exp(-m_2 t)}{m_2 - k_{\text{ep}}} \right). \end{aligned}$$

The derivation of this form is given in Appendix A.1.

Another population-averaged AIF has been proposed in Parker et al. (2006). Attempts have been made to measure the AIF noninvasively (Fritz-Hansen et al., 1996). The AIF can be measured noninvasively if there is an artery in the imaged field and if this artery is close to the tissue of interest. When measuring the input function from an artery is feasible, a deconvolution procedure is necessary in order to make the AIF comparable with an AIF from blood samples. As alternative, reference tissue approaches aim to simultaneously extract the AIF and kinetic parameters from DCE-MRI data (Yang et al., 2007; Fluckiger et al., 2009).

3.2.2 Two tissue and multi tissue compartment model (2Comp and qComp)

Compartment models are based on several simplifying assumptions. Therefore, a compartment model (like any other model) cannot describe full reality but can only offer an

approximation of the underlying physiologic processes. In the Tofts model, many potential factors of contrast medium diffusion are not explicitly taken into account but are seen as confounding factors. Factors that—besides blood flow and permeability—could have an effect on contrast agent distribution include "the rate of lymphatic drainage, the rate or degree of contrast agent mixing within compartments, and the degree of tracer diffusion in regions without an effective blood supply"(Parker and Buckley, 2005, p. 83). Furthermore, Parker and Buckley (2005) stress that in tissue voxels examined with DCE-MRI one deals with multiple membranes rather than with a single cell membrane. Therefore, one should be aware that the parameters estimated in a compartment model are always bulk parameters.

Tumor heterogeneity

In healthy tissue, observed tracer kinetics is often sufficiently well described by the Tofts or extended Tofts models despite the simplifying assumptions made. However, the microvasculature in tumors often shows a high degree of heterogeneity (Yang and Knopp, 2011) and the Tofts and extended Tofts models fail in describing its observed uptake dynamics (Schmid et al., 2009). In Port et al. (1999) models with up to three tumor compartments have been proposed for the analysis of contrast medium uptake in mammary tumors, and the different models have been compared using a likelihood-ratio criterion. Analysis was performed on a region of interest level, not per voxel. According to their results, more than one tumor compartment was needed in 85% of the carcinomas in order to describe the contrast medium kinetics. The authors conclude that kinetic heterogeneity within tumors is characteristic of malignant tumors.

The heterogeneity of curve shapes observed in tumors is judged to be diagnostically informative (Padhani et al., 2005), and therefore "analysis and presentation of imaging data needs to take into account the heterogeneity of tumor vascular characteristics" (Parker and Padhani, 2003, p. 356). Several authors claim that tumor heterogeneity is linked to the malignancy and aggressiveness of tumors and can be a biomarker for the failure of cancer treatments (Giesel et al., 2006; Wyss et al., 2007; Agrawal et al., 2009; Rose et al., 2009). However, the quantification of tissue heterogeneity is still an unsolved issue (Yang and Knopp, 2011). To this end, models with more compartments have been proposed. However, these models are based on a region of interest (ROI), that is, the CTC are summed up over the ROI and not used on a voxel level (Port et al., 1999; Brix et al., 2009; Sourbron and Buckley, 2011). By analyzing summed curves on a ROI level one loses information on the spatial structure contained in the data.

Another approach to assess tumor heterogeneity is the analysis of small ROIs, for example predefined semi-automatically using clustering methods (Liney et al., 1999; Mohajer et al., 2012). Though, as these approaches use an independent analysis for each of the small ROIs they still lose spatial information for the kinetic analysis. As an alternative, texture analysis (Karahaliou et al., 2010) and fractal approaches (Rose et al., 2009) have been used on DCE-MRI data in order to quantify tissue heterogeneity. However, both approaches do not account for the underlying kinetics of the contrast medium in the tissue.

Therefore, we aim to characterize tissue heterogeneity with the help of a two tissue compartment model. Applied on a voxel level, the proposed approach allows to assess the heterogeneity of contrast medium uptake in the tissue (see Chapters 4 and 5). In the following, we introduce the two tissue compartment model deterministically. Based on this, corresponding statistical models can be defined either on a voxel level or on a ROI level.

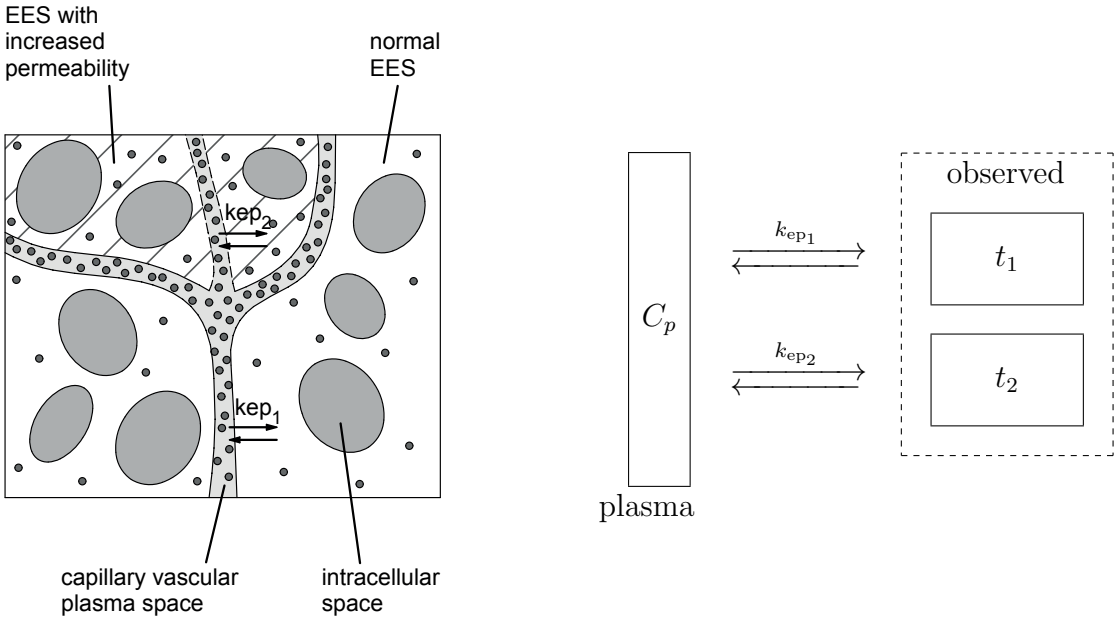

Figure 3.5: Sketch of physiological model and connectivity diagram of 2Comp model. Compartments used to model the diffusion of contrast medium (small circles) are sketched left: the concentration in the capillary vascular plasma space is $C_p(t)$, the concentration in normal EES is $C_{t_1}(t)$, and in EES with increased permeability (ruled) it is $C_{t_2}(t)$. Contrast medium does not enter the cells. The corresponding abstract two tissue compartment model is sketched right.

Two tissue compartment model (2Comp)

In order to adequately describe the uptake dynamics of contrast medium in tumor tissue, we propose a compartment model that allows for tissue heterogeneity. We assume that there may be two tissue compartments with different kinetic properties that exchange with plasma at distinct rates k_{ep_1} and k_{ep_2} (see Figure 3.5). Hence, we assume a model with two tissue compartments (2Comp) which can be interpreted as parts of the interstitial space with normal blood supply happening at rate k_{ep_1} and parts with increased (abnormal) blood supply happening at rate $k_{\mathrm{ep}_2} > k_{\mathrm{ep}_1}$. However, one should carefully interpret the kinetic parameters and be aware that those parameters are always bulk parameters describing an average exchange rate rather than the exchange rate of a single membrane.

This is because, at a voxel scale, one actually observes a system of multiple membranes (Parker and Buckley, 2005). The assumption of a single average exchange rate works well in healthy tissue, in heterogeneous tumor tissue it is too simplistic (Schmid, 2010). At tumor edges, for instance, one typically observes a fast contrast medium uptake followed by a slow washout. This leads to plateau shaped CTCs which cannot adequately be described by the Tofts model. CTCs of tumor edge voxels are, for example, depicted in Figure 4.5 b) and 5.3 c).

As $k_{\mathrm{ep}_k} = K_k^{\mathrm{trans}}/v_{t_k}$ (Tofts et al., 1999), the volume v_{t_k} of tissue t_k per unit volume of tissue can be computed as $v_{t_k} = K_k^{\mathrm{trans}}/k_{\mathrm{ep}_k}$. As described in Chapter 2, the implicit assumption in a compartment model is that the outflux of concentration of a certain compartment is proportional to the current concentration in this compartment, and the influx to this compartment is proportional to the current concentration in the source compartment. Therefore, here the changes in tissue concentrations are given by

$$\begin{aligned} v_{t_1} \tfrac{d}{dt} C_{t_1}(t) &= K_1^{\mathrm{trans}} \left[C_p(t) - C_{t_1}(t)\right] \\ v_{t_2} \tfrac{d}{dt} C_{t_2}(t) &= K_2^{\mathrm{trans}} \left[C_p(t) - C_{t_2}(t)\right]. \end{aligned} \tag{3.6}$$

The solution of these differential equations is given by

$$C_{t_k}(t) = C_p(t) * \frac{K_k^{\mathrm{trans}}}{v_{t_k}} \exp\left(-\frac{K_k^{\mathrm{trans}}}{v_{t_k}} t\right), \tag{3.7}$$

for k=1,2 denoting the different tissue compartments. Given the initial conditions $C_{t_1}(0) = C_{t_2}(0) = 0$, the total (observable) concentration is then given as $C_t = v_{t_1} C_{t_1} + v_{t_2} C_{t_2}$ by

$$C_t(t) = C_p(t) * \left(K_1^{\mathrm{trans}} \exp(-k_{\mathrm{ep}_1} t) + K_2^{\mathrm{trans}} \exp(-k_{\mathrm{ep}_2} t)\right). \tag{3.8}$$

Summarizing, the 2Comp model allows for tissue heterogeneity. With this model applied on a voxel level, CTCs in heterogeneous tissue can be more adequately described, especially at tumor margins, and tissue heterogeneity becomes accessible, see Chapters 4 and 5.

Relation to other compartment models

In the proposed 2Comp model, the observed concentration $C_t(t)$ is described by an impulse response function (sum of two exponentials) convolved with the AIF, see equation (3.8). In the two compartment exchange model (2CXM) proposed by Brix et al. (2009) the interstitial space and the interstitial plasma are modeled with separate compartments. This leads to a model with nested compartments (see Section 2.2). Though explained by different compartment designs, the impulse response of the 2CXM is also a sum of two exponentials. Hence, the 2Comp model and the 2CXM lead to the same nonlinear regression problem. We prefer to use the 2Comp model due to the charming fact that the impulse response is directly expressed by interpretable parameters K_1^{trans}, K_2^{trans}, k_{ep_1} and k_{ep_2}. In contrast, in the 2CXM the impulse response is expressed by auxiliary variables (called F_+, F_-, K_+ and K_-) which are complicated functions of interpretable quantities—see Lemma 3 of Sourbron and Buckley (2011).

For the case that the exchange rates are the same, $k_{\mathrm{ep}_1} = k_{\mathrm{ep}_2}$, or when one of the tissue volumes vanishes, $v_{t_1} = 0$ or $v_{t_2} = 0$, the impulse response reduces to a single exponential and the 2Comp model corresponds to the standard Tofts model (1Comp model).

For the case that one exchange rate becomes infinite, $k_{\mathrm{ep}_2} = \infty$, the observed concentration is of the form

$$C_t(t) = v_{t_2} C_p(t) + C_p(t) * K_1^{\mathrm{trans}} \exp(-k_{\mathrm{ep}_1} t). \tag{3.9}$$

In this case, the second tissue compartment takes the role of an interstitial plasma compartment, $C_{t_2} = C_p$, and the 2Comp model corresponds to the extended Tofts model.

Multi tissue compartment model (qComp)

As a further generalization of the extended Tofts model and its extensions described above, the multi-compartment model with q tissue compartments (qComp model) can be defined by

$$C_t(t) = v_p C_p(t) + \sum_{k=1}^{q} C_p(t) * K_k^{\mathrm{trans}} \exp(-k_{\mathrm{ep}_k} t). \tag{3.10}$$

With this model local physiological effects are even more flexibly modeled compared to the 2Comp model. The model allows for q tissue compartments and the inclusion of a plasma volume fraction in the observed space. The connectivity diagram of the qComp model is depicted in Figure 3.6.

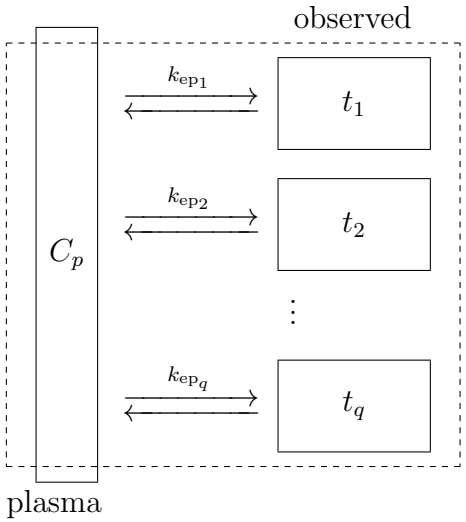

Figure 3.6: Connectivity diagram of qComp model

The idea is to allow for this very general model and to determine the adequate model complexity from the data. Models with more compartments have previously been used in the analysis of PET data; *e.g.*, in Gunn et al. (2002) a basis pursuit strategy has been used to determine the adequate number of compartments per voxel, however, not taking into

account the spatial structure of an image. In Steingoetter et al. (2011) a multi compartment model is used to model DCE-MRI data from simultaneously acquired CTCs for different ROIs (muscle, liver and tumor tissue). In Chapter 6, we analyze data from a DCE-MRI breast cancer study with the help of the proposed qComp model, and for each voxel we determine the number of tissue compartments q and decide whether the plasma volume term is actually needed.

3.3 DCE-MRI data from a breast cancer study

In the following, we introduce a DCE-MRI data set which is analyzed in Chapters 4–6 to evaluate the clinical use of proposed models and estimation approaches. This dataset was graciously provided by Dr. A.R. Padhani at Paul Strickland Scanner Center, Mount Vernon Hospital, Northwood, U.K.. It consists of DCE-MRI scans of breast cancer patients which were previously reported about (Ah-See et al., 2004; Schmid et al., 2005, 2006).

The dataset is part of a study aiming to assess the role of DCE-MRI in predicting success or failure of neoadjuvant chemotherapy. Data were acquired for twelve patients with primary breast cancer treated with chemotherapy. Informed consent was obtained from all patients and the study was acquired in accordance with recommendations given by Leach et al. (2005). Chemotherapy is expected to stop the process of angiogenesis, *i.e.*to lower the elevated blood flow to the tumor, and hence to lower the kinetic parameters K^{trans} and k_{ep}. Therefore, one attempts to predict the response to therapy with the help of DCE-MRI scans. Like this, one can react early and—in case of nonresponse—plan alternative treatment.

In total, the chemotherapy lasted six weeks. The first DCE-MRI scan was acquired shortly before the beginning of chemotherapy and is referred to as pre-treatment scan. A second scan was acquired after two weeks, termed post-treatment scan. The question under study was if success or failure of chemotherapy for a patient could be predicted using pre-treatment and post-treatment scans. After the second scan, the chemotherapy continued for four more weeks. At the end of therapy, tumors were removed and, as clinical evaluation, the response to therapy was evaluated histologically. Six of the twelve patients were identified as responders, the other six as nonresponders.

The scans were acquired with a 1.5 T Siemens MAGNETOM Symphony scanner, $TR = 11\ ms$ and $TE = 4.7\ ms$. The breasts were restrained during the acquisition and no motion was visible in the scans. Each scan consists of three slices of 230×256 voxels. The field of view for all scans was $260 \times 260 \times 8\ mm$ per slice, and thus the voxel dimension is $1.016 \times 1.016 \times 8\ mm$. A dose of $D = 0.1\ mmol/kg$ body weight Gd-DTPA was injected at the start of the fifth acquisition using a power injector. The first four scans were used to compute average T_{10} values. Regions of interest cover the tumor and surrounding normal tissue (Schmid et al., 2006).

Scans were acquired every 12.14 seconds over 36–46 equidistant time points after contrast medium injection, the total acquisition time was about ten minutes. Thus, the DCE-MRI data studied here have four dimensions—two dimensions within one image, the

third dimension for the slices, and the fourth dimension is time. However, only the central slice of the DCE-MR images is used for our analysis. Thus, we are dealing with a series of two-dimensional images over time. Figure 3.1 depicts a part of an DCE-MR imaging sequence taken from one of the breast cancer patients pre-treatment. This series shows the uptake dynamics of blood (and hence contrast medium) into the tissue over time. One can consider this either as a series of images, or as a concentration time series in each voxel.

We will analyze this data set on a voxel level. Hence, when implementing and evaluating estimation algorithms one has to deal with the described model complexity, *i.e.* for each scan concentration time curves of up to $230 \times 256 = 58,880$ voxels are to be described with the help of biologically meaningful parameters. For a voxelwise analysis of DCE-MRI typically parametric maps are used to visualize estimation results.

Chapter 4

Bayesian independent estimation per voxel

In Chapter 3, we have described how DCE-MRI data can be described with the help of compartment models linking observed concentration time curves (CTCs) to physiologically meaningful parameters. However, there are competing models, and the form and complexity of the compartment model is not known and should be seen as part of the scientific problem. On the one hand, knowledge about the true underlying biological or physiological processes has to be taken into account and should be represented in the model. On the other hand, the model has to match the observation made about reality in an experiment.

Often, scientists assume that certain underlying processes are more likely than others, and they can guess how reasonable parameter values might look like. The Bayesian theory offers a framework where prior knowledge can be explicitly taken into account and is brought in accordance with observations made. The posterior distribution obtained with a Bayesian approach provides valuable information on model fit and complexity as well as criteria for model selection.

We aim to decide how heterogeneous tissue in a DCE-MRI breast cancer study is and whether more complex models are needed than the standard Tofts model. Our first approach, presented in this chapter, is a Bayesian framework to estimate the parameters of the 1Comp and 2Comp models proposed in Section 3.2.2 independently per voxel. The proposed framework allows one to decide which model is more suitable per voxel based on the deviance information criterion (DIC). We analyze for which parameter constellations redundancy issues occur in a voxelwise 2Comp model based on simulation studies. We discuss how those redundancy issues can be detected by the effect they have on the posterior distribution obtained from MCMC sampling. Finally, we fit the voxelwise 1Comp and 2Comp models to data from the breast cancer study described in Section 3.3 and search for areas where the 2Comp model outperforms the 1Comp model. Our results show that this happens basically at tumor edges. However, due to redundancy issues the parameters of the voxelwise 2Comp model are not stably estimable for all voxels. Also, the use of the DIC is limited when dealing with multimodal posteriors due to redundancy issues.

This Chapter is mainly based on Kärcher and Schmid (2010). Sections 4.1.4 and 4.1.5

are in parts based on Sommer and Schmid (2012).

4.1 Methods

Using a compartment model, the theoretical form of the contrast medium concentration can be described with the help of an analytical function of kinetic parameters and time as explained in Chapter 3. Due to observation error, the observed contrast medium concentration $Y_{i,j}$ at time t_j, $j = 1, ..., T$ in voxel $i = 1, ..., N$ can hence be described by the theoretical concentration time curve C_t depending on the voxel-specific kinetic parameters $\boldsymbol{\phi}^i$ plus a stochastic error term $\epsilon_{i,j}$

$$Y_{i,j} = C_t(\boldsymbol{\phi}^i; t_j) + \epsilon_{i,j}. \tag{4.1}$$

Here, we assume Gaussian noise, more specifically $\epsilon_{i,j} \sim N(0, \sigma_i^2)$, such that the nonlinear regression problem can be expressed via

$$Y_{i,j} \sim N\left(C_t(\boldsymbol{\phi}^i; t_j), \sigma_i^2\right), \tag{4.2}$$

with σ_i^2 the voxel-specific variance of the Gaussian noise (Schmid et al., 2006). The form of C_t and the number of kinetic parameters is fixed, given a specific compartment model. Here, we compare the simple Tofts model (1Comp) with kinetic parameters $\boldsymbol{\phi}^i = (k_{\text{ep}}^i, K^{\text{trans},i})$ and the two tissue compartment model (2Comp) proposed in Section 3.2.2 with $\boldsymbol{\phi}^i = (K_1^{\text{trans},i}, K_2^{\text{trans},i}, k_{\text{ep}_1}^i, k_{\text{ep}_2}^i)$.

For each voxel, the parameters $\boldsymbol{\phi}^i$ of the 1Comp and 2Comp models are estimated independently using a Bayesian non-linear regression approach (Schmid et al., 2006; Orton et al., 2007). As the Bayesian approach allows to use prior knowledge on the biological parameters, it is more stable than a least squares approach. This is important as the number of parameters grows with model complexity. The Bayesian approach offers another advantage: the posterior distribution contains additional valuable information on the suitability of model fit and complexity. We exploit this by developing a model selection technique based on the DIC (Spiegelhalter et al., 2002) and propose to use the number of effective parameters to measure tissue heterogeneity.

In a Bayesian approach, probabilistic models have to be specified both for the data and the model parameters before observation (Gelman et al., 2004). Here, we choose the priors on the model parameters such that they reflect both uncertainty and physiological facts or assumptions. Inference is based on the posterior distribution which is typically calculated with the help of MCMC simulations (Gilks et al., 1996; Robert and Casella, 2004).

4.1.1 Bayesian 1Comp model voxelwise

The Tofts model with one tissue compartment was introduced in Section 3.2.1 and the form of the contrast medium concentration C_t in the nonlinear regression problem (4.1) is deterministically described by equation (3.3). In Schmid et al. (2006), the following

probabilistic model is used to incorporate prior knowledge on the kinetic parameters in this compartmental model. Gaussian priors on the logarithmic transforms of the kinetic parameters are assumed

$$\begin{aligned} \log(K^{\text{trans},i}) &\sim N(0,1) \\ \log(k^i_{\text{ep}}) &\sim N(0,1) \end{aligned}$$

for all voxels i independently. With this parametrization the rate constant and transfer constants, $K^{\text{trans},i}$ and k^i_{ep}, are insured to be positive, and with this prior they do not exceed values of 20 min^{-1} with 99,86% probability *a priori* (Schmid et al., 2006). This prior reflects our assumption for reasonable kinetic parameter values. A relatively flat inverse Gamma prior is used for the noise variance: $\sigma_i^2 \sim IG(c,d)$ with $c = 0.0001, d = 0.0001$. This model is used for comparison with the following model with two tissue compartments.

4.1.2 Bayesian 2Comp model voxelwise

In Section 3.2.2, we proposed a model with two tissue compartments as alternative to the Tofts model. For this model the form of C_t in the nonlinear regression problem (4.1) is deterministically described by equation (3.8). We choose the parametrization with fractional volumes and rate constants and denote $\boldsymbol{\theta}_i = (v^i_{t_1}, v^i_{t_2}, k^i_{\text{ep}_1}, k^i_{\text{ep}_2}, \sigma_i^2) \in \mathbb{R}^5_+$.

The parameters $\boldsymbol{\theta}_i$ are estimated for each voxel independently using a Bayesian approach (Schmid et al., 2006; Orton et al., 2007). We assume the following prior distributions for every voxel $i = 1, ..., N$:

$$v^i_{t_1}, v^i_{t_2} \sim Beta(a,b)$$

with $a = b = 1$, *i.e.,* a uniform distribution on $[0,1]$, as the volume of each compartment cannot exceed 100 percent. Note that $v^i_{t_j} \sim U[0,1]$ is equivalent to $K_j^{\text{trans},i} | k^i_{\text{ep}_j} \sim U[0, k^i_{\text{ep}_1}]$ as proposed in Orton et al. (2007). Using the fact that rate constants are non-negative, we assume

$$\log(k^i_{\text{ep}_2}) \sim N(0,1)$$

as in Schmid et al. (2006). In Section 3.2.2, we defined that the rate constant k^i_{ep} in the first compartment is smaller than in the second compartment. Therefore, we use the following prior

$$k^i_{\text{ep}_1} | k^i_{\text{ep}_2} \sim U\left[0, k^i_{\text{ep}_2}\right].$$

The Bayesian model is complete with an Inverse Gamma prior for the observation variance, $\sigma_i^2 \sim IG(c,d)$ with $c = 0.1, d = 0.001$.

4.1.3 Implementation

In order to obtain parameter estimates and the posterior probability density function (PDF), we use MCMC methods. By Bayes' theorem, the posterior density for the param-

eters in voxel i, $\boldsymbol{\theta}_i$ is given by

$$p(\boldsymbol{\theta}_i|C_t^i(t)) = \frac{p(C_t^i(t)|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)}{\int p(C_t^i(t)|\tilde{\boldsymbol{\theta}}_i)p(\tilde{\boldsymbol{\theta}}_i)d\tilde{\boldsymbol{\theta}}_i} \propto p(C_t^i(t)|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i), \tag{4.3}$$

where the joint density of the parameters $p(\boldsymbol{\theta}_i)$ is given by

$$\begin{aligned} p(\boldsymbol{\theta}_i) &\propto p\left(v_{t_1}^i\right) p\left(v_{t_2}^i\right) p\left(k_{\mathrm{ep}_1}^i|k_{\mathrm{ep}_2}^i\right) p\left(k_{\mathrm{ep}_2}^i\right) p\left(\tfrac{1}{\sigma_i^2}\right) \\ &= \mathbf{1}\left(v_{t_1}^i \in [0,1]\right) \\ &\times \mathbf{1}\left(v_{t_2}^i \in [0,1]\right) \\ &\times \exp\left(-\tfrac{1}{2}\left(\log(k_{\mathrm{ep}_2}^i)\right)^2\right) \\ &\times \tfrac{1}{k_{\mathrm{ep}_2}^i}\mathbf{1}\left(k_{\mathrm{ep}_1}^i \in \left[0, k_{\mathrm{ep}_2}^i\right]\right) \\ &\times \left(\tfrac{1}{\sigma_i^2}\right)^{c-1} \exp\left(-\tfrac{d}{\sigma_i^2}\right), \end{aligned} \tag{4.4}$$

and the likelihood $p(C_t^i(t)|\boldsymbol{\theta}_i)$ is

$$p(C_t^i(t)|\boldsymbol{\theta}_i) \propto \frac{1}{\sigma_i} \exp\left(-\frac{1}{2\sigma_i^2}\sum_{j=1}^{T}\left(C_i(t) - f\left(\boldsymbol{\theta}_i, t_j\right)\right)^2\right). \tag{4.5}$$

Let $\epsilon_{ij} = Y_{i,j} - C_t(\boldsymbol{\phi}^i, t_j)$ denote the random noise terms. Then, the full conditional of the inverse noise variance $\tau_\epsilon^i = \frac{1}{\sigma_i^2}$ is a Gamma-distribution

$$\tau_\epsilon^i|\cdot \sim Ga\left(c + \frac{T}{2}, d + \frac{1}{2}\sum_{j=1}^{T}\epsilon_{ij}^2\right). \tag{4.6}$$

Hence, we can use Gibbs updates to update τ_ϵ^i. For all other parameters, the full conditionals do not have a closed form to sample from. Thus, Metropolis-Hastings steps are performed. Parameters of the first compartment $(v_{t_1}^i, k_{\mathrm{ep}_1}^i)$ are updated in one block, the parameters of the second compartment $(v_{t_2}^i, k_{\mathrm{ep}_2}^i)$ are updated in a second block using random walk proposals. The proposal variance is tuned such that acceptance rates are about 40 %. For each voxel, we run 25.000 iterations and discard 12.000 iterations as burnin. For storage purposes samples are thinned; every tenth sample is saved.

We implemented the proposed 2Comp model extending the R-package `dcemriS4` (Whitcher and Schmid, 2009, 2011). The MCMC update steps are implemented in a C-procedure which is embedded in R. For efficiency, the MCMC procedure is run in parallel for several voxels.

4.1.4 Measure of model fit and complexity

Work has been done on model selection in compartment models used for DCE-MRI or PET analysis, however, not in a Bayesian framework. In Turkheimer et al. (2003), model

selection in PET studies is based on the Akaike information criterion (AIC), and model averaging is proposed to be based on Akaike weights. Similarly, Brix et al. (2009) base multimodel inference for DCE-MRI data on the AIC and on Akaike weights. Both in Turkheimer et al. (2003) and Brix et al. (2009), model choice is done on a ROI level, not on a per voxel level. In Ingrisch et al. (2010), the use of the AIC for model choice in DCE-MRI is evaluated, and model selection strongly depends on the length of acquisition time. Dependence of AIC based model selection on the sample size is a known problem and the use of a corrected AIC is recommended for small sample sizes (Burnham and Anderson, 2002). As alternative, Schwarz (1978) proposed the Bayesian information criterion (BIC) derived as an approximation of the logarithmized marginal likelihood of a model.

Similar to the AIC and the BIC, most information criteria share a similar form: They include a measure of lack of model fit plus a penalty term for increasing model size (Burnham and Anderson, 2002). Model size in this context is typically synonymous with the number of model parameters. In a Bayesian framework, however, the size of a model is not as clearly defined as the level of hierarchy may vary and still lead to the same marginal distributions. Therefore, Spiegelhalter et al. (2002) propose that the model complexity should also reflect the difficulty in estimation. They propose a measure of model complexity p_D which estimates the effective model complexity from the data. Based on the effective number of parameters p_D the DIC has been derived (Spiegelhalter et al., 2002). As, here, we base inference on a Bayesian hierarchical framework, we use the DIC to evaluate the model fit and to compare the 1Comp and 2Comp models per voxel. Furthermore, we use p_D maps to visually assess tissue heterogeneity.

The DIC is defined as the posterior median deviance plus the effective number of parameters p_D (Spiegelhalter et al., 2002)

$$\mathrm{DIC} = D(\bar{\boldsymbol{\theta}}) + p_D. \tag{4.7}$$

The deviance, defined as $D(\boldsymbol{\theta}) = -2l(\boldsymbol{\theta})$, with $l(\boldsymbol{\theta})$ the log-likelihood, measures the fit of the estimated parameters to the data. In contrast to other model selection criteria with penalty terms for higher model dimensions that are based on the *a priori* fixed number of model parameters, the DIC penalizes for the effective number of parameters which is calculated from the posterior distribution. The effective number of parameters p_D is calculated as difference of the posterior median of the deviance and the deviance evaluated at the posterior median value (Spiegelhalter et al., 2002)

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}). \tag{4.8}$$

However, as Spiegelhalter et al. (2002) point out, p_D can become small or even negative in cases where the posterior mean or posterior median is a poor estimator. This is certainly the case when dealing with multimodal posteriors due to parameter redundancy. Hence, we can implicitly detect redundancy issues by looking at the p_D.

Higher p_D values indicate an increased effective model complexity in the data observed, and hence we suggest using p_D to measure tissue heterogeneity. Here, we compute the p_D and the DIC for each voxel. The deviance in voxel i is $D(\boldsymbol{\theta}_i) = -2 * l(\boldsymbol{\theta}_i)$ where

$l(\boldsymbol{\theta}_i) = \log(p(C_t^i(t)|\boldsymbol{\theta}_i))$ is the log-likelihood function which is easily derived from equation (4.5). It is evaluated at the posterior median values of $\boldsymbol{\theta}_i$ in order to calculate the deviance of the median and it is evaluated at each sample value of $\boldsymbol{\theta}_i$ in order to calculate the median deviance.

4.1.5 Redundancy issues in the independent voxelwise model

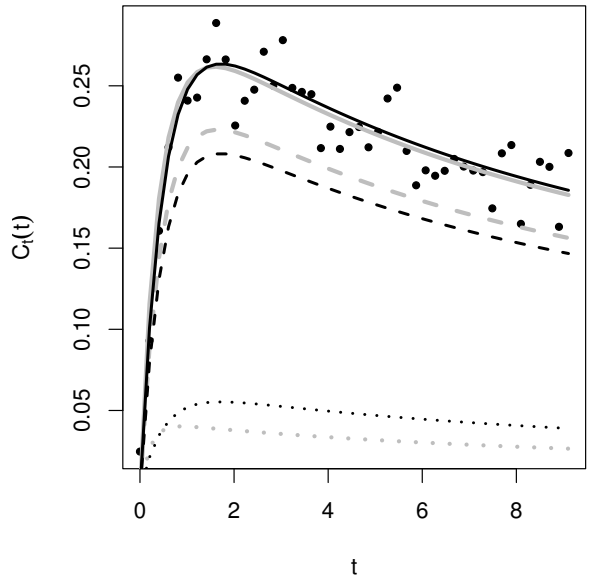

Figure 4.1: Similar CTCs for two different parameter vectors. Black line: CTC described by $k_{\text{ep}_1} = 2.07$, $k_{\text{ep}_2} = 2.07$, $K_1^{\text{trans}} = 0.55$, $K_2^{\text{trans}} = 0.15$ (can as well be described by only one compartment with $k_{\text{ep}_1} = 2.07$ and $K_1^{\text{trans}} = 0.7$). Gray line: CTC described by $k_{\text{ep}_1} = 2.19$, $k_{\text{ep}_2} = 5.02$, $K_1^{\text{trans}} = 0.62$, $K_2^{\text{trans}} = 0.24$. The dashed lines show the contribution of the first compartment and the dotted lines those of the second compartment.

The independent voxelwise 2Comp model leads to unstable estimates due to redundancy issues. Obviously, redundancy issues occur when the contribution of one compartment vanishes. However, they may as well occur when the exponential rates are too similar. As described in Section 2.3, there are theoretical results on parameter redundancy in sum of exponentials models (Seber and Wild, 1989). However, a generalization for the case of convolved exponentials is tricky. In Reich (1981) a redundancy measure was used to show that parameters in a sum of two exponentials model are highly redundant if the exponential rates differ by less than a factor of five. Even though this result does not directly transfer to the convolved exponentials given in equation (3.8), this result still helps to understand parameter redundancy in the 2Comp model. In Figure 4.1 we show an example for data simulated from a two-compartment model where the exponential rates differ by a factor of four. In this case, a solution from a compartment model with only one tissue compartment exists that fits the observed concentration reasonably well. Therefore, several quite distinct

parameter vectors describe very similar CTCs, meaning that parameters are redundant (as defined in equation (2.20)).

An advantage of Bayesian analysis is, of course, that posteriors can be calculated no matter if the parameters are redundant. From the form of the posterior one can identify how profound the information about certain parameters is. In a Bayesian framework, rather flat or multimodal posteriors are obtained in case of redundancy. From these, point estimates are difficult to obtain and will be unstable. In the analysis of DCE-MRI data, one deals with a large number of parameters due to the large number of voxels. Not being able to look at the posterior distribution of every single voxel, point estimates are necessary in order to visualize results with the help of parametric maps.

4.1.6 Simulation and data

In order to evaluate the proposed approach, CTCs for different parameter combinations were simulated. First, we simulated the 2Comp model, keeping $k_{\mathrm{ep}_2} = 2$ and $v_{t_2} = 0.2$ fixed and varying $k_{\mathrm{ep}_1} = 0.05, \ldots, 0.5$ and $v_{t_1} = 0.1, 0.2, 0.3, 0.4$ (simulation A). Secondly, we simulated the 2Comp model, keeping $k_{\mathrm{ep}_1} = 0.3$ and $v_{t_1} = 0.2$ fixed and varying $k_{\mathrm{ep}_2} = 1, \ldots, 10$ and $v_{t_1} = 0.1, 0.2, 0.3, 0.4$ (simulation B). For comparison, we simulated the 1Comp model as well, with $k_{ep} = 0.5, \ldots, 2.5$ and $v_t = 0.1, \ldots, 0.5$ (simulation C). Gaussian noise was added at different contrast to noise ratios (CNR = 5, 10, 15). For all simulated CTCs, the posterior PDFs of the 1Comp model and the 2Comp model were calculated.

To evaluate the clinical use of our approach, we use a subset of the DCE-MRI study on breast cancer described in Section 3.3.

4.2 Results

4.2.1 Simulation results

Figure 4.2 shows the true underlying CTCs, the simulated data points series (with CNR=10), and the fitted 1Comp- and 2Comp-curves for different parameter values. Figure 4.2 a) shows an example where the 2Comp model describes the simulated curve more adequately than the 1Comp model (fast uptake, slow washout). In this case, the DIC of the 2Comp model is smaller than for the 1Comp model, and the p_D of the 2Comp model exceeds the p_D of the 1Comp model: $p_{D_2} = 3.77$, $p_{D_1} = 1.65$. As the 2Comp model has more parameters, one can expect that $p_{D_2} > p_{D_1}$ holds. In c) the estimated posterior PDF of v_{t_1} for example a) is plotted. For this parameter constellation with distinct rate constants ($k_{\mathrm{ep}_1} = 0.3$, $k_{\mathrm{ep}_2} = 6$), model choice based on the DIC would correctly identify the 2Comp model as clearly superior to the 1Comp model ($\mathrm{DIC}_2 = 43.29 < \mathrm{DIC}_1 = 46.55$).

The situation is different for voxels with more similar rate constants. In Figure 4.2 b) the CTC simulated from the 2Comp model (with $k_{\mathrm{ep}_1} = 0.45$, $k_{\mathrm{ep}_2} = 2$) differs only slightly from a 1Comp model curve. There seems to be no need for a second compartment to adequately

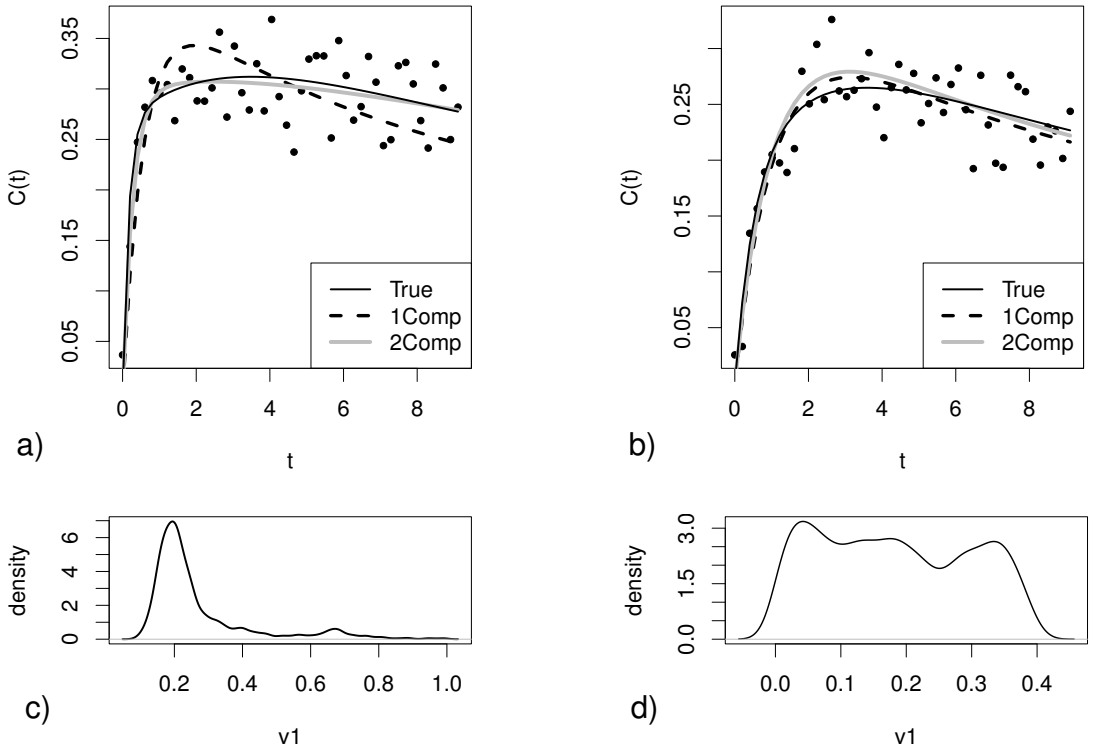

Figure 4.2: Examples of simulated CTCs and the fitted 1Comp- and 2Comp-curves a), b) with corresponding estimated posterior PDF c), d). Underlying parameters: a) $k_{\mathrm{ep}_1} = 0.3$, $k_{\mathrm{ep}_2} = 6$, $v_{t_1} = v_{t_2} = 0.2$; b) $k_{\mathrm{ep}_1} = 0.45$, $k_{\mathrm{ep}_2} = 2$, $v_{t_1} = v_{t_2} = 0.2$

describe the data. Figure 4.2 d) shows the estimated posterior PDF of v_{t_1}. Though the true values for the volume of the two compartments are $v_{t_2} = v_{t_1} = 0.2$, the posterior PDFs for these parameters are multimodal with maxima at parameter constellations that actually correspond to a situation where the volume of one compartment is near zero and the parameter constellation of the other compartment is similar to the 1Comp model. The samples for the volume parameters alternate between the cases where v_{t_1} is almost zero or v_{t_2} is almost zero, see Figure 4.3. For a multimodal PDF, estimates of the kinetic parameters are not reliable as the median does not provide a good point estimate in this case. Furthermore, the effective number of parameters p_D is underestimated. Here, the estimated number of effective parameters in the 2Comp model is even smaller than for the 1Comp model ($p_{D_2} = 0.37$, $p_{D_1} = 2.03$).

For a bimodal or multimodal posterior PDF, the estimation of p_D and DIC is not reliable (Spiegelhalter et al., 2002). Hence, model selection is done comparing the DIC of the 1Comp and 2Comp models and accounting for the cases where the posterior PDF is multimodal as in example 4.2 b). We use small p_D values or a large deviation of k_{ep_1} samples as indicators for multimodal posteriors. We define the following criterion for model selection: choose the 2Comp model if

(i) the standard deviation of the 2Comp model is not too large ($sd(k_{\mathrm{ep}_1})/k_{\mathrm{ep}_1} < 0.25$)

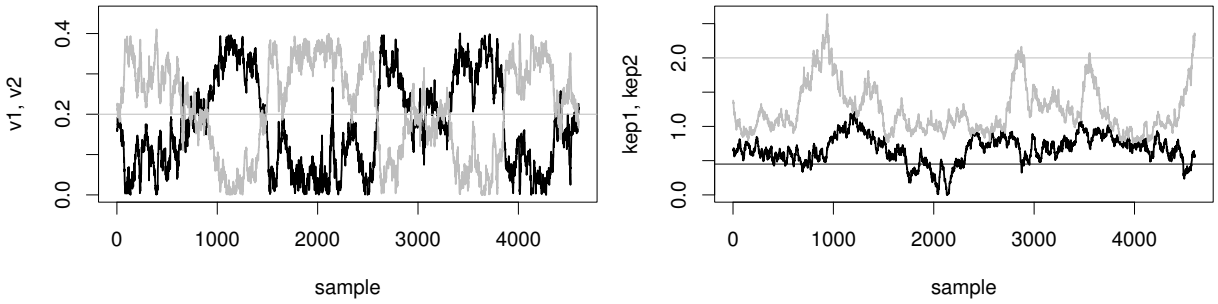

Figure 4.3: Thinned sampling paths for v_{t_1}, v_{t_2}, k_{ep_1}, and k_{ep_2} for data simulated from $k_{\text{ep}_1} = 0.45$, $k_{\text{ep}_2} = 2$, $v_{t_1} = v_{t_2} = 0.2$ (simulation A, CNR=10)

AND

(ii) the p_D of the 2Comp model exceeds typical values of the 1Comp model ($p_{D_2} > 1.5$)

AND

(iii) the DIC of the 2Comp model is smaller than the DIC of the 1Comp model ($\text{DIC}_2 < \text{DIC}_1$).

Our simulation results show that parameter redundancy has to be taken into account in parameter estimation and model selection. In Section 2.3, we have described how parameter redundancy can be classified in a sum of exponential model with the help of a redundancy measure. From this theory, parameters in a sum of two exponentials model are redundant if the exponential rates differ too little (less than a factor of five). Here, we obtain insight into parameter redundancy in the 2Comp model (sum of two convolved exponentials) by evaluating simulation studies with different parameter constellations.

Table 4.1 shows how often the 2Comp model was selected for simulation A, B and C. For every combination of parameters and at every noise level, there are 50 simulated CTCs. The median DIC and the median p_D are shown for the CTCs selected to be modeled better by the 2Comp model. Both the shape of the underlying curve (Figure 4.4) and the noise level (Table 4.1) influence how often the 2Comp model is selected. At lower noise levels the specificity for a model increases: The percentage of cases where the 2Comp model is preferable, the difference of the 1Comp and 2Comp model DIC, and the effective number of parameters p_D increase with increasing CNR (Table 4.1).

There are parameter combinations for which the 2Comp model is rarely selected even at a low noise level (Figure 4.4). For this parameter combinations the CTC of the 2Comp model can be approximated well by a 1Comp-curve as discussed in example 4.2 b). We state that this is the case in situations where the contribution of the first compartment is too small (*e.g.*, $v_{t_1} < v_{t_2}$) or the uptake behavior of both compartments is too similar ($k_{\text{ep}_1} \approx k_{\text{ep}_2}$). This result is in accordance with the theoretic results on redundancy in a sum of two exponentials described in Section 2.3 and in Appendix A.2.

Table 4.1: Results of model choice: cases where 2Comp model is selected, median DIC for the 2Comp (1Comp) model and median p_{D_2} (p_{D_1}) for these cases. Median DIC (2Comp) is significantly smaller than median DIC (1Comp) in all combinations (Wilcoxon rank sum test, p-values < 0.01).

Simulation	A	B	C
CNR	5	5	5
2Comp	49%	47%	33%
DIC	46.85 (48.53)	46.98 (48.29)	46.28 (48.03)
p_D	2.14 (1.87)	2.49 (1.77)	1.85 (1.88)
CNR	10	10	10
2Comp	50%	69%	30%
DIC	45.39 (47.12)	45.83 (47.47)	43.88 (45.55)
p_D	2.69 (1.88)	3.05 (1.77)	1.92 (1.83)
CNR	15	15	15
2Comp	56%	80%	25%
DIC	42.40 (45.59)	43.51 (46.87)	39.30 (40.81)
p_D	2.85 (1.88)	3.21 (1.77)	1.87 (1.86)

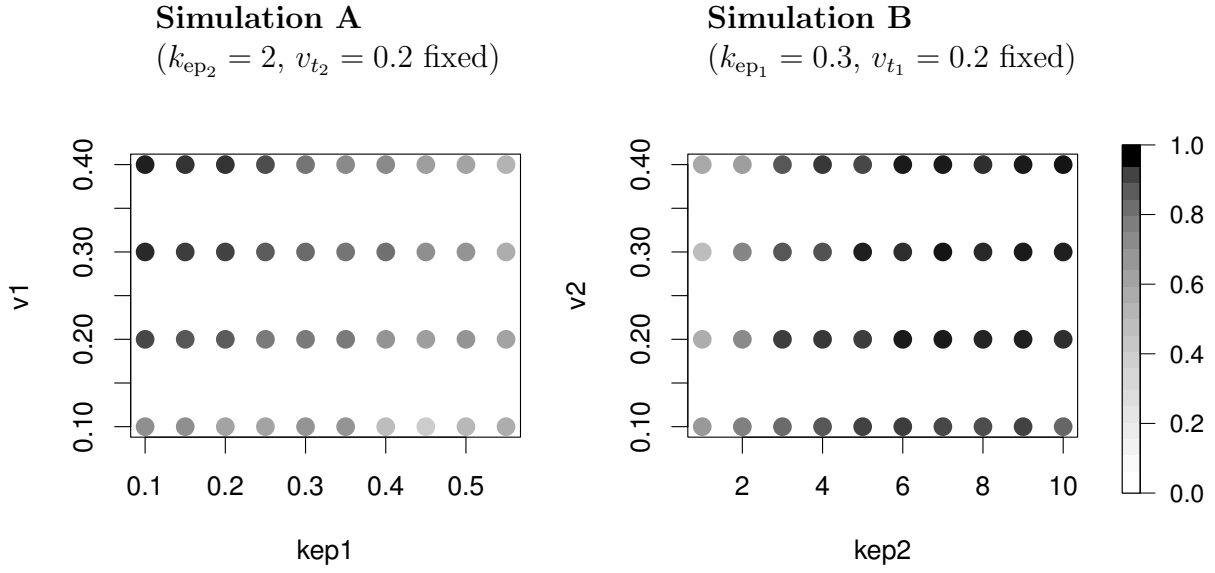

Figure 4.4: Selection of 2Comp model in percent for different combinations of kinetic parameters (simulation A and B, CNR=10)

Table 4.2: Model selection for twelve scans pre- and post-treatment: Median DIC, p_D, percentage of voxels where 2Comp model selected

Patient/ scan	DIC 1Comp	DIC 2Comp	p_D 1Comp	p_D 2Comp	2Comp
P1 pre	46.28	44.29	1.91	2.37	61.86%
P1 post	44.61	42.84	1.87	2.09	54.46%
P2 pre	47.76	46.07	1.89	2.34	53.85%
P2 post	44.68	42.60	1.92	2.28	51.86%
P3 pre	48.00	46.32	1.88	2.54	54.94%
P3 post	45.65	43.37	1.89	2.18	39.53%
P4 pre	45.98	44.03	1.90	2.20	40.27%
P4 post	36.16	33.39	1.87	2.22	41.30%
P5 pre	35.30	32.23	1.89	2.46	43.81%
P5 post	36.87	33.37	1.89	2.72	56.16%
P6 pre	36.71	34.56	1.91	2.30	42.33%
P6 post	35.17	32.74	1.90	2.29	41.77%

4.2.2 DCE-MRI breast cancer study

To evaluate the clinical value of the proposed method, we fitted the 1Comp and 2Comp models to CTCs of twelve DCE-MRI scans of the breast cancer study described in Section 3.3. Table 4.2 shows the percentage of voxels for which the 2Comp model was selected. For these voxels, the median DIC and the median p_D for the 1Comp and 2Comp models are listed. Other voxels were excluded due to multimodal posteriors. As described in the simulation study, for those voxels, the p_D and DIC values as well as parameter estimates are not reliable. For about 40–60 % of the voxels analyzed the 2Comp model is preferable.

In Figure 4.5, typical CTCs are shown for three voxels from the pre-treatment scan of patient 2. The concentration uptake in a voxel outside the tumor is rather slow (Figure 4.5 a)). For this voxel the fitted 1Comp- and 2Comp-curves look quite similar. Still, the 2Comp-model is preferable based on the DIC values ($\text{DIC}_2 = 45.20 < \text{DIC}_1 = 46.39$), and the estimated effective number of parameters in the 2Comp model is higher ($p_{D_2} = 2.49$, $p_{D_1} = 1.92$). For a voxel at the tumor edge, the CTC has a plateau shaped form (fast uptake, slow washout, see Figure 4.5 b)). For this voxel, the superiority of the fitted 2Comp-curve is obvious as the fitted 1Comp-curve cannot adequately describe the plateau shape. Furthermore, the estimated p_D for the 2Comp model is considerably higher ($p_{D_2} = 3.42$, $p_{D_1} = 1.97$, $\text{DIC}_2 = 47.73 < \text{DIC}_1 = 49.35$). For the voxel inside the tumor, there is a fast uptake but also a fast washout of the contrast medium (Figure 4.5 c)). For this voxel, the 1Comp model adequately describes the observed CTC. In contrast, the 2Comp model fails to adequately describe the CTC. This is because, due to redundancy issues,

the posterior marginal densities are multimodal leading to bad point estimates and hence to bad fits to the data. Here, a small p_D value indicates that the posterior density is multimodal in the 2Comp model and the DIC is not reliable ($p_{D_2} = 0.29$, $p_{D_1} = 2.02$, $\mathrm{DIC}_2 = 44.79 < \mathrm{DIC}_1 = 48.39$).

Figure 4.6 shows the model decision per voxel and Figure 4.7 the parameter maps for the pre-treatment scan of patient 2. The parameter estimates of the 2Comp model were plotted only for voxels where the 2Comp model was selected. Within the tumor, the 2Comp model is rarely selected. This indicates that the 1Comp model is usually appropriate to describe the uptake behavior inside the tumor. At the tumor margins, however, the 2Comp model outperforms the 1Comp model. For voxels at the tumor margins the DIC is smaller for the 2Comp model and the effective number of parameters p_D is significantly larger than for the 1Comp model (Figure 4.7 p_{D_2} - p_{D_1}). Also, the sum of squared errors (SSE) shows good improvement when using a second tissue compartment, especially at the tumor edges (Figure 4.6).

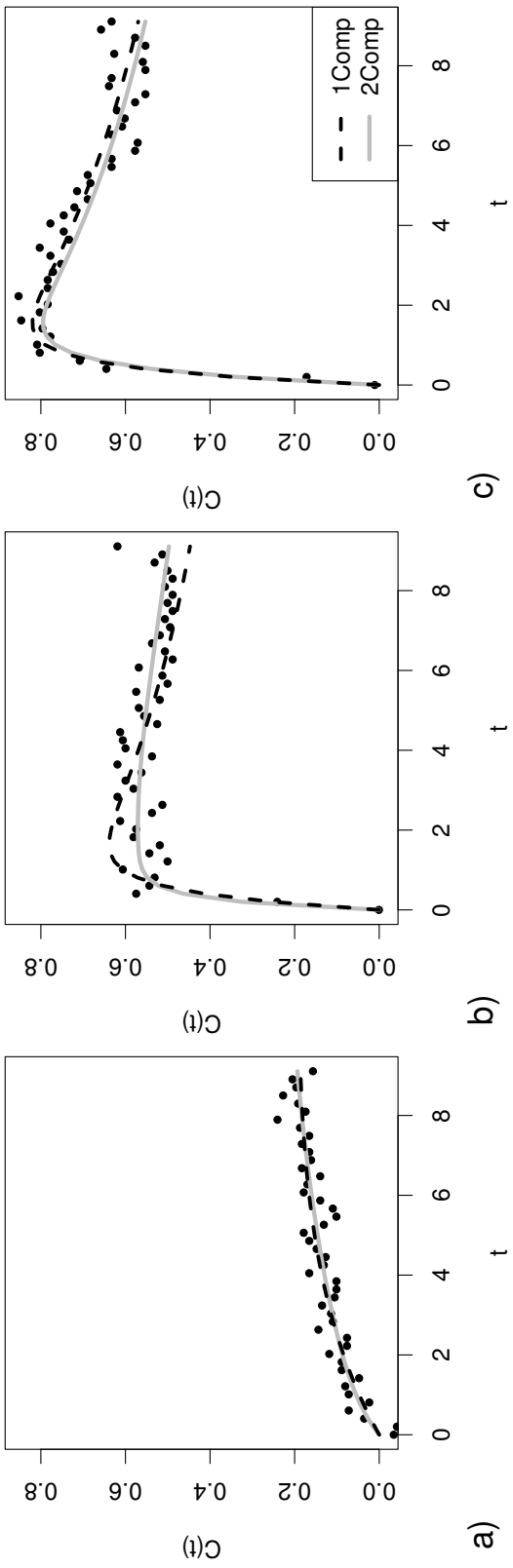

Figure 4.5: Underlying CTCs and the fitted 1Comp- and 2Comp-curves for a) voxel outside the tumor (15,15), b) voxel at the tumor edge (48,30), and c) voxel in the tumor center (51,42). All voxels were taken from the central slice of patient 2 pre-treatment.

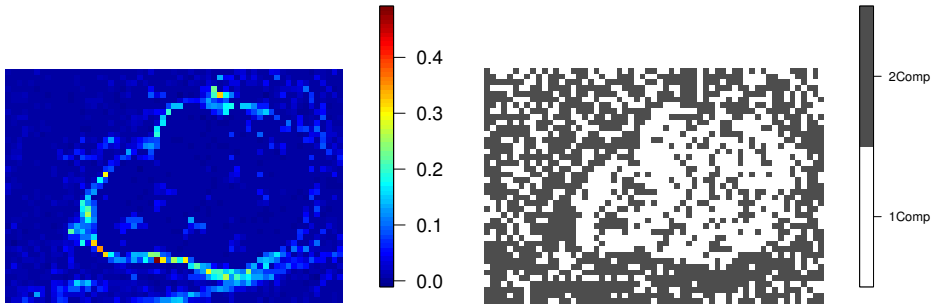

Figure 4.6: Differences in SSE and model choice for patient 2 pre-treatment

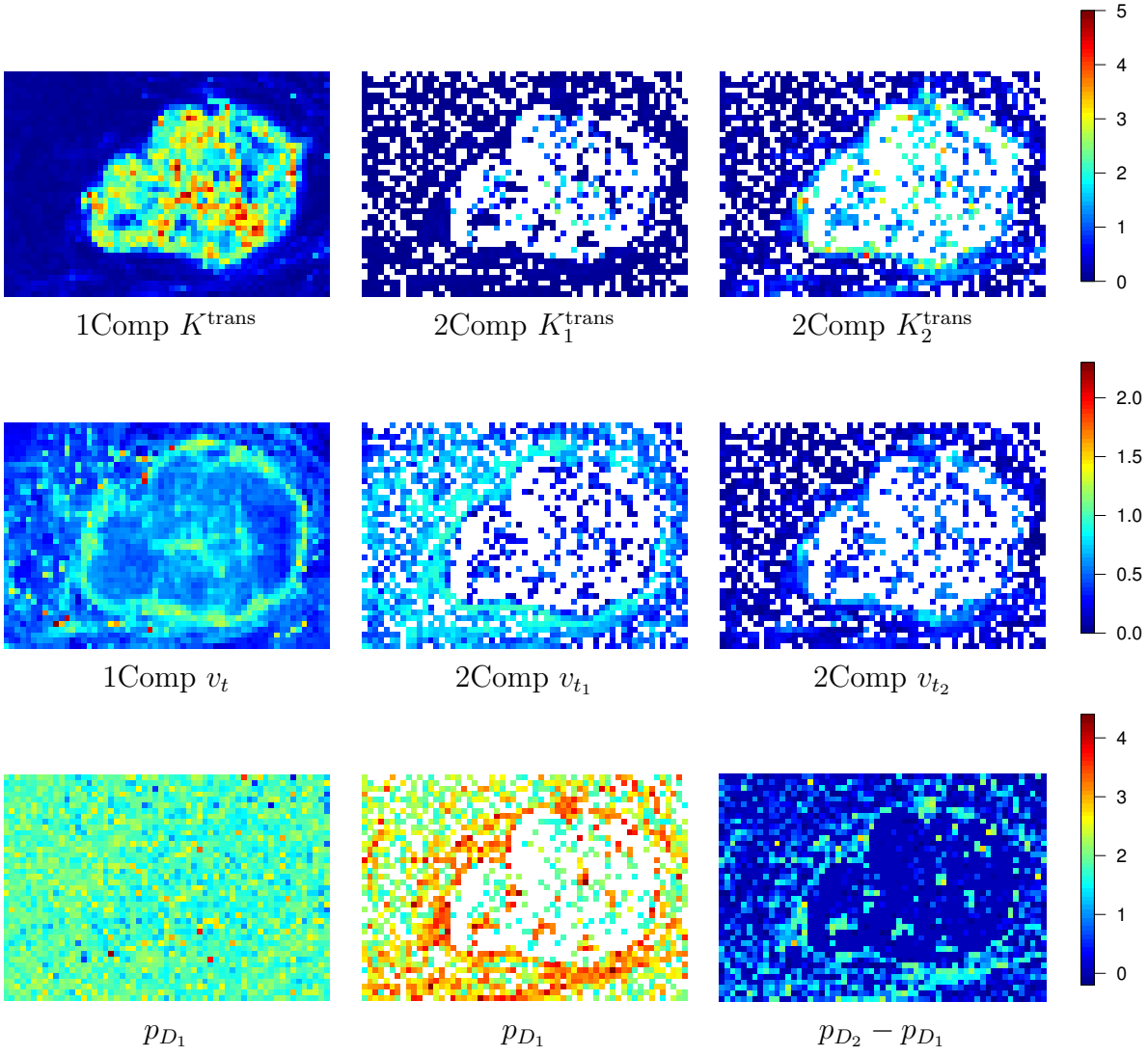

Figure 4.7: Parameter maps for patient 2 pre-treatment. Estimates which are not available are depicted in white.

4.3 Conclusions

We have fitted the 1Comp and 2Comp models introduced in Section 3.2 on a voxel level. Based on the DIC, we decided which model is more adequate for each voxel. We have found that the 2Comp model outperforms the 1Comp model regarding the contrast medium uptake behavior at tumor margins. Those results confirm that the model complexity of the proposed 2Comp model is actually needed to adequately describe concentration uptake behavior observed in tumorous tissue and that areas with increased tissue heterogeneity exist which cannot be modeled with a 1Comp model.

However, in other tissue areas like, for example, in the tumor core, a 1Comp model suffices to model the observed uptake behavior. In those areas, the 2Comp model is redundant and point estimates for the kinetic parameters, the number of effective parameters p_D, and the DIC are not reliable. Therefore, a statistical procedure would be useful that allows for two tissue compartments and sets one of the compartment volumes to zero for voxels where one tissue compartment suffices. Approaches with this wanted variable selection property are analyzed in Chapters 5–7.

The Bayesian framework is suitable to evaluate model fit and complexity and the DIC allows to compare the two models. Besides the DIC, the posterior distribution obtained from MCMC sampling provides useful information that a least squares approach cannot offer. This information has to be considered when deciding among different models. We have described how a Bayesian framework allows to analyze redundancy issues. Multimodal or flat PDFs lead to small or even negative p_D-values which can be used to indicate redundancy issues.

Here, we have modeled voxels independently from each other. Like this, the spatial information contained in an image is not accounted for. The voxelwise analysis of the 1Comp and 2Comp models in this chapter can be considered the basis needed for the extension to a spatial analysis as proposed in Chapter 5.

Chapter 5

Bayesian spatial regularization

In this chapter, we propose a spatial Bayesian framework for the 2Comp model proposed in Section 3.2.2 extending the approach considering all voxels independently from each other (Chapter 4). With the help of this framework, we determine the kinetic parameters on a voxel level, using prior biological knowledge on the parameters and accounting for the spatial structure of the image. The approach makes use of the intrinsic spatial information given by the voxel structure of the image. With the help of Gaussian Markov random fields as priors on the kinetic parameters, we incorporate our prior belief that the exchange rates between plasma and tissue observed in two neighboring voxels are likely to be similar. With this smoothness assumption the parameter space is spatially regularized.

We analyze to what extent this spatial regularization helps to avoid parameter redundancy and to obtain stable parameter point estimates in the 2Comp model. Choosing a full Bayesian approach, we obtain posteriors and point estimates running MCMC simulations. An advantage of the Bayesian framework is that the posterior can still be computed in the case of parameter redundancy; however, the redundant parameters will have multimodal marginal posteriors. This may lead to unstable parameter estimates and bad fits when analyzing voxels independently from each other as described in Chapter 4.

We find that assuming spatial smoothness on the exponential rates is an efficient way to regularize the parameter space and to make parameters of a 2Comp model identifiable. The proposed spatial 2Comp model has advantageous variable selection properties. In voxels which are well described by a 1Comp model, the contribution of one of the compartments is set to values close to zero. As a result, we obtain stable parameter estimates for all voxels. Hence, one can describe heterogeneity of the tissue without losing spatial information. The proposed approach is evaluated for simulated concentration time curves as well as for *in vivo* data from the breast cancer study introduced in Section 3.3.

This Chapter is based on Sommer and Schmid (2012).

5.1 Methods

Gaussian Markov random fields are a powerful tool for modeling spatial information. Applications of spatial Markov random fields, at their beginnings, often focused on agricultural field experiments (Besag, 1974). Later, spatial Markov random fields became popular in image analysis, *e.g.*, Besag (1986) discusses the use of Markov random fields for image restoration. Markov random fields have become especially important in Bayesian image analysis (Geman and Geman, 1986) and in Bayesian hierarchical models in general. Hidden Markov fields are, for instance, successfully used for image restoration (Marroquin et al., 2003). For an overview of applications, theoretical results, and efficient algorithms for Gaussian Markov random fields see Winkler (1995) and Rue and Held (2005).

In the context of DCE-MRI analysis, Gaussian Markov random field priors have previously been proposed and shown to be useful by Schmid et al. (2006) and Kelm et al. (2009) for the 1Comp model. With Gaussian Markov random field priors, one can incorporate smoothness assumptions on parameter maps. Here, we propose Gaussian Markov random field priors for the kinetic parameters of the 2Comp model.

As before, the observed contrast medium concentration $Y_{i,j}$ at time t_j, $j = 1, ..., T$ in voxel $i = 1, ..., N$ can be described by the theoretical concentration time curve $C_t(t)$ depending on the voxel-specific kinetic parameters $\boldsymbol{\phi}^i$ plus Gaussian noise, *i.e.*, $Y_{i,j} \sim N\left(C_t(\boldsymbol{\phi}^i; t_j), \sigma_i^2\right)$ (see equation (4.2)) with σ_i^2 the voxel-specific variance of the Gaussian noise (Schmid et al., 2006). The form of C_t and the number of kinetic parameters is fixed, given a specific compartment model. For the 2Comp model, the form of C_t is a sum of two convolved exponentials as described by equation (3.8). Parameters are estimated with the help of MCMC simulations. However, in contrast to Chapter 4, kinetic parameters of the different voxels $\boldsymbol{\phi}^i$ are no longer assumed to be independent. Hence, MCMC update schemes have to take into account potential dependencies in the parameters of different voxels.

5.1.1 Bayesian 2Comp model voxelwise

A voxelwise two tissue compartment model is used for comparison with a corresponding spatial model. In a voxelwise approach, the CTCs of all voxels are fitted independently from each other. Similar to Section 4.1.2, we define priors on the parameters. Here, we use slightly different prior distributions that are better comparable with the priors used in Schmid et al. (2006) for the 1Comp model. However, the type of prior information incorporated with this prior and the observed estimation behavior is very similar to the prior used in Section 4.1.2.

We use an exponential parametrization that insures the rate and transfer constants to be positive (see Schmid et al., 2006, and references therein): $\theta_k^i = \log\left(k_{\mathrm{ep}_k}^i\right)$, $\gamma_k^i = \log\left(K_k^{\mathrm{trans},i}\right)$, for $k = 1, 2$. Similar to the voxelwise Bayesian 1Comp model evaluated in Schmid et al. (2006), we impose Gaussian priors on the logarithmic rate constants θ_k^i

$$\theta_k^i | \tau_{\theta_k} \sim N\left(\mu_{\theta_k}, (\tau_{\theta_k})^{-1}\right)$$

and on the logarithmic transfer constants γ_k^i

$$\gamma_k^i|\tau_{\gamma_k} \sim N\left(\mu_{\gamma_k}, (\tau_{\gamma_k})^{-1}\right)$$

independently for all voxels $i = 1, \ldots, N$ with fixed precisions $\tau_{\theta_k} = \tau_{\gamma_k} = 1$ and $\mu_{\theta_1} = \mu_{\gamma_1} = \mu_{\gamma_2} = 0$, $\mu_{\theta_2} = \log(5)$. With this prior, all rate and transfer constants $k_{\mathrm{ep}_k}^i$ and $K_k^{\mathrm{trans},i}$ remain positive. Rate and transfer constants of the first compartment with *a priori* probability of 99.86% do not exceed 20 min^{-1}. The dynamics in the second compartment is assumed to be faster with *a priori* expected $k_{\mathrm{ep}_2}^i$ values of five.

5.1.2 Spatial Bayesian 2Comp model

In the following, we will introduce a spatial prior which accounts for the spatial information intrinsic in the DCE-MR images. This prior reflects our view about the physiology in tumorous tissue, and spatial regularization helps to solve the problem of redundancy.

We assume that the average permeability of capillaries is a rather persistent property given that the capillary walls are intact or amiss. However, the volume fractions of intact and disrupt capillaries (and hence normal or abnormal EES) recorded in a specific voxel may vary randomly. Therefore, in the proposed spatial model, we assume that rate constants $k_{\mathrm{ep}_k}^i$ vary smoothly in space and hence that the exchange properties of each tissue compartment are rather smooth. In contrast, the contribution of differently behaving compartments in each voxel is assumed to be quite flexible, meaning that the tissue volumes $v_{t_k}^i = K_k^{\mathrm{trans},i}/k_{\mathrm{ep}_k}^i$ may vary strongly from voxel to voxel. Then, the transfer rate as product of rate and volume $K_k^{\mathrm{trans},i} = k_{\mathrm{ep}_k}^i v_{t_k}^i$ inherits the spatial smoothness of $k_{\mathrm{ep}_k}^i$ but is less smooth due to varying $v_{t_k}^i$ values. The spatial smoothness of the kinetic parameters is modeled using a Gaussian Markov random field on its logarithmic transforms θ_k^i, γ_k^i (Rue and Held, 2005; Schmid et al., 2006).

We use a neighborhood structure where adjacent voxels of a two-dimensional image are neighbors, that is, each voxel has four neighbors unless it lies at the edge of the image. From this, a prior distribution can be defined by assuming a Gaussian distribution on the differences of neighboring logarithmic rate and transfer constants:

$$\theta_k^i - \theta_k^j|\tau_{\theta_k} \sim N\left(0, (\tau_{\theta_k})^{-1}\right) \text{ for } i \sim j$$

and

$$\gamma_k^i - \gamma_k^j|\tau_{\gamma_k} \sim N\left(0, (\tau_{\gamma_k})^{-1}\right) \text{ for } i \sim j.$$

Here, $i \sim j$ denotes that voxel i is a neighbor of voxel j.

This spatial prior on the logarithmic rates leads to smooth parameter maps of $k_{\mathrm{ep}_1}^i$ and $k_{\mathrm{ep}_2}^i$, $K_1^{\mathrm{trans},i}$ and $K_2^{\mathrm{trans},i}$. However, *a priori* we assume much smoother maps for $k_{\mathrm{ep}_1}^i$ and $k_{\mathrm{ep}_2}^i$ and less smooth maps for $K_1^{\mathrm{trans},i}$ and $K_2^{\mathrm{trans},i}$. Hence, we use Gamma priors on the precisions

$$\tau_{\theta_k} \sim Ga\left(a_{\theta_k}, b_{\theta_k}\right)$$

for $k = 1, 2$ with $a_{\theta_1} = a_{\theta_2} = 1000$ and $b_{\theta_1} = b_{\theta_2} = 1$ and

$$\tau_{\gamma_k} \sim Ga\left(a_{\gamma_k}, b_{\gamma_k}\right)$$

for $k = 1, 2$ with $a_{\gamma_1} = a_{\gamma_2} = 0.0001$ and $b_{\gamma_1} = b_{\gamma_2} = 0.01$.

Furthermore, we assume the noise variance to be the same for all voxels i, *i.e.* $\sigma_i^2 = \sigma^2$. For the observation variance, we assume an Inverse Gamma prior $\sigma_i^2 \sim IG(a, b)$ with a and b such that the *a priori* expected signal to noise ratio (SNR) corresponds to values typically observed in breast tumor DCE-MRI data (voxelwise SNR ranges from 10 to 20 for the data analyzed here). We choose this prior to be more informative with increasing number of voxels.

5.1.3 Implementation

The posterior distribution is obtained with Markov chain Monte Carlo methods (see also Section 4.1.3). The likelihood and full conditionals are calculated as follows.

The log-likelihood depends on the voxel-specific kinetic parameters $\boldsymbol{\phi}^i$ and the inverse noise variance $\tau_\epsilon = \frac{1}{\sigma^2}$:

$$l\left(\boldsymbol{\phi}^i, \tau_\epsilon\right) \quad = \quad \frac{T}{2}\log\left(2\pi\tau_\epsilon\right) - \frac{1}{2}\tau_\epsilon \sum_{j=1}^{T}\left(Y_{i,j} - C_t(\boldsymbol{\phi}^i, t_j)\right)^2.$$

In the spatial model, the full conditional distribution of the logarithmic rate constant in voxel i, θ_k^i, given the logarithmic rate constants of all other voxels, θ_k^{-i},

$$p(\theta_k^i|\theta_k^{-i}, \tau_{\theta_k}) \propto \exp\left(-\frac{\tau_{\theta_k}}{2}\sum_{j\in\partial(i)}\left(\theta_k^i - \theta_k^j\right)^2\right)$$

depends only on those of its direct neighbors for $k = 1, 2$. Here, $\partial(i)$ denotes the set of direct neighbors of voxel i. The full conditionals of the logarithmic transfer constants γ_k^i have the same form.

Let $\epsilon_{ij} = Y_{i,j} - C_t(\boldsymbol{\phi}^i, t_j)$ denote the random noise terms. Then, the full conditional of the inverse noise variance τ_ϵ in the full spatial model is a Gamma distribution

$$\tau_\epsilon|\cdot \sim Ga\left(a + \frac{NT}{2}, b + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{T}\epsilon_{ij}^2\right).$$

The full conditional of the precision τ_{θ_1} is

$$\tau_{\theta_1}|\cdot \sim Ga\left(a_{\theta_1} + \frac{N-1}{2}, b_{\theta_1} + \frac{1}{4}\sum_{i=1}^{N}\sum_{j\in\partial(i)}\left(\theta_1^i - \theta_1^j\right)^2\right).$$

Similarly for τ_{θ_2}, τ_{γ_1} and τ_{γ_2}.

We implemented the proposed spatial 2Comp model extending the R-package `dcemriS4` (Whitcher and Schmid, 2009, 2011). For each voxel, we simulate from the posterior of the model parameters with an MCMC algorithm (Gilks et al., 1996). Starting with random values, the voxels are subsequently updated in random order using the current parameter values of neighboring voxels where needed. More precisely, starting values are drawn from uniform distributions per voxel with $v_{t_1}^{start} \sim U[0,1]$, $v_{t_2}^{start} = 1 - v_{t_1}^{start}$, and $k_{\text{ep}_1}^{start} \sim U[0.1, 0.3]$, $k_{\text{ep}_2}^{start} \sim U[1.75, 5.25]$. The logarithmic rate and transfer constants θ_1^i, θ_2^i, γ_1^i, γ_2^i are updated with Metropolis-Hastings steps with random walk proposal. Gibbs update steps are used for the hyper parameters σ^2, $\tau_{\theta_k}^i$, and $\tau_{\gamma_k}^i$ as its full conditionals are Gamma distributions that can be sampled from directly.

The proposal variances of the random walk proposals are tuned such that the Metropolis-Hastings acceptance rates are approximately 20%. After a burn-in of 5,000 iterations, 50,000 iterations are performed. As samples of the parameters are calculated for up to 230×256 voxels, thinning is necessary for storage purposes, and only every tenth sample is saved. For parameter point estimation we use the median of the MCMC sample. As a convergence analysis on a voxel level is too extensive, we have checked convergence comparing the obtained results with results obtained with a considerably shorter MCMC-chain with different starting values (5,000 iterations after a burnin of 5,000, every third sample saved). We find that the estimated parameter maps obtained from both chains are visually not distinguishable and conclude that the MCMC chains have converged at burnin. Sampling paths and autocorrelations of the kinetic parameters are shown in the results (Figure 5.6).

5.1.4 Measure of model complexity

Similar to Chapter 4, the number of effective parameters p_D is used to assess the model complexity per voxel and hence the tissue heterogeneity as introduced in Section 4.1.4. Higher p_D values indicate an increased effective model complexity in the observed data, and hence we suggest using p_D to measure tissue heterogeneity. Here, although we are dealing with a joint model for all voxels, *i.e.*, a model for the whole image, we will also compute a voxelwise p_D using the deviance in each voxel. This allows us to visually assess the model complexity per voxel and hence the tissue heterogeneity.

5.2 Simulation study

5.2.1 Simulation setup

In order to evaluate the proposed voxelwise and spatial 2Comp models, we simulated a DCE-MR image of 25×25 voxels with different parameter combinations in a two tissue compartment model. The parameter configuration is given in Figure 5.1. For three blocks of different size, we simulated CTCs from a true 2Comp model, *i.e.*, a mixture of two

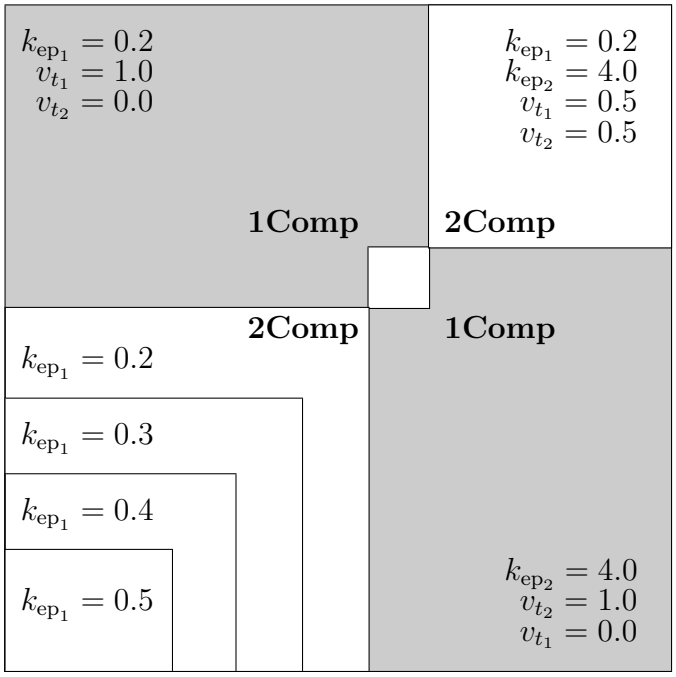

Figure 5.1: Sketch of simulation design

tissue compartments with very different exchange rates $k_{ep_1} = 0.2$ and $k_{ep_2} = 4$. Both compartments contribute equally with volumes $v_{t_1} = v_{t_2} = 0.5$. In the lower left block, the exchange rate k_{ep_1} varied smoothly from 0.2 in the middle to 0.5 at the corner.

For two blocks, we simulated CTCs from a 1Comp model. One of those blocks is described by a tissue compartment exchanging rather slowly with plasma at rate $k_{ep_1} = 0.2$ and having volume $v_{t_1} = 1$ (the fast exchanging compartment with $k_{ep_2} = 4$ has no contribution, *i.e.* $v_{t_2} = 0$). For the other block, the exchange with plasma is rather fast: $k_{ep_2} = 4$, $v_{t_2} = 1$ (the slow exchanging compartment $k_{ep_1} = 0.2$ has no contribution, *i.e.* $v_{t_1} = 0$).

Within each block, uniformly distributed noise $U[0.8, 1.2]$ was multiplied to the parameters k_{ep_1}, k_{ep_2}, K_1^{trans} and K_2^{trans} per voxel and the corresponding CTC was computed from these kinetic parameters. Gaussian noise was added to the simulated CTCs with standard deviation $\sigma = 0.05$. With this simulated data we try to rebuild typical behavior observed in real data with realistic signal to noise ratios and smooth parameter maps with some random variation.

5.2.2 Results

As discussed above, when fitting a model with two tissue compartments one often deals with identifiability issues. In these cases, the model is overparametrized and one observes unstable point estimates of all parameters. The Bayesian approach allows the evaluation of the posterior anyway; however, in the case of redundancy, the marginal posteriors typically

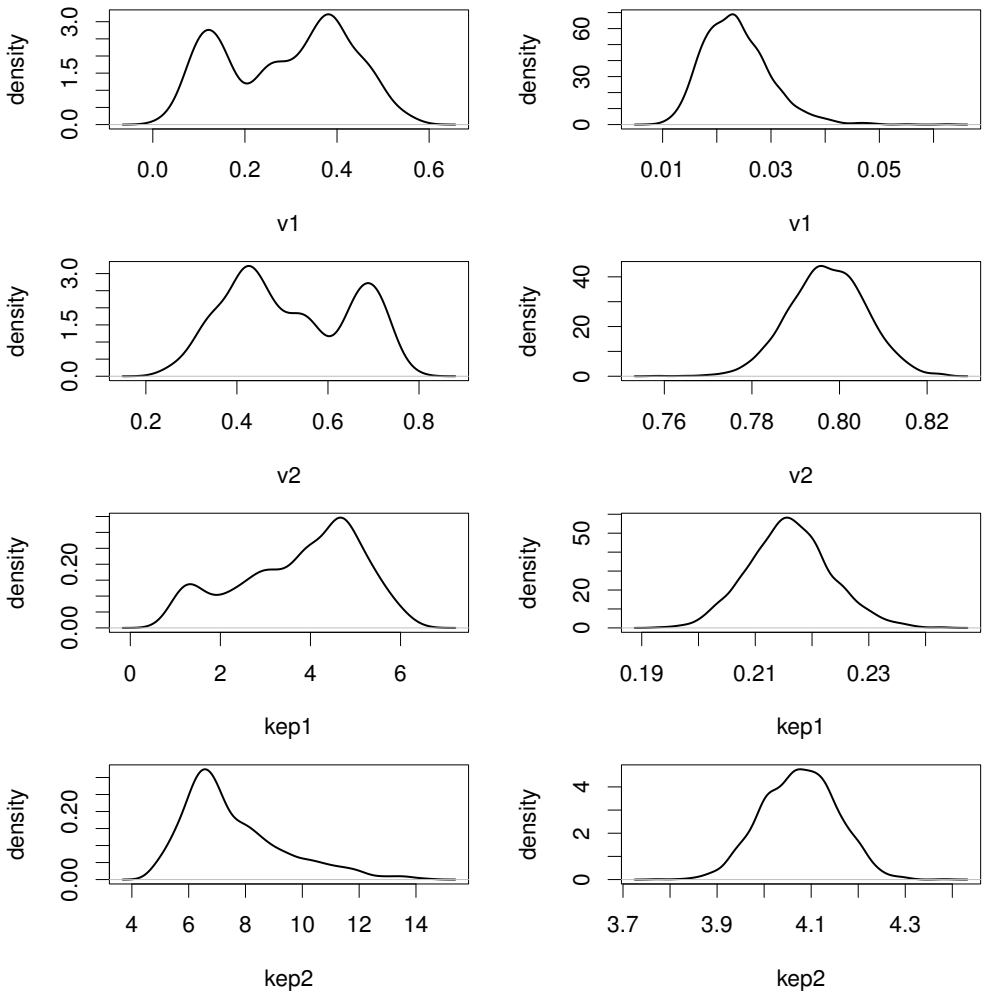

Figure 5.2: Posterior marginal densities for curve simulated from 1Comp. Left: Voxelwise 2Comp fitted. Right: Spatial 2Comp fitted

are multimodal.

For example, in the blocks simulated from a 1Comp model parameters are redundant, and hence point estimates are unstable in the lower right of the simulated image. For one of these voxels, Figure 5.2 depicts the marginal posteriors of volume fractions v_{t_k} and rate constants k_{ep_k}. With the voxelwise approach, the posteriors are multimodal, and hence there is no good point estimator for the kinetic parameters. In comparison, the spatial approach produces unimodal posteriors and good point estimates can be gained by computing the median of the MCMC sample. In contrast to the voxelwise model, the contribution of the two compartments are well separated (k_{ep_1} and k_{ep_2} samples are not too similar) and the estimated volume of the first compartment is close to zero.

Using the point estimates for the kinetic parameters, we obtain estimated CTCs and refer to them as fit. In Figure 5.3, we compare the fit of the 1Comp model, the voxelwise,

and the spatial 2Comp model. For a curve simulated from a 1Comp model (Figure 5.3 (a)), the fit of the spatial 2Comp model is similar to the fit of the 1Comp model. However, using the point estimates to obtain a CTC estimate, the voxelwise 2Comp model fails to adequately fit the curve due to redundancy issues. The sum of squared errors (SSE) is about 0.14 for the voxelwise 2Comp model and about 0.1 for the 1Comp model as well as for the spatial 2Comp model. For a curve simulated from a 2Comp model (Figure 5.3 (b)), both the spatial and voxelwise 2Comp models clearly outperform the 1Comp model with similar fits. Here, the SSE is about 0.12 for the spatial and voxelwise 2Comp models and 0.2 for the 1Comp model.

In Figure 5.4, the SSE per voxel is depicted for the 1Comp model and the voxelwise and the spatial 2Comp model. Considerable differences in SSE for the 1Comp model compared to both 2Comp models can be observed for the three blocks simulated from a true 2Comp model. The voxelwise and spatial 2Comp models have similar SSE, with increased SSE in the voxelwise model for voxels with multimodal posteriors. These differences cannot be distinguished at this scale and were shown for a specific curve above (Figure 5.3 (a)).

In order to visualize the model complexity on the voxel level, we depict p_D maps in Figure 5.4. Please note that for voxels with multimodal posteriors, the estimates of p_D are not meaningful. Hence, in the voxelwise 2Comp model estimated p_D values are often negative due to parameter redundancy. For the spatial 2Comp model, the p_D map visualizes the adaption of the model to the complexity in the tissue. Values range between 0.5 and 1 with median 0.68 in the 1Comp blocks. In the 2Comp blocks, p_D values between 0.8 and 1.6 with median 1.2 show increased tissue heterogeneity.

In Figure 5.5, we show the parameter maps for the point estimates of k_{ep_1}, k_{ep_2}, K_1^{trans}, and K_2^{trans} from the voxelwise and the spatial model as well as the true underlying parameter values. As the voxelwise approach leads to unstable point estimates, the estimation results differ strongly from the true underlying values. Especially for the voxels simulated from a 1Comp model, the voxelwise 2Comp model leads to unstable point estimates. For instance, for voxels in the lower right simulated from a true 1Comp model with $v_{t_1} = 0$, k_{ep_1} is overestimated, K_1^{trans} is overestimated, and consequently K_2^{trans} is underestimated. Compared to the voxelwise model, the spatial model leads to smooth parameter maps that can be interpreted more intuitively and to stable point estimates that better match the true underlying parameter values.

In the spatial model, the parameter maps for k_{ep_1} and k_{ep_2} are smooth and the estimates match the true underlying values quite well. There is some oversmoothing such that the higher k_{ep_1} values in the lower left corner are underestimated and as a consequence also the corresponding k_{ep_2} are underestimated. The estimates of K_1^{trans} and K_2^{trans} perfectly match the true underlying values. For the blocks simulated from a 1Comp model either the K_1^{trans} estimate or the K_2^{trans} estimate becomes zero. Like this, model redundancy is avoided and the posteriors are unimodal.

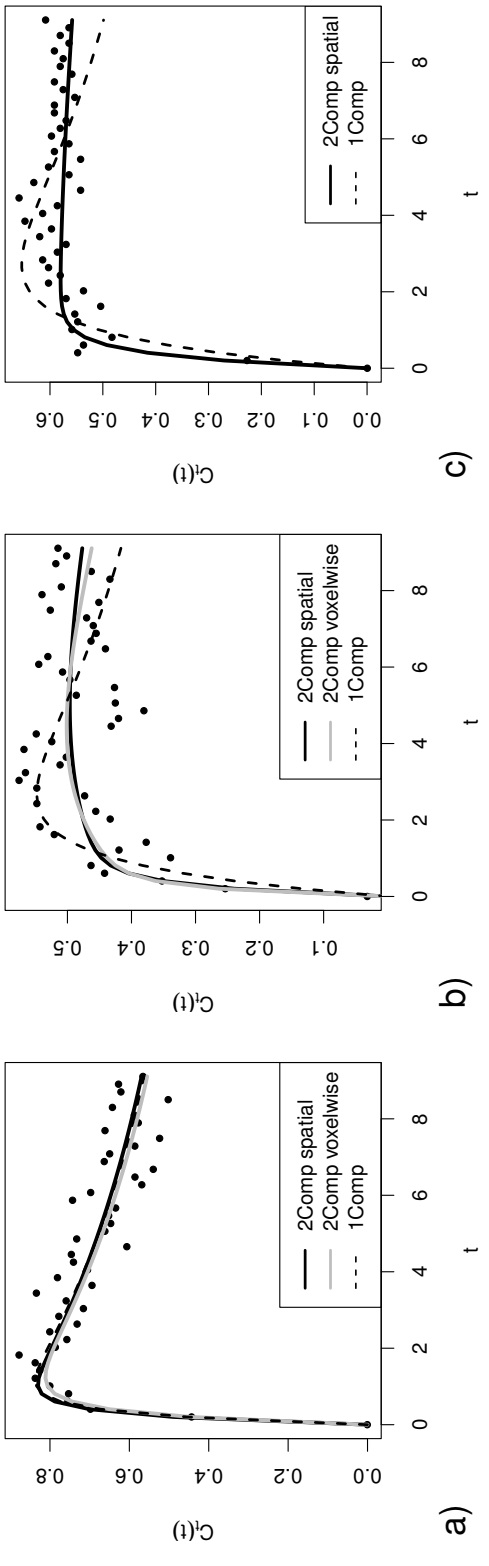

Figure 5.3: Examples of concentration time curves and corresponding curve fits for 1Comp model, spatial and voxelwise 2Comp models. (a) Data simulated from a 1Comp model (parameter constellation as in the lower right corner). (b) Data simulated from a 2Comp model (parameter constellation as in the upper right corner). (c) Concentration uptake observed in a tumor edge voxel.

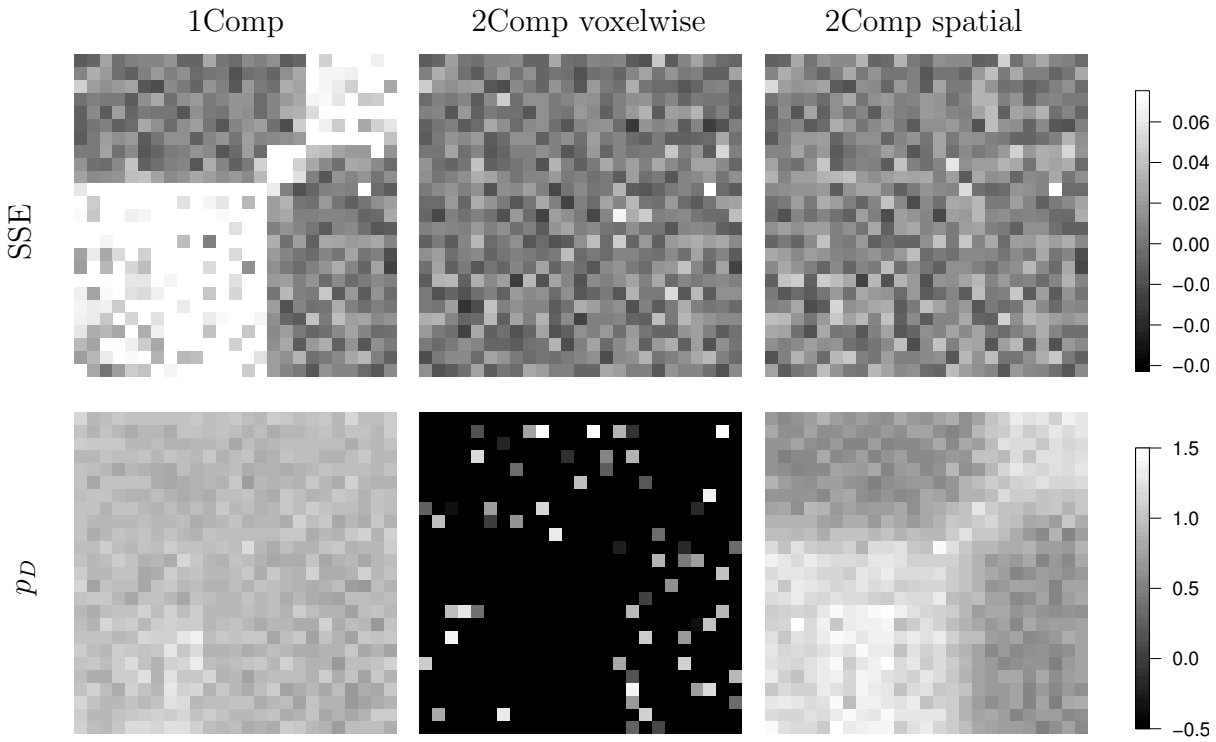

Figure 5.4: Evaluation of model fit: sum of squared errors (SSE) and p_D

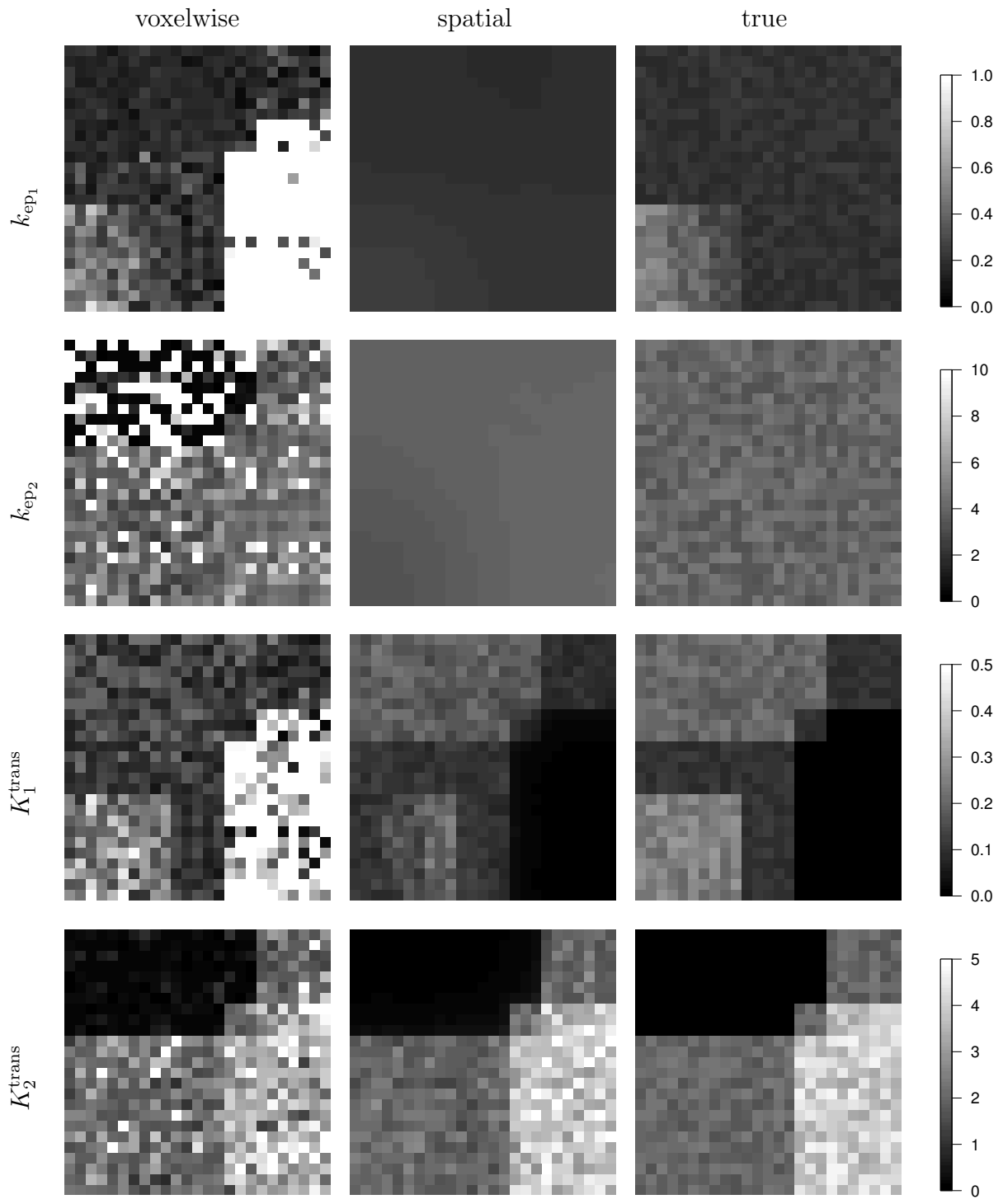

Figure 5.5: Parameter maps for simulation study: Voxelwise (left column) and spatial fit (center) of 2Comp model for true underlying values (right)

5.3 DCE-MRI breast cancer study

To evaluate the clinical use of our approach, we use the DCE-MRI study on breast cancer described in Section 3.3.

The sampling paths and corresponding autocorrelation plots for two voxels are depicted in Figure 5.6. The autocorrelation of k_{ep_1} and k_{ep_2} samples is stronger than for v_{t_1} and v_{t_2} due to the stronger smoothness assumption. For the global parameters updated with Gibbs steps (τ_{θ_k}, τ_{γ_k}, and σ^2) autocorrelations are smaller.

In Figures 5.7 and 5.8 the parameter maps from the spatial 2Comp model are shown for pre- and post-treatment scans of patients 2, 4, 5, and 6. Similar to the results of the simulation study and in accordance with the prior assumptions, the estimated parameter maps for the DCE-MR images are quite smooth for the exponential rates k_{ep_1} and k_{ep_2}, whereas the K_1^{trans} and K_2^{trans} estimates show more spatial variation. The contribution of the second compartment vanishes (K_2^{trans} close to zero) in healthy tissue. In those regions, the 1Comp model suffices to describe the observed uptake dynamics, meaning that the tissue is homogeneous there. Interestingly, tissue inside of the tumor is often homogeneous as well. Note that with the spatial prior meaningful parameter estimates are obtained for all voxels whereas in the voxelwise model no stable point estimates could be obtained for voxels inside the tumor, see Figure 4.7. The second compartment has nonzero contribution and improves the fit of observed CTCs at tumor margins and in parts of the surrounding tissue, see for example Figure 5.3 c). In those regions, the tissue is heterogeneous as both the slow and the fast exchanging compartments contribute to the uptake dynamics. Larger p_D values and improved fit compared to the 1Comp model reflect this heterogeneity.

For patients 4 and 5 (nonresponder to therapy) the parameter maps for the pre-treatment scan depict increased K_1^{trans} and K_2^{trans} values for a large tissue region. For patient 4, the kinetic parameters k_{ep_1}, k_{ep_2}, K_1^{trans} and K_2^{trans} have higher values post-treatment, but the tissue region with increased blood flow becomes smaller and more dense. Reduced tumor volume could easily be misinterpreted as treatment success. Here, the p_D map contains additional information that might help to assess treatment success. For patient 4 voxelwise p_D values are even higher in the post-treatment scan. For patient 5 both increased p_D values and increased tumor volume indicate treatment failure.

For patients 2 and 6 (responder to therapy) parameter maps of k_{ep_1} and k_{ep_2} are quite smooth. The contribution of one compartment—K_1^{trans} or K_2^{trans}—is close to zero inside of the tumor and in surrounding healthy tissue. Non-vanishing contribution of both compartments is observed at tumor margins and in surrounding tissue. After treatment, the number of voxels where the second compartment contributes decreases notably. For patients 2 and 6 tumor margins and extensions around the tumor are heterogeneous and better described with the aid of an additional second compartment. Both the p_D values and the size of the tissue region with increased p_D decrease after treatment.

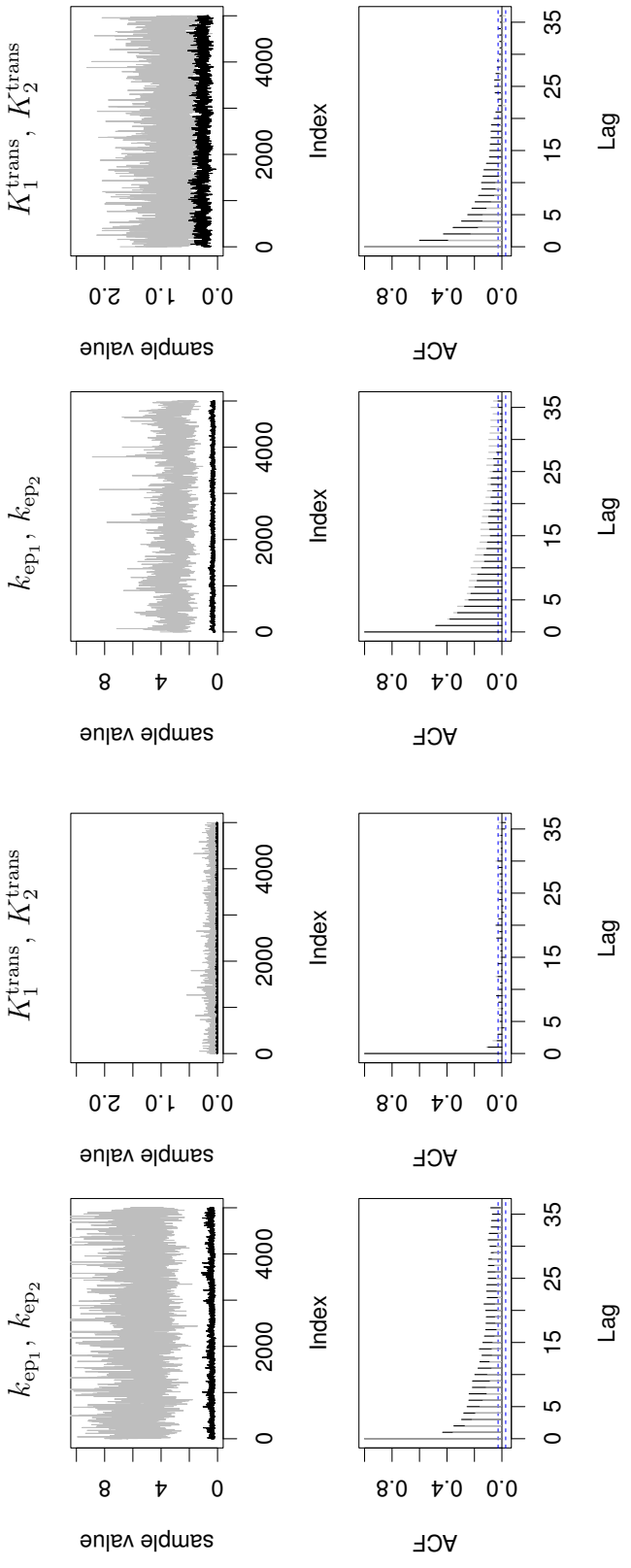

Figure 5.6: Thinned samples after burnin and corresponding autocorrelation functions (ACF) of kinetic parameters for a voxel inside tumor (40,75), and a voxel at tumor margin (21,86) taken from the pre-treatment scan of patient 6. Values for k_{ep_1} and K_1^{trans} are shown in black, for k_{ep_2} and K_2^{trans} in gray.

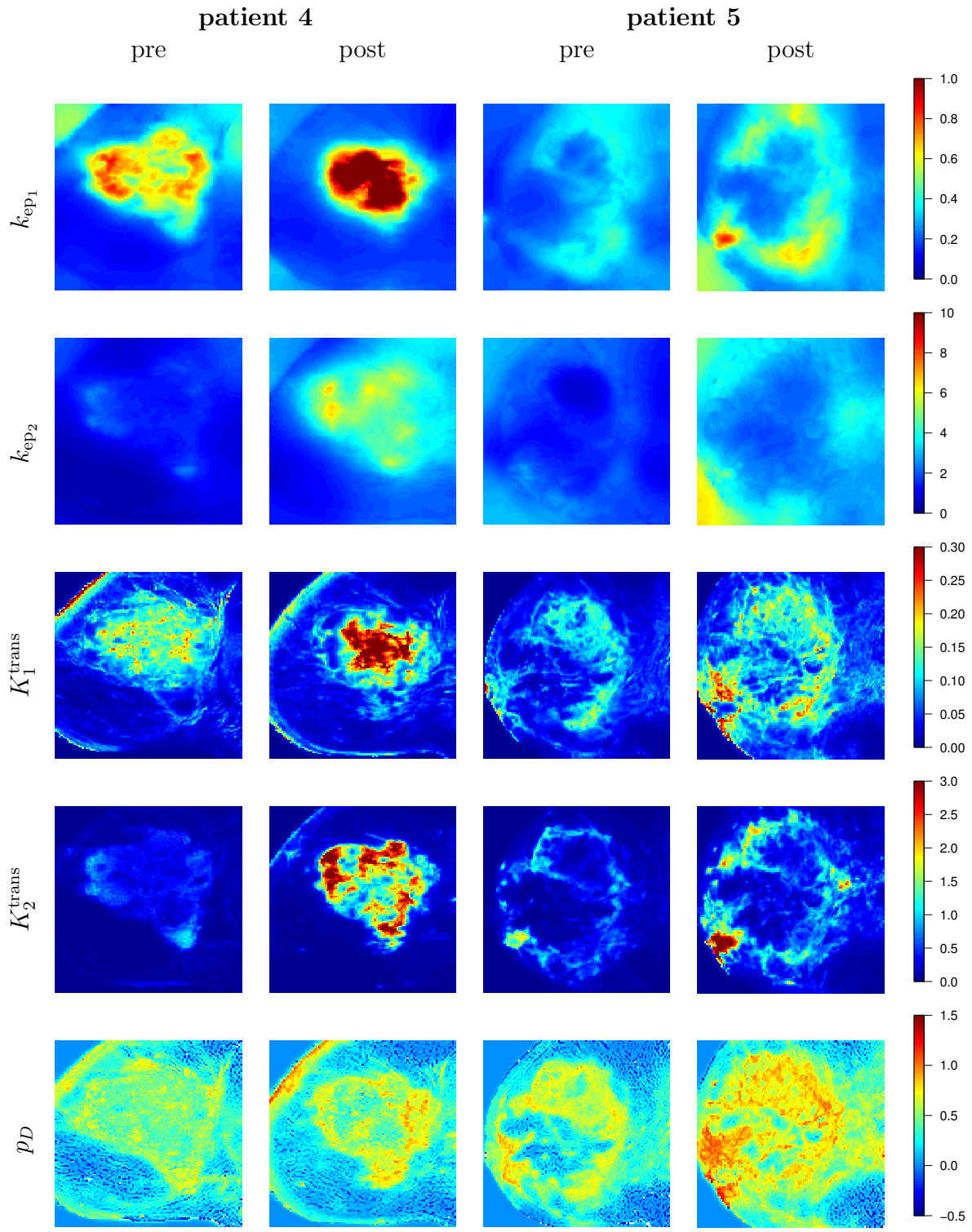

Figure 5.7: Spatial 2Comp model: parameter maps for patients 4 and 5 (nonresponders) pre- and post-treatment

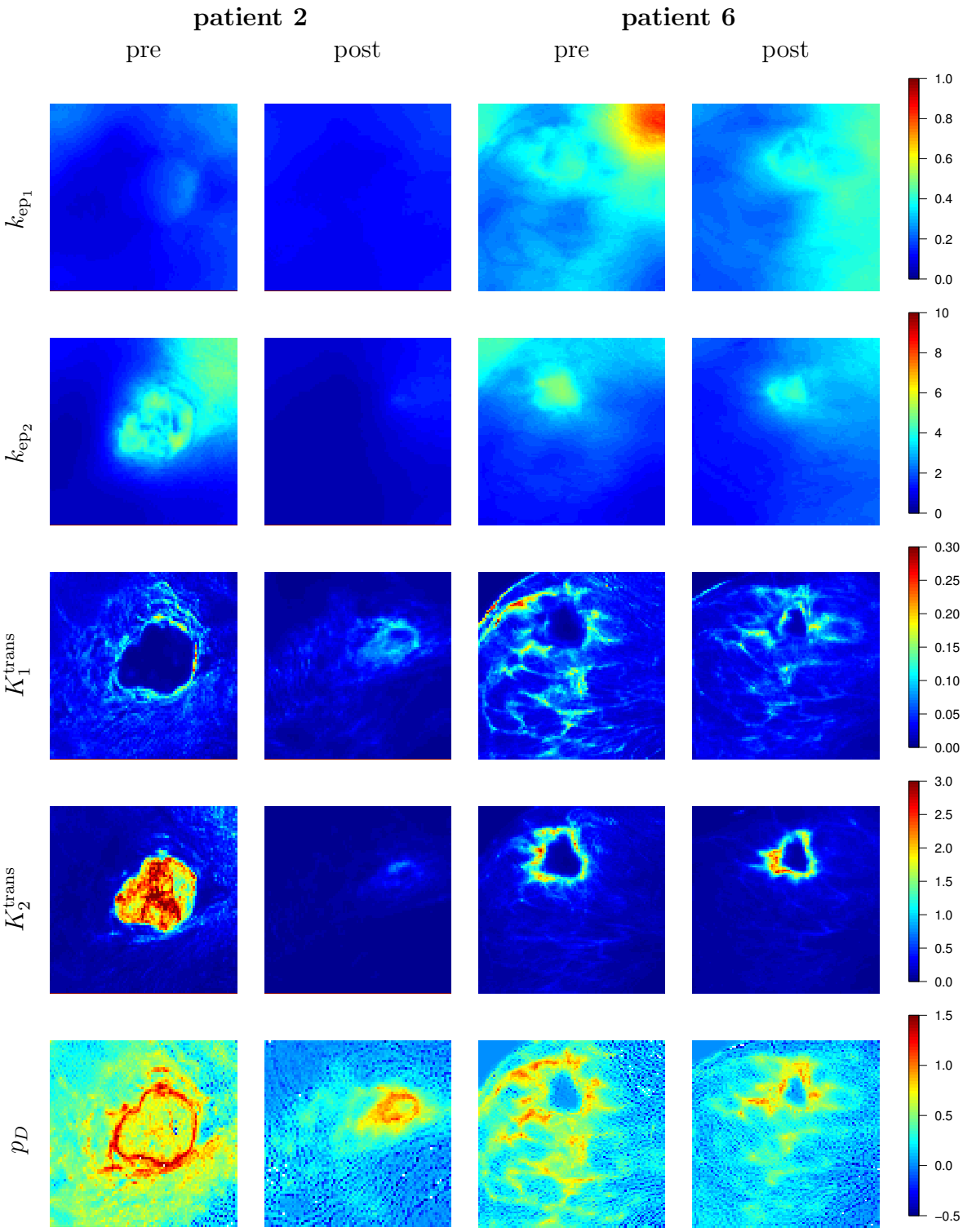

Figure 5.8: Spatial 2Comp model: parameter maps for patients 2 and 6 (responders) pre- and post-treatment

5.4 Conclusion and discussion

In this chapter, we have proposed a spatial Bayesian model for the estimation of kinetic parameters in a two tissue compartment model used for the analysis of DCE-MRI. Compared to existing alternatives, this approach allows to model tissue heterogeneity on a voxel level. We have developed suitable spatial priors for the kinetic parameters. On the one hand, those priors reflect smoothness assumptions which are biologically intuitive (maps of k_{ep} parameters are smoother than maps of K^{trans} parameters). On the other hand, those spatial priors regularize the parameter space and assure parameter identifiability. With this prior, a 2Comp model can be fitted at a voxel level and CTCs in heterogeneous tissue, especially at tumor margins, can be described better than with the standard 1Comp model. An important advantage of the proposed spatial regularization is that it implicates good variable selection properties. For CTCs that are adequately described by the 1Comp model, the estimates of one of the compartment volumes is close to zero. Like this and in contrast to a voxelwise approach, parameter point estimates are stable and easy to interpret.

Another important innovation of the proposed approach is the use of heterogeneity maps obtained when calculating the number of effective parameters p_D per voxel. We have proposed p_D as a measure that contains additional information about the heterogeneity of the tissue whereas the kinetic parameters contain information about the uptake dynamics only. We find it interesting that estimates of the effective number of parameters p_D rarely exceed values of 1.5, even for CTCs simulated from a 2Comp model with four kinetic parameters. Besides evaluating changes in the kinetic parameters, comparing p_D maps pre-treatment and post-treatment (*i.e.* after two weeks of chemotherapy) may help in predicting treatment response.

In our approach, we use an informative prior for the k_{ep} parameters implying a very smooth k_{ep} surface in order to handle the redundancy issue. That is, we use a compromise between the assumption of k_{ep} parameters having the same value in a region of interest and an independent voxel by voxel estimation of k_{ep} parameters. The prior on the smoothing parameter of the K^{trans} parameters has a large variance, implying a much lesser degree of smoothing. As can be seen in Figure 5.8, in practice the smoothing of the K^{trans} maps is very small. Therefore, the proposed method can still depict boundaries and sharp features in the tumor tissue. As an alternative for global smoothing, adaptive smoothing has been proposed for 1Comp models (Schmid et al., 2006) but is computationally very expensive (Brezger et al., 2007). Due to the redundancy issues in the 2Comp model, estimation of an adaptive spatial 2Comp model including a lot of additional local smoothing parameters seems not feasible.

We have used a neighborhood structure for two-dimensional images as only the central slice of the DCE-MRI data was used. When modeling several slices, this neighborhood can be extended to a three-dimensional setting. The strength of spatial smoothing, however, has to be diminished in the third direction as the spatial intervals between slices are large compared to the spatial intervals between voxels of the same slice.

In summary, we proposed and evaluated spatial regularization for two-compartment

models, allowing a more comprehensive insight into tissue perfusion, in particular in heterogeneous tissue. Spatial regularization allows us to overcome the redundancy issues by "borrowing strength" across the tissue of interest, and hence allows us to fit complex compartment models even on voxel level with low signal to noise ratio. Additional clinical studies should be performed to further explore the clinical potential of this model.

Chapter 6

Spatially penalized ML-estimation and model selection

As described in Chapter 3, competing compartment models of different complexities have been used for the quantitative analysis of DCE-MRI data. A model can never reflect full reality but helps to describe reality by means of some important factors, and it is not clear which factors have to be considered. Hence, model selection is an important statistical task. In this chapter, we present a spatial elastic net approach that allows to estimate the number of compartments for each voxel such that the model complexity is not fixed *a priori*.

A multi-compartment approach is considered which is translated into a restricted least squares model selection problem. We do so by using a set of basis functions for a given set of candidate rate constants. The form of the basis functions is derived from a kinetic model and thus describes the contribution of a specific compartment. Using a spatial elastic net estimator, we chose a sparse set of basis functions per voxel, and hence, rate constants of compartments. The spatial penalty takes into account the voxel structure of an image and performs better than a penalty treating voxels independently. We evaluate the proposed estimation method for simulated images and apply it to data from the breast cancer study introduced in Section 3.3.

This Chapter is based on Sommer et al. (2012).

6.1 Introduction

Quantitative characterization of contrast agent uptake can be done with model-driven methods or with data-driven methods. With data-driven methods like nonparametric regression using a spline basis, no *a priori* compartment-structure has to be defined (Schmid et al., 2009). Model-driven methods are based on pharmacokinetic models describing the exchange of contrast agent between different, well-mixed compartments. Such compartment models provide quantitative physiological parameters characterizing the amount and rate of capillary leakage (see Chapter 3).

However, several compartmental models with different, *a priori* fixed, numbers of compartments have been proposed, and it remains unclear which model to use. This is particularly the case when the imaged tissue is heterogeneous as often observed in cancerous tissue. Therefore, several authors propose more complex models to describe perfusion in tissue as described in Chapter 3. However, the adequate number of compartments might be different in different types of tissue and—as cancerous tissue is often heterogeneous—it might even vary over a field of voxels. That is why, in this chapter, we aim to estimate the number of compartments per voxel from the data. This is important as the degree of tissue heterogeneity itself is diagnostically informative.

To this end, a multi-compartment model is fitted using likelihood based regularization techniques. We use a bundle of exponential basis functions, each of which is derived from the differential equation describing the tracer uptake of a tissue compartment. Like this, corresponding coefficients remain interpretable and linked to physiological parameters. The coefficients are sparsely selected and estimated while penalizing for an increasing number of parameters. By selecting clusters of nonzero coefficients, the number of used compartments is implicitly selected as well. With this approach we combine the advantages of model-driven and data-driven methods, *i.e.*, a good fit to the data and biologically interpretable parameters.

An approach of sparse basis selection—not considering the spatial image structure though—has also been proposed for compartment models used in PET (Gunn et al., 2002). The basis pursuit approach proposed there corresponds to unrestricted lasso (Tibshirani, 1996) estimation. However, compared to the lasso, ridge regression (Hoerl and Kennard, 1970) has often been shown to produce better results in case of highly correlated covariates. Therefore, we use a restricted elastic net, combining the advantages of ridge and lasso estimation (Zou and Hastie, 2005).

Spatial information is frequently used in image processing, most prominently in neuroimaging (Gössl et al., 2001; Zou et al., 2004; Christensen and Yetkin, 2005). For DCE-MRI, several authors have proposed Bayesian hierarchical models (assuming a fixed number of compartments) to account for the spatial structure intrinsic in an image (Schmid et al., 2006; Kelm et al., 2009; Sommer and Schmid, 2012, see Chapter 5).

For our approach we need to (1) restrict the parameters to be non-negative to ensure the positiveness of the physiological parameters; (2) we include prior spatial information by assuming spatial smoothness of the parameters. We do so by penalizing quadratic differences of neighboring coefficients. With this spatially penalized maximum likelihood (ML) approach we use the intrinsic spatial information given by the voxel structure of the image.

The chapter is organized as follows. First, we introduce a linearized regression problem for the multi tissue compartment model introduced in Section 3.2.2. In Subsection 6.2.2 and 6.2.3 the proposed estimation techniques—the voxelwise elastic net and spatial elastic net—are introduced. Then, the proposed approach is evaluated for simulated data (Section 6.3) as well as for *in vivo* data from a breast cancer study (Section 6.4). Section 6.5 concludes.

6.2 Methods

6.2.1 Multi tissue compartment model

As a generalization of the Tofts model with one tissue compartment (3.4), the multi-compartment model with q tissue compartments (3.10) has been proposed in Section 3.2.2. With this multi-compartment model, tumor heterogeneity is taken into account on a voxel level. Due to measurement error, the observed concentration $Y_{i,t}$ at time points $t = 0, \ldots, T$ in voxel $i = 1, \ldots, N$ is modeled as:

$$\begin{aligned} Y_{i,t} &= C_t^i(t) + \epsilon_{i,t} \\ &= v_{p,i} C_p(t) + \sum_{k=1}^{q} K_{k,i}^{\text{trans}} \Psi_k(t) + \epsilon_{i,t}, \end{aligned} \tag{6.1}$$

where $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ is a Gaussian noise term and $\Psi_k(t) = C_p(t) * \exp(-k_{\text{ep}_k} t)$ are basis functions. Each tissue compartment k is characterized by how fast it exchanges with the plasma compartment, expressed by its rate constant k_{ep_k}. As candidate values we consider $\log(k_{\text{ep}_k}) \in \{-3, -2.9, -2.8, \ldots, 2.9, 3\}$, and the adequate values need to be selected. Moreover, each compartment is characterized by its transfer constant K_k^{trans}. The transfer constant is obtained by the product of the volume fraction v_k and the rate constant of the compartment and is hence non-negative: $K_k^{\text{trans}} = k_{\text{ep}_k} \cdot v_k \geq 0$. Thus, for voxel i, the vector $\boldsymbol{\theta}_i = (\theta_{i,0}, \ldots, \theta_{i,q})^T = (v_{p,i}, K_{i,1}^{\text{trans}}, \ldots, K_{i,q}^{\text{trans}})^T$ is unknown and has to be estimated. Wherever the estimated transfer constant is positive ($\hat{K}_{i,k}^{\text{trans}} > 0$), the corresponding compartment, respectively k_{ep_k} value, is selected. From (6.1) it can be seen that a linear regression problem with predictors $C_p(t), \Psi_k(t)$, $k = 1, \ldots, q$, is to be solved. Figure 6.1 depicts a subset of those predictors.

When estimating the parameter vector $\boldsymbol{\theta}_i$ with simple ML inference, under the assumption of independent Gaussian distributed observation errors $\epsilon_{i,t}$, the residual sum of squares $\sum_t (Y_{i,t} - \hat{C}_t^i(t))^2$ has to be minimized. However, here the $\boldsymbol{\theta}_i$ need to be non-negative, and hence the pure ML-estimate is

$$\hat{\boldsymbol{\theta}}_i^{\text{ML}} = \text{argmin}_{\boldsymbol{\theta}_i \geq 0} \left\{ \sum_t \left(Y_{i,t} - \boldsymbol{z}(t)^T \boldsymbol{\theta}_i \right)^2 \right\}, \tag{6.2}$$

with

$$\boldsymbol{z}(t) = D(C_p(t), \Psi_1(t), \ldots, \Psi_q(t))^T, \; t = 1, \ldots, T.$$

Since, however, we have a large number of $\boldsymbol{\theta}_i$ parameters and adjacent entries of $\boldsymbol{z}(t)$ are highly correlated (due to construction), pure ML-estimates are unstable or even not unique. Therefore, we use two penalized approaches.

6.2.2 Voxelwise regularized estimation

In order to stabilize the estimation of parameter vector $\boldsymbol{\theta}_i$, the log-likelihood is additively corrected by a penalty term $J(\boldsymbol{\theta}_i)$. More precisely, for a fixed voxel i, we use the estimator

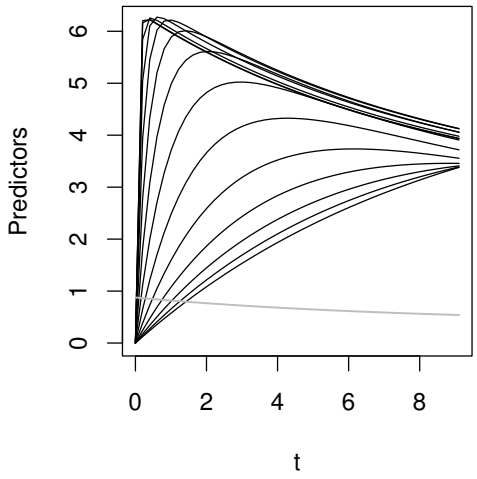

Figure 6.1: Subset of predictors (every fifth predictor is shown). Gray line: $C_p(t)$, black lines: $\Psi_1(t), \ldots, \Psi_q(t)$

$$\hat{\boldsymbol{\theta}}_i = \text{argmin}_{\boldsymbol{\theta}_i \geq 0} \left\{ \sum_t \left(Y_{i,t} - \boldsymbol{z}(t)^T \boldsymbol{\theta}_i \right)^2 + \lambda J(\boldsymbol{\theta}_i) \right\}, \tag{6.3}$$

with $\boldsymbol{z}(t)$ as given in (6.2). The strength of penalization is controlled by λ. The crucial point, however, is to choose an appropriate penalty $J(\boldsymbol{\theta}_i)$. For example, Vega-Hernandez et al. (2008) discussed the use of different penalties for solving the so-called inverse problem of the electroencephalography (EEG) in neuroscience.

A well established regularization technique which was constructed in particular for high-dimensional problems with highly correlated explanatory variables (as found in $\boldsymbol{z}$) is the so-called elastic net (Zou and Hastie, 2005), with penalty

$$J(\boldsymbol{\theta}_i) = \alpha \sum_{k=0}^{q} \theta_{i,k}^2 + (1-\alpha) \sum_{k=0}^{q} |\theta_{i,k}|. \tag{6.4}$$

Alternatively, the corresponding estimate can be written as

$$\hat{\boldsymbol{\theta}}_i = \text{argmin}_{\boldsymbol{\theta}_i \geq 0} \left\{ \sum_t \left(Y_{i,t} - \boldsymbol{z}(t)^T \boldsymbol{\theta}_i \right)^2 + \lambda \sum_{k=0}^{q} \theta_{i,k}^2 \right\},$$
$$\text{subject to } \sum_{k=0}^{q} |\theta_{i,k}| \leq s. \tag{6.5}$$

This optimization problem can be solved using quadratic programming methods, for example, the R package `quadprog` (Turlach and Weingessel, 2011). Before running the optimization algorithm, entries of $\boldsymbol{z}$ are scaled to have unit variance over time, because otherwise

$\boldsymbol{\theta}_k$ corresponding to entries of $\boldsymbol{z}$ with smaller variance would implicitly undergo higher penalization.

Seeking for a sparse solution in a high-dimensional predictor space, we use a two-stage estimation procedure that separates model selection from parameter shrinkage similar to the relaxed lasso (Meinshausen, 2007). In a first step, a set of basis functions is selected with the aid of restrictive L_1- and L_2-penalties. Due to the L_1-type penalty term in (6.5), coefficients from $\{\hat{\theta}_{i,0}, \ldots, \hat{\theta}_{i,q}\} = \{\hat{v_{p,i}}, \hat{K}^{\text{trans}}_{i,1}, \ldots, \hat{K}^{\text{trans}}_{i,q}\}$ may be set to zero (see *e.g.* Zou and Hastie (2005)) which means that corresponding arterial plasma or tissue compartments are excluded. As the basis functions are highly correlated, non-vanishing coefficients mostly appear in clusters. From each cluster, only the basis function with maximal contribution is selected to obtain a sparse basis. Like this, the selected predictors correspond to compartments the exchange rates of which are different enough to be biologically meaningful. In a second step, we refit the sparse model without penalization (pure ML-estimates).

Tuning parameters λ and s can, for example, be determined using the Bayesian information criterion (BIC) as described in Hastie et al. (2009). For given (λ, s) the BIC can be calculated as

$$\text{BIC}_{(\lambda,s)} = -2\log(L(\boldsymbol{Y}, \hat{\boldsymbol{\theta}}_{(\lambda,s)})) + p \cdot \log(N \cdot T). \tag{6.6}$$

Here, $p = \hat{q} + \mathbf{1}(\hat{v}_p > 0)$ corresponds to the number of selected predictors, and $L(\boldsymbol{Y}, \hat{\boldsymbol{\theta}}_{(\lambda,s)})$ is the Likelihood of the observed concentration $\boldsymbol{Y}$ evaluated at the elastic net estimates (6.5) of all voxels.

6.2.3 Spatially regularized estimation

So far, voxels have been treated separately, and for each voxel i parameters $\boldsymbol{\theta}_i$ have been fitted independently of each other. Since, however, there is some spatial structure across voxels, parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^T$ from all voxels should be estimated taking this structure into account. As done in Section 5.1.1, we assume a two-dimensional neighborhood structure where adjacent voxels are neighbors. That is, each voxel has four direct neighbors unless it lies at the edge of the image. Therefore, parameters for voxel i are now estimated by

$$\hat{\boldsymbol{\theta}}_i = \text{argmin}_{\boldsymbol{\theta}_i \geq 0} \left\{ \sum_t \left(Y_{i,t} - \boldsymbol{z}(t)^T \boldsymbol{\theta}_i \right)^2 + \lambda J_i(\boldsymbol{\theta}) \right\}, \tag{6.7}$$

with penalty

$$J_i(\boldsymbol{\theta}) = \alpha Q_i(\boldsymbol{\theta}) + (1-\alpha) \sum_{k=0}^{q} |\theta_{i,k}|. \tag{6.8}$$

Similar to the voxelwise regularization, the L_1-type penalty term only penalizes parameters of voxel i and favors some of the coefficients to be set to zero. The quadratic penalty term Q_i, however, enforces some spatial smoothness of parameters $\boldsymbol{\theta}$ by penalizing differences

of neighboring $\boldsymbol{\theta}_j$:

$$Q_i(\boldsymbol{\theta}) \;=\; \sum_{j\in\partial(i)} \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|^2 = \sum_{j\in\partial(i)} \sum_{k=0}^{q} (\theta_{i,k} - \theta_{j,k})^2 \,. \tag{6.9}$$

Here, $\partial(i)$ is the set of voxels that are direct neighbors of voxel i. Note that the penalty terms J_i and Q_i do depend only on parameters corresponding to neighboring voxels: $Q_i(\boldsymbol{\theta}) = Q(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\partial(i)})$ and $J_i(\boldsymbol{\theta}) = J(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{\partial(i)})$ with $\boldsymbol{\theta}_{\partial(i)} = (\theta_j)_{j\in\partial(i)}$.

Computational issues

For the computation of the spatially regularized estimates, we introduce pseudo-observations; see also Zou and Hastie (2005). More precisely, for voxel i, we define the vector of "response values"

$$\boldsymbol{y}^i = (Y_{i,0}, \ldots, Y_{i,T}, \sqrt{\lambda} \cdot \boldsymbol{\xi}^i)^T,$$

where $\boldsymbol{\xi}^i = \left(\boldsymbol{\theta}_j \right)_{j\in\partial(i)}$ are the parameter values of neighboring voxels used as pseudo-observations. The design matrix

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{Z} \\ \sqrt{\lambda} \cdot \boldsymbol{D}_i \end{pmatrix}$$

consists of

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{z}(0)^T \\ \boldsymbol{z}(1)^T \\ \vdots \\ \boldsymbol{z}(t)^T \end{pmatrix}$$

and

$$\boldsymbol{D}_i = \left.\begin{pmatrix} \boldsymbol{I} \\ \vdots \\ \boldsymbol{I} \end{pmatrix}\right\} |\partial(i)| \text{ times}$$

where $\boldsymbol{I}$ is the identity matrix of dimension $(q+1) \times (q+1)$. Thus, we have

$$\hat{\boldsymbol{\theta}}_i = \text{argmin}_{\boldsymbol{\theta}_i \geq 0} \left\{ (\boldsymbol{y}^i - \boldsymbol{X}\boldsymbol{\theta}_i)^T (\boldsymbol{y}^i - \boldsymbol{X}\boldsymbol{\theta}_i) \right\}, \quad \text{subject to } \sum_{k=0}^{q} |\theta_{i,k}| \leq s. \tag{6.10}$$

Since parameters for neighboring voxels, $\boldsymbol{\theta}_{\partial(i)}$, are unknown but needed for the calculation of $\boldsymbol{\xi}^i$, we plug in current estimates $\hat{\boldsymbol{\theta}}_{\partial(i)}$ and iterate this procedure over all voxels until convergence. For efficiency, we use a parallel update of voxels following a checkerboard pattern that uses conditional independence from other voxels given all neighboring voxels. As starting values, we use the estimates which describe the mean concentration time curve (CTC) over all voxels. Predictors are selected using the same two-step estimation procedure as for the voxelwise regularized estimation.

6.3 Simulation study

6.3.1 Simulation setup

To evaluate the proposed voxelwise and spatial elastic net, we simulated a DCE-MR image of 75×75 voxels. We chose different parameter combinations of the multi-compartment model with blocks simulated from one, two or three compartments ($q = 1, 2, 3$) with or without contribution of the plasma compartment ($v_p = 0.1$ or $v_p = 0$). In the left block, the true underlying parameters for the CTCs are $k_{\text{ep}_1} = 0.2$ and $K_1^{\text{trans}} = 0.2$; in the central block $k_{\text{ep}_1} = 0.2$, $k_{\text{ep}_2} = 4$, $K_1^{\text{trans}} = 0.1$, and $K_2^{\text{trans}} = 2$; in the right block $k_{\text{ep}_1} = 0.2$, $k_{\text{ep}_2} = 1$, $k_{\text{ep}_3} = 4$, $K_1^{\text{trans}} = 0.07$, $K_2^{\text{trans}} = 0.3$, and $K_3^{\text{trans}} = 1$. Gaussian noise was added to the CTCs for each voxel with standard deviation $\sigma = 0.05$. Figure 6.2 shows the simulated CTCs, and Figure 6.3 gives a more detailed sketch of the simulation design. The optimal penalization parameters λ^* and s^* for the proposed methods were chosen corresponding to lowest BIC values.

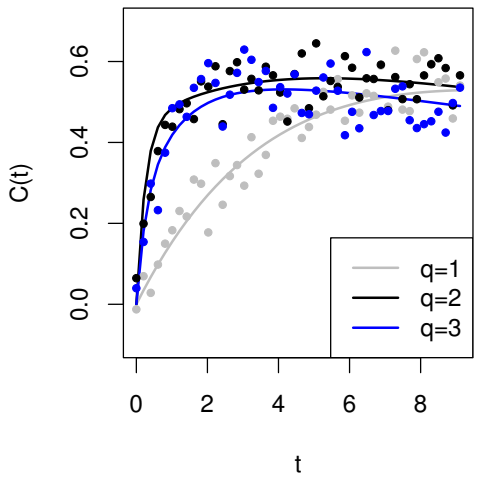

Figure 6.2: Simulated CTCs with underlying parameters $v_p = 0$, $k_{\text{ep}_1} = 0.2$, $k_{\text{ep}_2} = 1$, $k_{\text{ep}_3} = 4$. $q = 1$: $K_{\text{trans}_1} = 0.2$. $q = 2$: $K_{\text{trans}_1} = 0.1$, $K_{\text{trans}_3} = 2$. $q = 3$: $K_{\text{trans}_1} = 0.07$, $K_{\text{trans}_2} = 0.3$, $K_{\text{trans}_3} = 1$

$q = 1$ $v_p > 0$	$q = 2$ $v_p > 0$	$q = 3$ $v_p > 0$
$q = 1$ $v_p = 0$	$q = 2$ $v_p = 0$	$q = 3$ $v_p = 0$

Figure 6.3: Sketch of simulation design. The simulated image consists of 75×75 voxels divided into three main blocks ($q = 1, 2, 3$). Each main block is split up into a block with $v_p > 0$ or $v_p = 0.1$.

6.3.2 Results

Figure 6.4 summarizes the model selection for the simulated image comparing the extended Tofts model and the voxelwise and spatially regularized multi-compartment models; it shows the number q of selected basis functions Ψ_k and whether the AIF C_p was included as predictor. The extended Tofts model has a fixed number of compartments (one tissue and

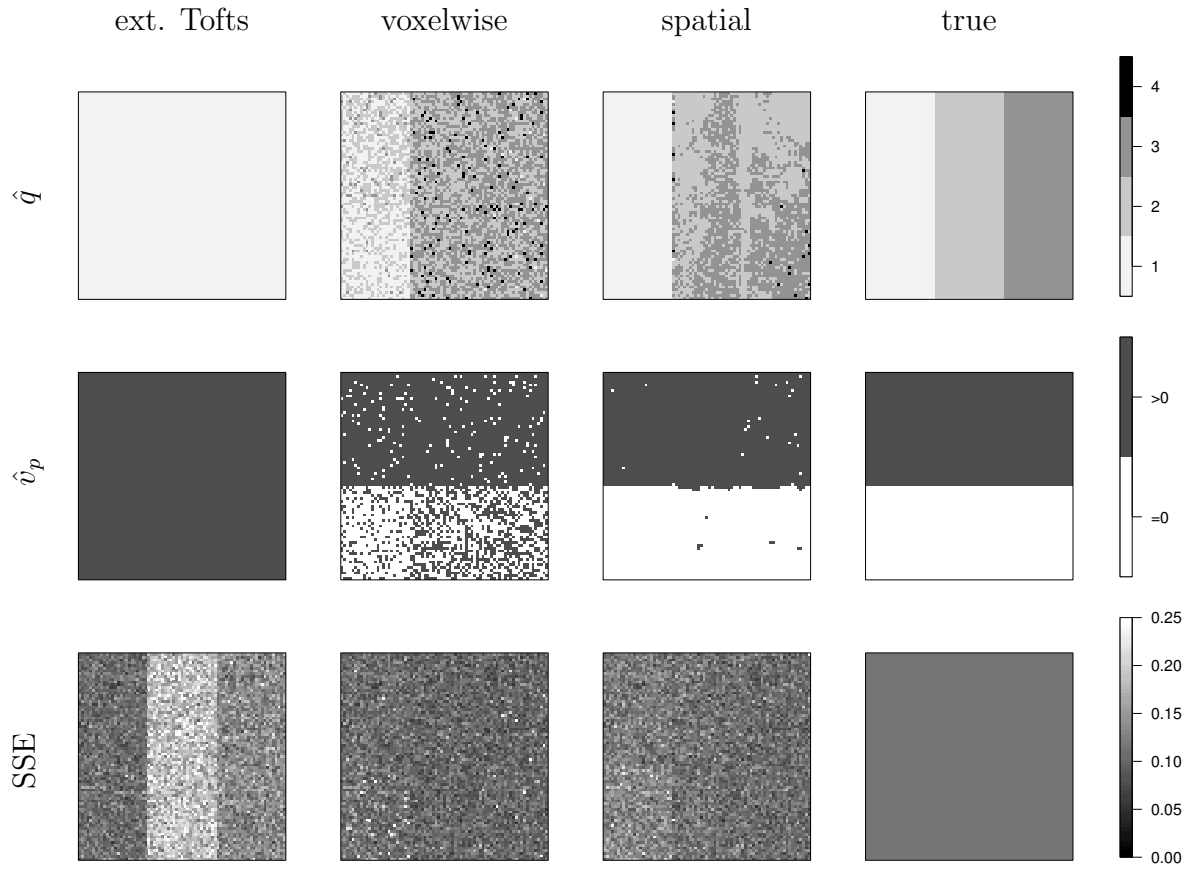

Figure 6.4: Results of simulation study: estimated number of tissue compartments $\hat{q}$, estimated inclusion of plasma compartment $\hat{v}_p$ and sum of squared errors (SSE)

one plasma compartment) for all voxels. With the voxelwise and spatial elastic net, one to four tissue compartments ($\hat{q}$) are selected as well as—optionally—a plasma compartment ($\hat{v}_p$). The spatial regularization helps to select the number of parameters more sparsely compared to the voxelwise model. The estimated model complexity corresponds very well to the true underlying model.

Figure 6.4 also depicts the sum of squared errors (SSE) per voxel. For the voxelwise and spatial elastic net, the SSE is at a similar level. In the left block, the spatial elastic net has a slightly increased SSE. For the extended Tofts model, the SSE is much higher in the second and third blocks, as the Tofts model does not account for the contribution of additional tissue compartments. This is why both restricted multi-compartment models have considerably lower BIC values compared to the extended Tofts model. In Table 6.1, the BIC as well as the average number of selected coefficients per voxel is shown. The BIC is lowest for the spatial elastic net.

Let's now discuss the model selection with the spatial elastic net in more detail. Figure 6.6 depicts the results of the spatial elastic net for three of the simulated voxels: one

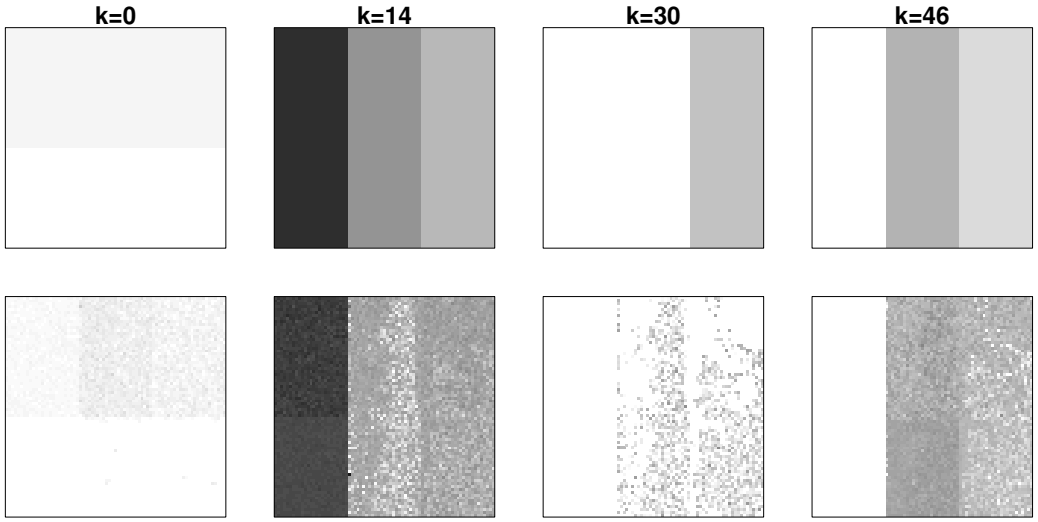

Figure 6.5: Maps of true underlying coefficients Ψ_k of the simulation study and corresponding estimated coefficients (spatially regularized)

from the left, center and right block respectively. For the voxel from the left block ($q = 1$), only one basis function is selected and the true estimated coefficient is close to the true underlying coefficient (Figure 6.6). The number of compartments q is correctly estimated to be one for almost all voxels in the left block (Figure 6.4). In the center block, q is sometimes overestimated—occasionally three instead of two basis functions are selected. This is mainly because the first basis function is often additionally selected. For a CTC simulated from three compartments (right block, $q = 3$), the contributing basis functions are too similar and—in the voxel depicted—only two basis functions are selected (Figure 6.6). Also, for other voxels in the right block, the estimate for q is often two, sometimes three. This result could be expected. In all of the three blocks, the fitted CTCs (gray) and the true underlying CTCs (blue) coincide (Figure 6.6). Model selection for the v_p-term matches the true underlying models: in the upper half the AIF is mostly included in the model, in the lower half it is mostly excluded (Figure 6.4). Furthermore, with the spatial regularization smooth parameter maps are obtained that match the true underlying parameters very well (see Figure 6.5).

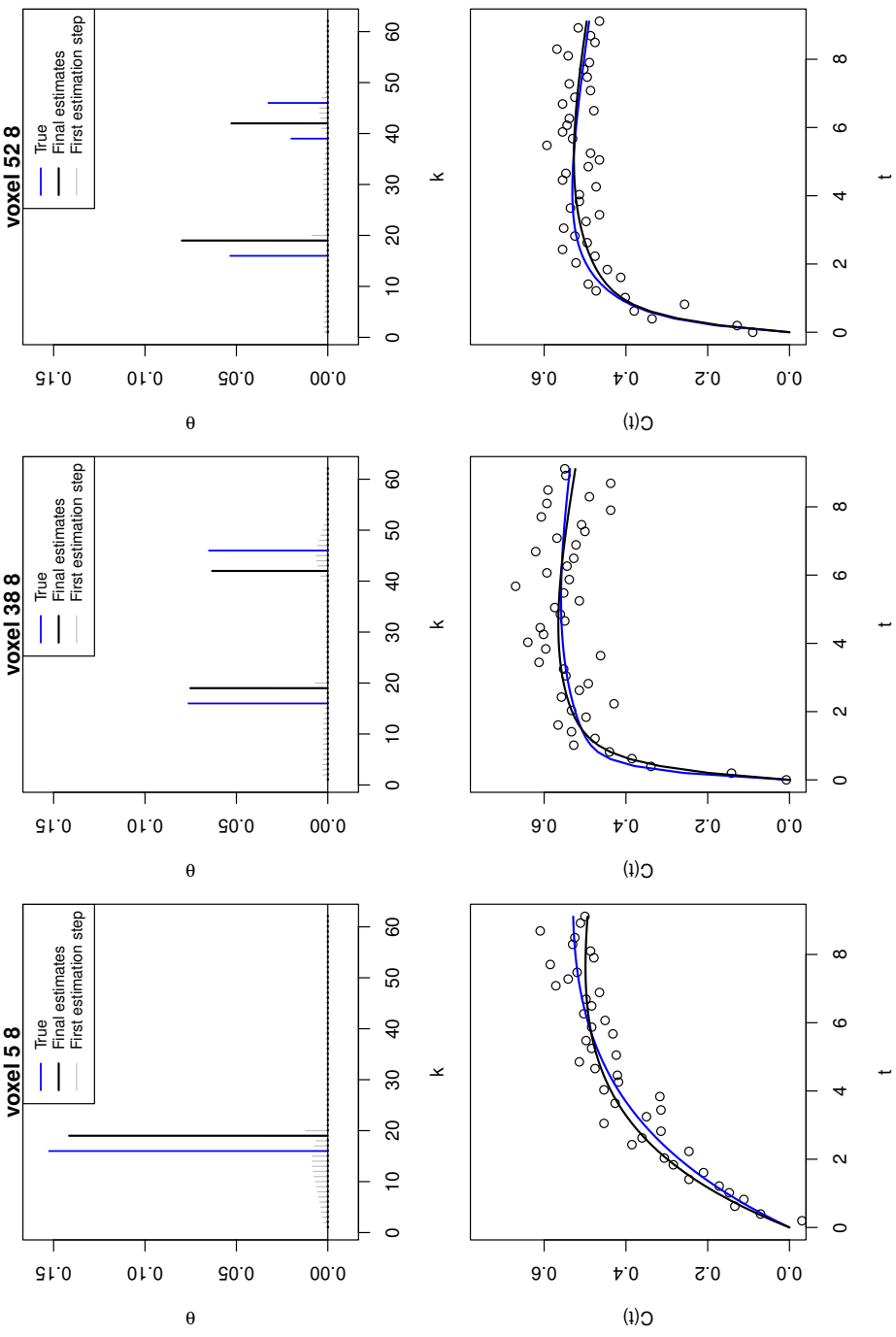

Figure 6.6: Results of spatially regularized estimation: Selection of basis functions and corresponding curve fits for three simulated voxels from the left ($q = 1$), center ($q = 2$) and right ($q = 3$) block

Table 6.1: Comparison of extended Tofts model and the voxelwise or spatially penalized multi-compartment model: $\text{BIC}_{(\lambda^*,s^*)}$ and average number of compartments per voxel ($\hat{q} + \mathbf{1}(\hat{v}_p > 0)$) for simulated image and scans from breast cancer study.

		extended Tofts	voxelwise	spatial
simulation	λ^*	–	0.1	200
	BIC	462,317	438,313	436,423
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.90	2.51
scan 1	λ^*	–	10^{-10}	10^{-10}
	BIC	758,745	764,613	717,703
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.22	1.80
scan 2	λ^*	–	10^{-10}	10^{-7}
	BIC	619,243	587,659	561,450
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	1.86	1.53
scan 3	λ^*	–	10^{-10}	10^{-10}
	BIC	539,332	535,391	508,484
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.20	1.86
scan 4	λ^*	–	10^{-10}	10^{-10}
	BIC	638,952	630,269	610,377
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.26	1.89
scan 5	λ^*	–	10^{-7}	10^{-10}
	BIC	296,972	287,248	276,943
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.30	2.05
scan 6	λ^*	–	10^{-10}	10^{-10}
	BIC	440,293	428,462	407,239
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.00	1.67
scan 7	λ^*	–	10^{-7}	10^{-10}
	BIC	866,187	805,909	781,456
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	1.78	1.59
scan 8	λ^*	–	10^{-7}	10^{-10}
	BIC	672,523	618,126	584,098
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.00	1.69
scan 9	λ^*	–	10^{-10}	10^{-10}
	BIC	855,938	796,163	763,135
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.23	1.77
scan 10	λ^*	–	10^{-10}	10^{-7}
	BIC	777,398	704,287	675,509
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.53	2.13
scan 11	λ^*	–	10^{-7}	10^{-10}
	BIC	567,411	567,171	534,742
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.35	2.02
scan 12	λ^*	–	10^{-7}	10^{-7}
	BIC	582,249	587,292	559,053
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.53	2.24

6.4 DCE-MRI breast cancer study

To evaluate the clinical use of our approach, we use a subset of the DCE-MRI breast cancer data set introduced in Section 3.3. Optimal penalization parameters for the voxelwise and spatial elastic net were $s^* = 5$ for all scans and $\lambda^* = 10^{-7}$ or $\lambda^* = 10^{-10}$, see Table 6.1. The dependence of the BIC on the penalty parameter λ is depicted in Figure 6.7 for one scan. For the voxelwise and spatially penalized multi-compartment models as well as for the extended Tofts model, the BIC and the average number of selected coefficients per voxel are listed in Table 6.1. The spatial elastic net has the lowest BIC for all scans. This indicates that this approach is suitable to sparsely select parameters in DCE-MR images and still allowing for increased model complexity where needed.

The estimation results of the spatial elastic net for the pre- and post-treatment scans of patient 6 are depicted in Figure 6.8. The estimated number of tissue compartments $\hat{q}$ and the inclusion of the AIF ($\mathbf{1}(\hat{v}_p > 0)$) are shown separately and sum up to the total number of compartments. The estimated number of tissue compartments $\hat{q}$ ranges between one and four and is two for a large number of voxels. In the pre-treatment scan, inside the tumor (top of the image), the tissue is relatively homogeneous ($\hat{q} = 1$ and $\hat{v}_p = 0$), whereas in surrounding tissue and especially at tumor margins, the tissue is more heterogeneous ($\hat{q} \geq 1$). In the post-treatment scan, tumor size is reduced, and the model complexity in the shrunk tumor has increased. From the difference in SSE compared to the extended Tofts model we find that the maximal benefit of additional compartments is at tumor margins. Here, the tissue is too heterogeneous to be adequately described by the extended Tofts model.

For most voxels, the AIF is not selected as predictor. It is selected more frequently at tissue borders, indicating a relevant amount of vascular space. If selected, the corresponding estimated v_p terms are rather small compared to contributions of other compartments: the median of non-zero v_p estimates ranges from 0.01 to 0.05 in the different scans. Summarizing, the average number of plasma and tissue compartments per voxel ($\hat{q}+\mathbf{1}(\hat{v}_p > 0)$)

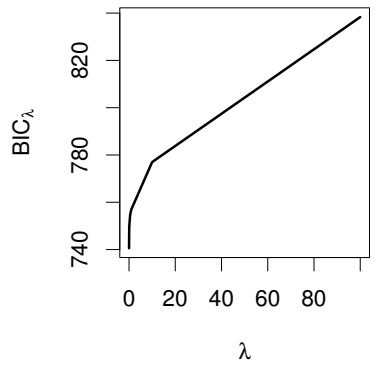

Figure 6.7: $\text{BIC}_{(\lambda,s)}$ in the spatial elastic net for $s = s^*$ and varying λ for the pre-treatment scan of patient 1

in a single scan ranges from 1.55 to 2.24 depending on the volume of homogeneous tissue, see Table 6.1.

The spatial elastic net estimates provide good fit to the observed contrast agent concentration in different tissue regions. For exemplary voxels inside the tumor, at the tumor edge, and in surrounding tissue the selection of basis functions and the final curve fits are shown in Figure 6.9. Exemplary voxels are taken from the pre-treatment scan of patient 2, see Figure 6.10 for precise location of the voxels. Parameter estimates for the entire scan are also shown in Figure 6.11.

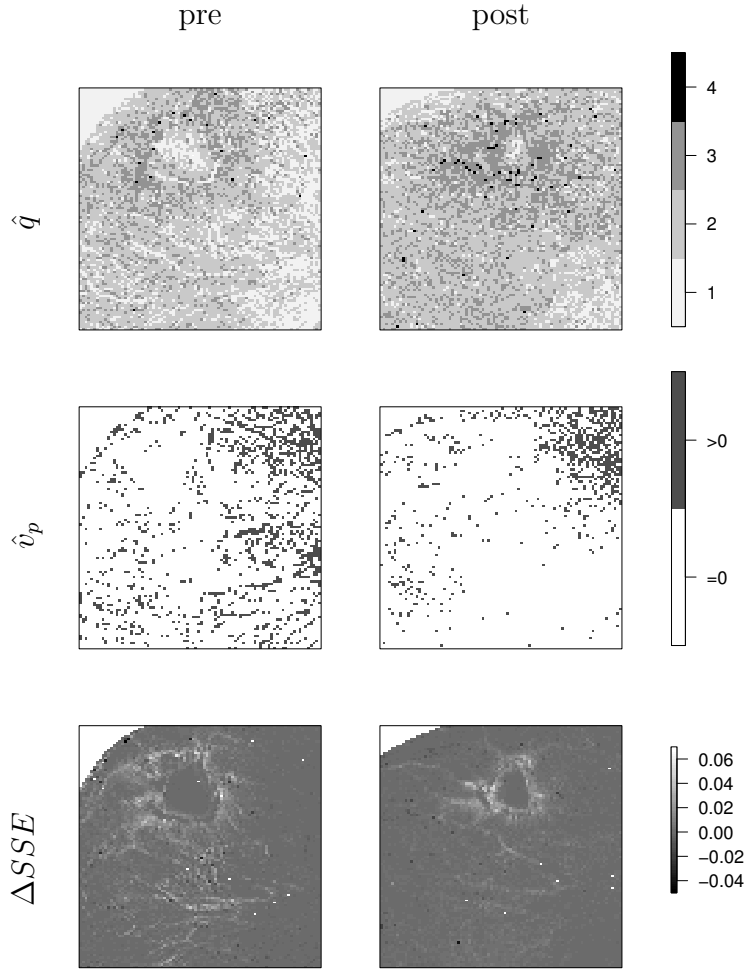

Figure 6.8: Results of spatially regularized estimation. Parameter maps for the mid-slice of patient 6 pre-treatment and post-treatment scans: estimated number of tissue compartments $\hat{q}$, estimated inclusion of plasma compartment $\hat{v}_p$ and difference in SSE ($\Delta SSE = SSE_{ext.Tofts} - SSE_{spatial}$)

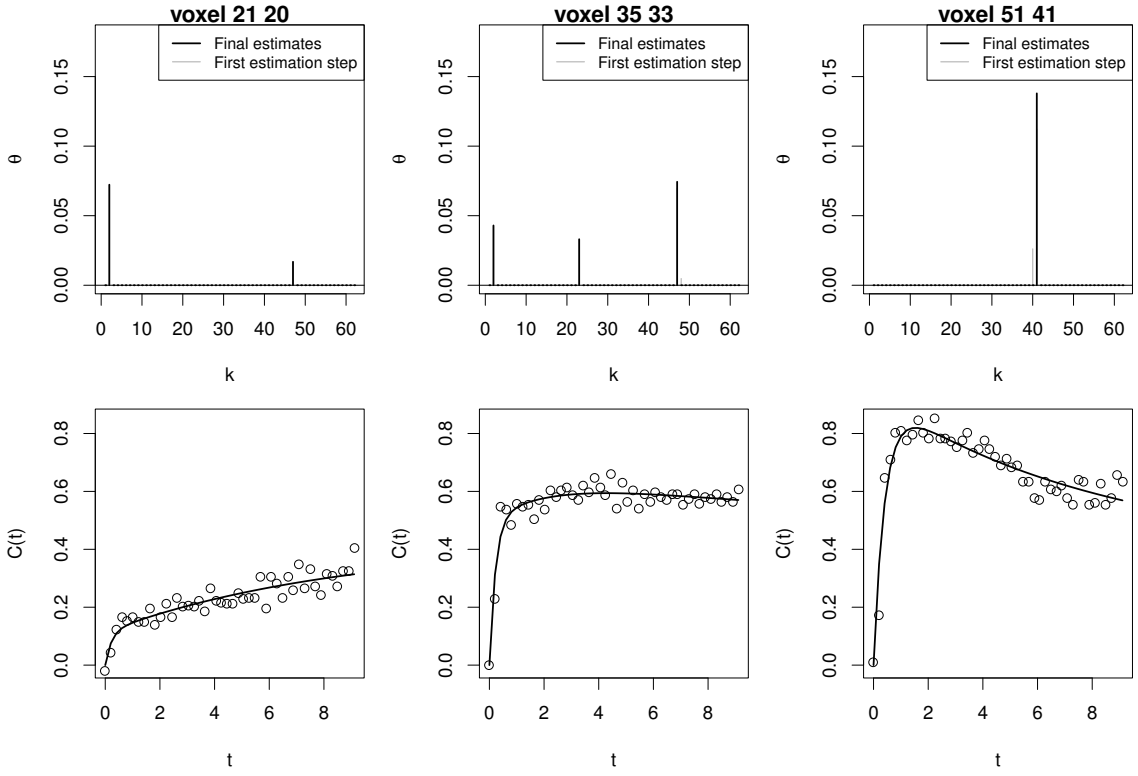

Figure 6.9: Results of spatially regularized estimation: Selection of basis functions and corresponding curve fits for a voxel in normal tissue $(21, 20)$, at the tumor edge $(35, 33)$ and inside the tumor $(51, 41)$; voxels from the mid-slice of patient 2 pre-treatment scan

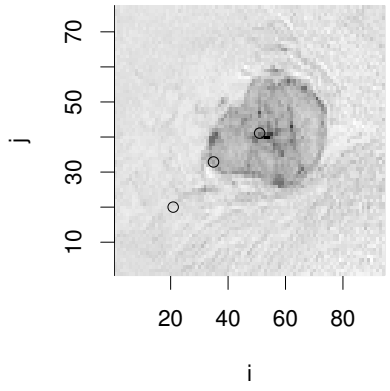

Figure 6.10: Concentration map at a fixed time point (patient 2 pre-treatment). Voxels $(21, 20)$, $(35, 33)$ and $(51, 41)$ are marked.

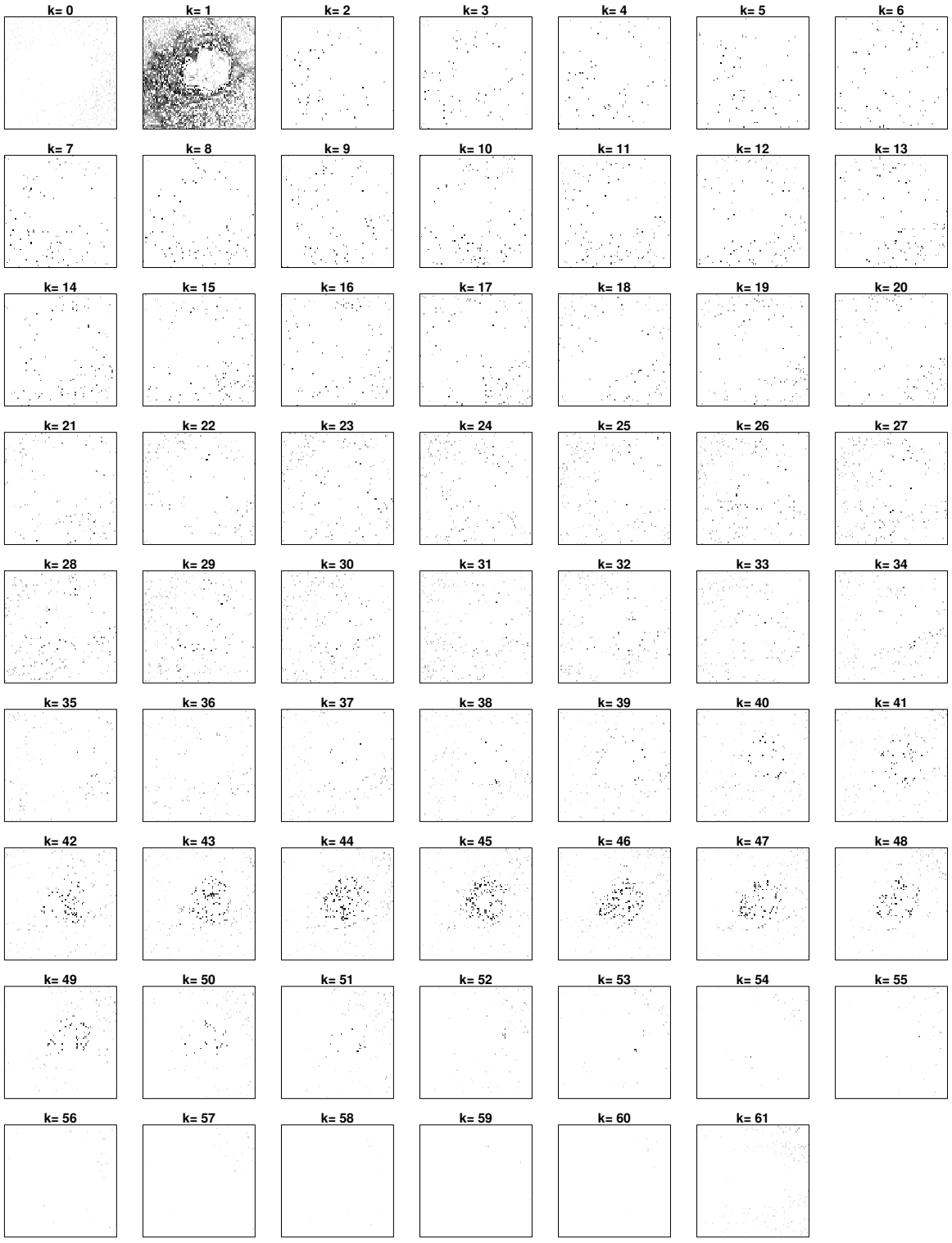

Figure 6.11: Maps of estimated coefficients Ψ_k, $k = 0, \ldots, 61$ (patient 2 pre-treatment)

6.5 Conclusion and discussion

Based on a multi-compartment model, we have proposed two penalized ML based approaches for data driven model choice and parameter estimation in DCE-MRI. Choosing basis functions based on the contribution of corresponding compartments, both approaches combine the advantages of data-driven and model-driven approaches, and parameters remain interpretable. The spatial elastic net which—as a newly developed feature—incorporates the spatial structure intrinsic in an image performed better than the voxelwise elastic net. Due to "borrowing strength" from neighboring pixels, the spatial elastic net is more robust compared to the independent voxelwise elastic net. Hence, the spatial elastic net provides a sparser solution with similar fit to the data.

With a simulation study and the analysis of twelve DCE-MRI scans we found that the spatially penalized multi-compartment model outperforms the commonly used extended Tofts model as well as the voxelwise elastic net. Our results indicate that additional model complexity is needed to adequately describe observed CTCs, especially at tumor margins and surroundings. These results are in accordance with those obtained for the spatial Bayesian model described in Chapter 5. As the number of contributing tissue compartments (and the plasma compartment) is estimated per voxel, important information about the tissue heterogeneity is gained that cannot be obtained with *a priori* fixed model architectures.

The spatial L_2-penalty is comparable with a Bayesian model with Gauss-Markov random field priors on the parameters as proposed in Chapter 5. The analogy of ridge estimates with the posterior mode of a simple normal Bayesian model with Gaussian prior is, for example, described in Hastie et al. (2009, p. 64). Besides the general differences in the viewpoints of Bayesian inference and ML inference, the penalized ML approach proposed here and the spatial Bayesian approach proposed in Chapter 5 differ in several aspects. First, the nonlinear regression problem arising from compartmental modeling is linearized with the help of basis functions, here. Hence, the exchange rates are not continuous variables but can take values on a predefined grid only. Considering this aspect, the linearized approach is less general then the approach analyzed in Chapter 5. In contrast, the linearized approach is more flexible considering the model architecture. Here, the number of compartments is not fixed *a priori* but is determined from the data.

Compared to the Bayesian approach, the spatial elastic net is easier to implement (*e.g.* no tuning of proposal variances is needed), easier to parallelize, and the calculation is much faster. As a main advantage, the degree of optimal spatial smoothing is determined from the data and does not have to be specified *a priori*. To this end, the iterative optimization algorithms has to be run for different values of the smoothing parameters λ and s. Even with this computationally expensive step the computation time is competitive with MCMC simulations run for the spatial Bayesian model.

Chapter 7

Bayesian model selection with reversible jump MCMC with application to FRAP experiments

In this chapter, we propose a Bayesian model selection approach based on reversible jump Markov chain Monte Carlo (RJMCMC) (Green, 1995). Basing inference on a hierarchical Bayesian framework including the model dimension as unknown parameter, the RJMCMC procedure allows for simultaneous parameter estimation and model selection. In Chapter 4, we have described how the choice between two competing models can be based on an information criterion. The RJMCMC approach is different, as the model dimension is treated as unknown parameter, *i.e.*, it is considered part of the parameter space. Like this, the number of compartments is determined along with other unknown parameters, and the uncertainty in model selection can be quantified. Here, we develop a RJMCMC procedure for compartment models. In contrast to the variable selection approach of Chapter 6, no approximation with a linear model needs to be done.

7.1 Introduction

The RJMCMC algorithm is a special case of a MCMC algorithm which allows to jump between different models. More precisely, in our case, it can either jump from a model with K compartments to a model with $K+1$ compartments (birth step) or to a model with $K-1$ compartments (death step). The number of compartments K is then estimated from how frequently the Markov chain takes which value (posterior distribution). To our knowledge, the RJMCMC algorithm has not been applied to nonlinear regression problems yet. We have adapted the RJMCMC algorithm to parameter estimation in a sum of exponentials model which arises in a certain class of compartment models. In contrast to many problems in linear regression models, there is no one-to-one transformation between models with different dimension K in nonlinear regression. Hence, the crucial point is to define adequate birth and death transformations. We have designed a mapping from a

model with K exponentials to a model with $K+1$ exponentials using an adequate variable transformation.

In this chapter, we focus on the special case (2.10) of the general compartment model (2.9) with constant concentration in the basis compartment. This compartment model is, for example, used in biochemical experiments that give insight to the binding behavior of molecules in living cells (Sprague and McNally, 2005). In such experiments, each molecule can adopt to different states: it can either be unbound (*i.e.* free) or it can be bound to one of several binding partners. The class of unbound molecules forms one compartment, the class of molecules bound to a binding partner forms a second compartment. However, there may be binding partners with different binding properties which form additional compartments (also called mobility classes, see Schneider (2009)). Such models are, for example, used in the analysis of fluorescence recovery after photobleaching (FRAP) experiments.

In FRAP experiments, one is confronted with yet another model selection problem in compartment models. The number of mobility classes is unknown, but is a very interesting parameter to know and to infer from the data. Therefore, we aim to determine the number of mobility classes along with other parameters of interest with the help of RJMCMC.

As described before, parameter estimation in nonlinear regression is challenging because of parameter redundancies. With the aid of a redundancy measure, Reich (1981) showed that there is high parameter redundancy in a sum of two exponentials if the exponential rates differ by less than a factor of five (see Section 2.3). From the definition of a mobility class as a class of binding partners with certain binding properties it is clear that different mobility classes should have well distinguishable binding properties. Hence, we propose a suitable regularizing prior on the exponential rates (Section 7.2). We *a priori* assume that the binding rates of two mobility classes in expectation differ by a factor of five and, hence, that the mobility classes are distinguishable. With this prior, we obtain robust estimation results. Estimation is furthermore stabilized by the automated model selection approach: In case of parameter redundancy, the Markov chain more likely jumps to a simpler model with identifiable parameters. This is an advantage over model selection based on information criteria which is done after and separately from parameter estimation.

In Section 7.2, we give a short introduction to the model equations and to the corresponding Bayesian nonlinear regression model. As main part of our work, we propose birth and death transformations for a RJMCMC algorithm, suitable for estimation and model selection in compartmental models (Section 7.2.4, Appendix B). We evaluate the performance of the proposed algorithm in a simulation study with different parameter constellations and different prior assumptions in Section 7.3, as well as in the analysis of FRAP data in Section 7.4. Finally, we summarize and discuss our results.

7.2 Methods

7.2.1 Compartment model

We assume a compartment model with constant concentration in the basis compartment. As derived in Section 2.2, the observed concentration can in this case be described by a sum of exponentials. We rewrite equation (2.10) and work with

$$\begin{aligned} C(t) &= v_B + \sum_{k=1}^{K} \frac{v_k b_k^{in}}{b_k^{out}} - \sum_{k=1}^{K} \frac{v_k b_k^{in}}{b_k^{out}} \exp(-b_k^{out} t) \\ &= 1 - a_0 - \sum_{k=1}^{K} a_k \exp(-b_k^{out} t), \end{aligned} \tag{7.1}$$

writing $a_k = \frac{v_k b_k^{in}}{b_k^{out}}$ and $v_B + \sum_{k=1}^{K} \frac{v_k b_k^{in}}{b_k^{out}} = 1 - a_0$. The parameters b_k^{in} and b_k^{out} are the transfer rates per volume unit whereas $a_k = v_k b_k^{in}$ can be interpreted as weighted transfer coefficients or as volume adjusted influx rates. Note that the outflux rates b_k^{out} can be directly obtained from (7.1), the influx rates, however, cannot. In the following, we will write b_k instead of b_k^{out} for simplicity. Here, the limit at equilibrium is assumed to be one. In FRAP experiments, this is ensured by normalization of the observed variable. With a_0 one allows for deviations from the limit one at equilibrium. Non-zero a_0 values can be due to erroneous pre-processing or can indicate the existence of a compartment with very slow diffusion, *i.e.*, very small outflux rates.

7.2.2 Bayesian nonlinear regression model

The unknown parameters $\theta = (a_0, \ldots, a_K, b_1, \ldots, b_K)$ and K in the compartmental model specified above can be inferred from the observed concentrations $C(t_i)$ at time points t_i using the nonlinear regression approach

$$C(t_i) = 1 - a_0 - \sum_{k=1}^{K} a_k \exp(-b_k t_i) + \epsilon_i, \text{ for } i = 1 \ldots T, \tag{7.2}$$

where $\epsilon_i \sim N(0, \sigma^2)$ are independent Gaussian noise terms. As hyper prior for the noise variance we assume an inverse Gamma distribution $\sigma^2 \sim IG(a, b)$.

For the number of compartments K we use a Poisson distribution with parameter $\mu > 0$, that is $p(K) \propto \frac{\mu^K}{K!}$ with $\mu = 0.5$. This prior sets more probability mass on models of lower dimension.

Considering that amplitudes shall all be positive and not too large, we propose a uniform prior of the form

$$a_k \sim U[0, 5],$$

for $k = 1, \ldots, K$. For the constant coefficient a_0 we allow for negative values and assume

$$a_0 \sim U[-0.5, 0.5].$$

The exchange rates b_k are known to be non-negative but not too large, and without loss of generality, rates are ordered: $0 \leq b_1 < b_2 < \ldots < b_K \leq 10$. The simplest way of modeling this prior knowledge is with the corresponding uniform distribution (joint density: $p(b_1, \ldots, b_K) \propto \mathbf{1}(0 \leq b_1 < b_2 < \ldots < b_K \leq 10)$).

In sum of exponential models with similar exponents b_k, these parameters are known to be not identifiable (Seber and Wild, 1989). Hence, parameter estimation is unstable when assuming a uniform prior on the exchange rates, as can be seen from the simulation study in Section 7.3. Furthermore, a model with compartments that have distinct binding properties is more plausible and better interpretable. Therefore, in the next section we will propose a regularizing prior which produces more stable results.

7.2.3 Regularizing prior on exponential rates

In order to avoid identifiability issues, we assume that the exchange rates b_k are not too similar. In addition, the positiveness and the sorting property specified above apply. A log-normal prior ensures non-negative exchange rates (Schmid et al., 2006) and allows to incorporate expected differences between them:

$$\begin{aligned} \log(b_1) &\sim N(\mu_1, \tau_1^2) \\ \log(b_2)|b_1 &\sim N(\mu_2, \tau_2^2) \\ &\vdots \\ \log(b_K)|b_{K-1} &\sim N(\mu_K, \tau_K^2). \end{aligned} \tag{7.3}$$

The hyper parameters μ_k have to be chosen so that redundancy can be avoided. This can be done using a redundancy measure in a model with two compartments (see Section 2.3). Reich (1981) has shown that in a model with a sum of two exponentials the parameter redundancy is high if the b_k parameter differ by less than a factor of five (Reich, 1981, p. 48): "If we accept 0.01 as a limit (which means a factor of 10 in the transformation from measurement error to parameter error), then parameters are redundant if decay constants differ by less than a factor of five." Therefore, we choose our prior such that the exchange rates b_k are expected to differ by a factor of five and use $\mu_1 = -2$, $\mu_k = \log(b_{k-1}) + \log(5)$ for $k = 2, \ldots, K$ and $\tau_1^2 = 2.5$, $\tau_2^2 = \tau_3^2 = \ldots = \tau_K^2 = 0.5$.

Details on regularizing prior

Choosing $\mu_1 = -2$ and $\tau_1^2 = 2.5$ we obtain $\mathrm{E}[b_1] \approx 0.5$ and $\mathrm{Var}[b_1] \approx 2.5$. In order for the expected exponentials to differ by a factor of five, *i.e.*, $\mathrm{E}[b_k|b_{k-1}] \approx 5b_{k-1}$, we choose the hyperparameters $\mu_k = \log(b_{k-1}) + \log(5)$ and $\tau_2^2 = \tau_3^2 = \ldots = \tau_K^2 = 0.5$. As

$$\begin{aligned} \mathrm{E}[b_k|b_{k-1}] &= \exp\left(\mu_k + \frac{\tau_k^2}{2}\right) \\ \mathrm{Var}[b_k|b_{k-1}] &= \exp\left(2\mu_k + \tau_k^2\right)\left(\exp\left(\tau_k^2\right) - 1\right), \end{aligned}$$

this is achieved. For example,

$$\mathrm{E}[b_2|b_1] = 5b_1 \exp\left(\frac{\tau_2^2}{2}\right) \approx 6.4 \cdot b_1 \text{ and}$$

$$\mathrm{Var}[b_2|b_1] = (5b_1)^2 \exp\left(\tau_2^2\right)\left(\exp\left(\tau_2^2\right) - 1\right) \approx 25 \cdot b_1^2 \cdot 1.07 \approx 26.7 \cdot b_1^2.$$

7.2.4 RJMCMC on number of compartments

As an extension of the MCMC algorithm (Gilks et al., 1996; Robert and Casella, 2004), the RJMCMC procedure was proposed for statistical problems where the dimension of the variable space is itself unknown (Green, 1995). Including the model dimension in the variable space, the RJMCMC procedure allows to accomplish model selection and model averaging. The main idea is simple: each model corresponds to a subspace of variables with a certain dimension. Jumps between those subspaces are performed additionally to ordinary MCMC update steps within each subspace. Before proposing RJMCMC jumps for sum of exponentials models, let us introduce to the RJMCMC procedure as proposed by Green (1995) and establish the notation. For an introduction to RJMCMC as a generalization of MCMC procedures see Green (2000).

Let θ be a parameter of dimension K. For a reversible jump step with change in dimension, a jump to a parameter θ^* of dimension K^* is proposed. For the transition of the variable space with dimension K to the variable space with dimension K^* a bijection g has to be specified. This bijection g maps (θ, u) to (θ^*, u^*) and vice versa. The random numbers u and u^* are auxiliary random variables. They have density q and are needed in order to map between spaces of different dimension. The bijective mapping g itself is completely deterministic. The proposal of a new state is random due to the auxiliary random variables u and u^*. In order for the Markov transition kernel to meet the detailed balance equation the dimensions must match, meaning that

$$dim(\theta) + dim(u) = dim(\theta^*) + dim(u^*).$$

As the dimension is part of the parameter space, we denote the unknown parameter as $\Theta = (\theta, K)$. The acceptance probability for a move from Θ to Θ^* for a Metropolis-Hastings step in general is

$$\alpha(\Theta, \Theta^*) = min\left\{1, \frac{\pi(\Theta^*)}{\pi(\Theta)} \frac{h(\Theta|\Theta^*)}{h(\Theta^*|\Theta)}\right\}$$

where h is the proposal density. We use

$$\pi(\Theta) = p(\theta, K) = p(\theta|K) \cdot p(K)$$

and the special way of proposing a new state by proposing auxiliary variables u and u^* and by applying the deterministic mapping g. Then, the acceptance probability of a jump with change in dimension is

$$\alpha(\Theta, \Theta^*) = min\left\{1, \frac{p(\theta^*|K^*)}{p(\theta|K)} \frac{p(K^*)}{p(K)} \frac{q(u^*)}{q(u)} det\left(\frac{\partial(\theta^*, u^*)}{\partial(\theta, u)}\right)\right\},$$

a special case of the general Metropolis-Hastings acceptance rate. Here, $det\left(\frac{\partial(\theta^*,u^*)}{\partial(\theta,u)}\right)$ denotes the determinant of the Jacobian matrix of the bijection g. Green (1995, 2000) gives the exact derivation of the acceptance probabilities by use of the detailed balance condition.

In RJMCMC procedures, the common way of changing the dimension are so called birth and death moves. Those birth and death moves are restricted to jumps between neighboring variable subspaces. In a birth step, one increases the dimension of the subspace and proposes a jump from K to $K^* = K + 1$. A death step is a move from K to $K^* = K - 1$. We have designed the following birth and death variable transformations for sum of exponential models of the form (7.2).

Birth move $(K \longrightarrow K+1)$:
Given the actual state $x = (a_1, ..., a_K,\ \ v_1, ...v_K, u_1, u_2)$, a new state $y = \big(a_1^*, ..., a_{K+1}^*, v_1^*, ..., v_{K+1}^*,\big)$ is proposed as follows:

Draw two random numbers in order to extend the variable space:

$$u = (u_1, u_2) \sim q(u_1, u_2) = U[0,1] \times U[0,1].$$

Draw an index $n \in \{0, ..., K\}$. Divide the amplitude a_n to the amplitudes a_n^* and a_{n+1}^* determining their proportion using u_1. Divide v_n into the new volumes v_n^* and v_{n+1}^* determining their proportion using u_2. The birth transformation $g_B(x)$ is given by:

$$\begin{array}{lcll} a_j^* & = & a_j & \text{for } j < n \\ a_n^* & = & u_1 a_n & \\ a_{n+1}^* & = & (1-u_1)a_n & \\ a_j^* & = & a_{j-1} & \text{for } j \geq n+2 \end{array}$$

and

$$\begin{array}{lcll} v_j^* & = & v_j & \text{for } j < n \\ v_n^* & = & u_2 v_n & \\ v_{n+1}^* & = & (1-u_2)v_n & \\ v_j^* & = & v_{j-1} & \text{for } j \geq n+2. \end{array}$$

Death move $(K \longrightarrow K-1)$:
Given the actual state $y = (a_1, ..., a_K, v_1, ...v_K)$, a new state $x = \big(a_1^*, \ldots, a_{K-1}^*, v_1^*, \ldots, v_{K-1}^*, u_1^*, u_2^*\big)$ is proposed as follows:

Draw an index $n \in \{1, ..., K\}$. Join the volume v_n and v_{n+1} to v_n^* and the amplitudes a_n and a_{n+1} to a_n^*. The death transformation $g_D(y)$ is given by:

$$\begin{array}{lcll} a_j^* & = & a_j & \text{for } j < n \\ a_n^* & = & a_n + a_{n+1} & \\ a_j^* & = & a_{j+1} & \text{for } j \geq n+1 \\ u_1^* & = & \frac{a_n}{a_n + a_{n+1}} & \end{array}$$

and

$$\begin{aligned} v_j^* &= v_j && \text{for } j < n \\ v_n^* &= v_n + v_{n+1} \\ v_j^* &= v_{j+1} && \text{for } j \geq n+1 \\ u_2^* &= \tfrac{v_n}{v_n + v_{n+1}}. \end{aligned}$$

The auxiliary random variables u_1^* and u_2^* are chosen such that $g_D(y)$ is the inverse transformation of $g_B(x)$. The determinants of the Jacobian matrices of birth or death transformations as derived in Appendix B.1 are

$$det(J_{birth}) = det\left(\frac{\partial y}{\partial x}\right) = a_n v_n$$

and

$$det(J_{death}) = \frac{1}{a_n^* v_n^*} = \frac{1}{(a_n + a_{n+1})(v_n + v_{n+1})}.$$

Birth and death mappings are designed such that the curve corresponding to the new proposed parameters is not too different from the curve represented by the actual parameters. For the proposed mapping, the curves have the same value at $t = 0$ ($a_n = a_n^* + a_{n+1}^*$) and the area under the curves are approximately the same ($v_n = v_n^* + v_{n+1}^*$). A derivation of those conditions is given in Appendix B.2.

7.2.5 Sampling

Ordinary MCMC update steps or reversible jump moves are performed, depending on whether the new proposed dimension K^* does or does not equal the current dimension K. First, the new dimension K is proposed using the following proposal kernel:

$$p(K_*|K) = \begin{cases} p_c \cdot p_{birth}, & K_* = K+1, K < K_{max} \\ p_c \cdot p_{death}, & K_* = K-1, K > 1 \\ 1 - p_c, & K_* = K \end{cases}$$

where p_c is the probability to propose a change in dimension and p_{birth} and p_{death} are the probabilities to propose birth or death steps. Here, we use $p_c = p_{birth} = p_{death} = 0.5$. With $p_{birth} = p_{death} = 0.5$, the kernel is symmetric for birth and death steps, except for $K = 1$ and $K = K_{max}$.

If there is a change in dimension ($K_* \neq K$), a birth or death step is performed as described before. If the dimension of the parameter stays the same ($K^* = K$), ordinary MCMC update steps are performed.

For the coefficients a_k, we can derive full conditionals in closed form

$$a_{k^*}|C, b, a_{-k^*} \sim N_{[0,5]}^{trunc}\left(\frac{m}{v}, \frac{1}{v}\right)$$

with $m = \sum_{i=1}^{T} \frac{\left\{C(t_i) - 1 + a_0 + \sum_{k \neq k^*} a_k \exp(-b_k t_i)\right\} \exp(-b_{k^*} t_i)}{\sigma^2}$ and $v = \sum_{i=1}^{T} \frac{\exp(-2b_{k^*} t_i)}{\sigma^2}$. With $N_{[\alpha,\beta]}^{trunc}$ we denote the truncated normal distribution with zero density for $a_{k^*} \notin [\alpha, \beta]$.

We can use Gibbs-update steps to update the coefficients a_k, as we can directly sample from the full conditionals. Similarly, σ^2 can directly be sampled from an inverse Gamma distribution. We cannot directly sample from the full conditional of b_k, and hence we use a Metropolis-Hastings update step for the exponents b_k. As proposal we use a log-normal distribution.

In cases where the model dimension is not clearly identifiable, the estimation of the model dimension K may become unstable and may depend on the starting values. Therefore, we use ten parallel Markov chains with random starting values.

We estimate the mean frequency f_k of samples with model dimension $K = k$ using all parallel chains. Furthermore, we estimate the model dimension K^* for each curve as mode of the dimension in all runs. Once the model dimension K^* is estimated, the exponential rates b_k are estimated using only the samples with model dimension $K = K^*$. Estimates for a_k are obtained correspondingly. We also calculate the corresponding 90 percent credibility intervals.

7.3 Simulation study

In order to evaluate the performance of the proposed RJMCMC algorithm, a simulation study was done. We simulated curves from a sum of two exponentials with equal amplitudes $a_1 = a_2 = 0.2$ and $a_0 = 0$. For the exponents b_k, the following parameter settings were used: $b_1 = 0.2$ and $b_2 = c \cdot b_1$ with $c = 1, 1.5, 2, 3, 5, 10$. For $c = 1$, the parameters b_1 and b_2 are equal. This corresponds to the case of only one exponential ($K = 1$), and we can as well write $a_1 = 0.4$, $a_2 = 0$ in this case. All other constellations correspond to a sum of two exponentials. However, for $c = 1.5, 2, 3$, the rates b_1 and b_2 are still quite similar, and parameters are highly redundant in this case as they differ less than a factor of five (see Section 2.3). Thus, for those parameter constellations, we would expect a model dimension somewhere between one and two ($1 < K < 2$). For $c = 10$, we have $b_1 = 0.2$ and $b_2 = 2$. As the exponents differ by a factor of ten, we consider this a model where the sum of two components should be clearly identifiable ($K = 2$). The curves for parameter constellations with $c = 1$, $c = 2$, and $c = 10$ are shown in Figure 7.1.

For each set of parameters we simulated ten curves, each from the corresponding sum of exponentials model with added Gaussian noise with standard deviation $\sigma = 0.08$, $\sigma = 0.02$, $\sigma = 0.008$, $\sigma = 0.004$. Time points where chosen equidistant between $t_0 = 0$ to $t_{150} = 15$ at a distance of 0.1 each. For each of the simulated curves the proposed reversible jump algorithm was performed with $70,000$ iterations for the parameter updates. The first $50,000$ samples were discarded as burnin and the samples were thinned for storage purposes (every third sample was saved).

7.3.1 Simulation results using a uniform prior on the exponential rates

First, we use a uniform prior on the exponential rates b_k as described in Section 7.2.2. Here, the mixing between models of different dimension is good. However, the model complexity is often overestimated.

For simulated curves with $c = 1$ (and hence true model dimension $K = 1$) and moderate noise ($\sigma = 0.004$ and $\sigma = 0.008$), about 90% of the RJMCMC-samples have model dimension $K = 1$. In Figure 7.2, we see a part of the samples for the parameter setting with $c = 1$. The posterior model dimension, hence, corresponds to the true model dimension in this case. However, for simulated curves with $c = 1.5$, $c = 2$ or $c = 3$ (and hence true model dimension $1 < K < 2$) and moderate noise, the posterior model dimension is $K = 2$ or even $K = 3$. The RJMCMC chain rarely jumps to the parameter space of dimension $K = 1$. For $c = 10$ and moderate noise, the mode of the posterior dimension is $K = 3$; see Figure 7.3.

The model dimension is overestimated for two reasons. First, some RJMCMC-samples become very small (*e.g.* $b_1 \approx 0$ and $a_1 \approx 0$), such that the contribution of the first compartment to the observed concentration vanishes, but nevertheless increases the estimated model dimension. Secondly, in many cases, two compartments have similar sample values for b_k (*e.g.* $b_2 \approx b_3$). The contribution of the two compartments is then not distinguishable due to redundancy issues. In this case, the estimates of the corresponding a_k have high variance and are highly dependent. As the posterior distribution of a_k is bimodal or even multimodal, parameter estimation becomes difficult.

With larger noise ($\sigma = 0.08$), the estimated frequencies are similar for the different values of c: $(f_1, f_2, f_3) \approx (0.50, 0.30, 0.15)$. The different models cannot be distinguished, as with more noise the differences of the parameter constellations can hardly be detected from the observed curves and the prior on the model dimension becomes a determining factor.

Due to the redundancy issues of the sum of exponentials model, the algorithm cannot distinguish between a model with one compartment and a model with two compartments and similar b_k. Hence, we propose a regularizing prior (7.3) so that the algorithm only selects compartments with distinguishable b_k. This also improves the interpretability of the model.

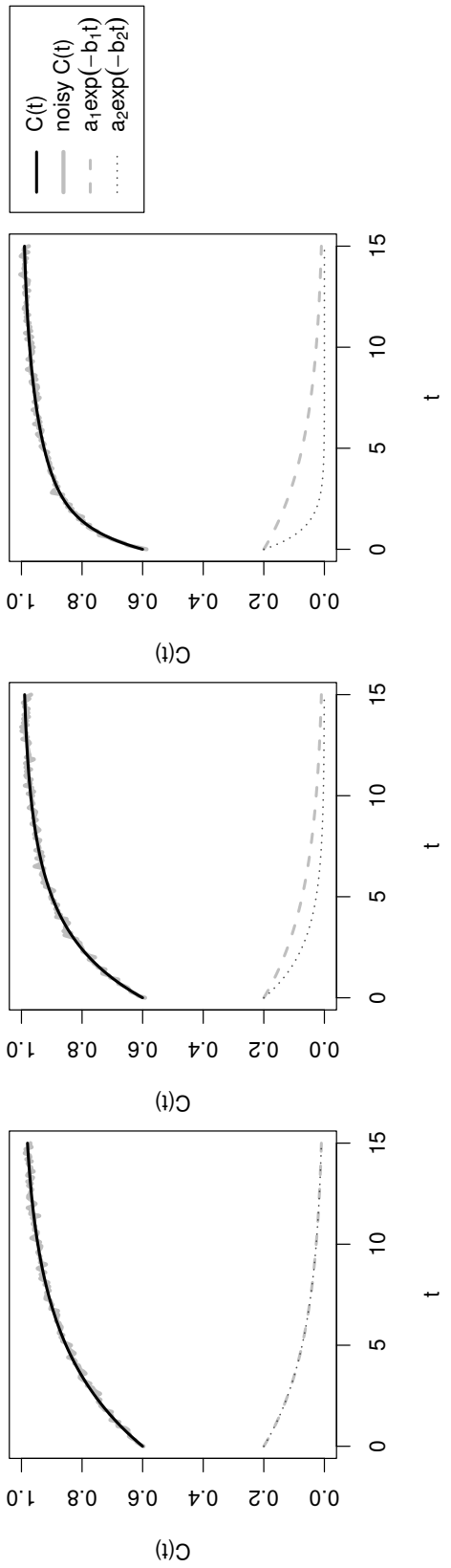

Figure 7.1: Simulated curves with noise (gray) and without noise (black) and contribution of single compartments for different parameter constellations: $b_1 = b_2 = 0.2$ (left), $b_1 = 0.2$ and $b_2 = 0.4$ (center), $b_1 = 0.2$ and $b_2 = 2$ (right). Noise level $\sigma = 0.008$.

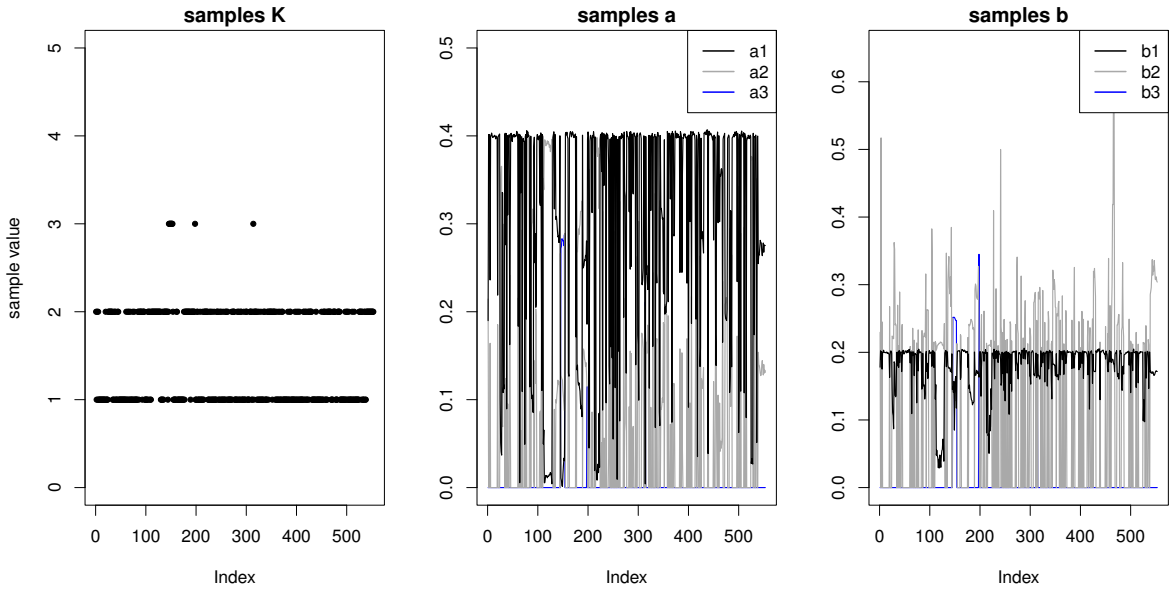

Figure 7.2: Samples for simulation with uniform prior and $b_1 = b_2 = 0.2$ ($c = 1$). For better visualization only a small part of thinned samples is shown.

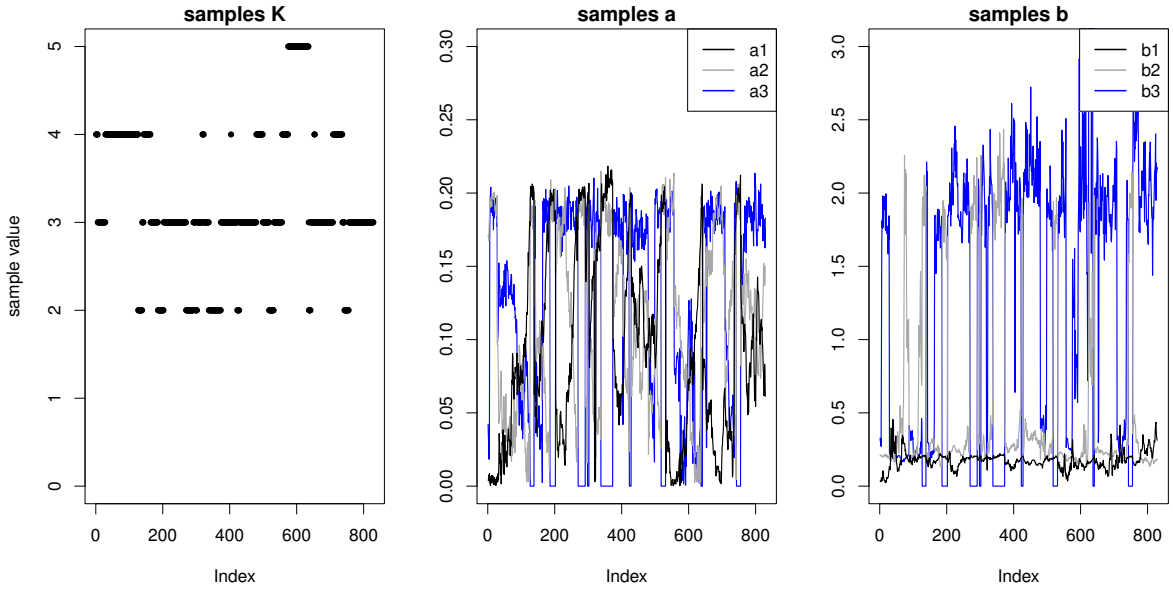

Figure 7.3: Samples for simulation with uniform prior and $b_1 = 0.2$, $b_2 = 2$ ($c = 10$). For better visualization only a small part of thinned samples is shown.

Table 7.1: Estimation results for regularizing prior and noise level $\sigma = 0.008$: median and 90% credible interval

		K	b_1	b_2	a_1	a_2
$(c = 1)$	estimates	$K = 1$	0.2	0	0.4	0
	credibility	1	(0.2,0.2)	(0,0)	(0.39,0.41)	(0,0)
$(c = 1.5)$	estimates	$K = 1$	0.24	0	0.4	0
	credibility	1	(0.24,0.25)	(0,0)	(0.39,0.4)	(0,0)
$(c = 2)$	estimates	$K = 2$	0.2	0.43	0.21	0.19
	credibility	0.67	(0.08,0.24)	(0.31,0.78)	(0.03,0.34)	(0.07,0.37)
$(c = 3)$	estimates	$K = 2$	0.2	0.62	0.21	0.2
	credibility	0.95	(0.16,0.23)	(0.48,0.85)	(0.13,0.27)	(0.14,0.27)
$(c = 5)$	estimates	$K = 2$	0.2	1.02	0.2	0.2
	credibility	0.98	(0.18,0.22)	(0.82,1.25)	(0.17,0.23)	(0.17,0.22)
$(c = 10)$	estimates	$K = 2$	0.2	2.03	0.2	0.2
	credibility	0.99	(0.19,0.22)	(1.75,2.38)	(0.19,0.21)	(0.18,0.22)

7.3.2 Simulation results using a regularizing prior on the exponential rates

Using the regularizing prior (7.3) we analyze the same simulated data set as before. Table 7.1 lists the estimated parameters for different parameter constellations at moderate noise level ($\sigma = 0.008$). For $c = 1$, the Markov chain remains in a one-dimensional model ($K = 1$) for all samples after burnin. The model dimension K and the parameters a_1 and b_1 are correctly estimated. For $c = 10$, the Markov chain remains in a two-dimensional model ($K = 2$) for 99% of the samples. The parameter estimates for K, a_1, a_2, b_1, and b_2 perfectly match the true underlying values.

For the parameter constellation with exponential rates differing only by a factor of two ($c = 2$, $1 < K < 2$), the model dimension is $K = 1$ for 33% of the samples and $K = 2$ for 67% percent. This result correctly displays a model with redundant parameters and model uncertainty. The estimates for b_1, b_2, a_1, and a_2 match the true underlying values. Credibility intervals for a_1 and a_2 are very broad in this case due to parameter redundancy.

With the regularizing prior, mixing between models of different dimensions is not as good as without regularization. In the model with uniform prior proposed in Section 7.3.1, a jump from model dimension $K = 1$ to $K = 2$ is more likely, as a jump from one compartment to two very similar compartments is not penalized by the regularizing prior. Estimated model dimensions for a single chain may depend on the starting values. Using parallel chains, however, we obtain appropriate estimates for the model dimension. Summarizing, the regularizing prior helps to avoid models with too similar compartments or compartments with little contribution. With its help, the model dimension is no longer overestimated and the parameter estimates become more stable.

Naturally, the estimated model dimension depends on the noise level. Figure 7.4 depicts the estimated model frequencies f_1, f_2, and f_3 for different parameter constellations ($c = 1, 1.5, 2, 3, 5, 10$) and different noise level ($\sigma = 0.004$, $\sigma = 0.008$, $\sigma = 0.02$). For values $c = 1$ and $c = 1.5$, the model dimension is $K = 1$ for all samples, as the rates b_1 and b_2 are too similar to be distinguished. With increasing c, the model dimension is $K = 2$ for more samples, indicating that the more complex model can still be identified. However, this is only the case if the noise is not too large. For large noise variance σ^2 the samples most often have model dimension $K = 1$. Even if the curve was generated from a more complex model, it cannot be distinguished from the simpler model if there is too much noise.

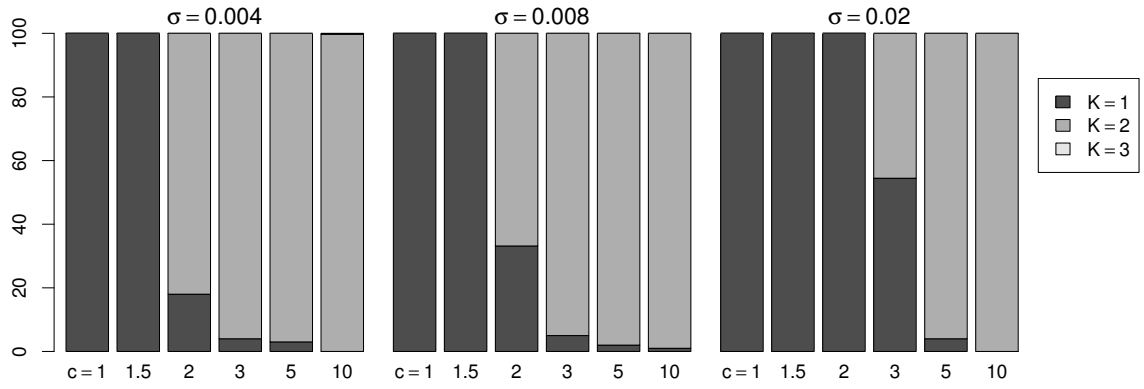

Figure 7.4: Model frequencies f_k in percent for different parameter constellations and and different noise level (simulation using regularizing prior)

7.4 Application: Analysis of molecular binding in a FRAP experiment

7.4.1 Biological background, experimental setup and motivation

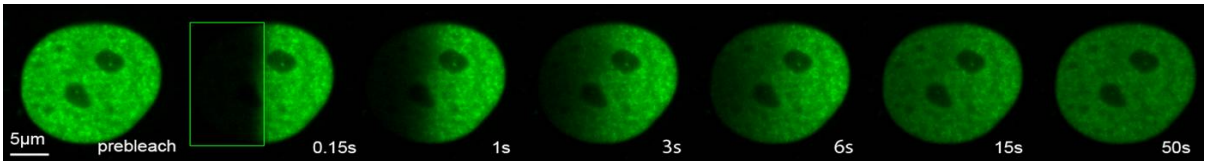

Figure 7.5: FRAP recovery over time after half of the nucleus was bleached, from Schneider (2009)

We use a data set from a FRAP experiment to show the practical relevance of the proposed approach. FRAP is a common application of compartmental models. The aim of FRAP experiments is to describe the dynamic behavior of molecules in a living cell nucleus and to determine number, rate, and strength of their binding interactions (Sprague and McNally, 2005; Phair et al., 2004). Here, DNA methyltransferase 1 (Dnmt1) was labeled with a green fluorescent protein (GFP). In order to make the protein's mobility visible, the fluorescent molecules in one half of the nucleus are permanently bleached. Then, the recovery of fluorescence in the bleached region is observed, see Figure 7.5. From this recovery, one can infer the diffusion and the binding of the molecule to chromatin.

A cell passes through different cell cycle phases consisting of the first gap phase (G1), the synthesis phase (S), the second gap phase (G2), and the mitosis phase (M). Normally, the DNA methylation pattern is maintained over many cell cycles. If the methylation pattern is not maintained, the risk that tumor cells are formed is increased. The binding patterns of Dnmt1 in different cell cycle phases is of interest, as it is important for the maintenance of DNA methylation patterns. For a more detailed description see Schneider (2009), Dargatz (2010), and McNally (2008).

The dynamic behavior of recovery in the bleached part is modeled via one perfectly mixed compartment of free, unbound molecules and one or several compartments of bound molecules (see Figure 7.6). Here, the free (unbound) molecules form the central compartment. The dynamics of the molecules' binding and unbinding is characterized by constant rates. The binding behavior is assumed to be the same in the bleached and the unbleached part of the cell. Only the fluorescence in the bleached part is used as observed variable. More precisely, one observes the mean gray value in the bleached section over time, here denoted as $frap(t)$. It can be calculated as fraction of unbleached molecules (in the bleached section) with respect to the total number of molecules (in the bleached section). The total amount of free molecules is assumed to be constant during observation. Thus, the observed

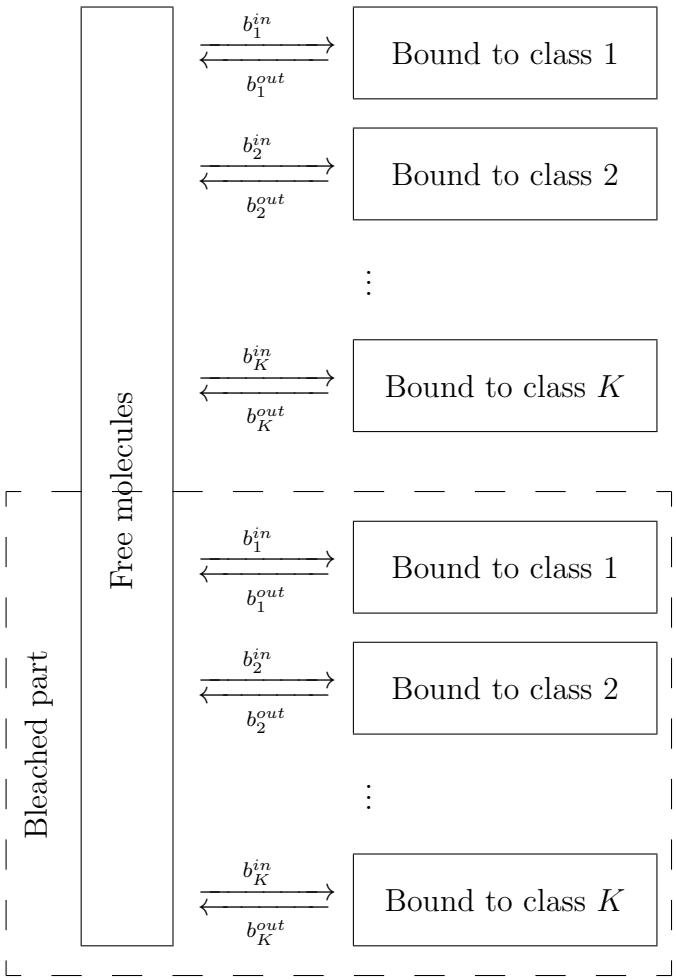

Figure 7.6: Compartment model for FRAP experiment with K binding classes. The observed variable is calculated only from the bleached part of the cell.

fraction of unbleached molecules in the region of interest $frap(t)$ is obtained as

$$frap(t) = 1 - a_0 - \sum_{k=1}^{K} a_k \exp(-b_k t),$$

where b_k is the off-rate of binding of the kth mobility class. This model is called diffusion-uncoupled FRAP (McNally, 2008) and assumes that the diffusion dynamics of unbound molecules is very fast compared to the binding dynamics and can, hence, be ignored.

The FRAP experiment was conducted at the LMU Biocenter Martinsried and was previously reported about (Schneider, 2009; Dargatz, 2010; Schneider et al., 2013). For biological processing and image processing details see Schneider (2009). The observed data were normalized in order to account for the decrease of fluorescence over time, and the gray value was normalized such that the $frap$-value was one after full recovery (Dargatz, 2010),

that is, after the two cell parts are again in equilibrium. Observations were gained every 0.15 seconds for 780 post-bleach frames. Here, we use the observed concentration curves of GFP-labeled Dnmt1 in G1 and early S phase of the cell cycle and observed curves of unconjugated GFP. For each class, ten cells were observed.

7.4.2 Results

The parameter estimates for GFP-labeled Dnmt1 in cell cycle phases G1 and early S and for unconjugated GFP are shown in Tables 7.2–7.4. For the G1 and the early S phase, the number of compartments is nearly always chosen to be $K = 2$. However, there are differences in the exponential rates for the different cell cycle phases. For the G1 phase, the smaller exponential decay rate b_1 is estimated to be about 0.15 and the larger rate b_2 to be about 0.5 to 1.2. The contribution of the second compartment, $a_2 \exp(-b_2 t)$, is rather small. Figure 7.7 depicts the observed concentration curve, the fitted curve, and the contributions of each compartment for one of the cells for unconjugated GFP and Dnmt1 in G1 and early S phase.

For the early S phase, there is one binding class with very strong binding leading to a slow exponential rate b_1 of about 0.07 as shown in Table 7.3. The second binding class has estimated exponential rates b_2 of about 0.15 to 0.4. Both mobility classes have considerable contributions (see Figure 7.7). Here, a_0 is positive in all cases but one. This indicates a third binding class with very slow binding.

For unconjugated GFP, the estimated number of mobility classes is $K = 1$ for most of the cells. As shown in Table 7.4, the estimates of the exponential rate vary from $b_1 = 0.9$ to $b_1 = 1.26$ over the ten cells. The diffusion of unconjugated GFP is very fast compared to the diffusion of GFP-tagged chromatin. This result is in agreement with what biologists would expect and even require as premise for FRAP experiments: "The premise of this experiment is that unconjugated GFP should not bind to cellular structures, and should simply diffuse." (McNally, 2008, p.340) The parameter a_0 is around 0 in all cells. In two cases, the 90% credibility interval only covers small negative values. Here, the normalization mentioned above may be erroneous. A typical time series observed for unconjugated GFP and the fitted sum of exponential curve are shown in Figure 7.7.

7.4.3 Sensitivity analysis

In order to check the sensitivity to the prior assumptions, we run the RJMCMC procedure with different priors on b_1:

(i) Prior as above, that is $\mu_1 = -2$ and $\tau_1^2 = 2.5$.

(ii) Prior with $\mu_1 = -5$ and $\tau_1^2 = 5$ such that $\mathrm{E}[b_1] \approx 0.08$ and $\mathrm{Var}[b_1] \approx 1$.

(iii) Prior with $\mu_1 = -5.5$ and $\tau_1^2 = 6$ such that $\mathrm{E}[b_1] \approx 0.08$ and $\mathrm{Var}[b_1] \approx 2.7$.

(iv) Prior with $\mu_1 = -1$ and $\tau_1^2 = 0.8$ such that $\mathrm{E}[b_1] \approx 0.54$ and $\mathrm{Var}[b_1] \approx 0.37$.

We found that the estimates of the model dimension K and all model parameters stay the same for different prior mean and variance of b_1. Only the width of the credibility intervals varies slightly (wider for larger variance or when the prior mean is in conflict with data). In Table 7.5, we show the estimates for K and b_1 for the G1 curves. Summarizing, we found that parameter estimates for the FRAP experiment are not very sensitive to the prior on b_1.

Table 7.2: Estimation results for FRAP curves for cell cycle G1: median and 90% credible interval

G1		K	b_1	b_2	a_1	a_2	a_0
1	estimates	$K=1$	0.17	0	0.89	0	0.004
	credibility	1	(0.17,0.18)	–	(0.87,0.9)	–	(0.001,0.006)
2	estimates	$K=2$	0.16	0.49	0.75	0.16	-0.003
	credibility	0.9	(0.14,0.17)	(0.27,1.1)	(0.5,0.83)	(0.09,0.4)	(-0.005,-0.002)
3	estimates	$K=2$	0.14	0.9	0.71	0.21	0.012
	credibility	0.9	(0.14,0.15)	(0.65,1.28)	(0.68,0.74)	(0.18,0.24)	(0.011,0.013)
4	estimates	$K=2$	0.17	0.83	0.69	0.21	0.018
	credibility	1	(0.14,0.18)	(0.39,1.75)	(0.52,0.75)	(0.15,0.35)	(0.016,0.02)
5	estimates	$K=2$	0.12	0.61	0.63	0.3	0.016
	credibility	0.95	(0.12,0.13)	(0.48,0.78)	(0.57,0.67)	(0.26,0.34)	(0.015,0.017)
6	estimates	$K=2$	0.12	0.49	0.64	0.28	-0.003
	credibility	1	(0.11,0.13)	(0.34,0.73)	(0.54,0.71)	(0.22,0.36)	(-0.004,-0.001)
7	estimates	$K=2$	0.12	0.45	0.6	0.3	0.014
	credibility	0.85	(0.11,0.13)	(0.33,0.63)	(0.5,0.67)	(0.24,0.38)	(0.013,0.015)
8	estimates	$K=2$	0.04	0.19	0.14	0.75	-0.015
	credibility	0.95	(0.03,0.05)	(0.18,0.21)	(0.11,0.18)	(0.71,0.78)	(-0.019,-0.011)
9	estimates	$K=2$	0.14	1.23	0.8	0.18	0.003
	credibility	0.8	(0.14,0.15)	(0.82,1.84)	(0.77,0.82)	(0.15,0.21)	(0.002,0.005)
10	estimates	$K=2$	0.13	0.73	0.68	0.28	-0.001
	credibility	0.8	(0.12,0.14)	(0.55,0.98)	(0.63,0.71)	(0.24,0.32)	(-0.002,0.000)

Table 7.3: Estimation results for FRAP curves for cell cycle early S: median and 90% credible interval

early S		K	b_1	b_2	a_1	a_2	a_0
1	estimates	$K=2$	0.08	0.39	0.7	0.19	0.035
	credibility	1	(0.08,0.08)	(0.29,0.53)	(0.66,0.73)	(0.16,0.22)	(0.034,0.037)
2	estimates	$K=2$	0.06	0.36	0.8	0.11	0.058
	credibility	0.95	(0.06,0.06)	(0.24,0.55)	(0.77,0.82)	(0.09,0.13)	(0.056,0.059)
3	estimates	$K=2$	0.07	0.45	0.78	0.18	0.018
	credibility	1	(0.07,0.07)	(0.36,0.56)	(0.75,0.8)	(0.15,0.2)	(0.016,0.019)
4	estimates	$K=2$	0.05	0.22	0.37	0.5	-0.005
	credibility	0.95	(0.05,0.06)	(0.19,0.26)	(0.3,0.43)	(0.44,0.56)	(-0.009,-0.003)
5	estimates	$K=2$	0.04	0.2	0.37	0.54	0.007
	credibility	0.95	(0.04,0.05)	(0.19,0.22)	(0.33,0.41)	(0.51,0.57)	(0.004,0.009)
6	estimates	$K=2$	0.06	0.25	0.6	0.3	0.039
	credibility	0.95	(0.06,0.06)	(0.22,0.31)	(0.56,0.64)	(0.27,0.34)	(0.038,0.041)
7	estimates	$K=2$	0.07	0.26	0.68	0.21	0.058
	credibility	0.95	(0.06,0.07)	(0.18,0.38)	(0.58,0.73)	(0.17,0.29)	(0.056,0.06)
8	estimates	$K=2$	0.09	0.31	0.66	0.29	0.036
	credibility	0.9	(0.08,0.09)	(0.24,0.39)	(0.57,0.72)	(0.23,0.37)	(0.034,0.037)
9	estimates	$K=2$	0.08	0.37	0.82	0.13	0.041
	credibility	1	(0.08,0.09)	(0.26,0.55)	(0.77,0.84)	(0.1,0.17)	(0.039,0.042)
10	estimates	$K=2$	0.07	0.32	0.76	0.16	0.045
	credibility	1	(0.07,0.07)	(0.22,0.46)	(0.71,0.79)	(0.13,0.2)	(0.044,0.047)

Table 7.4: Estimation results for FRAP curves for unconjugated GFP: median and 90% credible interval

GFP		K	b_1	b_2	a_1	a_2	a_0
1	estimates	$K=1$	0.94	0	0.83	0	-0.007
	credibility	1	(0.89,1)	–	(0.79,0.86)	–	(-0.009,-0.005)
2	estimates	$K=1$	1.26	0	0.92	0	0.005
	credibility	1	(1.19,1.34)	–	(0.88,0.96)	–	(0.003,0.007)
3	estimates	$K=1$	0.9	0	0.84	0	0.004
	credibility	1	(0.85,0.95)	–	(0.8,0.87)	–	(0.002,0.006)
4	estimates	$K=1$	0.94	0	0.91	0	0.008
	credibility	1	(0.89,0.99)	–	(0.87,0.94)	–	(0.006,0.01)
5	estimates	$K=1$	1.26	0	0.61	0	-0.011
	credibility	1	(1.15,1.39)	–	(0.57,0.66)	–	(-0.013,-0.009)
6	estimates	$K=1$	1.24	0	0.77	0	0.02
	credibility	1	(1.15,1.34)	–	(0.72,0.81)	–	(0.018,0.022)
7	estimates	$K=2$	0.2	1.47	0.08	0.79	-0.011
	credibility	0.95	(0.01,0.44)	(1.2,1.94)	(0.04,0.2)	(0.7,0.84)	(-0.077,-0.008)
8	estimates	$K=1$	1.01	0	0.94	0	-0.006
	credibility	1	(0.95,1.07)	–	(0.9,0.98)	–	(-0.008,-0.004)
9	estimates	$K=1$	0.73	0	0.89	0	0
	credibility	0.53	(0.7,0.77)	–	(0.86,0.93)	–	(-0.002,0.002)
10	estimates	$K=1$	1	0	0.83	0	0.012
	credibility	1	(0.94,1.06)	–	(0.8,0.87)	–	(0.010,0.015)

Table 7.5: Sensitivity analysis: Estimation results for G1 curves for different priors on b_1.

G1	Prior	(i)		(ii)		(iii)		(iv)	
		K	b_1	K	b_1	K	b_1	K	b_1
1	estimates	$K=1$	0.17	$K=1$	0.17	$K=1$	0.17	$K=1$	0.17
	credibility	1	(0.17,0.18)	1	(0.17,0.18)	1	(0.17,0.18)	1	(0.17,0.18)
2	estimates	$K=2$	0.16	$K=2$	0.16	$K=2$	0.16	$K=2$	0.16
	credibility	0.9	(0.14,0.17)	0.95	(0.12,0.17)	1	(0.13,0.17)	0.95	(0.12,0.17)
3	estimates	$K=2$	0.14	$K=2$	0.14	$K=2$	0.14	$K=2$	0.14
	credibility	0.9	(0.14,0.15)	1	(0.14,0.15)	1	(0.14,0.15)	0.95	(0.14,0.15)
4	estimates	$K=2$	0.17	$K=2$	0.17	$K=2$	0.17	$K=2$	0.17
	credibility	1	(0.14,0.18)	0.95	(0.15,0.18)	0.95	(0.14,0.18)	0.95	(0.15,0.18)
5	estimates	$K=2$	0.12	$K=2$	0.12	$K=2$	0.12	$K=2$	0.12
	credibility	0.95	(0.12,0.13)	0.95	(0.12,0.13)	0.9	(0.12,0.13)	0.85	(0.12,0.13)
6	estimates	$K=2$	0.12	$K=2$	0.12	$K=2$	0.12	$K=2$	0.12
	credibility	1	(0.11,0.13)	0.95	(0.11,0.13)	0.9	(0.11,0.13)	1	(0.11,0.13)
7	estimates	$K=2$	0.12	$K=2$	0.12	$K=2$	0.12	$K=2$	0.12
	credibility	0.85	(0.11,0.13)	0.95	(0.11,0.13)	0.95	(0.11,0.13)	0.95	(0.11,0.13)
8	estimates	$K=2$	0.04	$K=2$	0.04	$K=2$	0.04	$K=2$	0.04
	credibility	0.95	(0.03,0.05)	0.95	(0.03,0.05)	0.95	(0.03,0.05)	1	(0.03,0.05)
9	estimates	$K=2$	0.14	$K=2$	0.14	$K=2$	0.14	$K=2$	0.14
	credibility	0.8	(0.14,0.15)	0.9	(0.14,0.15)	0.9	(0.14,0.15)	0.85	(0.14,0.15)
10	estimates	$K=2$	0.13	$K=2$	0.13	$K=2$	0.13	$K=2$	0.13
	credibility	0.8	(0.12,0.14)	0.8	(0.12,0.13)	1	(0.12,0.13)	1	(0.12,0.13)

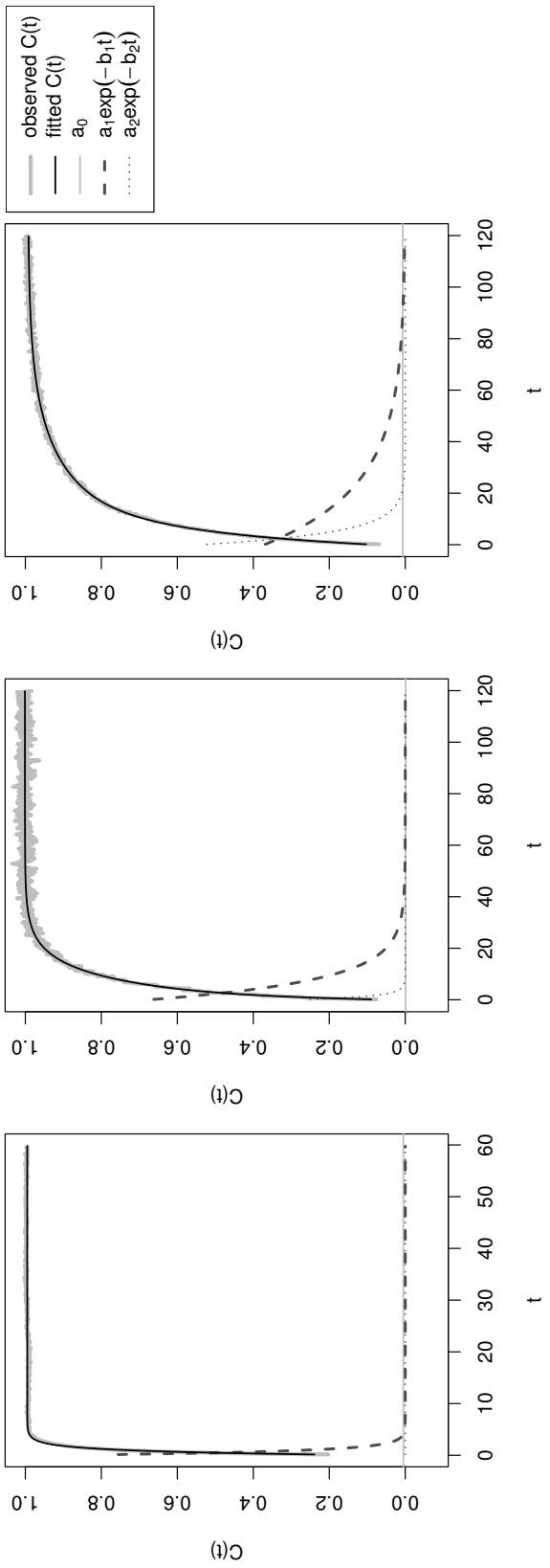

Figure 7.7: Original and estimated FRAP curves for unconjugated GFP (left), G1 phase (center), and early S phase (right)

7.5 Conclusion and discussion

We have proposed a RJMCMC algorithm for the nonlinear regression problem (7.2) arising from compartmental models. To this end, new birth and death transformations were developed. With this RJMCMC procedure, the number of compartments is determined from the observed data, the algorithm allowing to jump between models of different dimensions.

As the simulation study shows, a uniform prior on the exponents b_k leads to an overestimation of compartments. Models with lots of compartments are fitted with vanishing contribution of some of them or with some compartments having similar exchange rates and not being distinguishable. This can be explained by the redundancy of parameters in the sum of exponential model. Using a redundancy measure, we have designed a regularizing prior on the exchange rates. With this regularizing prior, the number of compartments is no longer overestimated as the fitted compartments are distinguishable. The algorithm chooses the right dimension for cases where the number of compartments is well-defined. In cases, where there is parameter redundancy and the number of compartments is not as clearly defined, model averaging (Hoeting et al., 1999) occurs.

For the FRAP experiments, for which the observed variable commonly is on a region of interest level (here the bleached part of the cell), we have proposed to use a regularizing prior on the binding rates. This makes sense also from a biological viewpoint as one is interested in finding binding classes that summarize binding partners with similar binding behavior. As the data observed in FRAP experiments are images, another option would be to perform the analysis on a pixel level—if the spatial resolution allows to do so. Then, a spatial regularization as proposed in Chapter 4 could be helpful.

The proposed algorithm allows to estimate the model complexity from the data. Of course, the estimated model complexity does not equal the "true" underlying physical model complexity. With the help of our simulated examples we have shown that a complex underlying model is not always detectable. Either, because there is parameter redundancy, which is a severe problem in nonlinear regression. Or, because there is too much noise. Thus, the model complexity is adequately estimated if non-redundant contributions are detected at reasonable noise.

Markov chain mixing is a problem in the RJMCMC algorithm when using the regularizing prior. This is due to the fact that there is no one-to-one mapping from a exponential curve to a sum of two exponentials. With the regularizing prior, mixing of the model dimension is poor. Though the algorithm provides the right dimension in cases where the number of compartments is clearly defined, the outcome depends on the starting value in cases where the model dimension is not as clearly defined as, *e.g.*, in parameter redundant models. We use parallel Markov chains to alleviate the dependence on starting values. Nevertheless, one should interpret the obtained model frequencies with care.

Constructing appropriate jump moves is known to be one of the main difficulties in the reversible jump methodology, even for simple linear nested regression models. Though RJMCMC is a generalization of the standard Metropolis-Hastings to general state spaces (Tierney, 1998), other challenges arise as for fixed dimensional MCMC. In fixed dimensional Metropolis-Hastings steps, one can always choose a random walk proposal and achieve

high acceptance rates by tuning the proposal's variance (Brooks et al., 2003). In variable dimensional MCMC, there is no corresponding natural choice of proposal that ensures high acceptance rates. Brooks et al. (2003) propose two methods to construct a more efficient reversible jump proposal: one uses the Taylor series expansion of the acceptance probabilities, the other one is a saturated space approach using data augmentation. Both methods help to adjust the proposal density of the auxiliary variables, here denoted as $q(u)$, and take the deterministic birth and death transformations, here $g_B()$ and $g_D()$, for granted. As there is no one-to-one mapping between sums of one or two exponentials, see Section 7.2.4, methods adjusting $q(u)$ will bring—if at all—little improvement but they will not cope with the intrinsic problems caused by the nonlinearity.

RJMCMC has been successfully applied to very distinct model selection problems, *e.g.* for DNA segmentation (Boys and Henderson, 2004) and for recovery/recapture data (King and Brooks, 2002). In those applications, the advantage of model selection with RJMCMC over model choice with information criteria is obvious as the number of competing models is large. For compartmental models, the number of competing models under consideration is rather small. In theory, model selection with RJMCMC has important advantages over model choice with information criteria, even for a small number of competing models: First, RJMCMC makes regularization possible and estimation more stable by enforcing simpler—and hence better interpretable—models for redundant parameter constellations. For model choice with information criteria, one would have to fit complex models, for which parameters may often not be estimable due to parameter redundancies. Second, with the model selection approach, the model uncertainty is captured and made visible with model frequencies and credibility intervals.

Summarizing, RJMCMC has theoretic merits that encouraged us to adapt the algorithm to the nonlinear regression problem arising in compartmental modeling. However, from our practical experience we cannot fully advise the use of RJMCMC for nonlinear regression problems.

Chapter 8

Summary and discussion

In this thesis, we have proposed and evaluated different approaches coping with the estimation and model selection in compartment models, after giving some theoretical background (Chapter 2) and motivating the need for more complex compartment models (Chapter 3). For the main application, a DCE-MRI breast cancer study, two hierarchical Bayesian models have been newly developed for parameter estimation in a two tissue compartment model on a voxel level: one with independent voxelwise priors (Chapter 4) and one with spatial priors on the kinetic parameters (Chapter 5). Allowing for two tissue compartments on a voxel level, the proposed models account for heterogeneity in tumor tissue. Comparing estimation results with the standard Tofts model with one tissue compartment, we found that additional model complexity is actually needed in tumorous tissue. For the DCE-MRI breast cancer data the 2Comp model outperforms the 1Comp model at tumor margins and in tumor surrounding tissue. Inside the tumor, the tissue is adequately described with the Tofts model.

We have analyzed how redundancy issues may occur in voxels of certain tumor regions, especially, inside the tumor. Using theoretical results on parameter redundancy and with the help of simulation studies, we have identified parameter constellations for which redundancy issues occur. For voxels with redundant parameters, we have observed flat or multimodal posterior distributions and small or negative p_D values. As sampling paths and the posterior distribution of parameters cannot be examined for every single voxel, parameter maps of the number of effective parameters, p_D, were suggested as helpful tool in evaluating redundancy issues and in measuring effective tissue heterogeneity (Chapters 4 and 5).

Evaluating simulation studies and *in vivo* DCE-MRI data, we have shown that assuming smoothness of the exponential rates is an effective—and from a biological perception reasonable—way of regularizing the parameter space in a two tissue compartment model (Chapter 5). With a spatial prior, implicit model selection is achieved as the contribution of one compartment is set to zero for voxels where the additional model complexity is not needed. Explicit model selection based on a model selection criterion like the DIC was also studied, however, one has to carefully take into account that p_D and DIC values are not meaningful for multimodal posteriors (see Chapter 4).

With the spatial elastic net, a completely different inference approach has been proposed in Chapter 6. For this penalized maximum likelihood based approach, the nonlinear regression model was linearized using a bundle of basis functions. We have developed a new, spatial version of the elastic net, penalizing for differences in parameters of neighboring voxels. Compared to an ordinary elastic net, with this spatially penalized approach, basis functions are selected more sparsely and, hence, model dimension and model parameters are robustly estimated per voxel from the data.

Finally, we have proposed a reversible jump MCMC algorithm for compartment models (Chapter 7). With this Bayesian model state approach, the number of compartments is flexibly treated as model parameter. For this approach, birth and death steps were designed which are suitable for a nonlinear regression problem. Furthermore, a regularizing prior was proposed to obtain compartments behaving different enough to be distinguishable. The usefulness of the proposed model space approach was discussed and analyzed in simulations and for data from a biochemical experiment. Due to the unfavorable mixing behavior, however, we were not completely satisfied with the performance of this approach.

Discussion on compartmental modeling

In compartmental modeling several assumptions are made. An assumption which is intrinsic in all compartment models is that compartments are homogeneous and well-mixed. If those assumptions are not fulfilled in a simple compartment model, more compartments with different kinetic properties may be needed in the model. When choosing the number of compartments from the data, one avoids strong divergence from those assumptions. Another assumption of the compartment models considered is that the rate of exchange remains constant over time. Though the exchange properties may vary in the long run, the assumption of constant exchange properties seems feasible for the short duration of the image acquisition. Furthermore, in compartment models one has to assume some initial condition. In DCE-MRI, for example, the concentration of contrast medium at time t_0, before injection of the contrast medium, is assumed to be zero. This assumption intuitively makes sense. However, as there may be some time delay from the time of injection to the first arrival of contrast medium in the imaged tissue, t_0 is not as clearly defined. Often, one pragmatically defines t_0—the enhancement onset time—to be the last time point before observing enhancement of contrast medium. As alternative, one can include a variable for the time delay in the model and, like this, account for the uncertainty in this parameter (Orton et al., 2007; Schmid et al., 2009).

Besides those intrinsic assumptions in compartment models, in the quantitative analysis of DCE-MRI, important assumptions are made regarding two important physical quantities: assumptions about the calculation of the contrast medium concentration and assumptions about the form of the arterial input function (AIF). First, one should be well aware that MRI cannot measure the concentration of contrast medium directly. MRI can only measure relaxation times. Only by assuming proportionality to the change in relaxation times with and without contrast medium, the concentration of contrast medium can be calculated (see equation (3.1) and Buckley and Parker (2005)). If the assumed relation

between contrast medium concentration and changes in relaxation times does not hold, the validity and interpretability of quantitative results will be strongly affected.

Second, in the analysis of the DCE-MRI data studied here, we had to assume a population based form of the AIF because the concentration of contrast medium in plasma could not be measured separately. Although common practice, this is a sensitive assumption. Methods that separately measure the AIF avoid this source of error (Fritz-Hansen et al., 1996). In case that separate measurement of the AIF is not possible, methods that simultaneously estimate the AIF along with other model parameters include this source of error and resulting uncertainties in the model, though not avoiding it (Yang et al., 2007; Fluckiger et al., 2009). Such approaches open a promising area for further research.

Alternative modeling approaches compete with compartmental modeling. One alternative to the compartmental modeling approaches considered in this thesis is modeling with Markov process models (Seber and Wild, 1989, p. 415 ff.). With the compartmental modeling approach used in this thesis, one assumes to observe a noisy version of some deterministic dynamic behavior. When modeling with Markov process models, the dynamic behavior of molecules is assumed to be stochastic and can be described by stochastic differential equations. For example, in Dargatz (2010) the state of a molecule in a FRAP experiment is modeled with the help of Markov jump processes and its diffusion approximation. Modeling with Markov processes is more flexible with regard to the error structure. The error in such a stochastic model may depend on the concentration and, hence, may vary with time. An error term varying with concentration and time is not captured in a deterministic dynamic model with additive Gaussian noise as assumed throughout this thesis. For the analyzed DCE-MRI and FRAP datasets we have, however, not experienced any difficulties with those restrictions on the error term.

Model-free approaches offer another alternative to modeling with compartment models. Such approaches assume a linear combination of nonlinear basis functions as predictor. Like this, one does not have to deal with a nonlinear regression problem. For example, Schmid et al. (2009) use Bayesian penalized B-splines for the analysis of DCE-MRI data. With model-free approaches, the observed data can be even more flexibly modeled. Those models do without any assumptions on the model architecture and on the number and properties of compartments. The disadvantage of those models is that there is no direct link between model parameters and biologically meaningful quantities. Using basis functions describing the behavior of a compartment, the elastic net approach presented in Chapter 6 offers a compromise between model-free and model-based approaches.

Discussion on methods

Different inference approaches and ways to deal with parameter estimation and model uncertainty have been considered in this thesis. Here, we highlight the differences of the approaches studied and discuss their strength and weaknesses.

First, the approaches differ in their flexibility with respect to the model dimension. While the spatial elastic net approach and the RJMCMC model state approach allow for a completely flexible model dimension ($K = 1, \ldots, K_{max}$), the model dimension is fixed for

the two tissue compartment model ($K = 2$). In the spatial two tissue compartment model, however, the model dimension is implicitly reduced in voxels where only one compartment is needed, as the estimated contribution of one compartment is close to zero. The contribution of a third tissue compartment would, however, be missed for a fixed model dimension of $K = 2$. Yet, we have experienced that the contribution of a third tissue compartment rarely brings considerable improvement of fit. When constructing simulation studies, it was not easy to find parameter constellations of a three tissue compartment model that cannot properly be described by a two tissue compartment model.

Second, the approaches differ in the way they deal with the nonlinearity of the model. For the elastic net approach, the nonlinear regression problem was linearized (see equation (6.1)). All other approaches proposed here work with the nonlinear regression problem without linearization. The main advantage of working with the linearized problem is that well established estimation and model selection techniques for linear models can be applied. However, in a linearized model, parameters have to be restricted to lie on a predefined grid. This may lead to imprecise estimation results, especially, if the grid is defined too coarse or if the grid covers an inappropriate range of parameter values. In the nonlinear regression model, in contrast, the parameters are continuous variables. However, estimation in nonlinear regression may be challenging due to redundancy issues, and there are few estimation methods which have been designed for the specific needs of estimation in those models. This thesis, therefore, contributes with evaluating and developing methods in this field.

Last, the approaches differ in the underlying principles employed for inference. While most of the methods presented in this thesis are based on Bayesian inference, the elastic net approach is based on the concept of penalized maximum likelihood estimation. There is an uncompromising debate between "frequentists" and "Bayesians" about if and in what way prior knowledge may be incorporated in a statistical model. Considering that for a large class of priors in Bayesian models there exists a one-to-one correspondence to some L_1 or L_2 penalty in the penalized maximum likelihood setting, this controversy is put into perspective. In case of a one-to-one correspondence, the point estimates and its dependence on priors respectively penalization are comparable. For instance, there are Bayesian alternatives for the common penalized maximum likelihood based approaches in the linear regression problem: The Bayesian regression with Gaussian prior corresponds to ridge regression. Similarly, the lasso and the elastic net have their Bayesian correspondents. In a Bayesian model, a Gaussian prior corresponds to a L_2-penalty and a Laplace prior to a L_1-penalty (Park and Casella, 2008; Li and Lin, 2010; Hastie et al., 2009).

What really is different for Bayesian and maximum likelihood based inference is the algorithms for parameter estimation and, most notably, the interpretation of estimation results. As discussed for the spatial elastic net approach in Chapter 6, the computational burden of parameter estimation in a penalized maximum likelihood based approach is considerably lower compared to a Bayesian approach. Generally, Bayesian inference requires MCMC simulations and is, hence, computationally more demanding. In structured additive regression models, MCMC sampling can be avoided as good approximations of the marginal posteriors can be gained using an integrated nested Laplace approximation (Rue

et al., 2009). However, for a nonlinear regression model such an approximation is not feasible.

Concerning the interpretation of estimation results, the inference approaches differ considerably. While least squares or maximum likelihood based inference typically provides point estimates and confidence regions for those estimates, Bayesian inference is based on the posterior distribution of the parameters. The posterior contains a lot of information that helps to understand how adequate and reliable parameter estimates are. When developing adequate regularizing priors for parameter redundant constellations within the scope of this thesis, looking at the sampling paths and marginal posteriors of the parameters was very helpful for getting an intuition if things are working and—if they are not—why they are not working.

The Bayesian approach is flexible enough to treat the model dimension itself as unknown parameter. Like this, the Bayesian approach allows to explicitly account for model uncertainty. When estimating the model complexity, the results of a penalized maximum likelihood based approach is a point estimate for the model dimension (maximizing the penalized likelihood), whereas, in a Bayesian model state approach one obtains the posterior model probabilities of all candidate models. Furthermore, the Bayesian approach offers a framework for model averaging (Hoeting et al., 1999). Correspondingly, in a maximum likelihood based approach, model uncertainties and model averaging can be accounted for using Akaike weights as proposed in Turkheimer et al. (2003) for model averaging in PET compartmental modeling. However, for models which are nonlinear in parameters, average model parameters—calculated as linear combination of model parameters as in Turkheimer et al. (2003)—are not meaningful. We would, therefore, not recommend this kind of model averaging in compartment models.

The methods proposed in this thesis may be helpful in different applications using compartmental modeling. Bayesian spatial regularization and spatially penalized maximum likelihood estimation offer a solution to parameter redundancy issues that can be applied in other fields as well, *e.g.*, in the quantitative analysis of PET and SPECT images. In PET and SPECT neuroreceptor imaging studies, the kinetics of ligand uptake in the brain is described with the help of compartment models (Slifstein and Laruelle, 2001). The data structure and the compartment models used in the analysis of PET and SPECT are comparable to those used in the analysis of DCE-MRI. When estimating receptor parameters, one copes with similar identifiability issues encountered in DCE-MRI analysis.

Discussion on application results

In applications using compartment models such as, for example, the analysis of DCE-MRI, commonly used models may be too simplistic to adequately describe the observed behavior. Thus, the need for more complex models arises. Generally, users prefer to use simple models they are familiar with, and they are not easily convinced to use more complex models. Therefore, the proposed 2Comp model for the analysis of DCE-MRI data is unlikely to being used by physicians in the day-to-day routine of cancer diagnosis and treatment in

the near future. We are, however, convinced that this thesis and the proposed models with increased model complexity give important insights to heterogeneity in tumorous tissue.

Analyzing the heterogeneity of enhancement patterns in tumors is considered the key factor for understanding the mechanisms of cancer growth and for the development of new treatment approaches (Yang and Knopp, 2011). In this thesis, we have proposed two ways of quantifying tissue heterogeneity. In the Bayesian framework, the number of effective parameters, p_D, measures model complexity and, hence, quantifies the degree of heterogeneity (Chapters 4 and 5). In the elastic net approach, the number of selected compartments reflects the degree of heterogeneity (Chapter 6). Both, p_D and the number of selected compartments, $\hat{q}$, were calculated per voxel and are easily visualized with parametric maps. Like this, a map of heterogeneity is obtained. Our results about tissue heterogeneity for both estimation approaches, though being conceptually very different, are consistent.

The results obtained for the DCE-MRI breast cancer study confirm conjectures about increased heterogeneity in cancerous tissue. Typically, there is a tumor rim showing increased heterogeneity. In the tumor core, in contrast, one often observes homogeneous, poorly perfused tissue. Such poorly perfused regions are assumed to develop during a phase of fast tumor growth, when tumor cells proliferate faster than angiogenesis (Yang and Knopp, 2011). With studies of DCE-MRI sequences acquired more frequently, perhaps even daily, the dynamics and pattern of tumor growth could be further exploited. Likewise, it would be interesting to further evaluate the potential of measuring tumor heterogeneity with more extensive clinical studies including patients with different kinds of benign and malignant tumors. Different heterogeneity patterns could then be used to classify tumor type and disease progression.

In the analysis of FRAP experiments, the estimation of the model complexity from the data may bring important insights about the binding behavior of proteins (Chapter 7). Like this, one can learn about the protein mobility in living cells and about how the binding behavior of proteins regulates cellular processes. In Chapter 7, inference was done on a region of interest level and, hence, estimated binding rates are average binding rates. Extending the approach to a spatial model and allowing to estimate binding rates and the number of binding partners on a voxel level would bring additional insight about the spatial distribution of binding partners in the nucleus. Such an extension is, however, challenging due to the low spatial resolution of FRAP images.

Technical progress will most likely bring improvement of spatial and temporal resolution of imaging devices, both for medical and biological imaging applications. Consequently, observed concentration time curves are likely to have higher signal to noise ratios in the future. Looking ahead, this development will generally facilitate inference in more and more complex compartment models and will, hence, allow to describe underlying processes in more detail. However, some of the estimation challenges which are due to parameter redundancy issues may occur even for low noise levels. Therefore, suitable regularization will remain an important issue even with increasing signal to noise ratios.

Appendix A

Calculating redundancy in sums of (convolved) exponentials

A.1 Convolved exponentials

A.1.1 Bi-exponential input function

Assuming a bi-exponential arterial input function

$$C_p(t) = a_1 \exp(-m_1 t) + a_2 \exp(-m_2 t)$$

and denoting

$$E(t) = \exp(-bt),$$

the convolution is

$$\begin{aligned}
(C_p * E)(t) &= \int_0^t C_p(\tau)\exp(-b(t-\tau))d\tau \\
&= \int_0^t \left(a_1\exp(-m_1\tau) + a_2\exp(-m_2\tau)\right)\exp(-bt)\exp(b\tau)d\tau \\
&= \exp(-bt)\int_0^t \left(a_1\exp((b-m_1)\tau) + a_2\exp((b-m_2)\tau)\right)d\tau \\
&= \exp(-bt)\left[\left(\tfrac{a_1}{(b-m_1)}\exp((b-m_1)\tau) + \tfrac{a_2}{(b-m_2)}\exp((b-m_2)\tau)\right)\right]_{\tau=0}^{t} \\
&= \exp(-bt)\left\{\tfrac{a_1}{(b-m_1)}\exp((b-m_1)t) + \tfrac{a_2}{(b-m_2)}\exp((b-m_2)t)\right. \\
&\quad \left.\cdots - \tfrac{a_1}{(b-m_1)} - \tfrac{a_2}{(b-m_2)}\right\} \\
&= \tfrac{a_1}{(b-m_1)}\left(\exp((b-m_1)t)\exp(-bt) - \exp(-bt)\right) \\
&+ \tfrac{a_2}{(b-m_2)}\left(\exp((b-m_2)t)\exp(-bt) - \exp(-bt)\right) \\
&= \tfrac{a_1}{b-m_1}\left(\exp(-m_1 t) - \exp(-bt)\right) + \tfrac{a_2}{b-m_2}\left(\exp(-m_2 t) - \exp(-bt)\right) \\
&= \tfrac{a_1}{m_1-b}\left(\exp(-bt) - \exp(-m_1 t)\right) + \tfrac{a_2}{m_2-b}\left(\exp(-bt) - \exp(-m_2 t)\right).
\end{aligned}$$

A.1.2 Constant input function

Assuming a constant input function, $C_B(t) \equiv 1$, and denoting $E(t) = \exp(-bt)$, the convolution with an exponential simplifies to

$$\begin{aligned}(C_B * E)(t) &= \int_0^t \exp(-b(t-\tau))d\tau \\ &= \exp(-bt)\int_0^t \exp(b\tau)d\tau \\ &= \exp(-bt)\left[\tfrac{1}{b}\exp(b\tau)\right]_0^t \\ &= \tfrac{1}{b} - \tfrac{1}{b}\exp(-bt).\end{aligned}$$

A.2 Redundancy measure in sum of exponentials

In Reich (1981) the sensitivity matrix $M_{\boldsymbol{\theta}}$, the scaled sensitivity matrix $R_{\boldsymbol{\theta}}$, and a closed form solution of its determinant $|R_{\boldsymbol{\theta}}|$ were calculated for a sum of two exponentials model in order to measure parameter redundancy. Here, we calculate the elements of $M_{\boldsymbol{\theta}}$ and $R_{\boldsymbol{\theta}}$ for a general sum of K exponentials model. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{2K}) = (a_1, \ldots, a_K, b_1, \ldots, b_K)$ and

$$f(t, \boldsymbol{\theta}) = \sum_{i=1}^{K} a_i \exp(-b_i t).$$

Then the matrices' elements can be calculated in four blocks:
For $1 \le i, j \le K$

$$\begin{aligned}m_{ij} &= \int_0^\infty \tfrac{\delta f}{\delta a_i}\tfrac{\delta f}{\delta a_j}dt \\ &= \int_0^\infty \exp(-(b_i+b_j)t)dt \\ &= \left[-\tfrac{1}{b_i+b_j}\exp(-(b_i+b_j)t)\right]_0^\infty \\ &= \tfrac{1}{b_i+b_j} \\ r_{ij} &= \tfrac{2\sqrt{b_i b_j}}{b_i+b_j}.\end{aligned}$$

For $1 \le i \le K$, $K+1 \le j \le 2K$ $(j' = j - K)$

$$\begin{aligned}m_{ij} &= \int_0^\infty \tfrac{\delta f}{\delta a_i}\tfrac{\delta f}{\delta b_{j'}}dt \\ &= \int_0^\infty \exp(-b_i t)a_{j'}(-t)\exp(-b_{j'}t)dt \\ &= -a_{j'}\int_0^\infty \exp(-(b_i+b_{j'})t)dt \\ &= -a_{j'}\left(\left[t\tfrac{\exp(-(b_i+b_{j'})t)}{-(b_i+b_{j'})}\right]_0^\infty + \tfrac{1}{(b_i+b_{j'})}\int_0^\infty \exp(-(b_i+b_{j'})t)dt\right) \\ &= -\tfrac{a_{j'}}{(b_i+b_{j'})^2} \\ r_{ij} &= -\tfrac{2b_{j'}\sqrt{2b_i b_{j'}}}{(b_i+b_{j'})^2}.\end{aligned}$$

For $K+1 \le i \le 2K$, $1 \le j \le K$

$$\begin{aligned}m_{ij} &= m_{ji} \\ r_{ij} &= r_{ji}.\end{aligned}$$

For $K+1 \leq i,j \leq 2K$ ($i^{'} = i-K$ and $j^{'} = j-K$)

$$\begin{aligned} m_{ij} &= \int_0^\infty \frac{\delta f}{\delta b_{i'}} \frac{\delta f}{\delta b_{j'}} dt \\ &= \int_0^\infty a_{i'}(-t)\exp(-b_{i'}t) a_{j'}(-t)\exp(-b_{j'}t) dt \\ &= a_{i'}a_{j'}\int_0^\infty t^2 \exp(-(b_{i'}+b_{j'})t) dt \\ &= \vdots \\ &= \frac{a_{i'}a_{j'}}{(b_{i'}+b_{j'})^3} \\ r_{ij} &= \left(\frac{2b_{i'}b_{j'}}{b_{i'}+b_{j'}}\right)^3 . \end{aligned}$$

For $K = 2$, elaborate matrix manipulations yield a closed form solution for the determinant of the scaled sensitivity matrix (Reich, 1981)

$$|R_{\boldsymbol{\theta}}| = \frac{1}{4}\left(\frac{1-\frac{b_2}{b_1}}{1+\frac{b_2}{b_1}}\right)^8 .$$

Reich (1981) proposes to judge parameters redundant using a threshold value $|R_{\boldsymbol{\theta}}|^{-1} \geq 100$. Here, this is equivalent to $\frac{b_2}{b_1} \geq 0.2$ or $b_1 \leq 5b_2$. This means that in a sum of two exponentials parameters are judged redundant if the decay rates differ by less than a factor of five. Interestingly, the redundancy measure does not depend on the parameters a_1 and a_2. Table A.1 shows $|R_{\boldsymbol{\theta}}|^{-1}$ for different parameter constellations in a sum of two exponentials model ($K = 2$).

For more complex models the derivation of a closed form solution for the determinant $|R_{\boldsymbol{\theta}}|$ is not feasible. In those cases, one can calculate the determinant numerically. We have implemented the calculation of $M_{\boldsymbol{\theta}}$, $R_{\boldsymbol{\theta}}$, and $|R_{\boldsymbol{\theta}}|^{-1}$ in R. Table A.2 displays the redundancy measure for a sum of three exponentials model ($K = 3$). Alike in the case of $K = 2$, the redundancy measure depends only on the exponential rates b_1, b_2, and b_3, on their relation among each other, and not on the values of a_1, a_2, and a_3. Given $b_1 = 0.05$ and $b_2 = 5$, parameters are judged redundant ($|R_{\boldsymbol{\theta}}|^{-1} \geq 100$) if b_3 and b_2 differ less then a factor of six. Given that $b_1 = 0.2$ and $b_2 = 5$, b_3 and b_2 have to differ by at least a factor of nine for the model parameters to be judged non-redundant.

Table A.1: Redundancy measure $|R_{\boldsymbol{\theta}}|^{-1}$ for $K = 2$, $b_2 = 5$ fixed and $b_1 = c * b_2$ varying. Using a threshold of 100 parameters are redundant for $b_1 \geq 1$ given that $b_2 = 5$.

b_1	$c = \frac{b_1}{b_2}$	$\lvert R_{\boldsymbol{\theta}}\rvert^{-1}$
5	1.00	Inf
4.5	0.90	67934252567.87
4	0.80	172186884.07
3.5	0.70	4252862.33
3	0.60	262144.00
2.5	0.50	26244.00
2	0.40	3514.59
1.5	0.30	566.01
1	0.20	102.52
0.5	0.10	19.92
0.05	0.01	4.69
0.005	0.00	4.06

Table A.2: Redundancy measure $|R_{\boldsymbol{\theta}}|^{-1}$ for $K = 3$, b_1, b_2 fixed and $b_3 = c * b_2$ varying. Left: $b_1 = 0.05$ and $b_2 = 5$. Right: $b_1 = 0.2$ and $b_2 = 5$.

$b_1 = 0.05$, $b_2 = 5$		
b_3	$c = \frac{b_3}{b_2}$	$\lvert R_{\boldsymbol{\theta}}\rvert^{-1}$
5	1	Inf
10	2	66725.73
15	3	2535.02
20	4	581.76
25	5	248.43
30	6	142.29
35	7	95.94
40	8	71.52
45	9	56.96
50	10	47.50
55	11	40.96
60	12	36.21

$b_1 = 0.2$, $b_2 = 5$		
b_3	$c = \frac{b_3}{b_2}$	$\lvert R_{\boldsymbol{\theta}}\rvert^{-1}$
5	1	Inf
10	2	137135.42
15	3	4809.30
20	4	1060.39
25	5	442.09
30	6	249.19
35	7	166.11
40	8	122.77
45	9	97.13
50	10	80.57
55	11	69.17
60	12	60.92

A.3 Redundancy measure in sum of convolved exponentials

The calculation of the sensitivity matrix $M_{\boldsymbol{\theta}}$ is much more complicated for a sum of convolved exponentials needed in the analysis of DCE-MRI (Chapters 3–6). This is because for convolved exponentials the partial derivatives and its integrals are more complex. Consider a model with one convolved exponential

$$f(t,\boldsymbol{\theta}) = vbC_p(t) * \exp(-bt) = vb\left(C_p * E\right)(t)$$

with $\boldsymbol{\theta} = (v,b)$ and $E(t) = \exp(-bt)$.

The partial derivatives of $f(t,\boldsymbol{\theta})$ are

$$\begin{aligned}\tfrac{df}{dv} &= \tfrac{d}{dv}\left(vbC_p(t) * \exp(-bt)\right)\\ &= bC_p(t) * \exp(-bt)\\ &= b\left(C_p * E\right)(t)\end{aligned}$$

and

$$\begin{aligned}\tfrac{df}{db} &= \tfrac{d}{db}\left(vbC_p(t) * \exp(-bt)\right)\\ &= vC_p(t) * \exp(-bt) + vb\underbrace{\frac{d}{db}\left(C_p(t) * \exp(-bt)\right)}_{:=I(t)}.\end{aligned}$$

With

$$\begin{aligned}I(t) &= \tfrac{d}{db}\left(C_p(t) * \exp(-bt)\right)\\ &= \tfrac{d}{db}\int_0^t C_p(\tau)\exp(-b(t-\tau))d\tau\\ &= \int_0^t C_p(\tau)\tfrac{d}{db}\exp(-b(t-\tau))d\tau \text{ (apply Leibniz rule)}\\ &= \int_0^t C_p(\tau)(-1)(t-\tau)\exp(-b(t-\tau))d\tau\\ &= \int_0^t C_p(\tau)(\tau-t)\exp(b(\tau-t))d\tau\\ &= \underbrace{\int_0^t C_p(\tau)\tau\exp(b(\tau-t))d\tau}_{I_1(t)} - \underbrace{t\left(C_p * E\right)(t)}_{I_2(t)}\\ &= I_1(t) - I_2(t)\end{aligned}$$

this simplifies to

$$\begin{aligned}\tfrac{df}{db} &= v\left(C_p * E\right)(t) + vbI(t)\\ &= v\left(C_p * E\right)(t) + vb\left[I_1(t) - I_2(t)\right]\\ &= v\left(C_p * E\right)(t) + vb\left[I_1(t) - t\left(C_p * E\right)(t)\right]\\ &= v\left(1-bt\right)\left(C_p * E\right)(t) + vbI_1(t).\end{aligned}$$

The product of the partial derivatives is

$$\begin{aligned}\tfrac{df}{dv}\tfrac{df}{db} &= b\left(C_p * E\right)(t)\left\{v\left(1-bt\right)\left(C_p * E\right)(t) + vbI_1(t)\right\}\\ &= bv\left(1-bt\right)\left[\left(C_p * E\right)(t)\right]^2 + vb^2\left(C_p * E\right)(t)I_1(t).\end{aligned}$$

Then, the elements of $M_{\boldsymbol{\theta}}$ can be expressed with the help of several complicated integrals

$$\begin{aligned} m_{11} &= \int_0^\infty \tfrac{df(t)}{dv}\tfrac{df(t)}{dv}dt \\ &= b^2 \int_0^\infty (C_p * E(t))^2\, dt \end{aligned}$$

and

$$\begin{aligned} m_{12} = m_{21} &= bv \underbrace{\int_0^\infty \left[(C_p * E)(t)\right]^2 dt}_{=m_{11}/b^2} \\ &- b^2 v \underbrace{\int_0^\infty t\left[(C_p * E)(t)\right]^2 dt}_{:=I_3} \\ &+ vb^2 \underbrace{\int_0^\infty (C_p * E)(t) I_1(t) dt}_{:=I_4} \\ &= \tfrac{v}{b} m_{11} - b^2 v I_3 + b^2 v I_4 \end{aligned}$$

and

$$\begin{aligned} m_{22} &= \int_0^\infty \tfrac{df(t)}{db}\tfrac{df(t)}{db}dt \\ &= \int_0^\infty \left\{v(1-bt)(C_p * E)(t) + vbI_1(t)\right\}^2 dt \\ &= v^2 \int_0^\infty (1-bt)^2 \left((C_p * E)(t)\right)^2 + 2b(1-bt)(C_p * E)(t) I_1(t) + b^2 I_1^2(t) dt \end{aligned}$$

$$\begin{aligned} &= v^2 \Bigg\{ \underbrace{\int_0^\infty \left((C_p * E)(t)\right)^2 dt}_{m_{11}/b^2} - 2b \underbrace{\int_0^\infty t\left((C_p * E)(t)\right)^2 dt}_{=I_3} \\ &+ b^2 \underbrace{\int_0^\infty t^2 \left((C_p * E)(t)\right)^2 dt}_{I_5} + 2b \underbrace{\int_0^\infty (C_p * E)(t) I_1(t) dt}_{I_4} \\ &- 2b^2 \underbrace{\int_0^\infty t (C_p * E)(t) I_1(t) dt}_{I_6} + b^2 \underbrace{\int_0^\infty I_1^2(t) dt}_{I_7} \Bigg\} \\ &= v^2 \left\{ m_{11}/b^2 - 2bI_3 + b^2 I_5 + 2bI_4 - 2b^2 I_6 + b^2 I_7 \right\}. \end{aligned}$$

Closed form solutions for all integrals needed to calculate m_{11}, m_{12}, m_{21}, and m_{22} are derived in Section A.3.1. Based on those, the calculation of $M_{\boldsymbol{\theta}}$, $R_{\boldsymbol{\theta}}$, and $|R_{\boldsymbol{\theta}}|$ was implemented in R. This allows to calculate the redundancy measure for a given arterial input function and given parameter values for v and b.

Our numerical calculations of the redundancy measure $|R_{\boldsymbol{\theta}}|^{-1}$ indicate that there is typically no parameter redundancy in a sum of one convolved exponential model. However, if b takes values very close to the parameters m_1, and m_2 of the arterial input function, parameter redundancy may occur. Similar to the case of a simple exponential, for the

convolved exponential the redundancy measure $|R_{\boldsymbol{\theta}}|^{-1}$ does not depend on values of v but only on values of the exponential rate b.

For a sum of two or more convolved exponentials the derivation of all partial derivatives and its integrals seems not feasible. From the results of a sum of two (simple) exponentials (Section A.2) we expect that similar or slightly increased redundancy issues occur for a sum of two convolved exponentials. This is also what we have experienced in simulation studies (see Chapter 4).

A.3.1 Calculation of integrals

Here, we calculate the integrals needed to obtain $M_{\boldsymbol{\theta}}$. All integrals can be expressed as functions of t, v, b and the parameters a_1, a_2, m_1, and m_2 of the arterial input function

$$C_p(t) = a_1 \exp(-m_1 t) + a_2 \exp(-m_2 t).$$

We will use

$$\begin{aligned} \int_0^\infty \exp(-\beta t)dt &= \left[-\frac{\exp(-\beta t)}{\beta}\right]_{t=0}^{\infty} \\ &= 0 - (-\frac{\exp(0)}{\beta}) = \frac{1}{\beta}, \end{aligned} \tag{A.1}$$

$$\begin{aligned} \int_0^\infty t\exp(-\beta t)dt &= \left[-\frac{\exp(-\beta t)(\beta t+1)}{\beta^2}\right]_{t=0}^{\infty} \\ &= 0 - (-\frac{\exp(0)(0+1)}{\beta^2}) = \frac{1}{\beta^2}, \end{aligned} \tag{A.2}$$

and

$$\begin{aligned} \int_0^\infty t^2\exp(-\beta t)dt &= \left[\frac{\exp(-\beta t)(-\beta t(\beta t+2)-2)}{\beta^3}\right]_{t=0}^{\infty} \\ &= 0 - (\frac{\exp(0)(0(0+2)-2)}{\beta^3}) = \frac{2}{\beta^3}. \end{aligned} \tag{A.3}$$

Let us first calculate

$$
\begin{aligned}
m_{11} &= \int_0^\infty \tfrac{df(t)}{dv}\tfrac{df(t)}{dv}dt \\
&= b^2\int_0^\infty \left(C_p * E(t)\right)^2 dt \\
&= b^2\int_0^\infty \Big(\tfrac{a_1}{b-m_1}\left(\exp(-m_1 t)-\exp(-bt)\right) \\
&+ \tfrac{a_2}{b-m_2}\left(\exp(-m_2 t)-\exp(-bt)\right)\Big)^2 dt \\
\\
&= b^2\int_0^\infty \left(\tfrac{a_1}{b-m_1}\right)^2\left(\exp(-m_1 t)-\exp(-bt)\right)^2 \\
&+ 2\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left(\exp(-m_1 t)-\exp(-bt)\right)\left(\exp(-m_2 t)-\exp(-bt)\right) \\
&+ \left(\tfrac{a_2}{b-m_2}\right)^2\left(\exp(-m_2 t)-\exp(-bt)\right)^2 dt \\
\\
&= b^2\int_0^\infty \left(\tfrac{a_1}{b-m_1}\right)^2\left(\exp(-2m_1 t)-2\exp(-(m_1+b)t)+\exp(-2bt)\right) \\
&+ 2\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left\{\exp(-(m_1+m_2)t)-\exp(-(m_1+b)t)\right. \\
&\quad \left.\cdots-\exp(-(m_2+b)t)+\exp(-2bt)\right\} \\
&+ \left(\tfrac{a_2}{b-m_2}\right)^2\left(\exp(-2m_2 t)-2\exp(-(m_2+b)t)-\exp(-2bt)\right)dt \\
\\
&= b^2\left\{\left(\tfrac{a_1}{b-m_1}\right)^2\left[-\tfrac{1}{2m_1}\exp(-2m_1 t)+\tfrac{2}{(m_1+b)}\exp(-(m_1+b)t)\right.\right. \\
&\quad \left.\cdots-\tfrac{1}{2b}\exp(-2bt)\right]_{t=0}^{\infty} \\
&+ 2\,\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left[-\tfrac{1}{(m_1+m_2)}\exp(-(m_1+m_2)t)+\tfrac{1}{(m_1+b)}\exp(-(m_1+b)t)\right. \\
&\quad \left.\cdots+\tfrac{1}{(m_2+b)}\exp(-(m_2+b)t)-\tfrac{1}{2b}\exp(-2bt)\right]_{t=0}^{\infty} \\
&+ \left(\tfrac{a_2}{b-m_2}\right)^2\left[-\tfrac{1}{2m_2}\exp(-2m_2 t)+\tfrac{2}{(m_2+b)}\exp(-(m_2+b)t)\right. \\
&\quad \left.\left.\cdots-\tfrac{1}{2b}\exp(-2bt)\right]_{t=0}^{\infty}\right\} \\
\\
&= b^2\left\{\left(\tfrac{a_1}{b-m_1}\right)^2\left[\tfrac{1}{2m_1}-\tfrac{2}{(m_1+b)}+\tfrac{1}{2b}\right]\right. \\
&+ 2\,\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left[\tfrac{1}{(m_1+m_2)}-\tfrac{1}{(m_1+b)}-\tfrac{1}{(m_2+b)}+\tfrac{1}{2b}\right] \\
&+ \left.\left(\tfrac{a_2}{b-m_2}\right)^2\left[\tfrac{1}{2m_2}-\tfrac{2}{(m_2+b)}+\tfrac{1}{2b}\right]\right\}.
\end{aligned}
$$

For the calculation of m_{12} and m_{21} we need to calculate

$$
\begin{aligned}
I_1 &= \exp(-bt)\int_0^t \tau C_p(\tau)\exp(b\tau)d\tau \\
&= \exp(-bt)\int_0^t \tau\left(a_1\exp(-m_1\tau)+a_2\exp(-m_2\tau)\right)\exp(b\tau)d\tau \\
&= \exp(-bt)\int_0^t \tau\left(a_1\exp((b-m_1)\tau)+a_2\exp((b-m_2)\tau)\right)d\tau \\
&= \exp(-bt)\left[\frac{a_1\exp((b-m_1)\tau)((b-m_1)\tau-1)}{(b-m_1)^2}+\frac{a_2\exp((b-m_2)\tau)((b-m_2)\tau-1)}{(b-m_2)^2}\right]_{\tau=0}^{t} \\
&= \exp(-bt)\left[\frac{a_1\exp((b-m_1)t)((b-m_1)t-1)}{(b-m_1)^2}+\frac{a_2\exp((b-m_2)t)((b-m_2)t-1)}{(b-m_2)^2}\right. \\
&\quad \left.\cdots+\frac{a_1}{(b-m_1)^2}+\frac{a_2}{(b-m_2)^2}\right] \\
&= t\left[\frac{a_1}{b-m_1}\exp(-m_1t)+\frac{a_2}{b-m_2}\exp(-m_2t)\right] \\
&+ \frac{a_1}{(b-m_1)^2}\left(\exp(-bt)-\exp(-m_1t)\right) \\
&+ \frac{a_2}{(b-m_2)^2}\left(\exp(-bt)-\exp(-m_2t)\right)
\end{aligned}
$$

using $\int t\exp(\beta t)dt = \frac{\exp(\beta t)(\beta t-1)}{\beta^2}$.

Next,

$$\begin{aligned}
I_3 &= \int_0^\infty t\left[(C_p * E)(t)\right]^2 dt \\
&= \int_0^\infty t\left(\tfrac{a_1}{b-m_1}\left(\exp(-m_1 t) - \exp(-bt)\right)\right. \\
&+ \left.\tfrac{a_2}{b-m_2}\left(\exp(-m_2 t) - \exp(-bt)\right)\right)^2 dt \\
&= \int_0^\infty t\left(\tfrac{a_1}{b-m_1}\right)^2 \left\{\exp(-m_1 t) - \exp(-bt)\right\}^2 \\
&+ 2t\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left\{\exp(-m_1 t) - \exp(-bt)\right\}\left\{\exp(-m_2 t) - \exp(-bt)\right\} \\
&+ t\left(\tfrac{a_2}{b-m_2}\right)^2 \left\{\exp(-m_2 t) - \exp(-bt)\right\}^2 dt \\
&= \int_0^\infty t\left(\tfrac{a_1}{b-m_1}\right)^2 \left\{\exp(-2m_1 t) - 2\exp(-(m_1+b)t) + \exp(-2bt)\right\} \\
&+ 2t\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left\{\exp(-(m_1+m_2)t) - \exp(-(m_1+b)t)\right. \\
&\quad \left.\cdots - \exp(-(m_2+b)t) + \exp(-2bt)\right\} \\
&+ t\left(\tfrac{a_2}{b-m_2}\right)^2 \left\{\exp(-2m_2 t) - 2\exp(-(m_2+b)t) - \exp(-2bt)\right\} dt \\
&= \left(\tfrac{a_1}{b-m_1}\right)^2 \left[\tfrac{1}{(2m_1)^2} - \tfrac{2}{(m_1+b)^2} + \tfrac{1}{(2b)^2}\right] \\
&+ 2\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left[\tfrac{1}{(m_1+m_2)^2} - \tfrac{1}{(m_1+b)^2} - \tfrac{1}{(m_2+b)^2} + \tfrac{1}{(2b)^2}\right] \\
&+ \left(\tfrac{a_2}{b-m_2}\right)^2 \left[\tfrac{1}{(2m_2)^2} - \tfrac{2}{(m_2+b)^2} + \tfrac{1}{(2b)^2}\right]
\end{aligned}$$

using equation (A.2) and correspondingly

$$\begin{aligned}
I_5 &= \int_0^\infty t^2\left[(C_p * E)(t)\right]^2 dt \\
&= \int_0^\infty t^2\left(\tfrac{a_1}{b-m_1}\left(\exp(-m_1 t) - \exp(-bt)\right)\right. \\
&+ \left.\tfrac{a_2}{b-m_2}\left(\exp(-m_2 t) - \exp(-bt)\right)\right)^2 dt \\
&= \ldots \text{ same steps as for } I_3 \\
&= \left(\tfrac{a_1}{b-m_1}\right)^2 \left[\tfrac{2}{(2m_1)^3} - \tfrac{4}{(m_1+b)^3} + \tfrac{2}{(2b)^3}\right] \\
&+ 2\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\left[\tfrac{2}{(m_1+m_2)^3} - \tfrac{2}{(m_1+b)^3} - \tfrac{2}{(m_2+b)^3} + \tfrac{2}{(2b)^3}\right] \\
&+ \left(\tfrac{a_2}{b-m_2}\right)^2 \left[\tfrac{2}{(2m_2)^3} - \tfrac{4}{(m_2+b)^3} + \tfrac{2}{(2b)^3}\right]
\end{aligned}$$

using equation (A.3).

Then, we calculate

$$\begin{aligned}
I_4 &= \int_0^\infty (C_p * E)(t) I_1(t) dt \\
&= \int_0^\infty P(t) dt
\end{aligned}$$

with

$$
\begin{aligned}
P(t) &= I_1(t) \cdot (C_p * E)(t) \\
&= \Big\{ t \left[\tfrac{a_1}{b-m_1} \exp(-m_1 t) + \tfrac{a_2}{b-m_2} \exp(-m_2 t) \right] \\
&+ \tfrac{a_1}{(b-m_1)^2} \left(\exp(-bt) - \exp(-m_1 t) \right) \\
&+ \tfrac{a_2}{(b-m_2)^2} \left(\exp(-bt) - \exp(-m_2 t) \right) \Big\} \\
&\cdot \Big\{ \tfrac{a_1}{b-m_1} \left(\exp(-m_1 t) - \exp(-bt) \right) + \tfrac{a_2}{b-m_2} \left(\exp(-m_2 t) - \exp(-bt) \right) \Big\} \\
&= t \left(\tfrac{a_1}{b-m_1} \right)^2 \exp(-m_1 t) \left(\exp(-m_1 t) - \exp(-bt) \right) \\
&+ t \tfrac{a_1}{b-m_1} \tfrac{a_2}{b-m_2} \exp(-m_1 t) \left(\exp(-m_2 t) - \exp(-bt) \right) \\
&+ t \tfrac{a_1}{b-m_1} \tfrac{a_2}{b-m_2} \exp(-m_2 t) \left(\exp(-m_1 t) - \exp(-bt) \right) \\
&+ t \left(\tfrac{a_2}{b-m_2} \right)^2 \exp(-m_2 t) \left(\exp(-m_2 t) - \exp(-bt) \right) \\
&- \tfrac{a_1^2}{(b-m_1)^3} \left(\exp(-m_1 t) - \exp(-bt) \right)^2 \\
&- \tfrac{a_1}{(b-m_1)^2} \tfrac{a_2}{b-m_2} \left(\exp(-m_1 t) - \exp(-bt) \right) \left(\exp(-m_2 t) - \exp(-bt) \right) \\
&- \tfrac{a_2}{(b-m_2)^2} \tfrac{a_1}{b-m_1} \left(\exp(-m_1 t) - \exp(-bt) \right) \left(\exp(-m_2 t) - \exp(-bt) \right) \\
&- \tfrac{a_2^2}{(b-m_2)^3} \left(\exp(-m_2 t) - \exp(-bt) \right)^2 \\
&= t \left(\tfrac{a_1}{b-m_1} \right)^2 \left(\exp(-2m_1 t) - \exp(-(m_1+b)t) \right) \\
&+ t \tfrac{a_1}{b-m_1} \tfrac{a_2}{b-m_2} \left(2\exp(-(m_1+m_2)t) - \exp(-(m_1+b)t) \right. \\
&\quad \left. \cdots - \exp(-(m_2+b)t) \right) \\
&+ t \left(\tfrac{a_2}{b-m_2} \right)^2 \left(\exp(-2m_2 t) - \exp(-(m_2+b)t) \right) \\
&- \tfrac{a_1^2}{(b-m_1)^3} \left(\exp(-2m_1 t) - 2\exp(-(m_1+b)t) + \exp(-2bt) \right) \\
&- \tfrac{a_1}{(b-m_1)^2} \tfrac{a_2}{b-m_2} \left(\exp(-(m_1+m_2)t) - \exp(-(m_1+b)t) \right. \\
&\quad \left. \cdots - \exp(-(m_2+b)t) + \exp(-2bt) \right) \\
&- \tfrac{a_2}{(b-m_2)^2} \tfrac{a_1}{b-m_1} \left(\exp(-(m_1+m_2)t) - \exp(-(m_1+b)t) \right. \\
&\quad \left. \cdots - \exp(-(m_2+b)t) + \exp(-2bt) \right) \\
&- \tfrac{a_2^2}{(b-m_2)^3} \left(\exp(-2m_2 t) - 2\exp(-(m_2+b)t) + \exp(-2bt) \right).
\end{aligned}
$$

Using equations (A.2) and (A.1) we obtain

$$\begin{aligned}
I_4 &= \int_0^\infty P(t)dt \\
&= \left(\frac{a_1}{b-m_1}\right)^2 \left(\frac{1}{(2m_1)^2} - \frac{1}{(m_1+b)^2}\right) \\
&+ \frac{a_1}{b-m_1}\frac{a_2}{b-m_2}\left(\frac{2}{(m_1+m_2)^2} - \frac{1}{(m_1+b)^2} - \frac{1}{(m_2+b)^2}\right) \\
&+ \left(\frac{a_2}{b-m_2}\right)^2 \left(\frac{1}{(2m_2)^2} - \frac{1}{(m_2+b)^2}\right) \\
&- \frac{a_1^2}{(b-m_1)^3}\left(\frac{1}{2m_1} - \frac{2}{m_1+b} + \frac{1}{2b}\right) \\
&- \frac{a_1}{(b-m_1)^2}\frac{a_2}{b-m_2}\left(\frac{1}{m_1+m_2} - \frac{1}{m_1+b} - \frac{1}{m_2+b} + \frac{1}{2b}\right) \\
&- \frac{a_2}{(b-m_2)^2}\frac{a_1}{b-m_1}\left(\frac{1}{m_1+m_2} - \frac{1}{m_1+b} - \frac{1}{m_2+b} + \frac{1}{2b}\right) \\
&- \frac{a_2^2}{(b-m_2)^3}\left(\frac{1}{2m_2} - \frac{2}{m_2+b} + \frac{1}{2b}\right).
\end{aligned}$$

Using equations (A.3) and (A.2) we obtain

$$\begin{aligned}
I_6 &= \int_0^\infty t\,(C_p * E)\,(t) I_1(t)dt \\
&= \int_0^\infty tP(t)dt \\
&= \left(\frac{a_1}{b-m_1}\right)^2 \left(\frac{2}{(2m_1)^3} - \frac{2}{(m_1+b)^3}\right) \\
&+ \frac{a_1}{b-m_1}\frac{a_2}{b-m_2}\left(\frac{4}{(m_1+m_2)^3} - \frac{2}{(m_1+b)^3} - \frac{2}{(m_2+b)^3}\right) \\
&+ \left(\frac{a_2}{b-m_2}\right)^2 \left(\frac{2}{(2m_2)^3} - \frac{2}{(m_2+b)^3}\right) \\
&- \frac{a_1^2}{(b-m_1)^3}\left(\frac{1}{(2m_1)^2} - \frac{2}{(m_1+b)^2} + \frac{1}{(2b)^2}\right) \\
&- \frac{a_1}{(b-m_1)^2}\frac{a_2}{b-m_2}\left(\frac{1}{(m_1+m_2)^2} - \frac{1}{(m_1+b)^2} - \frac{1}{(m_2+b)^2} + \frac{1}{(2b)^2}\right) \\
&- \frac{a_2}{(b-m_2)^2}\frac{a_1}{b-m_1}\left(\frac{1}{(m_1+m_2)^2} - \frac{1}{(m_1+b)^2} - \frac{1}{(m_2+b)^2} + \frac{1}{(2b)^2}\right) \\
&- \frac{a_2^2}{(b-m_2)^3}\left(\frac{1}{(2m_2)^2} - \frac{2}{(m_2+b)^2} + \frac{1}{(2b)^2}\right).
\end{aligned}$$

Finally, for the calculation of

$$I_7 = \int_0^\infty I_1^2(t)dt$$

we first need

$$
\begin{aligned}
I_1^2 &= \left\{ t\left[\tfrac{a_1}{b-m_1}\exp(-m_1 t)+\tfrac{a_2}{b-m_2}\exp(-m_2 t)\right] \right.\\
&+ \tfrac{a_1}{(b-m_1)^2}\left(\exp(-bt)-\exp(-m_1 t)\right)\\
&+ \left.\tfrac{a_2}{(b-m_2)^2}\left(\exp(-bt)-\exp(-m_2 t)\right)\right\}^2\\
&= t^2\left[\tfrac{a_1}{b-m_1}\exp(-m_1 t)+\tfrac{a_2}{b-m_2}\exp(-m_2 t)\right]^2\\
&+ \tfrac{a_1^2}{(b-m_1)^4}\left\{\exp(-bt)-\exp(-m_1 t)\right\}^2\\
&+ \tfrac{a_2^2}{(b-m_2)^4}\left\{\exp(-bt)-\exp(-m_2 t)\right\}^2\\
&+ 2t\left[\tfrac{a_1}{b-m_1}\exp(-m_1 t)+\tfrac{a_2}{b-m_2}\exp(-m_2 t)\right]\tfrac{a_1}{(b-m_1)^2}\left\{\exp(-bt)-\exp(-m_1 t)\right\}\\
&+ 2t\left[\tfrac{a_1}{b-m_1}\exp(-m_1 t)+\tfrac{a_2}{b-m_2}\exp(-m_2 t)\right]\tfrac{a_2}{(b-m_2)^2}\left\{\exp(-bt)-\exp(-m_2 t)\right\}\\
&+ 2\tfrac{a_1}{(b-m_1)^2}\left\{\exp(-bt)-\exp(-m_1 t)\right\}\tfrac{a_2}{(b-m_2)^2}\left\{\exp(-bt)-\exp(-m_2 t)\right\}\\
&= t^2\left[\left(\tfrac{a_1}{b-m_1}\right)^2\exp(-2m_1 t)+2\tfrac{a_1}{b-m_1}\tfrac{a_2}{b-m_2}\exp(-(m_1+m_2)t)\right.\\
&\quad \left.\cdots+\left(\tfrac{a_2}{b-m_2}\right)^2\exp(-2m_2 t)\right]\\
&+ \tfrac{a_1^2}{(b-m_1)^4}\left\{\exp(-2bt)-2\exp(-(b+m_1)t)+\exp(-2m_1 t)\right\}\\
&+ \tfrac{a_2^2}{(b-m_2)^4}\left\{\exp(-2bt)-2\exp(-(b+m_2)t)+\exp(-2m_2 t)\right\}\\
&+ 2t\tfrac{a_1^2}{(b-m_1)^3}\left[\exp(-(b+m_1)t)-\exp(-2m_1 t)\right]\\
&+ 2t\tfrac{a_1}{(b-m_1)^2}\tfrac{a_2}{b-m_2}\left[\exp(-(b+m_2)t)-\exp(-(m_1+m_2)t)\right]\\
&+ 2t\tfrac{a_2}{(b-m_2)^2}\tfrac{a_1}{b-m_1}\left[\exp(-(b+m_1)t)-\exp(-(m_1+m_2)t)\right]\\
&+ 2t\tfrac{a_2^2}{(b-m_2)^3}\left[\exp(-(b+m_2)t)-\exp(-2m_2 t)\right]\\
&+ 2\tfrac{a_1}{(b-m_1)^2}\tfrac{a_2}{(b-m_2)^2}\left\{\exp(-2bt)-\exp(-(b+m_2)t)\right.\\
&\quad \left.\cdots\cdots-\exp(-(b+m_1)t)+\exp(-(m_1+m_2)t)\right\}.
\end{aligned}
$$

Then, using equations (A.1–A.3) we obtain

$$
\begin{aligned}
I_7 &= \int_0^\infty I_1^2(t)dt \\
&= \left(\frac{a_1}{b-m_1}\right)^2 \frac{2}{(2m_1)^3} + 2\frac{a_1}{b-m_1}\frac{a_2}{b-m_2}\frac{2}{(m_1+m_2)^3} + \left(\frac{a_2}{b-m_2}\right)^2 \frac{2}{(2m_2)^3} \\
&+ \frac{a_1^2}{(b-m_1)^4}\left\{\frac{1}{2b} - \frac{2}{b+m_1} + \frac{1}{2m_1}\right\} \\
&+ \frac{a_2^2}{(b-m_2)^4}\left\{\frac{1}{2b} - \frac{2}{b+m_2} + \frac{1}{2m_2}\right\} \\
&+ 2\frac{a_1^2}{(b-m_1)^3}\left[\frac{1}{(b+m_1)^2} - \frac{1}{(2m_1)^2}\right] \\
&+ 2\frac{a_1}{(b-m_1)^2}\frac{a_2}{b-m_2}\left[\frac{1}{(b+m_2)^2} - \frac{1}{(m_1+m_2)^2}\right] \\
&+ 2\frac{a_2}{(b-m_2)^2}\frac{a_1}{b-m_1}\left[\frac{1}{(b+m_1)^2} - \frac{1}{(m_1+m_2)^2}\right] \\
&+ 2\frac{a_2^2}{(b-m_2)^3}\left[\frac{1}{(b+m_2)^2} - \frac{1}{(2m_2)^2}\right] \\
&+ 2\frac{a_1}{(b-m_1)^2}\frac{a_2}{(b-m_2)^2}\left\{\frac{1}{2b} - \frac{1}{b+m_2} - \frac{1}{b+m_1} + \frac{1}{m_1+m_2}\right\}.
\end{aligned}
$$

Appendix B

Details in reversible jump MCMC birth and death transformations

B.1 Determinants of Jacobian matrices for birth and death transformations

In order to determine the determinant of Jacobian matrix $J_{birth} = \frac{\partial(a_1^*,\ldots,a_K^*,v_1^*,\ldots,v_K^*)}{\partial(a_1,\ldots,a_{K-1},v_1,\ldots,v_{K-1},u_1,u_2)}$ for the birth transformation $g_B(x)$, it suffices to consider the parameters with changes (*i.e.* a_n^*, $a_{n+1}^*, \ldots$, v_n^*, $v_{n+1}^*, \ldots$). We thus consider the submatrix of the Jacobian J_{birth}^*

$$
\begin{aligned}
J_{birth}^* &= \frac{\partial\left(a_n^*,a_{n+1}^*,a_{n+2}^*,v_n^*,v_{n+1}^*,v_{n+2}^*\right)}{\partial(a_n,a_{n+1},v_n,v_{n+1},u_1,u_2)} \\
&= \begin{pmatrix}
u_1 & (1-u_1) & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & u_2 & (1-u_2) & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
a_n & -a_n & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & v_n & -v_n & 0
\end{pmatrix}.
\end{aligned}
$$

The determinant of this matrix is

$$
\begin{aligned}
det(J^*_{birth}) &= det\begin{pmatrix} u_1 & (1-u_1) & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & u_2 & (1-u_2) & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \\ a_n & -a_n & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & v_n & -v_n & 0 \end{pmatrix} \\
&= (-1)det\begin{pmatrix} u_1 & (1-u_1) & 0 & 0 \\ 0 & 0 & u_2 & (1-u_2) \\ a_n & -a_n & 0 & 0 \\ 0 & 0 & v_n & -v_n \end{pmatrix} \\
&= (-u_1)det\begin{pmatrix} 0 & u_2 & (1-u_2) \\ -a_n & 0 & 0 \\ 0 & v_n & -v_n \end{pmatrix} + (1-u_1)det\begin{pmatrix} 0 & u_2 & (1-u_2) \\ a_n & 0 & 0 \\ 0 & v_n & -v_n \end{pmatrix} \\
&= -u_1a_ndet\begin{pmatrix} u_2 & (1-u_2) \\ v_n & -v_n \end{pmatrix} + (1-u_1)(-a_n)det\begin{pmatrix} u_2 & (1-u_2) \\ v_n & -v_n \end{pmatrix} \\
&= (-u_1a_n - a_n + u_1a_n)(-u_2v_n - (1-u_2)v_n) \\
&= a_n(u_2v_n + v_n - u_2v_n) \\
&= a_nv_n
\end{aligned}
$$

and $det(J_{birth}) = det(J^*_{birth})$. The determinant for the Jacobian of the death transformation $g_D(y)$ is

$$
\begin{aligned}
J_{death} &= \tfrac{1}{a^*_nv^*_n} \\
&= \tfrac{1}{(a_n+a_{n+1})(v_n+v_{n+1})}.
\end{aligned}
$$

B.2 Derivation of birth and death transformation

Here, we derive mapping conditions for jumps from $K = 1$ to $K^* = 2$ which can be generalized to jumps from K to $K^* = K + 1$. In a model of one exponential ($K = 1$)

$$C(t_i) = 1 - a_1 \exp(-b_1t_i) + \epsilon_i, \text{ for } i = 1 \dots T,$$

we want to propose parameters $(a^*_1, a^*_2, b^*_1, b^*_2)$ for a model with two exponentials ($K^* = 2$)

$$C(t_i) = 1 - a^*_1 \exp(-b^*_1t_i) - a^*_2 \exp(-b^*_2t_i) + \epsilon_i, \text{ for } i = 1 \dots T.$$

Assuming that the model with one exponential well describes the observed curve $C(t_i)$, we would like to propose $(a^*_1, a^*_2, b^*_1, b^*_2)$ such that the new fitted curve is similar to the actual fitted curve. Ideally, the two curves coincide:

$$a_1 \exp(-b_1t_i) = a^*_1 \exp(-b^*_1t_i) + a^*_2 \exp(-b^*_2t_i) \text{ for all } t_i,$$

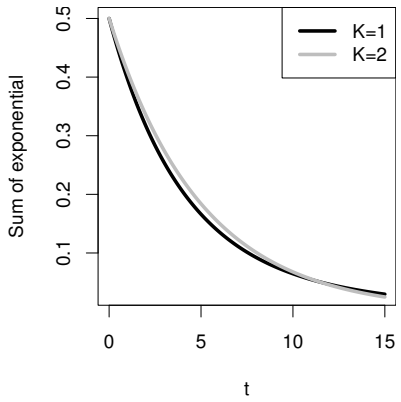

Figure B.1: Curves corresponding to parameters with model dimension $K = 1$ (black) and $K = 2$ (gray). Underlying parameters for $K = 1$: $a_1 = 0.5$, $u_1 = 0.2$, $b_1 = 0.2$, $u_2 = 0.4$; for $K = 2$: $a_1 = 0.1$, $a_2 = 0.4$, $b_1 = 0.1$, $b_2 = 0.27$.

for $i = 1 \dots T$. This equation does not have solutions except the trivial ones, where either ($a_1^* = 0$ and $a_2^* = a_1$) or ($a_2^* = 0$ and $a_1^* = a_1$) or ($b_1 = b_1^* = b_2^*$ and $a_1 = a_1^* + a_2^*$).

For $t_i = 0$, we obtain the following condition for mapping the amplitudes:

$$a_1 = a_1^* + a_2^*. \tag{B.1}$$

Integrating over the interval of the measurement times, we obtain

$$\int_0^T a_1 \exp(-b_1 t)dt = \int_0^T a_1^* \exp(-b_1^* t) + a_2^* \exp(-b_2^* t)dt$$

and hence

$$-\frac{a_1}{b_1}\left(1 - \exp(-b_1 T)\right) = -\sum_{i=1}^{2} \frac{a_i^*}{b_i^*}\left(1 - \exp(-b_i^* T)\right).$$

Assuming that T is large enough and that therefore

$$\exp(-b_1 T), \exp(-b_1^* T), \exp(-b_2^* T) \cong 0,$$

we obtain the following requirement

$$\frac{a_1}{b_1} = \frac{a_1^*}{b_1^*} + \frac{a_2^*}{b_2^*}$$

depending on the ratios of amplitudes over decay rates. Writing these ratios as volume fractions $v = \frac{a}{b}$, we gain

$$v_1 = v_1^* + v_2^*. \tag{B.2}$$

This is a plausible requirement for the mapping as the volumes in both models add up to the same value. We use both requirements (B.1) and (B.2) to construct a mapping between models of different dimensions. Figure B.1 shows how the proposed birth step maps a curve described by one exponential to a curve described by a sum of two exponentials.

Bibliography

Agrawal, G., M.-Y. Su, O. Nalcioglu, S. A. Feig, and J.-H. Chen (2009). Significance of breast lesion descriptors in the ACR BI-RADS MRI lexicon. *Cancer 115*(7), 1363–1380.

Ah-See, M.-L. W., A. Makris, N. J. Taylor, R. J. Burcombe, M. Harrison, J. J. Stirling, P. I. Richman, M. O. Leach, and A. R. Padhani (2004). Does vascular imaging with MRI predict response to neoadjuvant chemotherapy in primary breast cancer? *Journal of Clinical Oncology (Meeting Abstracts) 22*(14S), 582.

Ahearn, T., R. Staff, T. Redpath, and S. Semple (2005). The use of the Levenberg-Marquardt curve-fitting algorithm in pharmacokinetic modelling of DCE-MRI. *Physics in Medicine and Biology 50*, N85–N92.

Bates, D. M. and D. G. Watts (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 48*(3), 259–302.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society Series B (Statistical Methodology) 36*, 192–236.

Boys, R. J. and D. A. Henderson (2004). A Bayesian approach to DNA sequence segmentation. *Biometrics 60*(3), 573–581.

Brezger, A., L. Fahrmeir, and A. Hennerfeind (2007). Adaptive Gaussian Markov random fields with applications in human brain mapping. *Journal of the Royal Statistical Society. Series C (Applied) 56*(3), 327–345.

Brix, G., W. Semmler, R. Port, L. R. Schad, G. Layer, and W. J. Lorenz (1991). Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. *Journal of computer assisted tomography 15*(4), 621–628. Case Reports.

Brix, G., S. Zwick, F. Kiessling, and J. Griebel (2009). Pharmacokinetic analysis of tissue microcirculation using nested models: Multimodel inference and parameter identifiability. *Medical Physics 36*(7), 2923–2933.

Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump MCMC proposal distributions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 65*(1), 3–55.

Buckley, D. and G. Parker (2005). Measuring contrast agent concentration in T_1-weighted dynamic contrast-enhanced MRI. In *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*, pp. 69–80. Berlin, Heidelberg, New York: Springer.

Buckley, D. L. (2002). Uncertainty in the analysis of tracer kinetics using dynamic contrast-enhanced T1-weighted MRI. *Magnetic Resonance in Medicine 47*(3), 601–606.

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

Caravan, P. (2006). Strategies for increasing the sensitivity of gadolinium based MRI contrast agents. *Chemical Society Reviews 35*, 512–523.

Christensen, W. F. and F. Z. Yetkin (2005). Spatio-temporal analysis of auditory cortex activation as detected with silent event related fMRI. *Statistics in Medicine 24*(16), 2539–2556.

Crone, C. (1963). The permeability of capillaries in various organs as determined by use of the indicator diffusion method. *Acta Physiologica Scandinavica 58*(4), 292–305.

Dargatz, C. (2010). *Bayesian Inference for Diffusion Processes with Applications in Life Sciences*. Dissertation, Ludwig-Maximilians-Universität München.

Donaldson, S. B., C. M. L. West, S. E. Davidson, B. M. Carrington, G. Hutchison, A. P. Jones, S. P. Sourbron, and D. L. Buckley (2010). A comparison of tracer kinetic models for T1-weighted dynamic contrast-enhanced MRI: Application in carcinoma of the cervix. *Magnetic Resonance in Medicine 63*(3), 691–700.

Dössel, O. (2000). *Bildgebende Verfahren in der Medizin: von der Technik zur medizinischen Anwendung*. Belin, Heidelberg, New York: Springer-Verlag.

Edelmann, R. D. and S. Warach (1993). Magnetic resonance imaging. *The New England Journal of Medicine 328*(10), 708–716.

Eriksson, E. (1971). Compartment models and reservoir theory. *Annual Review of Ecology and Systematics 2*, 67–84.

Fluckiger, J. U., M. C. Schabel, and E. V. R. DiBella (2009). Model-based blind estimation of kinetic parameters in dynamic contrast enhanced (DCE)-MRI. *Magnetic Resonance in Medicine 62*(6), 1477–1486.

Fritz-Hansen, T., E. Rostrup, K. B. Larsson, L. Søndergaard, P. Ring, and O. Henriksen (1996). Measurement of the arterial concentration of Gd-DTPA using MRI: A step toward quantitative perfusion imaging. *Magnetic Resonance in Medicine 36*, 225–231.

Gelman, A., J. B. Carlin, D. B. Rubin, and H. S. Stern (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall.

Geman, D. and S. Geman (1986). Bayesian image analysis. In E. Bienenstock, F. Fogelman, and G. Welsbuch (Eds.), *Disordered Systems and Biological Organization*, pp. 301–319. Berlin: Springer-Verlag.

Gertheiss, J., J. C. Kärcher, and V. J. Schmid (2010). Analysis of DCE-MRI data using a nonnegative Elastic Net. *Institut für Statistik, Ludwig-Maximilians-Universität München: Technical Reports 90*.

Giesel, F. L., H. Bischoff, H. von Tengg-Kobligk, M.-A. Weber, C. M. Zechmann, H.-U. Kauczor, and M. V. Knopp (2006). Dynamic contrast-enhanced MRI of malignant pleural mesothelioma: A feasibility study of noninvasive assessment, therapeutic follow-up, and possible predictor of improved outcome. *CHEST Journal 129*(6), 1570–1576.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996). *Markov chain Monte Carlo in Practice*. London: Chapman & Hall.

Goh, V. and A. Padhani (2006). Imaging tumor angiogenesis: functional assessment using MDCT or MRI? *Abdominal Imaging 31*(2), 194–199.

Gössl, C., D. P. Auer, and L. Fahrmeir (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics 57*(2), 554–562.

Green, P. (2000). A primer on Markov chain Monte Carlo. In O. Barndorff-Nielsen, D. Cox, and C. Klüppelberg (Eds.), *Complex Stochastic Systems*, pp. 1–62. London: Chapman & Hall.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Gunn, R. N., S. R. Gunn, F. E. Turkheimer, J. A. D. Aston, and V. J. Cunningham (2002). Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. *Journal of Cerebral Blood Flow and Metabolism 22*, 1425–1439.

Hastie, T., R. Tibshirani, and J. H. Friedman (2009). *The Elements of Statistical Learning* (2nd ed.). New York: Springer.

Herbst, P. G. (1963). Organizational commitment: A decision process model. *Acta Sociologica 7*(1), 34–46.

Heuser, H. (1995). *Gewöhnliche Differentialgleichungen* (3rd ed.). Stuttgart: B.G. Teubner.

Hodgson, R. J., P. O'Connor, and R. Moots (2008). MRI of rheumatoid arthritis: image quantitation for the assessment of disease activity, progression and response to therapy. *Rheumatology 47*(1), 13–21.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*, 55–67.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–401.

Ingrisch, M., S. Sourbron, M. F. Reiser, and M. Peller (2010). Model selection in dynamic contrast enhanced MRI: The Akaike Information Criterion. In O. Dössel and W. C. Schlegel (Eds.), *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, Volume 25/4 of *IFMBE Proceedings*, pp. 356–358. Berlin, Heidelberg: Springer.

Jacquez, J. A. (1972). *Compartmental analysis in biology and medicine.* Amsterdam: Elsevier.

Johnson, J. A. and T. A. Wilson (1966). A model for capillary exchange. *American journal of physiology 210*(6), 1299–303.

Karahaliou, A., K. Vassiou, N. S. Arikidis, S. Skiadopoulos, T. Kanavou, and L. Costaridou (2010). Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. *British Journal of Radiology 83*(988), 296–309.

Kärcher, J. C. and V. J. Schmid (2010). Two Tissue Compartment Model in DCE-MRI: A Bayesian Approach. In *IEEE International Symposium on Biomedical Imaging. From Nano to Macro*, 3, pp. 724–727.

Kelm, B. M., B. H. Menze, O. Nix, C. M. Zechmann, and F. A. Hamprecht (2009). Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge. *IEEE Transactions on Medical Imaging 28*(10), 1534–1547.

Kety, S. (1960). Blood-tissue exchange methods. Theory of blood-tissue exchange and its applications to measurement of blood flow. *Methods in Medical Research 8*, 223–227.

King, R. and S. P. Brooks (2002). Model selection for integrated recovery/recapture data. *Biometrics 58*(4), 841–851.

Larsson, H. B. W., M. Stubgaard, J. L. Frederiksen, M. Jensen, O. Henriksen, and O. B. Paulson (1990). 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), 117–131.

Leach, M., K. Brindle, J. Evelhoch, J. Griffiths, M. Horsman, A. Jackson, G. Jayson, I. Judson, M. Knopp, R. Maxwell, and Others (2005). The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. *British journal of cancer 92*(9), 1599–1610.

Li, Q. and N. Lin (2010). The Bayesian Elastic Net. *Bayesian Analysis 5*(1), 151–170.

Liney, G. P., P. Gibbs, C. Hayes, M. O. Leach, and L. W. Turnbull (1999). Dynamic contrast-enhanced MRI in the differentiation of breast tumors: User-defined versus semi-automated region-of-interest analysis. *Journal of Magnetic Resonance Imaging 10*(6), 945–949.

Marroquin, J. L., E. A. Santana, and S. Botello (2003). Hidden Markov Measure Field Models for Image Segmentation. *IEEE Transaction on pattern analysis and machine intelligence 25*(11).

McMahon, C. J., B. N. Bloch, R. E. Lenkinski, and N. M. Rofsky (2009). Dynamic contrast-enhanced MR imaging in the evaluation of patients with prostate cancer. *Magnetic Resonance Imaging Clinics of North America 17*(2), 363–383.

McNally, J. G. (2008). Quantitative FRAP in analysis of molecular binding dynamics in vivo. In F. S. Kevin (Ed.), *Methods in Cell Biology*, Volume 85, pp. 329–351. Academic Press.

Meinshausen, N. (2007). Relaxed Lasso. *Computational Statistics & Data Analysis 52*(1), 374–393.

Mitchell, D. G. (1997). MR imaging contrast agents: what's in a name? *Journal of Magnetic Resonance Imaging 7*(1), 1–4.

Mohajer, M., V. J. Schmid, N. A. Engels, P. B. Noel, E. Rummeny, and K.-H. Englmeier (2012). Stepwise heterogeneity analysis of breast tumors in perfusion DCE-MRI datasets. In R. C. Molthen and J. B. Weaver (Eds.), *Medical Imaging 2012: Biomedical Applications in Molecular, Structural, and Functional Imaging*, pp. 83171K.

O'Connor, J. P. B., A. Jackson, G. J. M. Parker, and G. C. Jayson (2007). DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. *British Journal of Cancer 96*(2), 189–195.

O'Hara, R. B. and M. J. Sillanpää (2009). A review of Bayesian variable selection methods: What, how, and which? *Bayesian Analysis 4*(1), 85–118.

Orton, M. R., D. J. Collins, S. Walker-Samuel, J. A. d'Arcy, D. J. Hawkes, D. Atkinson, and M. O. Leach (2007). Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time. *Physics in Medicine and Biology 52*, 2393–2408.

Padhani, A. R., M.-L. Ah-See, and A. Makris (2005). MRI in the detection and management of breast cancer. *Expert Review of Anticancer Therapy 5*(2), 239–252.

Park, T. and G. Casella (2008). The Bayesian Lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Parker, G. and D. Buckley (2005). Tracer kinetic modelling for T_1-weighted DCE-MRI. In *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*, pp. 81–92. Berlin, Heidelberg, New York: Springer.

Parker, G. J., C. Roberts, A. Macdonald, G. A. Buonaccorsi, S. Cheung, D. L. Buckley, A. Jackson, Y. Watson, K. Davies, and G. C. Jayson (2006). Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magnetic Resonance in Medicine 56*(5), 993–1000.

Parker, G. J. M. and A. R. Padhani (2003). T1-w DCE-MRI: T1-weighted dynamic contrast-enhanced MRI. In P. Tofts (Ed.), *Quantitative MRI of the Brain*, pp. 341–364. Chichester, England: Wiley.

Phair, R. D., P. Scaffidi, C. Elbi, J. Vecerová, A. Dey, K. Ozato, D. T. Brown, G. Hager, M. Bustin, and T. Misteli (2004). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Molecular and Cellular Biology 24*, 6393–402.

Port, R. E., M. V. Knopp, U. Hoffmann, S. Milker-Zabel, and G. Brix (1999). Multicompartment analysis of gadolinium chelate kinetics: blood-tissue exchange in mammary tumors as monitored by dynamic MR imaging. *Journal of Magnetic Resonance Imaging 10*, 233–241.

Ratkowsky, D. (1990). *Handbook of Nonlinear Regression Models*. New York: Marcel Dekker.

Reich, J. G. (1981). *On parameter redundancy in curve fitting of kinetic data*. Kinetic data analysis: Design and analysis of enzyme and pharmacokinetic experiments. New York: Plenum Press.

Renkin, E. M. (1959). Transport of potassium-42 from blood to tissue in isolated mammalian skeletal muscles. *The American journal of physiology 197*, 1205–1210.

Robert, C. P. (2001). *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*. New York: Springer-Verlag.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (2 ed.). Springer Texts in Statistics. New York: Springer.

Rose, C. J., S. J. Mills, J. P. B. O'Connor, G. A. Buonaccorsi, C. Roberts, Y. Watson, S. Cheung, S. Zhao, B. Whitcher, A. Jackson, and G. J. M. Parker (2009). Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps. *Magnetic Resonance in Medicine 62*(2), 488–499.

Rosen, B. R., J. W. Belliveau, J. M. Vevea, and T. J. Brady (1990). Perfusion imaging with NMR contrast agents. *Magnetic Resonance in Medicine 14*(2), 249–265.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications (Monographs on Statistics and Applied Probability)*. London: Chapman & Hall.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Schmid, V., B. Whitcher, A. R. Padhani, and G.-Z. Yang (2009). Quantitative analysis of dynamic contrast-enhanced MR images based on Bayesian P-Splines. *IEEE Transactions on Medical Imaging 28*, 789–798.

Schmid, V. J. (2010). Kinetic models for cancer imaging. In H. R. Arabnia (Ed.), *Advances in Computational Biology*, pp. 549–558. New York, Heidelberg, Dordrecht, London: Springer.

Schmid, V. J. (2011). Voxel-based adaptive spatio-temporal modelling of perfusion cardiovascular MRI. *IEEE Transactions on Medical Imaging 30*(7), 1305–1313.

Schmid, V. J., B. Whitcher, A. Padhani, N. Taylor, and G.-Z. Yang (2006). Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Transactions on Medical Imaging 25*(12), 1627–1636.

Schmid, V. J., B. Whitcher, G. Z. Yang, N. J. Taylor, and A. R. Padhani (2005). Statistical analysis of pharmacokinetical models in dynamic contrast-enhanced magnetic resonance imaging. In J. Duncan and G. Gerig (Eds.), *Lecture Notes in Computer Science*, Number 3750, pp. 886–893. Berlin: Springer.

Schmitt, M., M. Viallon, M. Thelen, and W. G. Schreiber (2002). Quantification of myocardial blood flow and blood flow reserve in the presence of arterial dispersion: A simulation study. *Magnetic Resonance in Medicine 47*(4), 787–793.

Schneider, K., C. Fuchs, A. Dobay, A. Rottach, W. Qin, J. Álvarez Castro, M. Nalaskowski, V. Schmid, H. Leonhardt, and L. Schermelleh (2013). Dissection of cell cycle dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Research journal (under review)*.

Schneider, K. M. (2009). *Analysis of cell cycle dependent kinetics of Dnmt1 by FRAP and kinetic modelling*. Diplomarbeit, Ludwig-Maximilians-Universität München.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Seber, G. A. F. and C. J. Wild (1989). *Nonlinear Regression*. New York: Wiley.

Slifstein, M. and M. Laruelle (2001). Models and methods for derivation of in vivo neuroreceptor parameters with PET and SPECT reversible radiotracers. *Nuclear Medicine and Biology 28*(5), 595–608.

Sommer, J. C., J. Gertheiss, and V. J. Schmid (2012). Spatially regularized estimation for the analysis of DCE-MRI data. *Institut für Statistik, Ludwig-Maximilians-Universität München: Technical Reports 132*.

Sommer, J. C. and V. J. Schmid (2012). Spatial two tissue compartment model for DCE-MRI. *arXiv* (1209.0901v1).

Sourbron, S., M. Ingrisch, A. Siefert, M. Reiser, and K. Herrmann (2009). Quantification of cerebral blood flow, cerebral blood volume, and blood-brain-barrier leakage with DCE-MRI. *Magnetic Resonance in Medicine 62*(1), 205–217.

Sourbron, S. P. and D. L. Buckley (2011). On the scope and interpretation of the Tofts models for DCE-MRI. *Magnetic Resonance in Medicine 66*(3), 735–745.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. v. d. Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*, 583–639.

Sprague, B. L. and J. G. McNally (2005). FRAP analysis of binding: proper and fitting. *Trends in Cell Biology 15*(2), 84–91.

St. Lawrence, K. S. and T.-Y. Lee (1998). An adiabatic approximation to the tissue homogeneity model for water exchange in the brain: I. theoretical derivation. *Journal of Cerebral Blood Flow & Metabolism 18*, 1365–1377.

Steingoetter, A., D. Menne, and R. F. Braren (2011). Assessing antiangiogenic therapy response by DCE-MRI: Development of a physiology driven multi-compartment model using population pharmacometrics. *PLoS ONE 6*(10), e26366.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 58*, 267–288.

Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *The Annals of Applied Probability 8*(1), 1–9.

Tofts, P. S. (1997). Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *Journal of Magnetic Resonance Imaging 7*(1), 91–101.

Tofts, P. S., G. Brix, D. L. Buckley, J. L. Evelhoch, E. Henderson, M. V. Knopp, H. B. W. Larsson, T.-Y. Lee, N. A. Mayr, G. J. M. Parker, R. E. Port, J. Taylor, and R. Weiskoff (1999). Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: Standardized quantities and symbols. *Journal of Magnetic Resonance Imaging 10*, 223–232.

Tofts, P. S. and A. Kermode (1991). Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging - 1. fundamental concepts. *Magnetic Resonance in Medicine 17*, 357–367.

Turkheimer, F. E., R. Hinz, and V. J. Cunningham (2003). On the undecidability among kinetic models: From model selection to model averaging. *Journal of Cerebral Blood Flow & Metabolism 23*(4), 490–498.

Turlach, B. A. and A. Weingessel (2011). *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-4.

Vega-Hernandez, M., E. Martinez-Montes, J. M. Sanchez-Bornot, A. Lage-Castellanos, and P. A. Valdes-Sosa (2008). Penalized Least Squares Methods for Solving the EEG Inverse Problem. *Statistica Sinica 18*, 1535–1551.

Weinmann, H., M. Laniado, and W. Mützel (1984). Pharmokinetics of Gd-DTPA/dimeglumine after intravenous injection into healthy volunteers. *Physiological Chemistry & Physics & Medical NMR 16*, 167–72.

Whitcher, B. and V. J. Schmid (2009). *dcemriS4: A Package for Medical Image Analysis.* R package version 0.10.5.

Whitcher, B. and V. J. Schmid (2011). Quantitative Analysis of Dynamic Contrast-Enhanced and Diffusion-Weighted Magnetic Resonance Imaging for Oncology in R. *Journal of Statistical Software 44*(5), 1–29.

Winkler, G. (1995). *Image analysis, random fields and dynamic Monte Carlo methods*, Volume 27. Berlin, Heidelberg: Springer.

Wintermark, M., M. Sesay, E. Barbier, K. Borbély, W. P. Dillon, J. D. Eastwood, T. C. Glenn, C. B. Grandin, S. Pedraza, J.-F. Soustiel, T. Nariai, G. Zaharchuk, J.-M. Caillé, V. Dousset, and H. Yonas (2005). Comparative overview of brain perfusion imaging techniques. *Stroke 36*(9), e83–e99.

Wyss, M. T., S. Hofer, M. Hefti, E. Bärtschi, C. Uhlmann, V. Treyer, and U. Roelcke (2007). Spatial heterogeneity of low-grade gliomas at the capillary level: A PET study on tumor blood flow and amino acid uptake. *Journal of Nuclear Medicine 48*(7), 1047–1052.

Yang, C., G. S. Karczmar, M. Medved, and W. M. Stadler (2007). Multiple reference tissue method for contrast agent arterial input function estimation. *Magnetic Resonance in Medicine 58*(6), 1266–1275.

Yang, X. and M. V. Knopp (2011). Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: a review. *Journal of biomedicine & biotechnology 2011*, 1–12.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B 67*, 301–320.

Zou, K. H., W. M. Wells, R. Kikinis, and S. K. Warfield (2004). Three validation metrics for automated probabilistic image segmentation of brain tumours. *Statistics in Medicine 23*(8), 1259–1282.

Eidesstattliche Versicherung

(Gemäß Promotionsordnung vom 12.07.11, §8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 26. Februar 2013 Julia C. Sommer