

Bayesian Data Analysis for Magnetic Resonance Fingerprinting

vorgelegt von
M. Sc.
Gerburg Selma Metzner

an der Fakultät V - Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktorin der Naturwissenschaften
- Dr.rer.nat. -
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Marc Kraft
Gutachter: Prof. Dr. Tobias Schäffter
Gutachterin: Prof. Dr. Katja Ickstadt
Gutachter: Dr. Clemens Elster

Tag der wissenschaftlichen Aussprache: 28. September 2021

Berlin 2021

Abstract

Magnetic resonance imaging (MRI) is a medical imaging technique which is widely used in clinical practice. It is non-invasive and provides a good tissue contrast. However, usually only qualitative images are obtained. In quantitative MRI (qMRI) biological tissue properties are measured which enhances the reliability of diagnostics. Standard methods in qMRI require long acquisition times and usually just measure a single parameter. Magnetic Resonance Fingerprinting (MRF) is a recent approach to qMRI that allows for the simultaneous estimation of the tissue-related parameters within short acquisition time. The main idea of MRF is to conduct a series of measurements that are highly undersampled in the Fourier domain and perform a template matching between approximately reconstructed magnetization courses and modeled magnetization courses stored in a pre-computed dictionary.

The goal of this thesis is to apply Bayesian statistics to further enhance the data analysis of MRF. Advantages of a Bayesian approach include the possibility to incorporate available prior knowledge and to obtain a posterior distribution for the sought parameters. The posterior can be used to assign uncertainties and to make probability statements. This can be particularly useful when assessing diagnostics or therapy monitoring. The first contribution of this thesis is a Bayesian uncertainty quantification for the dictionary-based MRF estimates. The data analysis of the original MRF approach is shown to be equivalent to a maximum likelihood estimation for a particular statistical model, and a Bayesian inference is developed based on this model. Analytical expressions for the posterior are derived and numerical techniques utilizing the pre-computed dictionary lead to a fast probability characterization.

The second contribution of this thesis is the development of a Bayesian inference for MRF data based on direct modeling in the Fourier domain. The advantage is that significantly better estimates can be achieved when aliasing errors are the dominant uncertainty source of the dictionary-based MRF data analysis. However, the challenge is that a large-scale regression problem is faced. A general class of large-scale regression problems together with several classes of (improper) prior distributions is considered and theoretical properties of the posterior such as the existence of moments are explored. These results apply to MRF but can also be used for many other large-scale regression problems. Spatial smoothness of the parameters can be modeled through Gaussian Markov Random Field or so-called partition priors, and the potential advantage of such prior information is explored for MRF.

The results of this thesis demonstrate that the Bayesian inference developed for the original, dictionary-based MRF approach yields both a reliable uncertainty quantification and the possibility to make probability statements about the sought parameters for the first time. Furthermore, when aliasing errors are the dominating source of uncertainty for the dictionary-based MRF approach, the developed large-scale Bayesian inference can substantially improve the estimation. Additionally, the inclusion of valuable prior information can improve estimation quality for MRF substantially. The enhancement of data analysis for MRF developed in this thesis is expected to support MRF and its potential future application in clinical practice.

Zusammenfassung

Die Magnetresonanztomographie ist ein nichtinvasives, medizinisches Bildgebungsverfahren, das einen guten Gewebekontrast liefert und umfangreich in der klinischen Praxis eingesetzt wird. Damit werden jedoch üblicherweise nur qualitative Bilder erzeugt. In der quantitativen Magnetresonanztomographie werden fundamentale biologische Gewebeeigenschaften quantitativ gemessen, womit die Verlässlichkeit der Diagnostik verbessert wird. Standardmethoden haben lange Aufnahmezeiten und messen gewöhnlich nur einen Parameter. Magnetic Resonance Fingerprinting (MRF) ist eine neuartige Methode der quantitativen Magnetresonanztomographie, mit der es möglich ist in kurzer Zeit mehrere Parameter gleichzeitig zu messen. Der Hauptgedanke von MRF ist die Aufnahme einer Reihe von Messungen, die hochunterabgetastet im Fourier-Raum sind, um dann approximativ rekonstruierte Magnetisierungsverläufe und modellierte Magnetisierungsverläufe, die in einer Bibliothek gespeichert sind, zu vergleichen.

Ziel dieser Dissertation ist eine Weiterentwicklung der Datenanalyse von MRF mit Hilfe Bayesscher Statistik. Die Vorteile der Bayesschen Inferenz beinhalten die Möglichkeit verfügbares Vorwissen einzubinden sowie eine a-posteriori Wahrscheinlichkeitsverteilung der gesuchten Parameter zu gewinnen. Letztere kann genutzt werden, um Unsicherheiten zu bestimmen und Wahrscheinlichkeitsaussagen zu treffen. Das kann besonders hilfreich sein in der Diagnostik und in der Behandlungsüberwachung. Der erste Beitrag dieser Dissertation ist eine Bayessche Unsicherheitsbestimmung der bibliotheksbasierten MRF-Schätzer. Es wird gezeigt, dass die Datenanalyse der originalen MRF-Methode äquivalent ist zu einer Maximum-Likelihood-Schätzung für ein bestimmtes statistisches Modell; eine Bayessche Inferenz basierend auf diesem Modell wird entwickelt. Analytische Ausdrücke für die a-posteriori Verteilung werden hergeleitet und numerische Techniken, die auf der vorab berechneten Bibliothek basieren, ermöglichen eine schnelle Wahrscheinlichkeitscharakterisierung.

Der zweite Beitrag dieser Dissertation ist die Entwicklung einer Bayesschen Inferenz für direkt im Fourier-Raum modellierte Daten. Der Vorteil ist, dass wesentlich bessere Schätzer erreicht werden können, wenn Aliasing-Fehler die dominante Unsicherheitsquelle sind. Die Herausforderung dieser Methode ist das resultierende hochdimensionale Regressionsproblem. Eine allgemeine Klasse hochdimensionaler Regressionsprobleme zusammen mit verschiedenen a-priori Wahrscheinlichkeitsverteilungen wird betrachtet und theoretische Eigenschaften der a-posteriori Verteilung werden untersucht. Die Resultate können auf MRF angewendet werden, aber auch für andere hochdimensionale Probleme genutzt werden. Räumliche Glattheit der Parameter kann durch Gaussian Markov Random Field oder sogenannte Partitionspriorien modelliert werden. Für MRF wird der potentielle Nutzen dieses Vorwissens diskutiert.

Die Resultate dieser Dissertation verdeutlichen, dass die Bayessche Inferenz für die originale MRF-Methode erstmalig sowohl eine verlässliche Unsicherheitsbestimmung als auch die Möglichkeit von Wahrscheinlichkeitsaussagen über die gesuchten Parameter liefert. Des Weiteren kann die hochdimensionale Bayessche Inferenz die Schätzer wesentlich verbessern, wenn Aliasing-Fehler die dominante Unsicherheitsquelle sind. Das Einbinden von Vorwissen kann die Schätzer zusätzlich wesentlich verbessern. Die hier dargestellte Weiterentwicklung

der MRF-Datenanalyse soll MRF und sein mögliches Einsetzen in der klinischen Praxis unterstützen.

Acknowledgement

This thesis would not have been possible without the help of many people, whom I would like to thank at this point.

First of all, I thank Dr. Clemens Elster for his tremendous support during my PhD. Not only could I always rely on his expertise in statistical data analysis and his many years of scientific experience, but his patience, his willingness to answer questions at any time, and his thoughts on the problems we worked on contributed to a large extent to the development of this thesis.

Furthermore, I would like to thank Dr. Christoph Kolbitsch. When needed, I could always count on his broad expertise in the field of quantitative MR imaging. In particular, his gift to simplify complicated issues helped me as a mathematician to gain deeper understanding of MR.

I thank Prof. Dr. Tobias Schäffter for supervising my work and always being interested in the state of my research.

I thank Prof. Dr. Katja Ickstadt for taking the time to be an advisor of this thesis.

I am particularly grateful for all my colleagues at PTB. Thank you Lara, Katy, Franko, Judith, Manuel, Tobias, Jörg, Gerd, and Manuel for making my PhD time so enjoyable. Our numerous lunch discussions, meetings, and telephone calls helped to broaden my perspective on the many applications of statistical data analysis. I especially want to thank Dr. Gerd Wübbeler for his invaluable help throughout the thesis.

A PhD thesis is not only the result of research, but also of a writing process. I thank Dr. Marcus Gerhold for proofreading my work. Both his mathematical expertise and his feel for language helped a lot to improve this work.

Last but not least, I thank my family and friends, who have accompanied me on my way here. Your confidence, understanding and support always helped me to keep on going with my PhD.

Table of Contents

Title Page	i
Abstract	iii
Zusammenfassung	vi
Acknowledgement	vii
List of Figures	xi
Abbreviations	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Objective and Aims	3
1.3 Thesis Structure and Overview	4
1.4 List of Publications	5
2 Magnetic Resonance Imaging and Magnetic Resonance Fingerprinting	7
2.1 Basic Principles of Magnetic Resonance Imaging	7
2.2 Concept of Magnetic Resonance Fingerprinting	12
2.3 Research Areas in Magnetic Resonance Fingerprinting	15
2.4 Phantom and Simulations	17
3 Bayesian Statistics	19
3.1 Motivation of Bayesian Statistics	19
3.2 Bayesian Inference	20
3.3 Bayesian Computation	26
4 Bayesian Uncertainty Quantification for Dictionary-Based MRF	33
4.1 Methods	34
4.1.1 Dictionary Matching	34
4.1.2 Bayesian Inference for Dictionary Matching	38
4.2 Experiments and Simulations	42
4.3 Results	43
4.3.1 Phantom and In Vivo Measurements	44
4.3.2 Simulated Data	46

TABLE OF CONTENTS

4.4	Discussion	49
4.5	Summary	54
5	Large-Scale Bayesian Modeling	55
5.1	Related Work	56
5.2	Model	57
5.2.1	Statistical Model and Assumptions	57
5.2.2	Priors	57
5.3	Properties of Posteriors and Numerical Calculations	60
5.3.1	Properties of Posteriors	60
5.3.2	Numerical Calculations	65
5.4	Application to Magnetic Resonance Fingerprinting	67
5.4.1	Physical Model	67
5.4.2	Simulation Setting	69
5.4.3	Employed Priors and Marginal Posteriors	69
5.4.4	Results and Discussion	70
5.5	Summary	73
6	Comparison of Dictionary Matching and Bayesian Modeling	75
6.1	Impact of Noise	76
6.1.1	Data	76
6.1.2	Results	78
6.1.3	Discussion	81
6.2	Impact of Sampling Scheme	82
6.2.1	Data	82
6.2.2	Results	83
6.2.3	Discussion	85
6.3	Impact of Errors in the Physical Model	87
6.3.1	Data	88
6.3.2	Results	88
6.3.3	Discussion	89
6.4	Computational Aspects	90
6.5	Summary	91
7	Conclusion and Outlook	93
	Appendix A Physical Model of a bSSFP Sequence	97
	Appendix B Physical Model of a FISP Sequence	99
	Appendix C Matlab Code for the Bayesian Uncertainty Quantification	101
	References	105

List of Figures

2.1	Explanation of T_1 and T_2 relaxation times	9
2.2	Connection of Fourier data and magnetization image	11
2.3	Example of different sampling schemes	12
2.4	Concept of MRF	13
2.5	Reference phantom	17
2.6	Simulated reference phantom	18
2.7	Simulated brain data	18
3.1	Pseudo code of the Metropolis-Hastings algorithm	29
3.2	Pseudo code of the Hamiltonian Monte Carlo algorithm	30
4.1	Magnetization images of the phantom data	44
4.2	Residuals of the magnetization of the phantom data	44
4.3	Standard deviation and mean uncertainties of the tubes of the phantom data	45
4.4	Standard deviation and mean uncertainties for a restricted dictionary of the phantom data	45
4.5	Standard deviation and mean uncertainties for data with 5 radial lines of the phantom data	46
4.6	Comparison of reference measurement and dictionary matching solution of the phantom data	47
4.7	Dictionary matching solution, uncertainties and marginal distribution of the phantom data	48
4.8	Dictionary matching solution and estimates of the cardiac in vivo scan	49
4.9	Standard deviation and mean uncertainties of the tubes of the simulated data	50
4.10	Comparison of reference measurement and dictionary matching solution of the simulated data	50
4.11	Probability map of the phantom data	53
5.1	Illustration of partition prior	59
5.2	Implementation of the trust-region algorithm used to calculate the MAP	66
5.3	Values of tissue related parameters for a slice of the human brain	68
5.4	Undersampled Fourier data	68
5.5	Reconstructed images and reconstruction errors for the noninformative prior	71
5.6	Reconstructed images and reconstruction errors for the informative prior and width of 95% credible intervals	72

LIST OF FIGURES

5.7	Root-mean-square errors of the reconstructed images for the different choices of priors	72
6.1	Results for the bSSFP data set	78
6.2	Results for the FISP data set	79
6.3	Example of residuals in k -space for bSSFP and FISP data set	80
6.4	Results for the bSSFP data set with lower SNR	80
6.5	Example of residuals in k -space for bSSFP data set with lower SNR	81
6.6	Different sampling schemes for a comparison of the impact of the k -space modeling	82
6.7	Dictionary matching solutions and error maps of T_1 and T_2 for the three different sampling schemes	84
6.8	MAP estimates of the Bayesian k -space modeling approach and error maps of T_1 and T_2 for the three different sampling schemes	84
6.9	The average RMSE over the nine tubes	85
6.10	Residuals for the different sampling schemes	86
6.11	Results for simulated data with a model error in the flip angles	89
6.12	The average RMSE of several simulated data with a model error in the flip angles	90

Abbreviations

2D	two-dimensional
3D	three-dimensional
ABC	approximate Bayesian computation
bSSFP	balanced steady-state free precession
CI	credible interval
CRB	Cramér–Rao bound
ECG	electrocardiogram
EM	expectation-maximization
EPG	Extended Phase Graph
EPI	echo-planar imaging
FISP	fast imaging with steady-state precession
FoV	field of view
GMRF	Gaussian Markov Random Field
HMC	Hamiltonian Monte Carlo
HPD	highest posterior density
IGMRF	intrinsic Gaussian Markov Random Field
INLA	Integrated Nested Laplace Approximation
MAP	maximum a posteriori
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
ML	maximum likelihood
MRF	Magnetic Resonance Fingerprinting
MRI	magnetic resonance imaging
NUTS	No-U-Turn-Sampler

Abbreviations

PDF	probability density function
PSF	point spread function
qMRI	quantitative magnetic resonance imaging
RMSE	root-mean-square error
SNR	signal-to-noise ratio
TE	echo time
TI	inversion time
TR	repetition time
VB	Variational Bayes

1

Introduction

1.1 Motivation

Medical imaging is a widely used tool in clinical practice. Images of the inner body can help to give fast and accurate diagnoses of diseases. Additionally, images can also be utilized for monitoring therapy. The oldest medical imaging technique is X-ray radiography; other techniques include ultrasound, X-ray computed tomography, endoscopy and magnetic resonance imaging (MRI). Each modality has its advantages and disadvantages and the right procedure has to be chosen dependent on the medical problem.

After its introduction in 1973 (Lauterbur (1973)), MRI has become commonly used in clinical practice due to its unique benefits which include a superior soft tissue contrast, good spatial resolution and insights into physiological processes. Additionally, MRI has no need to apply ionizing radiation in contrast to, e.g., X-ray computed tomography and is therefore non-invasive. MRI is applied for diagnosing many pathologies such as neurological disorders, heart diseases or cancer (Frisoni et al. (2010); Cawley et al. (2009); Menezes et al. (2014)). Usually, MRI produces images which are only qualitative in their contrast. This contrast is achieved by manipulating nuclear magnetization in the human body and exploiting the sensitivity of MRI to various physical and physiological related tissue properties. Nonetheless, a quantitative determination of these properties from the qualitative images is hardly possible.

An advantage of truly quantitative MRI measuring fundamental biological tissue properties is the possibility to compare parameter values from different subjects obtained at possibly different sites as well as to compare results from one subject over time. Since qualitative MR images also depend on the scanner hardware applying the same scanning sequence might produce different results at different scanners. Another advantage of quantitative MRI is the possibility to become more independent from visual inspection by physicians and/or to reduce their workload. Although highly trained physicians with a broad expertise in varying contrast and signal levels are thoroughly capable to extract information of qualitative images, quantitative information could be superior for the reliability of diagnostics or therapy monitoring. It would be helpful to give them assistance by using supplementary automatic

diagnostic tools. Databases collecting normal and abnormal tissue properties, possibly together with machine learning techniques (Cuocolo et al. (2019); Salvatore et al. (2015); Leiner et al. (2019)), could then assist physicians in diagnosing. The ability to quantify the tissue-related parameters makes it also possible to avoid the acquisition of multiple, differently weighted MR images of the same position since weighted images with any contrast could be generated from the quantitative parameters retrospectively.

Standard quantitative measurements of the diagnostically important relaxation times T_1 and T_2 suffer from long acquisition times which make them clinically unfeasible. Many efforts have been made to shorten the measurement time, leading to a variety of techniques, such as DESPOT1/DESPOT2 (Deoni et al. (2005)) and applying echo planar imaging (Clare and Jezzard (2001)). Nonetheless, these methods commonly require two separate scans for T_1 and T_2 estimation and are often highly sensitive to system imperfections, e.g. in the experimental conditions (Ma et al. (2013)).

Magnetic Resonance Fingerprinting (MRF) (Ma et al. (2013)) is a novel technique in quantitative MRI that attempts to overcome the above mentioned limitations of qualitative measurements. It allows for the simultaneous estimation of tissue-related parameters within short scanning time. The main idea is the acquisition of highly undersampled Fourier data leading to noisy, temporal magnetization images. The individual magnetization courses of each voxel will then be compared to a pre-computed dictionary which is storing a huge range of possible signal evolution. Due to the intrinsic high undersampling, the acquisition of MRF data is fast assuring its clinical feasibility. MRF has already been applied to patients with prostate cancer, brain tumors, epilepsy and stroke diseases (Yu et al. (2017); Badve et al. (2017); Liao et al. (2018); Lemasson et al. (2016)). The set of quantifiable parameters is dependent on the MRF sequence which should be designed in a way that it is sensitive to the sought parameters. Beside the quantification of the relaxation times T_1 and T_2 , it has also been possible to estimate, e.g., the static magnetic field B_0 , the radiofrequency field B_1 and blood flow velocities (Körzdörfer et al. (2019); Buonincontri and Sawiak (2016); Flassbeck et al. (2019)).

MRF provides a quantitative estimation of tissue-related parameters. Nonetheless, uncertainties that reliably characterize the accuracy of the estimates are essential to assess the significance of observed differences in single results which is particularly relevant in applications such as therapy monitoring. Additionally, probability statements could help in cases where differences between multiple scans of a single subject have been observed in order to analyze if the differences are significant and therefore indicate pathological changes of the underlying tissue. The interpretability of quantitative MRI and in particular MRF results would drastically improve if a reliable uncertainty quantification was available.

From a statistical point of view the analysis of MRF data represents a large-scale, nonlinear regression task. As there are more and more large-scale data sets available, solving such problems gains much interest in the statistical community. Applications range from the analysis of genome-wide association studies (Smith and Fahrmeir (2007)), logistic regression of text categorization (Genkin et al. (2007)) or Bayesian model averaging for functional magnetic resonance imaging (Smith and Fahrmeir (2007); Lee et al. (2014)). Bayesian statistics is a field in statistics in which the probability is interpreted to express one's degree of belief. In

contrast to frequentist statistics, Bayesian statistics can account for prior knowledge about the quantity of interest which can be based on, e.g., previous data or expert knowledge. As there is often some sort of knowledge available for, e.g., physical parameters, the possibility of its incorporation presents an advantage over frequentist statistics. In Bayesian statistics, all unknowns will be modeled as random variables. Probability statements can hence be made conditional on the observed data in terms of a posterior distribution which leads to a natural uncertainty characterization of the results. The ability to calculate probability statements about the estimates improves the interpretability of the outcome. Bayesian methods have already been carried out for MRI, e.g. for denoising images (Awate and Whitaker (2007)) and image segmentation (Xia et al. (2016); Marroquín et al. (2002)). Bayesian inference for large-scale problems can be challenging as the calculation of the result poses a difficult task. Standard procedures such as Markov chain Monte Carlo are prohibitive due to the small step sizes required when the dimensionality scales up (Beskos et al. (2009)).

In this thesis, a Bayesian data analysis for MRF will be developed. Usually, MRF only provides estimates for the tissue-related parameters. A Bayesian framework for the original, dictionary-based MRF approach shall be established which yields such an uncertainty quantification for the first time and is additionally capable of including prior knowledge about the sought parameters. This statistical model will rely on the same assumption as the original, dictionary-based MRF approach, namely that the aliasing artifacts in the approximately reconstructed magnetization courses due to the usage of an imperfect Fourier transformation should follow a Gaussian distribution. Furthermore, a second Bayesian statistical model shall be developed in which the data will be modeled in the Fourier domain directly which will avoid the aliasing artifacts. It will lead to a large-scale, nonlinear regression task which will be challenging to solve. The influence of different prior distributions, e.g. a Gaussian Markov Random Field prior which can account for spatial dependencies between the parameters, shall be investigated. Finally, the original, dictionary-based MRF approach shall be compared to the Bayesian Fourier space modeling approach with respect to the accuracy of the estimates dependent on the level of noise, the MRF sampling scheme and computational complexity as well as in terms of their sensitivity to errors in the physical model.

1.2 Objective and Aims

The goal of this work is to apply Bayesian statistics to further enhance the data analysis of Magnetic Resonance Fingerprinting.

The main objectives of this thesis can be summarized as follows:

- **Objective 1:** *Bayesian uncertainty quantification of the dictionary-based MRF estimates.* The dictionary-based MRF approach usually restricts itself to the calculation of estimates for the tissue-related parameters. We aim to develop a novel and full probability characterization for these estimates. For the first time it will thus be possible to assign uncertainties to the dictionary-based MRF estimates. Instruments from Bayesian statistics shall be utilized leading to probability distributions of the relaxation times in every voxel. These uncertainties shall be validated and applied to phantom and in vivo data.

- **Objective 2: *Bayesian inference for a large-scale k -space MRF model with different prior distributions for the tissue-related parameters.*** The dictionary-based MRF approach introduces aliasing errors in the approximately reconstructed magnetization courses due to the usage of an imperfect inverse Fourier transformation. Previous work already applied a maximum likelihood estimation to a statistical model in k -space but no Bayesian framework was utilized so far. We aim to develop a general Bayesian inference for large-scale regression problems. MRF with a statistical model for the data in the Fourier domain is an example of such a problem and avoids the usage of an imperfect inverse Fourier transformation. We also aim to investigate the influence of different prior distributions for the propriety of the resulting posterior distributions as well as for the accuracy of the estimates of the tissue-related parameters.
- **Objective 3: *Comparison of the dictionary matching and the Bayesian k -space modeling approach.*** The dictionary-based MRF approach is easy to implement and can be calculated fast. The maximum a posteriori estimation of the Bayesian k -space modeling approach on the contrary leads to a high-dimensional and non-convex optimization task which is time consuming and can be challenging to solve. We aim to compare these two approaches in terms of the accuracy of the estimates dependent on the level of noise, the MRF sampling scheme and computational complexity as well as in terms of their sensitivity to errors in the physical model. We also aim to give guidance on whether it is worth to compute the maximum a posteriori estimate of the Bayesian k -space modeling approach or whether it is sufficient to only calculate the dictionary-based MRF estimate.

1.3 Thesis Structure and Overview

Chapter 2 (Magnetic Resonance Imaging and Magnetic Resonance Fingerprinting) provides an introduction to the basic principles of MRI and the concept of MRF. The fundamental physics of MR imaging are briefly recalled to understand the MR measurement process and lay out the parameters of interest. This is followed by a detailed explanation of the framework of dictionary-based MRF. In the end a review of the current state-of-the-art in MRF is presented.

Chapter 3 (Bayesian Statistics) describes the principles of Bayesian statistics as far as it is needed for the thesis. After a small motivation, the key elements of Bayesian statistics are presented. Advantages over frequentist statistics as well as their connection are discussed and important definitions are made. A special prior distribution is introduced which will be used in a later chapter. The calculation of Bayesian results usually requires numerical techniques some of which are reviewed in this chapter. In addition, the challenges for the task of MRF due to its high-dimensionality are discussed and appropriate methods are introduced.

Chapter 4 (Bayesian Uncertainty Quantification for Dictionary-Based MRF) presents a Bayesian uncertainty quantification which is a novelty as it reliably characterizes the uncertainties for the original, dictionary-based MRF approach for the first time. Initially, a statistical analysis of the original, dictionary-based MRF approach is conducted leading to

a statistical model for the approximately reconstructed magnetizations. Together with prior distributions for the dictionary-based MRF estimates, this leads to a joint posterior distribution of the relaxation times. Analytical expressions for the joint posterior can be derived and numerical quadrature techniques yield probability distributions for the two relaxation times in every voxel. The calculation of these distributions is based on the pre-computed dictionary; it is hence easy to employ and can be carried out fast. Applicability to simulated, phantom and in vivo data is shown.

Chapter 5 (Large-Scale Bayesian Modeling) introduces a general statistical model for a class of nonlinear, large-scale regression problems and several prior distributions. The propriety of the resulting posterior distributions is explored as well as the existence of moments. MRF is an example of such a problem if a statistical model is considered in which the data are directly modeled in the Fourier domain. Results of the estimation of the maximum a posteriori estimates of the different posterior distributions are shown for simulated data.

Chapter 6 (Comparison of Dictionary Matching and Bayesian Modeling) compares the dictionary-based MRF and Bayesian modeling approach with respect to the accuracy of the estimates dependent on the level of noise, the MRF sampling scheme, the computational complexity and lastly in terms of their sensitivity to errors in the physical model.

1.4 List of Publications

During my PhD studies, the following work has been published.

Peer-reviewed journals

- [1] Metzner, S., Wübbeler, G. and Elster, C., (2019). Approximate large-scale Bayesian spatial modeling with application to quantitative magnetic resonance imaging. *AStA (Advances in Statistical Analysis)*, <https://doi.org/10.1007/s10182-018-00334-0>.
- [2] Metzner, S., Wübbeler, G., Flassbeck, S., Gatefait, C., Kolbitsch, C. and Elster, C., (2021). Bayesian uncertainty quantification for magnetic resonance fingerprinting. *Phys. Med. Biol. (Physics in Medicine & Biology)*, <https://doi.org/10.1088/1361-6560/abeae7>.

Conference/Workshop Talks

- [3] Wübbeler, G., Metzner, S. and Elster, C., (2016). Quantitative Magnetic Resonance Imaging: A Regression Approach. In: *MATHMET (International Workshop on Mathematics and Statistics for Metrology)*.
- [4] Metzner, S., Wübbeler, G. and Elster, C., (2017). A spatial statistical framework for Magnetic Resonance Fingerprinting. In: *Spatial Statistics*.
- [5] Elster, C., Lehnert, J., Metzner, S. and Wübbeler, G., (2019). Application of Gaussian Markov random field priors for Bayesian spatial modeling. In: *MATHMET (International Workshop on Mathematics and Statistics for Metrology)*.

Conference/Workshop Posters

- [6] Metzner, S., Wübbeler, G., Kolbitsch, C., Schaeffter, T. and Elster, C., (2017). A Bayesian procedure for Magnetic Resonance Fingerprinting. In: *ISMRM (International Society for Magnetic Resonance in Medicine) Workshop on Magnetic Resonance Fingerprinting*.
- [7] Metzner, S., Wübbeler, G. and Elster, C., (2017). Large-Scale Bayesian Inference for Magnetic Resonance Fingerprinting. In: *ISBA (International Society for Bayesian Analysis)*.

Further publications

- [8] Satzer, B., Richter, U., Lippmann, U., Metzner, S., Notni, G. and Gross, H., (2015). Using the 3D-SMS for finding starting configurations in imaging systems with freeform surfaces. *Proc. SPIE (Society of Photo-Optical Instrumentation Engineers) 9626.*, <https://doi.org/10.1117/12.2191240>.

2

Magnetic Resonance Imaging and Magnetic Resonance Fingerprinting

This chapter gives an overview of Magnetic Resonance Imaging (MRI) and Magnetic Resonance Fingerprinting (MRF). In the first part of this chapter basic principles of MRI are presented. For a more detailed review the reader is referred to Brown et al. (2014); Vlaardingerbroek and Boer (2013); Nishimura (1996); Weishaupt et al. (2008) and Abragam (1983).

In the second part of this chapter, MRF as a fast, quantitative imaging modality is described. MRF was published in 2013 by Dan Ma and others (Ma et al. (2013)) and received growing interest in the MR community. On the one hand, it is based on a pulse sequence which is sensitive to the parameters of interest (e.g. T_1 and T_2). On the other hand, the Fourier domain will be undersampled leading to a very fast acquisition. A state-of-the-art in MRF is presented subsequently. Finally, reference values for the considered data sets throughout this thesis are shown.

2.1 Basic Principles of Magnetic Resonance Imaging

Magnetic Resonance Imaging is a powerful, non-invasive medical imaging technique. It is widely used in clinical practice to provide anatomical and functional information from the human body which can then be used to diagnose and monitor medical conditions. Diagnostic applications of MRI include heart and liver diseases, cancer, arthritis and many more.

The nucleus is the source of the MR signal. Atoms with an odd number of protons and/or odd numbers of neutrons possess a property called spin. In a visual interpretation that means that the proton spins around its axis. The atom ^1H has one proton and due to its high abundance in the human body, clinical imaging is mostly based on hydrogen. The spin has two important characteristics, the angular momentum J and the magnetic moment μ . Both are proportional, i.e. $\mu = \gamma J$ and γ denotes the gyromagnetic ratio. The angular momentum can be interpreted as a spinning top trying to maintain its orientation along the rotational axis. The spin also behaves like a small magnet which is called the magnetic moment. For

2. Magnetic Resonance Imaging and Magnetic Resonance Fingerprinting

MRI, it is now important what is happening when the magnetic moments are brought into a strong, external magnetic field B_0 . The magnetic moments will then orientate spin-up or spin-down to the field with a certain angle. They precess with the angular frequency which is proportional to the strength of the magnetic field, the so-called Larmor frequency ω_0 :

$$\omega_0 = \gamma B_0,$$

where γ denotes the gyromagnetic ratio and B_0 the strength of the magnetic field. The quantity γ is dependent on the nucleus and is, e.g., equal to 42.58 MHz/T for hydrogen.

The magnetic field B_0 shall be, without loss of generality, directed in the z -direction. The magnetic moments of the spins can either be oriented parallel (spin-up) or anti-parallel (spin-down) to B_0 . These two states differ by an energy difference:

$$\Delta E = \gamma \hbar B_0,$$

for $\hbar = h/(2\pi)$ and h denotes the Planck constant. If as many spins are anti-parallel as are parallel the net magnetization M_z would be 0. Due to the Boltzmann distribution, the spin-up state is commonly more populated than the spin-down state in equilibrium (at room temperature). This leads to an equilibrium magnetization $M_0 > 0$ which is oriented along B_0 . An electromagnetic wave can be used to induce transitions between spin-up and spin-down states (excitation). After this wave, the system transitions back into the equilibrium state (relaxation). In order to excite the spin system, the frequency of the electromagnetic wave has to be the Larmor frequency. This is the reason for the name *magnetic resonance*.

After the excitation with the radiofrequency pulse the magnetization returns to its equilibrium. The transverse magnetization M_{xy} decays and the longitudinal magnetization M_z will restore to its initial magnetization M_0 . The time until (1-1/e) (around 63%) of the longitudinal magnetization is recovered is called the T_1 (longitudinal, spin-lattice) relaxation. Here, e denotes Euler's number. The following differential equation describes the relationship, cf. Figure 2.1:

$$\frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1}.$$

The solution is given by

$$M_z = M_0 + (M_z(0) - M_0) \exp\left(\frac{-t}{T_1}\right).$$

The T_2 relaxation time on the other hand describes the loss of coherence due to spin-spin interactions. After the excitation through a radiofrequency pulse the spins are in the same phase, i.e. the spins precess synchronously in the xy -plane. The time until M_{xy} is reduced to 1/e (around 37%) of the initial transverse magnetization is called the T_2 (transverse, spin-spin) relaxation. The differential equation below describes the relationship of T_2 and M_{xy} , cf. Figure 2.1:

$$\frac{dM_{xy}}{dt} = -\frac{M_{xy}}{T_2}.$$

The solution is given by

$$M_{xy} = M_{xy}(0) \exp\left(\frac{-t}{T_2}\right).$$

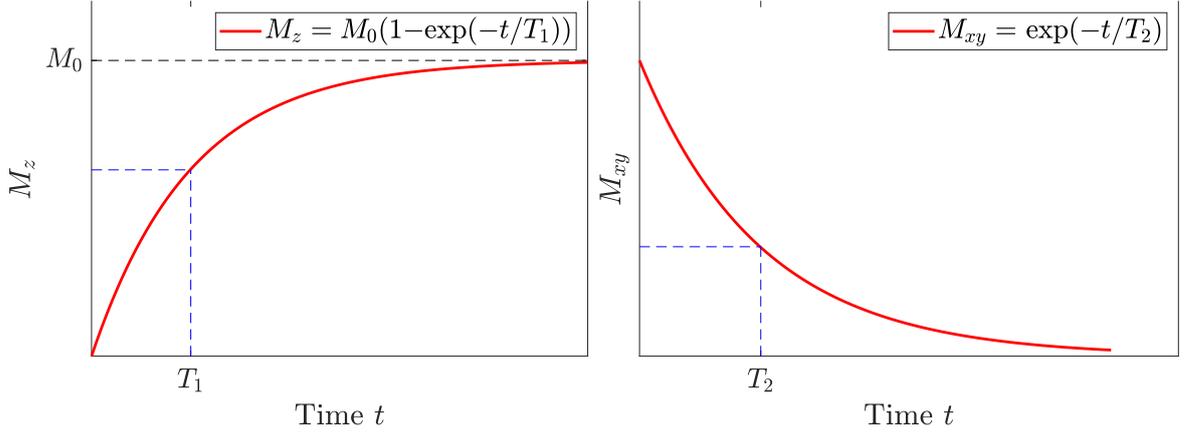


Figure 2.1: Left: The relaxation time T_1 corresponds to the time when $(1-1/e)$ of the longitudinal magnetization M_z is restored. Right: The relaxation time T_2 indicates the time when the transverse magnetization M_{xy} is decreased to $1/e$ of its initial value.

Table 2.1: Approximate values of T_1 and T_2 for different tissue types at 1.5 Tesla taken from Brown et al. (2014).

Tissue	T_1 in ms	T_2 in ms
Gray matter	950	100
White matter	600	80
Muscle	900	50
Cerebrospinal Fluid	4500	2200
Fat	250	60

In Table 2.1 approximate T_1 and T_2 values are shown for different tissue types at 1.5 Tesla (Brown et al. (2014)). Note that T_2 is always smaller than T_1 .

After introducing the relaxation times, it is now possible to derive the dynamic of the whole magnetization. The Bloch equations characterize the temporal development of the magnetization $M(t)$ in the presence of magnetic fields $B(t)$. They are given as the following coupled differential equation (Nishimura (1996)):

$$\frac{dM}{dt} = M \times \gamma B - \frac{M_x l + M_y m}{T_2} - \frac{(M_z - M_0)n}{T_1}, \quad (2.1)$$

where M_0 is the equilibrium sample magnetization due to the static magnetic field B_0 and l, m, n are the unit-vectors in the x, y, z -directions, respectively. In MR imaging, B consists of three different types of magnetic fields: the static magnetic field B_0 , the radiofrequency fields $B_1(t)$ and the gradient fields $G(t)$. $B_1(t)$ is used for excitation and $G(t)$ consists of three gradients for each direction used for the spatial localization. Both are time-varying other than the static, external field B_0 .

The transverse magnetization is now modeled as a complex number by $M := M_x + iM_y$ with initial condition $M^0 = M_x^0 + iM_y^0$.

Without loss of generality, we will restrict the following formulas to 2D (via a slice-selection of a plane centered at $z = z_0$ with thickness Δz). When a sample is measured, in addition to the time t also the position $r = (x, y)$ has to be taken into account. The signal at time t

induced in the receiver coils is the sum over the entire magnetization in a certain volume:

$$s_r(t) = k \int_x \int_y M(r, t) dx dy, \quad k > 0. \quad (2.2)$$

In (2.2) the equation for the measured signal at time t is stated. For an inhomogeneous sample the signal of each volume element (voxel) is of interest. A spatial encoding is therefore necessary. The main idea is that the Larmor frequency is a spatial function of the magnetic field strength. Three magnetic field gradients for each direction are required in order to make the homogeneous main field B_0 change linearly. As explained before, only spins which Larmor frequencies matching the B_1 field will be excited. The rest will not be affected. Starting with the slice-selection, a gradient in the z -direction will be applied (together with the radiofrequency pulse). Hence, every slice in the magnetic field experiences a different magnetic field strength and only those spins having the same frequency as the radiofrequency pulse will be excited. In order to encode the x and y position a phase and a frequency encoding gradient are needed. The phase encoding gradient is applied along an axis, e.g. the y -axis. It is switched on after the excitation of the spins. The spins precess in the xy -plane. The y -gradient changes the Larmor frequency along the axis during a time T_y , resulting in a phase shift of the spins after a certain time. Turning the gradient off, the spins will return to their precession but having a phase shift relative to each other as a function of y . Each line within the slice can now be identified by its unique phase. The frequency encoding gradient is applied in the x -direction. The gradient changes the Larmor frequency and the spins precess slower on the one side and faster on the other. Each column of the slice is therefore identified with a specific frequency. Altogether, every voxel now has a unique phase and frequency. The signal is measured during the frequency encoding. In order to get all the x and y positions, this process of applying phase and frequency encoding has to be repeated N_y times. In the case of a Cartesian sampling (cf. Figure 2.3) a phase and a frequency encoding gradient are applied one after another and thus one line in the x -direction is measured. This is repeated N_y times for different values of the y -gradient resulting in N_y different k -space lines.

In a setting with a time-varying gradient, equation (2.2) now becomes

$$s_r(t) = \int_x \int_y m(x, y) \exp(-i\omega_0 t) \exp\left(-i\gamma \int_0^t G(\tau) r d\tau\right) dx dy.$$

Here, $m(x, y)$ corresponds to the integral of $M^0(x, y, z)$ over the selected slice with respect to z . This equation is strongly connected to the Fourier transformation. By ignoring the term $\exp(-i\omega_0 t)$, the measured signal can be computed according to

$$s(t) \propto \int_x \int_y m(x, y) \exp(-i2\pi[k_x(t)x + k_y(t)y]) dx dy \quad (2.3)$$

$$= \mathcal{F}(m(k_x(t), k_y(t))), \quad (2.4)$$

for $k_x(t) = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau$ and $k_y(t) = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau$ (cf. Nishimura (1996)). For every time point t the signal $s(k_x(t), k_y(t))$ is a 2D-Fourier transformation of the magnetization $m(x, y)$ in spatial frequency space. The Fourier domain is called k -space. In Figure 2.2 an example of

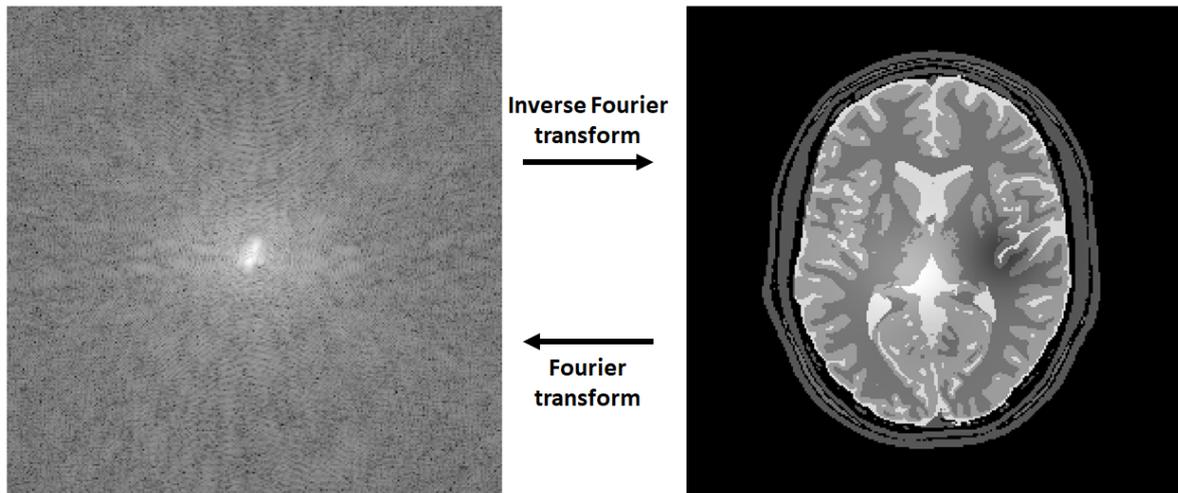


Figure 2.2: Exemplary image of Fourier data and its corresponding magnetization image. Note that the logarithm of the Fourier data is plotted.

a magnetization image and the corresponding Fourier data is presented. The formula in 3D can be derived similarly leading to a 3D-Fourier representation.

In general, an inverse Fourier transformation can be used to reconstruct the image from the acquired k -space data. For a Cartesian sampling the inverse fast Fourier transform (Cooley et al. (1969)) can be used. If the Nyquist criterion is fulfilled the image can be reconstructed artifact-free (except for Gibbs artifacts that occur because of the finite number of samples). In order to achieve faster imaging, it is possible to undersample the k -space, i.e. to sample with a density lower than the Nyquist rate. In Figure 2.3 the Cartesian red lines correspond to a Cartesian undersampling scheme. An undersampling in the k -space will lead to aliasing artifacts in the image domain. Mathematically, aliasing can be analyzed by the point spread function (PSF) (Sarty (2003)) which is a general concept describing imaging systems. The PSF captures aliasing by any signal which is not a delta peak in the center. In addition, a broadening of the central peak can also describe artifacts. The same number of acquired k -space data can lead to very different aliasing artifacts and the appearance of aliasing is dependent on how the samples are taken. There are different sampling schemes in the Fourier domain. Figure 2.3 shows a Cartesian, a radial and a spiral sampling scheme.

The usage of the fast Fourier transform requires that the k -space data are sampled on a rectangular grid. It cannot be directly applied to non-Cartesian sampling (such as spiral and radial, cf. Figure 2.3). For these trajectories it is possible to interpolate the acquired k -space data onto a Cartesian grid and then use the inverse fast Fourier transform. Since the interpolation is not exact, the inverse Fourier transformation will be imperfect.

In addition to the sampling scheme, an MR sequence is dependent on scanning parameters, namely the repetition time TR , echo time TE and the flip angle α . The repetition time is the interval between two consecutive excitations of the same slice. The echo time is the time between the center of the radiofrequency pulse and the formation of the echo, i.e. when the net area of the readout gradient is 0. The flip angle describes the rotation of the net magnetization M_z by the radiofrequency pulse.

Overall, an MR sequence is an ensemble of radiofrequency pulses and gradient fields. The effect on a single magnetization vector can be described by the Bloch equation (2.1). Note

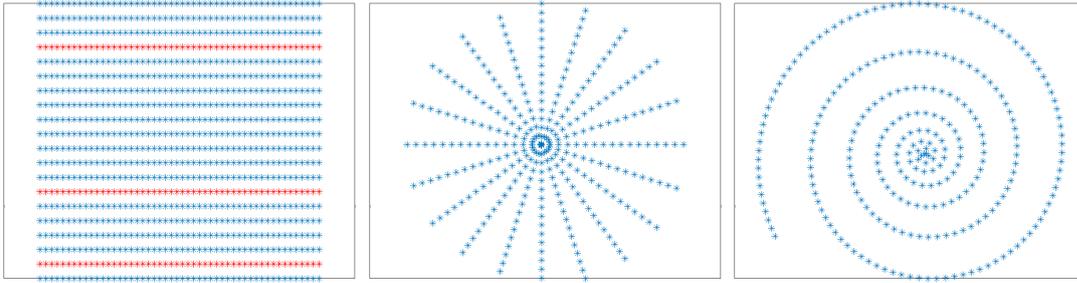


Figure 2.3: Different sampling schemes in k -space. From left to right: a Cartesian, a radial and a spiral trajectory.

that an excitation pulse corresponds to a rotation of the magnetization vector by α (Scheffler and Lehnardt (2003)). In addition, T_1 and T_2 relaxation occurs. For special sequences, it can be shown that under certain conditions the magnetization reaches a steady-state after a number of repetition times. There are different steady-state sequences which differ in the steady-state they reached. Dependent on its steady-state a sequence can be designed in a way such that it is sensitive to the parameters of interest (e.g. T_1). For example for a spin echo sequence, a short TR and a short TE will lead to a T_1 -weighted image, whereas a long TE and a long TR produces a T_2 -weighted image.

In this thesis, we will consider two different steady-state sequences, a balanced steady-state free precession (bSSFP) and a fast imaging with steady-state precession (FISP) sequence. In Appendix A and Appendix B their magnetization models are derived. Both have the advantage that they are sensitive to T_1 and T_2 . A bSSFP sequence ensures that the net gradient area after each TR is 0. This leads to a very high signal amplitude. Nevertheless, this also makes it very sensitive to B_0 off-resonance effects which can lead to banding artifacts. FISP overcomes this problem using an unbalanced gradient moment in each TR but at the cost of a lower signal amplitude.

2.2 Concept of Magnetic Resonance Fingerprinting

After introducing the basic concept of MRI, we will now consider the framework of Magnetic Resonance Fingerprinting (MRF). In MR usually only qualitative images are acquired. Nevertheless, for diagnostic purposes the quantitative values of the tissue-related parameters (e.g. the relaxation times T_1 and T_2) can also be of interest. An estimation of these parameters may result in improved diagnostic insights (Tofts (2005)). The gold standard for quantitative T_1 mapping is an acquisition of multiple time points along the signal recovery curve after an inversion or a saturation pulse. For T_2 a Carr-Purcell-Meiboom-Gill spin echo sequence is the gold standard where the T_2 decay is measured at different echo times (Jiang et al. (2015)). These techniques provide very accurate estimates of T_1 and T_2 but take too long to apply in clinical practice. Accelerated T_1 and T_2 mapping sequences have been proposed for use in patients. Nevertheless, commonly two separate scans are required for T_1 and T_2 . In 2013 Dan Ma and others (Ma et al. (2013)) proposed MRF, a fast quantitative MR technique which provides T_1 and T_2 maps in a single fast acquisition. We will now consider the basic principle; recent improvements and applications are subsequently presented.

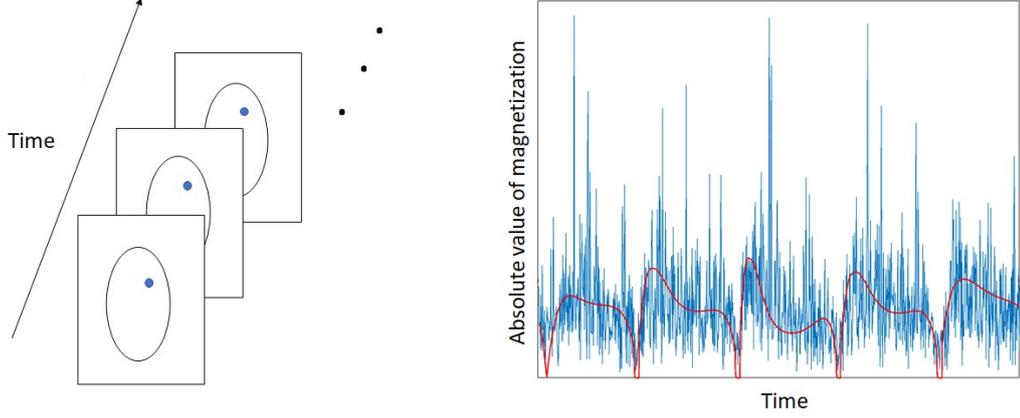


Figure 2.4: Concept of MRF: sequence of images which are obtained by applying a pseudo-inverse to the undersampled measurements in Fourier space (left) and temporal development of magnetization at a single voxel (blue, right). The red curve corresponds to the matched dictionary entry, the so-called "fingerprint". Tissue parameters at each voxel are extracted from the fingerprint. Figure from Metzner et al. (2021).

Let N denote the number of considered voxels in the selected slice. The proton spin density ρ and the sought parameters θ (e.g. T_1 and T_2) are vectors of length N . From a physical point of view, the proton spin density is a real number. In order to account for phase changes in the measured data (e.g. due to susceptibility effects), it is common in MR to model ρ in terms of a complex number. We will consider both cases in the following.

Let $m_l = m_l(\theta, \rho), l = 1, \dots, L$ denote the $N \times 1$ complex-valued vector of magnetizations at the N voxels after applying the l -th excitation pulse, and let m^i denote the $L \times 1$ complex-valued vector of magnetizations at the i -th voxel after applying all L pulses, cf. Figure 2.4.

The sequence of magnetizations at the i -th voxel, m^i , depends only on the unknown tissue parameters at that voxel (and the known parameters of the applied pulse sequence such as TR and TE), i.e. $m^i = m^i(\theta^i, \rho^i)$. Note that the model $m_l = m_l(\theta, \rho)$ depends linearly on ρ .

The measured data from the MR scanner consist of the complex-valued vectors

$$z_l = \mathcal{F}_l(m_l), \quad l = 1, \dots, L, \quad (2.5)$$

and the $n_l \times 1$ Fourier data z_l are modeled as the Fourier transformations of the magnetization vectors m_l .

One key principle of MRF is taking a sequence with an undersampling pattern in k -space. Only small parts of the Fourier domain are measured in every image, but such that for every image different points are acquired. Hence, the mapping \mathcal{F}_l has to be indexed by l . The number of measurements n_l is therefore much smaller than the number of unknowns N and thus the magnetization m_l cannot be reconstructed exactly.

For MRF, the magnetizations $m_l, l = 1, \dots, L$, are approximately reconstructed in a first step from (2.5) through

$$\hat{m}_l = \mathcal{F}_l^\dagger(z_l), \quad l = 1, \dots, L, \quad (2.6)$$

where \mathcal{F}_l^\dagger denotes a pseudo-inverse of the Fourier transform mapping \mathcal{F}_l in (2.5). From the sequence of estimated magnetizations $\hat{m}_1, \dots, \hat{m}_L$, the estimated course of magnetization \hat{m}^i for each voxel i is then obtained and subsequently denoted by

$$\hat{m}^i = ((\hat{m}_1)_i, \dots, (\hat{m}_L)_i)^T, \quad (2.7)$$

cf. Figure 2.4. The estimates are generally not accurate as they contain aliasing errors caused by the imperfect inversion \mathcal{F}_l^\dagger .

In a second step a dictionary $\{D_k\}_{k=1}^K$ with possible magnetization courses is computed (e.g. via the Bloch equations (2.1)). Let \tilde{m} denote the magnetization model m but without the linear factor ρ . A set of potential $\{\theta_k\}_{k=1}^K$ is constructed. For each of those values the $L \times 1$ magnetization $\tilde{m}(\theta_k)$ for $k = 1, \dots, K$ is computed and stored in the $L \times K$ dictionary

$$D_k = \tilde{m}(\theta_k). \quad (2.8)$$

The magnetization at each voxel depends on the tissue parameters of only that voxel. Hence, in a third step a voxel-wise comparison of the reconstructed magnetizations (2.7) with every entry of the dictionary is performed. For each voxel, θ_k is determined as that parameter of the magnetization model for which the corresponding dictionary entry shows highest correlation with the time series obtained from the measured data.

In more detail, for a real proton density $\rho \in \mathbb{R}^N$ the dictionary matching solution (cf. Davies et al. (2014)) $\hat{\theta}_i$ in voxel i is chosen to be the entry \hat{k}_i of the dictionary $\{D_k\}_{k=1}^K$ which maximizes

$$\hat{k}_i = \operatorname{argmax}_k \frac{\operatorname{Re}(\langle \hat{m}^i, D_k \rangle)}{\|D_k\|},$$

where $\langle \cdot, \cdot \rangle$ denotes the complex scalar product. The proton density is calculated according to

$$\hat{\rho}_i = \frac{\operatorname{Re}(\langle \hat{m}^i, D_{\hat{k}_i} \rangle)}{\|D_{\hat{k}_i}\|^2}.$$

For $\rho \in \mathbb{C}^N$, it is

$$\hat{k}_i = \operatorname{argmax}_k \frac{|\langle \hat{m}^i, D_k \rangle|}{\|D_k\|}$$

and

$$\hat{\rho}_i = \frac{\langle \hat{m}^i, D_{\hat{k}_i} \rangle}{\|D_{\hat{k}_i}\|^2}.$$

We will refer to this as the *dictionary matching* method or the *dictionary-based MRF* approach. Note that the dictionary-based MRF only works if a model for the magnetization is available. This physical model will be an approximation of the underlying physics and the quality of the results depend on the quality of the physical model.

A second key principle of MRF relies in the specific sequence choice. As described above, two common choices for MRF are bSSFP and FISP sequences. Usually, an additional inversion pulse at the beginning of these sequences is applied for MRF in order to further improve the T_1 sensitivity. Note that the magnetization model m_l changes accordingly to the sequence.

2.3 Research Areas in Magnetic Resonance Fingerprinting

Many different applications and modifications of MRF haven been presented since first introducing MRF in 2013. An overview of the developments are given e.g. by Panda et al. (2017) and Bipin Mehta et al. (2019). In the following, we will present basic ideas and applications without claim of completeness.

The original MRF paper (Ma et al. (2013)) uses an inversion recovery bSSFP sequence with varying flip angles. This sequence is sensitive to T_1 , T_2 and the off-resonance frequency δf . After each radiofrequency pulse, a variable density spiral is used for the readout. Repetition times and flip angles are pseudorandomly varied. Since a steady-state is not established in this bSSFP sequence, Assländer et al. (2017) applied a pseudo steady-state free precession sequence. According to the flip angle pattern, TE and TR are adjusted in order to restore this spin echo-like signal behavior. The bSSFP sequence is sensitive to T_1 and T_2 , but also to the off-resonance frequency δf . Off-resonance effects can lead to banding artifacts. To avoid these, a FISP sequence (with an inversion pulse and varying flip angles), which is insensitive to δf is presented for MRF by Jiang et al. (2015). Note, that the signal-to-noise ratio is decreased when applying this sequence. They applied a sinusoidal variation of the flip angles, random repetition times as well as a spiral readout. Rieger et al. (2017) used a rapid echo-planar imaging (EPI) approach. In many applications a spiral readout is performed, whereas Rieger et al. (2017) and Koolstra et al. (2019) employed a Cartesian and Flassbeck et al. (2019) a radial trajectory.

The resolution of the dictionary plays an important role in MRF. The larger the dictionary the more time is needed for the matching process. McGivney et al. (2014) compressed the dictionary using a singular value decomposition in the time domain which, in their case, led to an acceleration of the matching step by a factor of around 4. Even these low rank dictionaries can still be too large to store. A compression in the property domain where low rank approximations of the dictionary are calculated online and therefore do not need to be stored was presented by Yang et al. (2018). They used a randomized singular value decomposition and fitted the dictionary with polynomials in its low rank space. Another possibility is a fast group matching algorithm as proposed by Cauley et al. (2015). Dictionary entries with similar signal evolutions are put into groups. In the matching process, the signal evolution of each voxel is placed into one of these groups and then only in this part of the dictionary a pattern matching is carried out.

Recently, many machine learning techniques are applied to MRF to overcome the discrete nature of the dictionary. Cohen et al. (2018) used a TensorFlow framework which was trained on simulated data. A two-step deep learning model which also considers the spatial association among neighboring voxels was proposed by Fang et al. (2019). Virtue et al. (2017) took a complex neural net where the MRF signals are mapped onto the sought parameters.

MRF can be applied to several regions of the body. Ma et al. (2013) showed results for quantitative brain images. The abdomen was measured by Chen et al. (2016), the heart by Hamilton et al. (2017). A non-electrocardiogram (non-ECG) triggered cardiac MRF was proposed by Jaubert et al. (2020). MRF was also applied to patients with prostate cancer, brain tumors, epilepsy and stroke diseases (cf. Yu et al. (2017); Badve et al. (2017); Liao et al. (2018); Lemasson et al. (2016)).

The concept of MRF is versatile in the way that in addition to T_1 , T_2 and ρ also further parameters can be estimated as long as it is possible to establish a physical model for all of the sought parameters. Flassbeck et al. (2019) calculated blood flow velocities, Christen et al. (2014) microvascular characteristics and Körzdörfer et al. (2019) computed the magnetic field. Due to imperfections in the slice profile and local changes in the B_1 field, errors in the quantitative maps can occur. Buonincontri and Sawiak (2016) estimated the B_1 field and Ma et al. (2017) B_1 field and the slice profile.

The concept of MRF can be expanded to a 3D acquisition, e.g. by Liao et al. (2017) and Ma et al. (2018) using additional encoding along the 3rd dimension.

In most applications one voxel is only supposed to contain one tissue type with its associated tissue parameters. This is a simplification since it is possible that multiple tissues can exist in one voxel, especially in regions where one tissue type omit into another. This issue is called the partial volume effect and was discussed for MRF by, e.g., Deshmane et al. (2019) and Tang et al. (2018). McGivney et al. (2018) addressed a Bayesian framework incorporating prior knowledge about the weights for the different tissues in a voxel to this problem. In their approach multiple dictionary entries are allowed to contribute to the signal and a Bayesian statistical model is assumed for the corresponding weights. A prior which promotes sparsity is chosen and the maximum a posteriori estimate of the posterior distribution for the weights of the dictionary entries is calculated.

The MRF sequence is dependent on many experimental related parameters, such as the echo times, repetition times, flip angles and the number of acquired images. Zhao et al. (2018) used the Cramér-Rao bound in order to design an optimal experiment. The Cramér-Rao bound is a lower bound on the covariance of an unbiased estimator. Three different ways to judge the encoding capability were presented by Sommer et al. (2017). One of their measures for the encoding capability was a Monte Carlo method which could accurately predict the influence of different flip angle patterns. To simplify the Monte Carlo approach, an easy quality factor that states the efficiency of the MRF sequence was proposed by Kara et al. (2019). This factor is a function of T_1 and T_2 and it was shown to predict their errors accurately.

The dictionary-based approach is easy to implement and leads to good results. Nonetheless, a pseudo-inverse Fourier transformation has to be applied onto the measured k -space data. Due to the undersampling, severe aliasing artifacts arise. A first attempt into direct k -space modeling avoiding this was made by Davies et al. (2014). They introduced a compressed sensing framework and proposed an iterative projected Landweber algorithm for the estimation of the parameters. By proving the restricted isometry property convergence to the sparse solution was shown in the noiseless case. Low-rank methods that take advantage of the MR signal structure are proposed by, e.g., Doneva et al. (2017) and Mazor et al. (2018). The former reconstruct the missing k -space samples by exploiting the fact that the MR images are low-rank in the

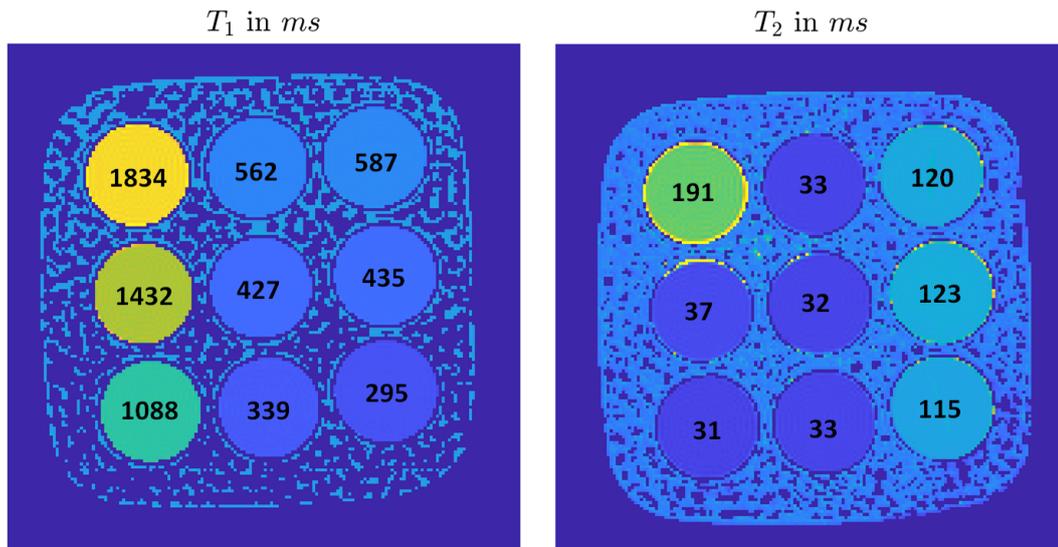


Figure 2.5: Results of T_1 and T_2 reference measurements of the phantom. The mean value of each tube is displayed.

temporal domain. A statistical maximum likelihood estimation based on variable splitting, the alternating direction method of multipliers, and the variable projection method was proposed by Zhao et al. (2016). The first iteration in their algorithm coincides with the usual dictionary matching method and the results improve with every further step. Wübbeler and Elster (2017) also applied a large-scale optimization method solving the maximum likelihood estimation. A trust-region algorithm which uses a sparse approximation of the Hessian is used and uncertainties associated with the estimates can be computed.

2.4 Phantom and Simulations

Throughout this thesis, phantom data as well as simulated data are used in order to verify the proposed methods. They are introduced in the following and images of the ground truth of T_1 and T_2 are shown.

In Chapter 4 a phantom consisting of nine tubes will be considered. Reference values were determined by standard quantitative MR measurements (details are given in Chapter 4). The ranges of the T_1 and T_2 values cover realistic human applications and are displayed in Figure 2.5.

Additionally, a simulated phantom with similar values for the nine tubes was created. The T_1 and T_2 values in each tube are constant and displayed in Figure 2.6. The simulated phantom will be used in Chapter 4 and 6.

In Chapter 5 and 6 simulated brain data of a slice of the human brain was taken from BrainWeb (2018). The values from BrainWeb (2018) were constant for each tissue type. In order to be more realistic the T_1 and T_2 maps were altered by adding four randomly chosen Gaussian peaks. Each peak has a standard deviation of 16 voxels and an absolute height equal to 10% of the value of the corresponding tissue map at the center of the Gaussian peak. In Figure 2.7 the resulting T_1 and T_2 values are shown.

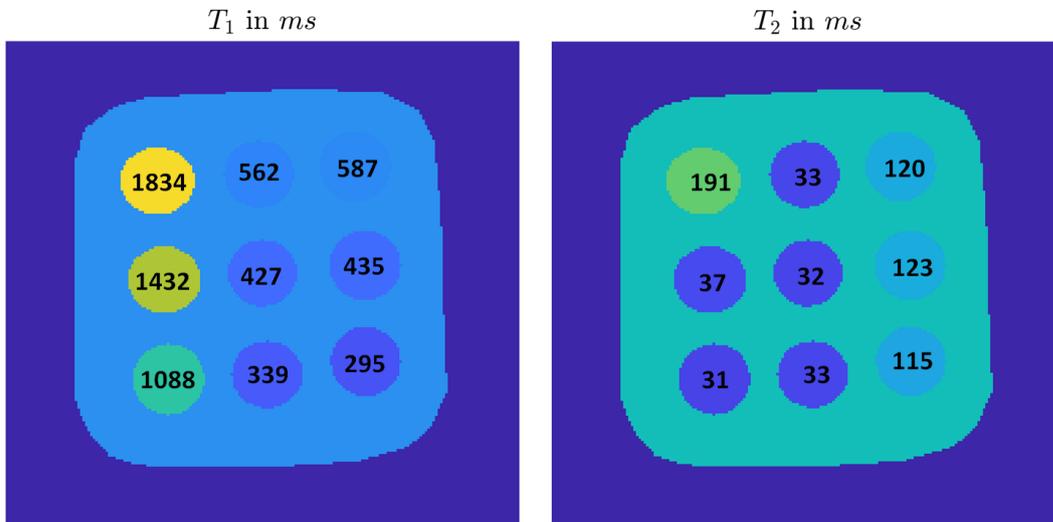


Figure 2.6: T_1 and T_2 values of the tubes of the simulated phantom. The displayed values are constant throughout each tube.

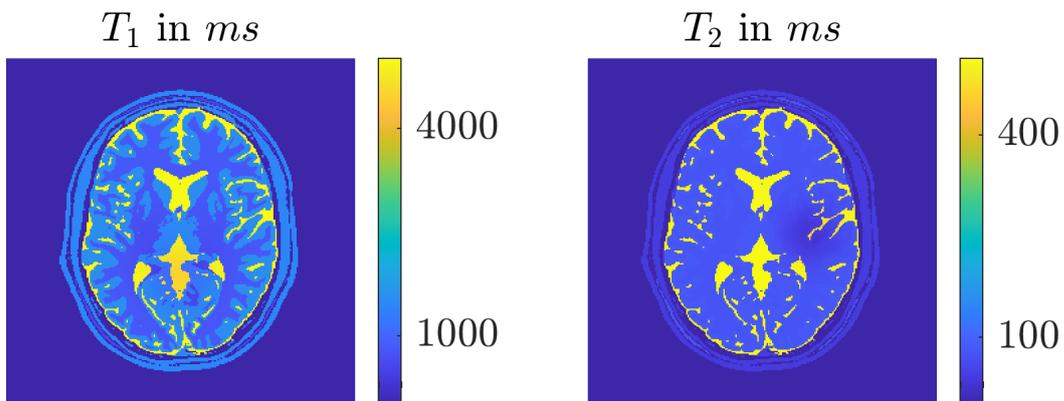


Figure 2.7: T_1 and T_2 values of simulated brain data set.

3

Bayesian Statistics

This chapter provides principles of Bayesian statistics which will play an important role in this thesis. Bayesian statistics is a field of statistics in which probability is interpreted to express one's degree of belief. A main advantage over frequentist statistics is the possibility to incorporate prior knowledge about the quantity of interest in terms of its prior distribution. A second important advantage is that probability statements can be made conditional on the data, i.e. after they have been observed. This is not possible for methods from frequentist statistics. There are numerous text books covering different aspects of Bayesian statistics. Basic results from this chapter are taken from Gelman et al. (2013); Bernardo and Smith (2009) and Robert (2007).

In a first part, some motivation towards Bayesian statistics is presented closing with Bayes' theorem. Afterwards, relevant concepts and definitions for Bayesian inference are introduced together with a description of Gaussian Markov Random Field (GMRF) priors. The GMRF prior will be used in Chapter 5 to a priori model spatial dependencies between MRF parameters. The Chapter will then be finished with the main aspects of Bayesian computations. Due to the high dimensional application MRF, a focus will be set on Bayesian computation for large-scale problems. Numerical integration, Monte Carlo methods, Laplace approximation and Markov Chain Monte Carlo methods will be presented in the context of Bayesian inference.

3.1 Motivation of Bayesian Statistics

In explorative statistics data are described, e.g., in terms of summary statistics such as their mean or standard deviation. Statistical inference, on the other hand, draws conclusions from data. For example, it infers a population mean given a (small) sample from it. Inferences drawn from a finite sample are necessarily uncertain, and this uncertainty needs to be characterized in the result of an inference. Bayesian statistics allows to accompany an inference with probability statements, e.g. a probability distribution for the population mean given a random sample from it.

In frequentist statistics conclusions are only drawn from the observed data whereas Bayesian statistics accounts for prior knowledge about the quantity of interest in addition. Prior

knowledge can be based on previous data, expert knowledge or other sources, and is quantified in a (prior) probability distribution. The probability in Bayesian statistics therefore expresses a degree of belief. All unknowns, including not observable (latent) parameters are modeled as random variables. It is then possible to make probability statements about latent variables conditional on the data, i.e. after the data have been observed. This feature as well as the potential to include prior knowledge into the analysis is not possible in frequentist statistics.

A key element in Bayesian statistics is the eponymous Bayes' theorem to update one's prior belief in view of the data. For two events A and B with $\mathbb{P}(B) > 0$, the probability of A under the condition that B has occurred can be expressed through the probability of B under A :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \quad (3.1)$$

Here, the probability of event A under the condition that event B is true is denoted by $\mathbb{P}(A|B)$. Bayes' theorem is a consequence of probability calculus, and it is likewise valid for frequentist statistics. The difference is that in frequentist statistics latent parameters are treated as fixed, non-random, while they are modeled as random variables in Bayesian statistics. (Therefore, frequentist statistics does not allow to apply Bayes theorem for latent parameters.) Based on equation (3.1) probability statements about unobservable quantities can be made as it is presented subsequently.

We will make the following notations throughout the thesis. Since the parameters in this thesis are continuous, the function π will always denote a probability density function (PDF). The function $\pi(x)$ will refer to a PDF for x while $\pi(y)$ to a PDF for y . The conditional PDF for the parameter θ given the data y is denoted by $\pi(\theta|y)$. Probability density functions are often simply termed distributions and all (latent) parameters are modeled as random variables that represent our state of knowledge about their unknown true values. The marginal distribution of θ is given as $\pi(\theta)$. The probability of a certain event (e.g. the event that θ is greater than a) is defined as $\mathbb{P}(\theta > a) = \int_{\theta > a} \pi(\theta)d\theta$. If the parameter θ is normally distributed with mean μ and variance σ^2 we will write $\theta \sim N(\mu, \sigma^2)$. If the parameter λ is distributed according to the inverse gamma distribution with shape parameter α and scale parameter β we will write $\lambda \sim IG(\alpha, \beta)$. Furthermore, if no limits of an integral are indicated explicitly then these limits extend over the whole domain of the corresponding variables.

3.2 Bayesian Inference

In Bayesian statistics probability statements can be made about the parameter θ conditional on the observed data y in terms of the posterior distribution $\pi(\theta|y)$. The model which expresses the joint probability distribution of θ and y is given by

$$\pi(\theta, y) = \pi(\theta)\pi(y|\theta). \quad (3.2)$$

We will refer to $\pi(\theta)$ as the *prior distribution* and $\pi(y|\theta)$ as the *sampling distribution*. With the help of Bayes' theorem, the *posterior distribution* $\pi(\theta|y)$ is given as

$$\pi(\theta|y) = \frac{\pi(\theta, y)}{\pi(y)} = \frac{\pi(\theta)\pi(y|\theta)}{\pi(y)}, \quad (3.3)$$

where the so-called *evidence* is defined as $\pi(y) = \int \pi(\theta)\pi(y|\theta)d\theta$. Since $\pi(y)$ is not dependent on the parameter θ the posterior is proportional to

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta). \quad (3.4)$$

In Bayesian statistics the *maximum a posteriori* (MAP) estimate is often calculated which is the mode of the posterior. Since the MAP is the same when maximizing the right hand side of (3.3) or (3.4), it is not necessary to calculate $\pi(y)$. This is also true for Markov Chain Monte Carlo methods used to draw random samples from the posterior, see below.

The two equations (3.3) or (3.4) capture the essence of Bayesian inference. The sampling distribution $\pi(y|\theta)$ viewed as a function of θ is called the *likelihood function* and may be denoted by $l(\theta; y)$. Bayes' theorem then takes the form $\pi(\theta|y) \propto \pi(\theta)l(\theta; y)$, i.e. the posterior is proportional to the product of prior distribution and the likelihood function. Note that the data have an influence on the posterior only through the likelihood but not through the prior.

Along the lines of Gelman et al. (2013) the process of Bayesian data analysis can be divided into three steps:

1. setting up a full probability model,
2. calculating the posterior distribution and interpreting the results, and carrying out a
3. sensitivity analysis of the model.

The last item refers to a very important part: a sensitivity analysis of the full probability model. In the first item, a statistical model for the data is chosen as well as a suitable prior. The influence of the choice of the statistical model and, particularly, of the prior, have to be evaluated. Techniques like a sensitivity analysis are applied in order to verify the sensitivity to model inaccuracies (see e.g. Roos et al. (2015); Gelman et al. (1996); Oakley and O'Hagan (2004)).

The posterior probability density contains all the information about the parameter θ . If θ is of dimension greater than one, it may be important to calculate the *marginal densities* of the posterior.

Definition 3.2.1 (Marginal posterior density). *Let $\pi(\theta_1, \theta_2|y)$ be the joint probability posterior density function of θ_1 and θ_2 . The marginal posterior density of θ_1 is given as*

$$\pi(\theta_1|y) = \int \pi(\theta_1, \theta_2|y)d\theta_2.$$

A complete summary of the results would be to deliver the posterior distribution of θ . Sometimes it is more appropriate and easier to only provide summary statistics. Summaries of the location can be the mean, median and/or mode of the posterior, summaries of the variation

can include standard deviation and credible intervals. In the following the main definitions of these quantities will be stated.

Definition 3.2.2 (Posterior mean). *The mean \mathbb{E} of the posterior distribution $\pi(\theta|y)$ is given as*

$$\mathbb{E}(\theta|y) = \int \theta \pi(\theta|y) d\theta.$$

Definition 3.2.3 (MAP estimate). *The maximum a posteriori (MAP) estimate $\hat{\theta}$ of the posterior distribution $\pi(\theta|y)$ is given as the mode of the posterior distribution:*

$$\hat{\theta} = \arg \max_{\theta} \pi(\theta|y).$$

The concept of the MAP estimate is related to the maximum likelihood (ML) estimate which is defined as the mode of the likelihood function. When a uniform prior distribution (e.g. a constant function) is used, the MAP and the ML estimate coincide. The MAP can be computed either analytically if the mode can be expressed in closed form, or numerically using an optimization algorithm.

Definition 3.2.4 (Posterior covariance and posterior variance). *The covariance Cov of the posterior distribution $\pi(\theta|y)$ is given as*

$$Cov(\theta|y) = \int (\theta - \mathbb{E}(\theta|y))(\theta - \mathbb{E}(\theta|y))^T \pi(\theta|y) d\theta.$$

For a single parameter θ_1 , the covariance matrix reduces to the posterior variance Var of θ_1 , namely

$$Var(\theta_1|y) = \int (\theta_1 - \mathbb{E}(\theta_1|y))^2 \pi(\theta_1|y) d\theta_1.$$

The uncertainty characterized by the posterior distribution is often summarized in terms of a credible region. For univariate parameters, a credible region usually is a credible interval (CI) in which the unobserved parameter values fall with a certain probability. The highest posterior density (HPD) region (cf. Berger (2013)) is an interval or a union of disjunct intervals defined in such a way, that all values within the region have a higher probability density than points outside. Note that the following definition is also suitable for multivariate parameter θ .

Definition 3.2.5 (HPD). *Let $\theta \in \Theta$. The $(1 - \alpha)$ HPD credible set of the posterior distribution $\pi(\theta|y)$ is given as a subset C of Θ of the form*

$$C = \{\theta \in \Theta : \pi(\theta|y) \geq k(\alpha)\},$$

where $k(\alpha)$ is the largest number such that

$$\int_{\theta: \pi(\theta|y) \geq k} \pi(\theta|y) d\theta = 1 - \alpha.$$

The HPD will be an interval when the posterior is unimodal and the parameter θ is univariate, for a multimodal posterior the HPD region might be unconnected.

After setting up the model for the parameters and the data, a prior has to be chosen. The prior expresses one's prior belief about the parameters and should be elicited as carefully

as possible. The choice of a suitable prior ranges from noninformative to informative. An informative prior expresses explicit knowledge about the parameters. This knowledge can arise, e.g., from prior measurements. It can also include physical knowledge of the parameters, such as smoothing constraints or the restriction to positive values.

Another class of relevant priors are conjugate prior distributions.

Definition 3.2.6 (Conjugate prior distribution). *Let \mathcal{F} be a class of sampling distributions and \mathcal{P} be a class of prior distributions. A prior distribution is called a conjugated prior distribution for the class \mathcal{F} of sampling distributions if*

$$\pi(\theta|y) \in \mathcal{P} \text{ for all } \pi(\cdot|\theta) \in \mathcal{F} \text{ and } \pi(\cdot) \in \mathcal{P}.$$

Example 3.2.1 (Example of conjugate prior distribution). *Assume a binomial distribution, i.e.*

$$\pi(y|\theta) \propto \theta^y(1-\theta)^{n-y}.$$

When a beta distributed prior is taken with

$$\pi(\theta|\alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

the posterior can be calculated according to

$$\begin{aligned} \pi(\theta|y) &= \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}. \end{aligned}$$

Hence, the posterior is also beta distributed with parameters $y + \alpha$ and $n - y + \beta$.

The choice of a conjugate prior has the advantage that the posterior is given in closed form and results can easily be calculated.

When no prior knowledge is available, the use of a noninformative prior can be reasonable. Research of what precisely constitutes a noninformative prior is still an active field of research. Since a detailed discussion is out of scope for this thesis, we want to mention Kass and Wasserman (1996) which provides a comprehensive review on the automatic choice of a prior.

In contrast to informative priors noninformative priors often are *improper*. A *proper* prior refers to a distribution which can be normalized, whereas this is not the case for an improper prior. When using a proper prior, the posterior will also be proper. This must not be true for an improper prior. Hence, it is always necessary to check the propriety of the posterior when using an improper prior.

An important theorem in asymptotic Bayesian statistics is the Bernstein-von Mises theorem. This theorem links Bayesian and frequentist statistics by stating important facts about the asymptotic properties of Bayesian point estimators (such as the posterior mean or the MAP) and the posterior distribution. Details can be found in Van der Vaart (2000). A light version of the Bernstein-von Mises theorem is given in the following.

Theorem 3.2.1 (Light version of the Bernstein-von Mises theorem). *Let θ_0 be the true population parameter and $y = y_1, \dots, y_n$ be a random sample from the population. Under*

certain assumptions about the sampling distribution $l(\theta; y)$ and support of the prior (e.g. the prior should not exclude θ_0), the posterior distribution $\pi(\theta|y)$ satisfies

$$\pi(\theta|y) \rightarrow N(\theta_0, (nI(\theta_0))^{-1}) \text{ as } n \rightarrow \infty, \quad (3.5)$$

where I denotes the Fisher information matrix

$$\begin{aligned} I_{ij}(\theta_0) &= -\mathbb{E}_\theta \left(\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(l(\theta; y)) \right]_{|\theta=\theta_0} \right) \\ &= - \int \pi(y|\theta) \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(l(\theta; y)) \right]_{|\theta=\theta_0} dy. \end{aligned}$$

In (3.5) the posterior converges in distribution to the normal distribution.

According to the Bernstein-von Mises theorem the posterior distribution becomes arbitrarily concentrated around the true value as n tends to infinity (cf. (3.5)). Thus, Bayesian estimation asymptotically does not depend on the choice of the prior distribution (for finite samples, Bayesian inference typically depends heavily on the prior). "Apparently, for an increasing number of observations one's prior beliefs are erased (or corrected) by the observations" (Van der Vaart (2000)). The theorem also justifies from a frequentist point of view the usage of Bayesian credible sets such as the HPD region defined in (3.2.5) when the posterior is dominated by the likelihood.

We will now introduce a special prior, namely the Gaussian Markov Random Field (GMRF) prior which is used in Chapter 5. More detailed information can be found in Rue and Held (2005). The idea is to define a prior for spatially distributed parameters. In the case of MRF, the parameters model some physical property on a 2D image at different voxels. These voxels are spatially connected, i.e. every voxel has neighboring voxels. A GMRF prior can account for these spatial dependencies between different voxels.

A GMRF is basically a multivariate normal distribution where the precision matrix (the inverse of the covariance matrix) fulfills certain properties.

Definition 3.2.7 (GMRF according to Rue and Held (2005)). *A random vector $\theta \in \mathbb{R}^n$ is called a GMRF with respect to a labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean μ and precision matrix $Q \succ 0$ (i.e. Q is positive definite) if and only if its density has the form*

$$\pi(\theta) = (2\pi)^{-n/2} |Q|^{1/2} \exp \left(-\frac{1}{2} (\theta - \mu)^T Q (\theta - \mu) \right)$$

and

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \mathcal{E} \text{ for all } i \neq j.$$

A labeled graph is a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = 1, \dots, n$. \mathcal{V} denotes the set of nodes in the graph and \mathcal{E} the set of edges i, j , where $i, j \in \mathcal{V}$ and $i \neq j$.

The existence of an edge between two nodes can therefore be derived from the precision matrix. A non-zero entry in the precision matrix means that the corresponding two nodes are connected. This is especially interesting in spatial applications where the quantity of interest models some spatial parameter.

The next theorem states a useful result for the precision matrix:

Theorem 3.2.2 (Rue and Held (2005)). *Let $\theta \in \mathbb{R}^n$ be normally distributed with mean μ and precision matrix Q . Let θ_{-ij} denote a vector with the same entries as θ but without θ_i and θ_j . Then $\pi(\theta_i, \theta_j | \theta_{-ij}) = \pi(\theta_i | \theta_{-ij})\pi(\theta_j | \theta_{-ij})$ (i.e. θ_i and θ_j are conditionally independent given θ_{-ij}) if and only if $Q_{ij} = 0$ for $i \neq j$.*

Hence, the zero entries of Q determine if θ_i and θ_j are conditionally independent given the rest. In contrast to that, zeros in the covariance matrix of a normal distribution express marginal independence.

Applications of a GMRF used as prior distribution range from image classification (Nishii and Eguchi (2006)) to image segmentation (Schmidt et al. (2013)) and magnetic resonance imaging (Schmid et al. (2006)). In many applications the precision matrix is sparse which prepares the ground for fast computations of GMRFs. A sparse precision matrix will not lead to a sparse covariance matrix in general. On the contrary, usually the covariance matrix will be dense in that case. It is therefore useful to represent GMRFs with their precision matrix. Additionally, modeling spatial smoothness can be achieved with a sparse precision matrix but not with a sparse covariance matrix. Setting up a prior for spatial smoothness is thus conveniently done in terms of the precision matrix, and its sparseness is helpful for subsequent calculations.

An intrinsic GMRF (IGMRF) is a special case of a GMRF:

Definition 3.2.8 (IGMRF of first order according to Rue and Held (2005)). *Let Q be an $n \times n$ symmetric positive semidefinite matrix with rank $n - 1 > 0$. Then $\theta \in \mathbb{R}^n$ is an IGMRF of first order with parameters (μ, Q) if its density is given as*

$$\pi(\theta) = (2\pi)^{\frac{-n+1}{2}} (|Q|^*)^{1/2} \exp\left(-\frac{1}{2}(\theta - \mu)^T Q(\theta - \mu)\right)$$

and $\sum_j Q_{ij} = 0$.

Since Q has no full rank, the generalized determinant $|Q|^*$ is defined as the product of all nonzero eigenvalues of Q . The parameters μ and Q are still denoted as mean and precision although the above expression does no longer represent a proper distribution.

Example 3.2.2 (IGMRF of first order). *Consider an image which is divided into N pixels. In some cases, it is likely that the values of neighboring pixels are somehow connected. This can be expressed via a IGMRF of first order. Let the precision matrix Q be defined as*

$$Q_{ij} = \begin{cases} -1, & i \sim j \\ n_i, & i = j \\ 0, & \text{otherwise} \end{cases} . \quad (3.6)$$

The equivalence relation \sim indicates that two pixels i and j are neighboring pixels. The number of neighboring pixels of pixel i is denoted by n_i .

Theorem 3.2.3. *The precision matrix Q in (3.6) is symmetric, positive semi-definite with a rank of $n - 1$.*

Proof. The equivalence relation is symmetric, therefore also is Q .

Q is positive semi-definite, because

$$\begin{aligned}
 \theta^T Q \theta &= \sum_{r=1}^n \theta_r^2 n_r - \theta_r \sum_{s \sim r} \theta_s \\
 &= \sum_{r=1}^n \sum_{s \sim r} \theta_r^2 - \sum_{r=1}^n \sum_{s \sim r} \theta_r \theta_s \\
 &= \sum_{r=1}^n \sum_{s \sim r, s > r} \theta_r^2 + \sum_{r=1}^n \sum_{s \sim r, s > r} \theta_s^2 - 2 \sum_{r=1}^n \sum_{s \sim r, s > r} \theta_r \theta_s \\
 &= \sum_{r=1}^n \sum_{s \sim r, s > r} (\theta_r - \theta_s)^2 \geq 0.
 \end{aligned}$$

It is $Q(1, \dots, 1)^T = (0 \dots 0)^T$, therefore the rank of Q is lower than or equal to $n - 1$. Let $\tilde{\theta}$ be another vector s.t. $Q\tilde{\theta} = 0$, then also $\tilde{\theta}^T Q \tilde{\theta} = 0$, therefore $\sum_{r=1}^n \sum_{s \sim r, s > r} (\tilde{\theta}_r - \tilde{\theta}_s)^2 = 0$. That is only possible if $\tilde{\theta}_r = \tilde{\theta}_s$ for all r, s which is again a vector proportional to $(1 \dots 1)^T$, therefore the rank of Q is greater equal $n - 1$. Altogether, the rank equals $n - 1$. \square

It follows that

$$(\theta)_i | (\theta)_{-i} \sim N \left(\frac{1}{n_i} \sum_{k \sim i} (\theta)_k, \frac{1}{n_i} \right). \quad (3.7)$$

Thus, the conditional distribution corresponds to a normal distribution with expectation equal to the mean of the effects of neighboring pixels and precision proportional to the number of neighboring voxels. Equation (3.7) therefore clarifies why and how the IGMRF from Example 3.2.2 can be used to model spatial dependencies (such as smoothness). The above relationship might be useful in many applications and gives an explanation on how such an IGMRF can be used to model the prior knowledge of smoothness between neighboring pixels.

3.3 Bayesian Computation

After setting up the model and choosing a prior distribution the posterior distribution and, if necessary, marginal distributions have to be calculated. In some cases this is possible analytically. In addition, quantities like the mean value should be computed. The integral in the definition of the posterior mean (3.2.2) has as many dimensions as the parameter itself which therefore might be complicated to evaluate especially in high dimensions as in the case of MRF. Thus, for difficult models and especially in high dimensions appropriate numerical methods have to be applied. In the following such methods are presented including approximate computing techniques.

Deterministic numerical integration

Numerical integration approximates the function at finitely many positions. If infinitely many positions were used, the integral could be exactly calculated. The standard procedure for deterministic numerical integration are quadrature rules for fixed knots $\theta_1, \dots, \theta_N$ and weights

w_1, \dots, w_N . For ease of notation a univariate θ is considered:

$$\int f(\theta) d\theta \approx \frac{1}{N} \sum_{n=1}^N w_n f(\theta^n).$$

The weights should fit to the volume space of the points θ^n . Basic methods include the midpoint, the trapezoidal or the Simpson's rule. In Chapter 4 the trapezoidal rule is used for a single parameter to approximate integrals on a grid.

Example 3.3.1 (Trapezoidal rule). *The trapezoidal rule approximates the region under the graph of the function f by a trapezoid. For a univariate θ , the integral can be calculated via*

$$\begin{aligned} \int_a^b f(\theta) d\theta &\approx \frac{b-a}{2N} \sum_{n=1}^N (f(\theta_n) + f(\theta_{n+1})) \\ &= \frac{b-a}{2N} (f(\theta_1) + 2f(\theta_2) + \dots + 2f(\theta_N) + f(\theta_{N+1})). \end{aligned}$$

This is implemented in the Matlab[®] function `trapz`.

Davis and Rabinowitz (2007) covers basic and more advanced techniques using, e.g., orthogonal polynomials. Note, that in high dimensions the *curse of dimensionality* occurs, whereby the number of knots N have to be intractably large in order to achieve good approximation rates.

Monte Carlo

The Monte Carlo (MC) method (Metropolis and Ulam (1949); Müller-Gronbach et al. (2012)) is an example of a procedure for direct simulation. The basic idea can be explained through an experiment which can be used to approximately calculate the number π .

Example 3.3.2 (Calculation of π). *Draw N random numbers which are uniformly distributed on $[0, 1]^2$. The ratio of numbers lying in the unit circle to N is approximately given as $\pi/4$. Hence, choosing random numbers in the unit square and determine the relative ratio of points lying in the unit circle gives an approximation of $\pi/4$. Note that the accuracy increases the more samples were drawn.*

The classical MC experiment which approximates the integral of a function defined on $[0, 1]^d$ uses independent, uniformly distributed random variables θ^n :

$$\int_{[0,1]^d} f(\theta) d\theta \approx \frac{1}{N} \sum_{n=1}^N f(\theta^n).$$

Deterministic procedures require a number of samples that grows exponentially with the dimension d in order to secure a prescribed error bound. The result from MC methods does not depend on the dimension d (Müller-Gronbach et al. (2012)). Nonetheless, random sampling may not account for important features of the function f . Problem specific sampling distributions may lead to improvements of the performance, especially in high dimensions. Note also that MC methods yield, e.g., moments of the posterior, but neither a random sample

for the posterior itself, nor a probability such as $\mathbb{P}(\theta \in A)$. A simple variant yielding random samples is importance sampling (Tokdar and Kass (2010)).

Markov Chain Monte Carlo

A widely used tool to generate samples from a probability distribution is called Markov Chain Monte Carlo (MCMC). MCMC uses Markov chains with a stationary distribution that is proportional to the desired distribution. For simplicity, only finite state spaces Z are considered in the following. We refer to Roberts et al. (2004) in which similar results for non-countable state spaces are presented. When using non-countable state spaces, the definition of a transition matrix can no longer be applied and has to be replaced by an abstract object called a transition kernel. Hence, concepts such as irreducibility and aperiodicity are more challenging to define. However, similar results such as the Ergodic theorem 3.3.1 can be proven.

Definition 3.3.1 (Markov chain). *A time discrete stochastic process $\{X_t\}_{t \in \mathbb{N}}$ with finite state space Z is called finite Markov chain, if*

$$\mathbb{P}(\{X_{t+1} = z_{t+1}\} | \{(X_0, \dots, X_t) = (z_0, \dots, z_t)\}) = \mathbb{P}(\{X_{t+1} = z_{t+1}\} | \{X_t = z_t\}) \quad (3.8)$$

for all $z_i \in Z$

Hence, a Markov chain is a sequence of random variables for which the distribution at time t only depends on the values at time $t - 1$.

Let $q_{z,z'}(t) := \mathbb{P}(X_{t+1} = z' | X_t = z)$ be the *transition probability*. If the transition probability is independent of time t (i.e. $\mathbb{P}(X_{m+1} = z' | X_m = z) = \mathbb{P}(X_{n+1} = z' | X_n = z) = q_{z,z'}$ for $\mathbb{P}(\{X_n = z\})\mathbb{P}(\{X_m = z\}) > 0$), the Markov chain is said to be *homogeneous*. The matrix $Q = \{q_{z,z'}\}_{z,z' \in Z} \in \mathbb{R}^{Z \times Z}$ is called *transition matrix* if $q_{z,z'} \geq 0$ and $\sum_{z' \in Z} q_{z,z'} = 1$.

Definition 3.3.2 (stationary, irreducible, aperiodic). *A vector of probabilities $\mu \in \mathbb{R}^Z$ is called stationary distribution of the matrix Q if $\mu = \mu Q$.*

A transition matrix $Q \in \mathbb{R}^{Z \times Z}$ is said to be irreducible if for all $z, z' \in Z$ there exists a number $n \in \mathbb{N}$ with $(Q^n)_{z,z'} > 0$.

Q is aperiodic if for all $z \in Z$ the greatest common divisor of $\{(Q^n)_{z,z} > 0\}_{n > 0}$ is equal to one.

An example of a stationary distribution is given via the following specific random walk.

Example 3.3.3 (Random walk). *Let $Z = \{1, \dots, k\}$, $p \in (0, 1)$ and $q_{z,z+1} = p$ for $z < k$, $q_{z,z-1} = 1 - p$ for $z > 1$, $q_{k,1} = p$, $q_{1,k} = 1 - p$ and otherwise $q_{z,z'} = 0$. In the case of $k = 4$, the transition matrix is given as*

$$Q = \begin{pmatrix} 0 & p & 0 & 1-p \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ p & 0 & 1-p & 0 \end{pmatrix}.$$

Here, the matrix Q has stationary distribution $\mu = (1/k, \dots, 1/k)$.

An irreducible Markov chain refers to a chain in which there exists a path over n edges for all pairs of states. If a system does not return to the same state at fixed intervals, it is called aperiodic.

With these definitions it is now possible to state the Ergodic theorem which assures convergence of the Markov chain to the desired distribution μ (Müller-Gronbach et al. (2012)).

Theorem 3.3.1 (Ergodic theorem). *Let $\{X_n\}_{n \in \mathbb{N}}$ be an irreducible, aperiodic and homogeneous Markov chain with stationary distribution μ . For every function $f : Z \rightarrow \mathbb{R}$ it is almost surely*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(X_k) = \sum_{z \in Z} \mu_z f(z). \quad (3.9)$$

In Bayesian inference, the state space Z and the posterior distribution are known (up to a multiplicative constant). An MCMC algorithm delivers an irreducible, aperiodic and homogeneous Markov chain with stationary distribution equal to the posterior distribution. Hence, MCMC methods use the fact that it is often easier to construct a Markov chain with stationary distribution equal to the posterior distribution rather than calculating the posterior directly. In the end, MCMC delivers samples from the posterior distribution which can then be used to calculate integrals such as the mean and the variance. However, these samples are not independent and identically distributed.

A very popular MCMC method is the Metropolis-Hastings algorithm (Metropolis et al. (1953)) which uses a proposal distribution. Dongarra and Sullivan (2000) listed it as one of the "Top Ten Algorithms of the 20th Century". A sample will be drawn according to the proposal distribution and with a certain probability this sample will be either accepted or rejected. This ensures the convergence to the desired target density which only needs to be known up to a multiplicative constant. In detail, a proof will include two steps. First, the algorithm uses an irreducible, aperiodic Markov chain which is not transient. Second, the stationary distribution of the Markov chain equals the target distribution (cf. Gelman et al. (2013)).

- 1: **procedure** METROPOLIS-HASTINGS ALGORITHM
- 2: Let $\pi(\theta)$ be the target density up to a multiplicative constant, $\theta^{(0)}$ be the starting value and $q(\theta|\theta^{(k)})$ be a proposal distribution
- 3: **for** $k = 0 : k_{max}$ **do**
- 4: Sample $\theta^* \sim q(\theta|\theta^{(k)})$
- 5: Calculate the acceptance probability $\alpha(\theta^{(k)}, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*)q(\theta^{(k)}|\theta^*)}{\pi(\theta^{(k)})q(\theta^*|\theta^{(k)})} \right\}$
- 6: Set $\theta^{(k+1)} = \theta^*$ with probability $\alpha(\theta^{(k)}, \theta^*)$, otherwise $\theta^{(k+1)} = \theta^{(k)}$
- 7: **end for**
- 8: **end procedure**

Figure 3.1: Pseudo code of the Metropolis-Hastings algorithm.

Note that a suitable choice of the proposal distribution has to be made by the user which can sometimes be challenging. Additionally, the algorithm in Figure 3.1 randomly walks through the target space. This can lead to very inefficient sampling with very low acceptance rates. For high-dimensional target distributions the local random walk would request an enormous amount of samples in order to cover the whole target space making it practically unapplicable (Beskos et al. (2009)).

Hamiltonian Monte Carlo (HMC) suppresses this purely random walk behavior by introducing an auxiliary variable scheme that transforms the sampling problem into a problem of simulating Hamiltonian dynamics (Hoffman and Gelman (2014)). The gradient of the log-posterior is used to avoid trial steps in undesirable directions. The sampler can thus move much faster through the target distribution. It is also called Hybrid Monte Carlo and was first introduced by Duane et al. (1987). HMC requires the gradient of the log-posterior as well as the choice of two parameters (the step size and the number of steps for how long the Hamiltonian system should run). A pseudo code is given in Figure 3.2. Advanced techniques as the No-U-Turn Sampler (NUTS), which eliminates the need of choosing the number-of-steps parameter is presented by Hoffman and Gelman (2014). HMC can be applied to high-dimensional posterior distributions, as shown in, e.g., Betancourt and Girolami (2015); Wei et al. (2015) and Gutmann and Hyvärinen (2012).

```

1: procedure HAMILTONAIN MONTE CARLO ALGORITHM
2:   Let  $\mathcal{L}$  be the logarithm of the target density up to a multiplicative constant,  $\theta^{(0)}$  be
   the starting value,  $\epsilon$  be the step size and  $L$  be the number of steps
3:   for  $k = 0 : k_{max}$  do
4:     Sample  $r^* \sim N(0, I)$ , where  $I$  denotes the identity matrix of appropriate dimension
5:     Set  $\theta^{(k+1)} = \theta^{(k)}$ ,  $\tilde{\theta} = \theta^{(k)}$  and  $\tilde{r} = r^*$ 
6:     for  $i = 0 : L$  do
7:       Set  $\tilde{r} = \tilde{r} + (\epsilon/2)\nabla_{\theta}\mathcal{L}(\tilde{\theta})$ 
8:       Set  $\tilde{\theta} = \tilde{\theta} + \epsilon\tilde{r}$ 
9:       Set  $\tilde{r} = \tilde{r} + (\epsilon/2)\nabla_{\theta}\mathcal{L}(\tilde{\theta})$ 
10:    end for
11:    Calculate the acceptance probability  $\alpha = \min \left\{ 1, \frac{\exp(\mathcal{L}(\tilde{\theta}) - 1/2\tilde{r}^T\tilde{r})}{\exp(\mathcal{L}(\theta^{(k)}) - 1/2(r^*)^T r^*)} \right\}$ 
12:    Set  $\theta^{(k+1)} = \tilde{\theta}$  and  $r^{(k+1)} = -\tilde{r}$  with probability  $\alpha$ 
13:  end for
14: end procedure

```

Figure 3.2: Pseudo code of the Hamiltonian Monte Carlo algorithm.

Laplace approximation

An easy way to approximate a unimodal posterior distribution is to use a multivariate normal distribution in the mode of the posterior. Let $g(\theta)$ denote the logarithm of the posterior distribution $\pi(\theta|y)$ up to a multiplicative constant and $\hat{\theta} = \arg \max_{\theta} g(\theta)$ denote the mode of g . A second-order Taylor series expansion around $\theta = \hat{\theta}$ is given as

$$g(\theta) \approx g(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^T Q_{\theta}(\hat{\theta})(\theta - \hat{\theta}), \quad (3.10)$$

where

$$Q_{\theta}(\hat{\theta}) = -\frac{\partial^2 \log \pi(\theta|y)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}}. \quad (3.11)$$

The variable θ is therefore approximately multivariate normal distributed:

$$\theta \sim N(\hat{\theta}, Q_{\theta}(\hat{\theta})^{-1}). \quad (3.12)$$

The covariance matrix can be either calculated analytically or numerically. The method of a Laplace approximation works fine in cases where the posterior is similar to a multivariate normal distribution, especially when a large number of observations are made (cf. the Bernstein-von Mises theorem 3.2.1). Laplace approximation of the posterior is a simple method to gain information about an otherwise complicated and possibly unfeasible posterior distribution. It is commonly used in practice (Riihimäki et al. (2014); Long et al. (2013)) and more advanced techniques such as the Integrated Nested Laplace Approximation (INLA) (Rue et al. (2009)) are based on it.

Other methods

There are further methods for Bayesian computation. One example is Variational Bayes (VB) in which the complex posterior distribution is approximated by a so-called variational distribution (Fox and Roberts (2012)). In most cases, the Kullback-Leibler divergence is used as a measure between the approximation and its target.

Approximate Bayesian computation (ABC) (Sunnåker et al. (2013)) covers a range of methods, all based on the idea to bypass the evaluation of the likelihood function and to use a (physical) model instead. The ABC rejection algorithm starts by drawing a sample according to the prior distribution. Then, a data set is computed and compared to the observed data. If these two data sets are close, the sample will be accepted, otherwise it will be rejected.

The Integrated Nested Laplace Approximation (INLA) (Rue et al. (2009)) approach is a tool for approximate Bayesian inference for latent Gaussian models. INLA uses Laplace approximations for high dimensional integrals and can produce fast and accurate results.

The expectation-maximization (EM) (Moon (1996)) algorithm is an iterative optimization method in order to find the MAP estimate of the posterior distribution. The algorithm alternates between the expectation and the maximization step. First, the expectation of the log-likelihood function using the current estimate is calculated. Then, the expectation function will be maximized leading to the next estimate for the parameters.

4

Bayesian Uncertainty Quantification for Dictionary-Based MRF

The dictionary matching method introduced by Ma et al. (2013) restricts itself to the calculation of estimates for the sought parameters. However, in order to assess the significance of differences of single qMRI results, which can be useful for, e.g. therapy monitoring, it is essential to also assign uncertainties that reliably characterize the accuracy of the calculated estimates (Polders et al. (2012)). Although statistical approaches (Zhao et al. (2016); Wübbeler and Elster (2017)) that fit the data directly in the Fourier domain allow for such an uncertainty quantification, their practical applicability is challenged due to the involved large-scale optimization task (cf. Chapter 5).

In this chapter our contribution is a *Bayesian uncertainty quantification of the dictionary-based MRF estimates*. The approach yields posterior probability distributions for the tissue-related parameters at all voxels allowing to make probability statements about these parameters to be made conditional on the observed data. Uncertainties associated with the estimates obtained by the dictionary matching method can then be calculated. The proposed Bayesian approach employs a noninformative prior (Kass and Wasserman (1996)) for the spin density and for the variance of the data combined with informative priors for the two relaxation times that utilizes as prior knowledge the ranges chosen for the relaxation times of the dictionary. The results are calculated using numerical quadrature on the basis of analytical expressions derived for the marginal posterior of the relaxation times. These calculations can effectively be carried out by using the same pre-computed dictionary as required for the calculations of the estimates and are hence fast.

The practicability of the approach is validated by analyzing simulated and real MRF measurements of a phantom that consists of several tubes with defined T_1 and T_2 values; applicability to an in vivo measurement is demonstrated as well.

After reviewing the dictionary matching method from a statistical perspective, the proposed Bayesian uncertainty quantification is developed in Section 4.1. This is followed by a description of the experiment for recording MRF data in Section 4.2. Results achieved by the proposed

Bayesian uncertainty quantification for numerical simulations, phantom experiments and an in vivo measurement are then presented in Section 4.3.

Most of the work presented in this chapter is published in Metzner et al. (2021).

4.1 Methods

As explained in Chapter 2, the acquired data from the MR scanner are samples in the Fourier domain of the image space. The magnetization in the image space which can be calculated via an inverse Fourier transformation depends on the tissue-related parameters, namely relaxation times T_1, T_2 and proton spin density ρ , as well as on the MR sequence parameters (e.g. echo time, repetition time, inversion time). Recall that in MRF a sequence of pulses with varying sequence parameters is applied, and due to the undersampling of the data in the Fourier domain the whole MRF measurement can be carried out fast. The dynamics of the magnetization in the image space depends on the whole sequence of pulses and can be modeled through a three-dimensional linear discrete time system, cf. Appendix B. Relevant parameters that are controlled in the design of the sequence of pulses are flip angles, repetition times, inversion time and echo times. The dependency of the dynamics of the magnetization on these controlled parameters is not indicated in the following.

We will use the same notations as in Chapter 2. Recall that N denotes the number of voxels in the selected slice, and $\theta_1 = (T_1^1, \dots, T_1^N)^T$, $\theta_2 = (T_2^1, \dots, T_2^N)^T$, $\theta_3 = (\rho^1, \dots, \rho^N)^T$ the sought vectors of relaxation times T_1, T_2 and proton spin density ρ , respectively, for all N voxels. The $N \times 1$ complex-valued vector of magnetizations is denoted by $m_l = m_l(\theta_1, \theta_2, \theta_3)$, $l = 1, \dots, L$, at the N voxels after applying the l -th excitation pulse, and m^i denotes the $L \times 1$ complex-valued vector of magnetizations at the i -th voxel after applying all L pulses.

Contrary to the Fourier domain where the data depends on all tissue parameters in parallel, the sequence of magnetizations in the image space at the i -th voxel, m^i , depends only on the unknown tissue parameters at that voxel (and the known parameters of the applied pulse sequence, which we will suppress), i.e. $m^i = m^i(T_1^i, T_2^i, \rho^i)$. Furthermore, the magnetization $m_l = m_l(\theta_1, \theta_2, \theta_3)$ depends linearly on θ_3 , which will be used in the following Bayesian treatment.

The measured data in the Fourier domain consist of the complex-valued vectors

$$z_l = \mathcal{F}_l(m_l), \quad l = 1, \dots, L, \quad (4.1)$$

where \mathcal{F}_l corresponds to a Fourier transform.

4.1.1 Dictionary Matching

The principles of the dictionary matching method according to Ma et al. (2013) have already been described in Chapter 2. The (approximately) reconstructed magnetizations from (4.1) are denoted by

$$\hat{m}_l = \mathcal{F}_l^\dagger(z_l), \quad l = 1, \dots, L, \quad (4.2)$$

where \mathcal{F}_l^\dagger is a pseudo-inverse of the Fourier transform mapping \mathcal{F}_l in (4.1). From the sequence of estimated magnetizations, $\hat{m}_1, \dots, \hat{m}_L$, the estimated course of magnetization \hat{m}^i for each

voxel i is denoted by

$$\widehat{m}^i = ((\widehat{m}_1)_i, \dots, (\widehat{m}_L)_i)^T. \quad (4.3)$$

Estimates of the tissue parameters can then be obtained by fitting the modeled dynamics of the magnetization of each voxel to the observed approximate dynamics (4.3) at that voxel. In Ma et al. (2013) this fitting process is done using a pre-computed dictionary of (normalized) courses of magnetizations on a discrete set of possible values of the sought relaxation times. The dictionary entry which shows maximum correlation to the observed dynamics (4.2) is chosen as best match. The spin density entering as a linear factor into the model for the magnetization is determined subsequently in a way such that the results fits best to the data.

In the next theorem we will show that this fitting can be achieved likewise without pre-computing a dictionary.

Theorem 4.1.1. *The dictionary matching method yields estimates $((\widehat{\theta}_1)_i, (\widehat{\theta}_2)_i, (\widehat{\theta}_3)_i) = (\widehat{T}_1^i, \widehat{T}_2^i, \widehat{\rho}^i)$ at the i -th voxel which are given by minimizing*

$$\|\widehat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\|^2 \quad (4.4)$$

with respect to T_1^i, T_2^i, ρ^i , provided that the resolution of the dictionary is made arbitrarily fine.

In other words, the dictionary matching method basically fits the modeled dynamics of the magnetization independently at each voxel to the observed estimates (4.3) at that voxel. In (4.4) $\|\cdot\|$ stands for the usual 2-norm in \mathbb{C}^L .

The following corollary is a straightforward consequence of Theorem 4.1.1.

Corollary 4.1.2. *If a homoscedastic Gaussian sampling model is assumed in which the complex-valued data are modeled as realizations of*

$$\widehat{m}^i = m^i(T_1^i, T_2^i, \rho^i) + \epsilon^i + i\tilde{\epsilon}^i, \quad \epsilon^i, \tilde{\epsilon}^i \sim N(0, \sigma^2 I_L), \quad (4.5)$$

and all $\epsilon^i, \tilde{\epsilon}^i$ are assumed independent, then the dictionary matching estimate coincides with the maximum likelihood estimate of the statistical model (4.5).

From a statistical point of view, the dictionary matching method hence yields the maximum likelihood estimator if a homoscedastic Gaussian model is assumed for the errors.

Proof (of Corollary 4.1.2). The likelihood function of the statistical model (4.5) is given as

$$\begin{aligned} l(T_1^i, T_2^i, \rho^i, \sigma; \widehat{m}^i) &\propto \frac{1}{\sigma^{2L}} \exp\left(-\frac{1}{2\sigma^2} \left(\widehat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\right)^H I_L \left(\widehat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\right)\right) \\ &= \frac{1}{\sigma^{2L}} \exp\left(-\frac{1}{2\sigma^2} \|\widehat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\|^2\right). \end{aligned}$$

The maximum likelihood estimator is calculated according to

$$\begin{aligned}
 (\hat{T}_1^i, \hat{T}_2^i, \hat{\rho}^i, \hat{\sigma}) &= \operatorname{argmax}_{T_1, T_2, \rho, \sigma} l(T_1^i, T_2^i, \rho^i, \sigma; \hat{m}^i) \\
 &= \operatorname{argmax}_{T_1, T_2, \rho, \sigma} \log(l(T_1^i, T_2^i, \rho^i, \sigma; \hat{m}^i)) \\
 &= \operatorname{argmax}_{T_1, T_2, \rho, \sigma} -\log(\sigma^{2L}) - \frac{1}{2\sigma^2} \|\hat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\|^2 \\
 &= \operatorname{argmin}_{T_1, T_2, \rho, \sigma} \log(\sigma^{2L}) + \frac{1}{2\sigma^2} \|\hat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\|^2
 \end{aligned}$$

For fixed σ the maximum likelihood estimator for $(\hat{T}_1^i, \hat{T}_2^i, \hat{\rho}^i)$ is then given as

$$(\hat{T}_1^i, \hat{T}_2^i, \hat{\rho}^i) = \operatorname{argmin}_{T_1, T_2, \rho} \|\hat{m}^i - m^i(T_1^i, T_2^i, \rho^i)\|^2$$

which equals the dictionary matching solution according to Theorem 4.1.1. \square

Proof (of Theorem 4.1.1). As explained in Chapter 2, the dictionary matching solution for \hat{T}_1^i and \hat{T}_2^i in voxel i is chosen to be the entry \hat{k}_i of the dictionary $\{D_k\}_{k=1}^K$ which maximizes

$$\hat{k}_i = \operatorname{argmax}_k \frac{|\langle \hat{m}^i, D_k \rangle|}{\|D_k\|}. \quad (4.6)$$

The complex proton density is calculated subsequently as

$$\hat{\rho}_i = \frac{\langle \hat{m}^i, D_{\hat{k}_i} \rangle}{\|D_{\hat{k}_i}\|^2}. \quad (4.7)$$

The reference to a specific voxel i is omitted in the following as the treatment is the same for all voxels.

The calculation of the dictionary is based on a physical model $m(T_1, T_2, \rho)$. We will denote $\tilde{m}(T_1, T_2)$ as the model without the linear factor ρ , i.e. $m(T_1, T_2, \rho) = \rho \tilde{m}(T_1, T_2)$. Equations (4.6) and (4.7) are therefore equivalent to

$$\begin{aligned}
 (\hat{T}_1, \hat{T}_2) &= \operatorname{argmax}_{T_1, T_2} \frac{|\langle \hat{m}, \tilde{m}(T_1, T_2) \rangle|}{\|\tilde{m}(T_1, T_2)\|} \\
 \hat{\rho} &= \frac{\langle \hat{m}, \tilde{m}(\hat{T}_1, \hat{T}_2) \rangle}{\|\tilde{m}(\hat{T}_1, \hat{T}_2)\|^2},
 \end{aligned}$$

under the assumption that the resolution of the dictionary is infinitely small.

Let \tilde{m}_R and \tilde{m}_I denote the real and imaginary part of $\tilde{m}(T_1, T_2)$, and \hat{m}_R and \hat{m}_I real and imaginary part of \hat{m} respectively. It follows

$$\begin{aligned}
 (\hat{T}_1, \hat{T}_2) &= \operatorname{argmax}_{T_1, T_2} \frac{\sqrt{(\hat{m}_R^T \tilde{m}_R + \hat{m}_I^T \tilde{m}_I)^2 + (-\hat{m}_R^T \tilde{m}_I + \hat{m}_I^T \tilde{m}_R)^2}}{\sqrt{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I}} \\
 &= \operatorname{argmax}_{T_1, T_2} \frac{\sqrt{h}}{\sqrt{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I}} \\
 &= \operatorname{argmax}_{T_1, T_2} \frac{h}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I}, \quad (4.8)
 \end{aligned}$$

for $h = (\hat{m}_R^T \tilde{m}_R)^2 + (\hat{m}_I^T \tilde{m}_I)^2 + (\hat{m}_R^T \tilde{m}_I)^2 + (\hat{m}_I^T \tilde{m}_R)^2 + 2(\hat{m}_R^T \tilde{m}_R \hat{m}_I^T \tilde{m}_I) - 2(\hat{m}_R^T \tilde{m}_I \hat{m}_I^T \tilde{m}_R)$ and

$$\begin{aligned} \hat{\rho} &= \frac{\langle \hat{m}_R + i\hat{m}_I, \tilde{m}_R + i\tilde{m}_I \rangle}{\|\tilde{m}_R + i\tilde{m}_I\|^2} \\ &= \frac{\hat{m}_R^T \tilde{m}_R + \hat{m}_I^T \tilde{m}_I}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I} + i \frac{\hat{m}_I^T \tilde{m}_R - \hat{m}_R^T \tilde{m}_I}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I}. \end{aligned} \quad (4.9)$$

Minimizing (4.4) with respect to μ leads to

$$\begin{aligned} \hat{\mu} &= \operatorname{argmin}_{\mu} \|\hat{m} - M\mu\|^2 \\ &= \operatorname{argmin}_{\mu} -2\hat{m}^T M\mu + \mu^T M^T M\mu \\ &= (M^T M)^{-1} M^T \hat{m} \\ &= \left(\begin{pmatrix} \tilde{m}_R^T & \tilde{m}_I^T \\ -\tilde{m}_I^T & \tilde{m}_R^T \end{pmatrix} \begin{pmatrix} \tilde{m}_R & -\tilde{m}_I \\ \tilde{m}_I & \tilde{m}_R \end{pmatrix} \right)^{-1} \begin{pmatrix} \tilde{m}_R^T & \tilde{m}_I^T \\ -\tilde{m}_I^T & \tilde{m}_R^T \end{pmatrix} \begin{pmatrix} \hat{m}_R \\ \hat{m}_I \end{pmatrix} \\ &= \frac{1}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I} \begin{pmatrix} \hat{m}_R^T \tilde{m}_R + \hat{m}_I^T \tilde{m}_I \\ -\hat{m}_R^T \tilde{m}_I + \hat{m}_I^T \tilde{m}_R \end{pmatrix}, \end{aligned}$$

which corresponds to the dictionary matching solution 4.9.

Let $c = \left(\hat{m}_R^T \tilde{m}_R + \hat{m}_I^T \tilde{m}_I, -\hat{m}_R^T \tilde{m}_I + \hat{m}_I^T \tilde{m}_R \right)^T$; minimizing (4.4) with respect to T_1 and T_2 yields

$$\begin{aligned} \operatorname{argmin}_{T_1, T_2} \|\hat{m} - M\hat{\mu}\|^2 &= \operatorname{argmin}_{T_1, T_2} -2 \langle \hat{m}, M\hat{\mu} \rangle + \langle M\hat{\mu}, M\hat{\mu} \rangle \\ &= \operatorname{argmax}_{T_1, T_2} 2 \langle \hat{m}, M\hat{\mu} \rangle - \langle M\hat{\mu}, M\hat{\mu} \rangle \\ &= \operatorname{argmax}_{T_1, T_2} 2\hat{m}^T M\hat{\mu} - \hat{\mu}^T M^T M\hat{\mu} \\ &= \operatorname{argmax}_{T_1, T_2} \frac{2}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I} c^T c \\ &\quad - \frac{1}{(\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I)^2} c^T \begin{pmatrix} \tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I & 0 \\ 0 & \tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I \end{pmatrix} c \\ &= \operatorname{argmax}_{T_1, T_2} \frac{1}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I} c^T c \\ &= \operatorname{argmax}_{T_1, T_2} \frac{h}{\tilde{m}_R^T \tilde{m}_R + \tilde{m}_I^T \tilde{m}_I}, \end{aligned}$$

which corresponds to the dictionary matching solution 4.8.

Hence, the dictionary matching method yields estimates which are given by minimizing (4.4) provided that the resolution of the dictionary is infinitely small. \square

To ease notations, we suppressed (e.g. in (4.1)) and will suppress the presence of measurements using multiple receiver coils (as used in our experiments). This simplification does not affect our treatment according to the following relationship. Let $\{C_i\}_{i=1}^{n_c}$ denote the n_c different $\sqrt{N} \times \sqrt{N}$ coil sensitivity matrices. We define $\tilde{C}_i = \operatorname{diag}(\{(C_i)_j\}_{j=1}^N) \in \mathbb{R}^{N \times N}$. Equation (4.1) will then become

$$z_{li} = \mathcal{F}_l(\tilde{C}_i m_i), \quad l = 1, \dots, L. \quad (4.10)$$

Let $\overline{\tilde{C}_i}$ denote the complex conjugate matrix of \tilde{C}_i . According to Roemer et al. (1990) an optimal combination of the single coil images $\tilde{C}_i m_l$ which maximizes the signal-to-noise ratio is realized through

$$\hat{m}_l = \left(\sum_{i=1}^{n_c} \overline{\tilde{C}_i} \tilde{C}_i \right)^{-1} \sum_{i=1}^{n_c} \overline{\tilde{C}_i} \mathcal{F}_l^\dagger(z_{li}), \quad l = 1, \dots, L. \quad (4.11)$$

Equation (4.2) will therefore change to equation (4.11) and the magnetizations can be estimated using a weighted sum of single coil reconstructions.

4.1.2 Bayesian Inference for Dictionary Matching

In the following, a Bayesian inference is carried out to quantify the uncertainty associated with the tissue-related parameters calculated by the dictionary matching method for the first time. The Bayesian inference is based on a statistical model (4.15) for the observed magnetizations in (4.3) as well as on the choice of prior distributions. Some additional notations are introduced subsequently.

Let the $2L \times 1$ real-valued vector \hat{y}^i denote real part and imaginary part of the observed estimates \hat{m}^i of the course of magnetizations at voxel i , i.e.

$$\hat{y}^i = (\text{Re}(\hat{m}^i)^T, \text{Im}(\hat{m}^i)^T)^T, \quad (4.12)$$

and the 2×1 real-valued vector μ^i the corresponding proton spin density,

$$\mu^i = (\text{Re}(\rho^i), \text{Im}(\rho^i))^T. \quad (4.13)$$

Furthermore, let the $2L \times 2$ real-valued matrix $M^i = M^i(T_1^i, T_2^i)$ be defined in a way such that

$$m^i(T_1^i, T_2^i, \rho^i) = M^i \mu^i. \quad (4.14)$$

The matrix M therefore contains real and imaginary part of the *modeled* course of magnetizations $m^i(T_1^i, T_2^i, \rho^i)$ at voxel i .

Theorem 4.1.3. *The matrix M is given as*

$$M^i = \begin{pmatrix} \text{Re}(\tilde{m}^i(T_1^i, T_2^i)) & -\text{Im}(\tilde{m}^i(T_1^i, T_2^i)) \\ \text{Im}(\tilde{m}^i(T_1^i, T_2^i)) & \text{Re}(\tilde{m}^i(T_1^i, T_2^i)) \end{pmatrix},$$

where again \tilde{m}^i denotes the magnetization m^i without the linear factor ρ^i .

Proof. The modeled magnetization course is given as $m^i(T_1^i, T_2^i, \rho^i) = (\text{Re}(\rho^i) + i \text{Im}(\rho^i))(\text{Re}(\tilde{m}^i) + i \text{Im}(\tilde{m}^i))$.

It follows that the real and imaginary part of m^i are $\text{Re}(m^i(T_1^i, T_2^i, \rho^i)) = \text{Re}(\rho^i) \text{Re}(\tilde{m}^i) - \text{Im}(\rho^i) \text{Im}(\tilde{m}^i)$ and $\text{Im}(m^i(T_1^i, T_2^i, \rho^i)) = \text{Im}(\rho^i) \text{Re}(\tilde{m}^i) + \text{Re}(\rho^i) \text{Im}(\tilde{m}^i)$. M^i is defined through

$$(\text{Re}(m^i(T_1^i, T_2^i, \rho^i))^T, \text{Im}(m^i(T_1^i, T_2^i, \rho^i))^T)^T = M^i (\text{Re}(\rho^i), \text{Im}(\rho^i))^T.$$

It follows that

$$M^i = \begin{pmatrix} \operatorname{Re}(\tilde{m}^i(T_1^i, T_2^i)) & -\operatorname{Im}(\tilde{m}^i(T_1^i, T_2^i)) \\ \operatorname{Im}(\tilde{m}^i(T_1^i, T_2^i)) & \operatorname{Re}(\tilde{m}^i(T_1^i, T_2^i)) \end{pmatrix}.$$

□

The dictionary matching method as well as the subsequent statistical analysis can be carried out independently at each voxel. We will hence omit the reference to a particular voxel i in the following, e.g., we use $M = M(T_1, T_2)$ to denote $M^i = M^i(T_1^i, T_2^i)$ for voxel i .

Statistical Model

The dictionary matching method treats the errors in the observed course of magnetization in (4.3) as noise (cf. Corollary 4.1.2). We will therefore assume a normal distribution with zero mean and variance σ^2 in which the errors are modeled as independently and identically distributed. Note that this is the same statistical as introduced in Corollary 4.1.2.

In detail, we assume

$$y \sim N\left(M\mu, \sigma^2 \mathbf{I}_{2L}\right), \quad (4.15)$$

where $M = M(T_1, T_2)$ depends on T_1 and T_2 only, μ denotes real and imaginary part of the spin density ρ , and \mathbf{I}_{2L} stands for the identity matrix of dimension $2L$. The observed data \hat{y} are hence modeled as a realization of a Gaussian distribution with mean vector $M\mu$ and covariance matrix $\sigma^2 \mathbf{I}_{2L}$. The statistical model (4.15) is a homoscedastic Gaussian model with a mean vector that depends linearly on the spin density μ but nonlinearly on the relaxation times T_1 and T_2 . Note that the variance parameter σ^2 in (4.15) is allowed to differ for different voxels.

Bayesian Inference

As described in Chapter 3, a Bayesian inference combines the prior knowledge about the unknowns with the information contained in the data resulting in the posterior probability distribution which is obtained via Bayes' theorem according to

$$\pi(T_1, T_2, \mu, \sigma^2 | y) \propto \pi(T_1, T_2, \mu, \sigma^2) l(T_1, T_2, \mu, \sigma^2; y). \quad (4.16)$$

In (4.16) the prior distribution about the unknowns T_1, T_2, μ, σ^2 is denoted by $\pi(T_1, T_2, \mu, \sigma^2)$ and $l(T_1, T_2, \mu, \sigma^2; y)$ refers to the likelihood function which is determined through the statistical model (4.15) for the data.

As a prior distribution we choose

$$\pi(T_1, T_2, \mu, \sigma^2) = \pi(T_1, T_2) \pi(\mu) \pi(\sigma^2) \propto 1_\Omega(T_1, T_2) / \sigma^2, \quad (4.17)$$

where Ω is defined as the same domain for T_1 and T_2 as used to create the dictionary, and the indicator function $1_\Omega(T_1, T_2)$ equals 1 for $(T_1, T_2) \in \Omega$ and 0 otherwise. The prior $\pi(T_1, T_2) \propto 1_\Omega(T_1, T_2)$ for the relaxation times thus represents the prior belief about the range of possible values for T_1 and T_2 . For the spin density which is acting as a location parameter in the likelihood we assume a standard noninformative prior used in such a case, cf. Kass

and Wasserman (1996), namely $\pi(\mu) \propto 1$. Similarly, a standard noninformative prior for the scale parameter σ^2 is chosen as $\pi(\sigma^2) \propto 1/\sigma^2$. For the important choice of a suitable prior distribution we therefore combine the available prior knowledge about the relaxation times with standard noninformative priors for the proton spin density and for the variance σ^2 in (4.15).

Marginal posterior distributions such as, e.g., $\pi(T_1|y)$ can be obtained by integrating out the remaining unknowns from the joint posterior $\pi(T_1, T_2, \mu, \sigma^2|y)$ in (4.16).

The uncertainty of, e.g., T_1 will be denoted by $u(T_1)$ and is given as the square root of the expected quadratic loss of the dictionary-based MRF estimate \hat{T}_1 :

$$\begin{aligned} u^2(\hat{T}_1) &= \int \pi(T_1|y)(T_1 - \hat{T}_1)^2 dT_1 \\ &= \text{Var}(\pi(T_1|y)) + \left(\mathbb{E}(T_1|y) - \hat{T}_1\right)^2 \end{aligned} \quad (4.18)$$

It is also possible to determine a 95% credible interval I which satisfies

$$0.95 = \int_I \pi(T_1|y) dT_1. \quad (4.19)$$

Note that relation (4.19) does not uniquely determine an interval I , and an additional condition needs to be posed, for example the interval of minimal length satisfying (4.19) or the choice of a probabilistically symmetric interval (c.f. Chapter 3). Another possibility is to use the marginal posterior to make probability statements, for example to determine the probability that T_1 exceeds a certain limit \bar{T}_1 according to

$$P(T_1 > \bar{T}_1|y) = \int_{\bar{T}_1}^{\infty} \pi(T_1|y) dT_1. \quad (4.20)$$

The next theorem will show that albeit the prior (4.17) is improper, i.e. cannot be normalized, propriety of the marginal joint posterior for T_1, T_2 can be ensured.

Theorem 4.1.4. *The marginal joint posterior for T_1, T_2 is given as*

$$\pi(T_1, T_2|y) \propto \frac{1}{\sqrt{\det(M^T M)}} \times \frac{\mathbf{1}_{\Omega}(T_1, T_2)}{\left(\|y - M\mu(T_1, T_2)\|^2\right)^{L-1}} \quad (4.21)$$

with

$$\mu(T_1, T_2) = \left[M^T M\right]^{-1} M^T y \quad (4.22)$$

and $M = M(T_1, T_2)$. It exists and is proper if $\det(M^T M) > \delta_1 > 0$ and $\min_{T_1, T_2} \|y - M\hat{\mu}\|^2 > \delta_2 > 0$.

Proof. In order to derive the posterior distribution, we make two prior statements: According to the PDF of a multivariate t-distribution, we have for $x \in \mathbb{R}^p, \mu \in \mathbb{R}^p, \nu > 1$ and a symmetric, positive-definite $\Sigma \in \mathbb{R}^{p \times p}$:

$$\int \frac{1}{\det(\Sigma)^{1/2}} \left(1 + \frac{1}{\nu}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)^{-(\nu+p)/2} dx \propto 1$$

and therefore

$$\int \left(1 + \frac{1}{\nu}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)^{-(\nu+p)/2} dx \propto \det(\Sigma)^{1/2}.$$

For $A \in \mathbb{R}^{2 \times 2}$ and $a \in \mathbb{R}$ it is

$$\begin{aligned} \det \left(\left(\frac{1}{a^2} A \right)^{-1} \right)^{1/2} &= \det \left(\frac{1}{a^2} A \right)^{-1/2} = \left(\left(\frac{1}{a^2} \right)^2 \det(A) \right)^{-1/2} \\ &= a^2 \det(A)^{-1/2}. \end{aligned}$$

We then get for $\hat{\mu} = \mu(T_1, T_2) = [M^T M]^{-1} M^T y$:

$$\begin{aligned} \pi(T_1, T_2 | y) &\propto \int_{-\infty}^{\infty} \int_0^{\infty} \pi(T_1, T_2, \mu, \sigma^2 | y) d\sigma^2 d\mu \\ &\propto \int_{-\infty}^{\infty} \int_0^{\infty} 1_{\Omega}(T_1, T_2) \frac{1}{\sigma^{2L+2}} \exp\left(-\frac{1}{2\sigma^2} \|y - M\mu\|^2\right) d\sigma^2 d\mu \\ &= 1_{\Omega}(T_1, T_2) \int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\tau^{L+1}} \exp\left(-\frac{1}{2\tau} \|y - M\mu\|^2\right) d\tau d\mu \\ &= 1_{\Omega}(T_1, T_2) \int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{(u\|y - M\mu\|^2)^{L+1}} \|y - M\mu\|^2 \exp\left(-\frac{1}{2u}\right) du d\mu \\ &= 1_{\Omega}(T_1, T_2) \int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{(u^{L+1})} (\|y - M\mu\|^2)^{-L} \exp\left(-\frac{1}{2u}\right) du d\mu \\ &\propto 1_{\Omega}(T_1, T_2) \int_{-\infty}^{\infty} \frac{1}{(\|y - M\mu\|^2)^L} d\mu \\ &= 1_{\Omega}(T_1, T_2) \int_{-\infty}^{\infty} \frac{1}{(\|y - M\hat{\mu}\|^2 + (\mu - \hat{\mu})^T M^T M (\mu - \hat{\mu}))^L} d\mu \\ &= 1_{\Omega}(T_1, T_2) \frac{1}{(\|y - M\hat{\mu}\|^2)^L} \int_{-\infty}^{\infty} \frac{1}{(1 + (\mu - \hat{\mu})^T \frac{M^T M}{\|y - M\hat{\mu}\|^2} (\mu - \hat{\mu}))^L} d\mu \quad (4.23) \\ &\propto 1_{\Omega}(T_1, T_2) \frac{1}{(\|y - M\hat{\mu}\|^2)^L} \|y - M\hat{\mu}\|^2 \frac{1}{\det(M^T M)^{1/2}} \\ &= 1_{\Omega}(T_1, T_2) \frac{1}{(\|y - M\hat{\mu}\|^2)^{L-1}} \frac{1}{\det(M^T M)^{1/2}}. \quad (4.24) \end{aligned}$$

Since we assume $\det(M^T M) > \delta_1 > 0$ and $\min_{T_1, T_2} \|y - M\hat{\mu}\|^2 > \delta_2 > 0$ and because Ω is compact, the marginal posterior in (4.24) is proper and hence the joint posterior exists. \square

Theorem 4.1.5. *The conditional posterior for μ is given as*

$$\mu | T_1, T_2, y \sim t_{2L-2} \left(\mu(T_1, T_2), \frac{\|y - M\mu(T_1, T_2)\|^2}{2L-2} [M^T M]^{-1} \right), \quad (4.25)$$

i.e. the conditional posterior distribution for μ given T_1, T_2 is a scaled and shifted 2D t -distribution with $2L - 2$ degrees of freedom.

Proof. A similar derivation as in the proof of Theorem 4.1.4, by applying a marginalization w.r.t. σ^2 , yields the expression over which is integrated in (4.23). The statement of the theorem then follows easily. \square

Since the parameter $L \gg 1$, the following Gaussian distribution is a good approximation to (4.25)

$$\mu|T_1, T_2, y \sim N\left(\mu(T_1, T_2), \frac{\|y - M\mu(T_1, T_2)\|^2}{2L - 2} [M^T M]^{-1}\right), \quad (4.26)$$

where, again, $M = M(T_1, T_2)$. The marginal posterior for μ is then obtained as

$$\pi(\mu|y) = \int \pi(\mu|T_1, T_2, y)\pi(T_1, T_2|y)dT_1dT_2. \quad (4.27)$$

Expressions (4.21) and (4.26) allow for a practical calculation of results using 2-d approximate numerical quadrature, which can be efficiently carried out on the dictionary itself. Once the marginal posterior distributions for T_1, T_2, μ have been determined, uncertainties such as (4.18), or probability statements as in (4.20) can be determined by again using numerical quadrature.

In Appendix C the MATLAB[®] source code is shown which utilizes the built-in MATLAB[®] function `trapz` for numerical quadrature as described in Chapter 3. The required input consists of the dictionary used for the dictionary matching method, together with the observed sequence of magnetizations in (4.3) for each voxel as well as estimates for T_1 and T_2 . Note that it is possible to use other estimates than those obtained by the dictionary matching method. The output of the source code are the discretized marginal posteriors for T_1 and T_2 and the uncertainties for T_1 and T_2 calculated according to (4.18). Due to the use of the pre-computed dictionary, calculations can be carried out fast. The computation time naturally depends on the size of the dictionary and on the number of voxels of interest. The calculation time of the uncertainties shown here was about 90 minutes, whereas calculating the estimates by the dictionary matching method took about 6 minutes. The computation time scales linearly with the size of the dictionary. The employed software together with exemplary MRF data is provided in Metzner and Wübbeler (2021).

4.2 Experiments and Simulations

The phantom data was acquired on a 3 Tesla Verio MR Scanner (Siemens Healthineers, Erlangen, Germany) with an 8-channel head coil. The in-plane resolution of the MRF scan was $0.78 \times 0.78 \text{ mm}^2$ with a slice thickness of 5.0 mm and a field of view (FoV) of $300 \times 300 \text{ mm}^2$. A fast imaging with steady-state precession (FISP) sequence was used starting with an inversion pulse. A radial readout was performed, each with 384 data points per radial line. Repetition times (TR) were chosen to be 12.54 ms , the inversion time (TI) as 42 ms , echo times (TE) as 7 ms . The flip angle follows the magnitude of a sinusoidal curve with a maximum of 70 degree (as already proposed in Flassbeck et al. (2019)). A total number of 1000 radial lines were acquired and for the parameter estimation one image per radial line was reconstructed by using the pseudo-inverse Fourier transformation (Greengard and Lee (2004)). In Figure 4.1 two reconstructed magnetization images are shown as well as the sum of all the reconstructions.

Reference values for the T_1 relaxation time (cf. Chapter 2)) were determined by a Cartesian inversion recovery spin echo sequence acquisition with 7 inversion times ($TI = 25\text{--}2400 \text{ ms}$, $TE/TR = 12/8000 \text{ ms}$). Reference values for the T_2 relaxation time (cf. Chapter 2))

were determined using a Cartesian spin echo sequence with multiple echo times ($TE = 24\text{-}1000\text{ ms}$, $TR = 3000\text{ ms}$). The in-plane resolution was $0.83 \times 0.83\text{ mm}^2$ with a slice thickness of 5.0 mm and a FoV of $159 \times 144\text{ mm}^2$ for both acquisitions. To determine the reference T_1 and T_2 values, a 2-parameter nonlinear least squares fitting algorithm was used.

The simulation closely follows the phantom measurement. All scanning parameters (TR , TE , TI , flip angles), k -space locations for the radial points and complex coil sensitivity maps are set to the same values as used in the phantom measurements. A total number of 1000 radial lines were simulated, and for the dictionary matching method parameter estimation was carried out using one image per radial line. We assigned similar values for T_1 , T_2 and the complex proton density such that they match those obtained in the analysis of the phantom data (cf. Chapter 2)). The magnetization during the simulated data acquisition was then modeled by the same FISP model as before (cf. Appendix B). Subsequently, the modeled magnetization was transformed into the Fourier domain where Gaussian noise with similar signal-to-noise ratio as for the real scanner data was added with a standard deviation of 3.43×10^{-6} (the maximum value of the real and the imaginary part of the simulated magnetizations are of the order 10^{-3}).

The in vivo data acquisition was performed on the same scanner using a 32-channel cardiac coil (Siemens Healthineers, Erlangen, Germany). For the in vivo data, a cardiac scan was realized in a short-axis orientation utilizing a FISP sequence with a total acquisition time of 12 s (allowing for an acquisition during a single breathhold). A radial readout was performed, each with 320 data points per radial line. Repetition times were chosen to be 8.2 ms , the inversion time as 21 ms , echo times as 4 ms , and the flip angle follows the magnitude of a sinusoidal curve with a maximum of 70 degree. Here, a total number of 1500 radial lines were acquired and for the parameter estimation one image per radial line was reconstructed using the pseudo-inverse Fourier transformation. The resolution of the in vivo scan was $1.3 \times 1.3\text{ mm}^2$ with a FOV of $320 \times 320\text{ mm}^2$ and a slice thickness of 8 mm . A bandwidth time product of 8 was chosen (SINC pulse length = 1920 ms) in order to reduce the impact of the slice profile. ECG triggering was used to start the acquisition during diastole. After that the data were acquired continuously without further triggering.

4.3 Results

Real and simulated MRF measurements of the phantom (see Figure 4.1) as well as a cardiac in vivo measurement were conducted and analyzed by the proposed uncertainty quantification for the dictionary matching method. Note that throughout this thesis the reconstructed magnetization was not normalized. The dictionary used for the following results is based on a FISP sequence and calculated according to Appendix B for $T_1 = (200, 205, \dots, 3000)\text{ ms}$ and $T_2 = (20, 22, \dots, 300)\text{ ms}$ assuming 201 isochromats in each voxel. The phantom considered here consists of 9 different tubes and the values for T_1 and T_2 within each tube can be assumed to be constant. The ranges of the dictionary were chosen in a way such that the variation of T_1 and T_2 between the tubes is covered, see Figure 4.3.

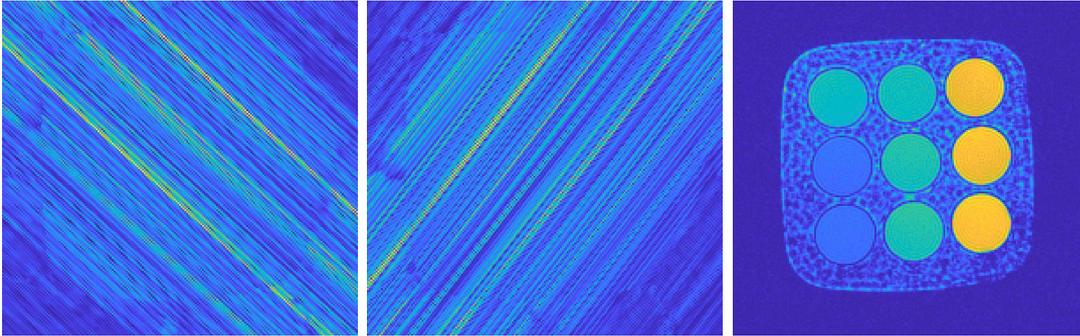


Figure 4.1: Left and Middle: two single reconstructed magnetization images. Right: Plot of the sum of all reconstructed magnetization images showing the phantom consisting of 9 different tubes. Figure from Metzner et al. (2021).

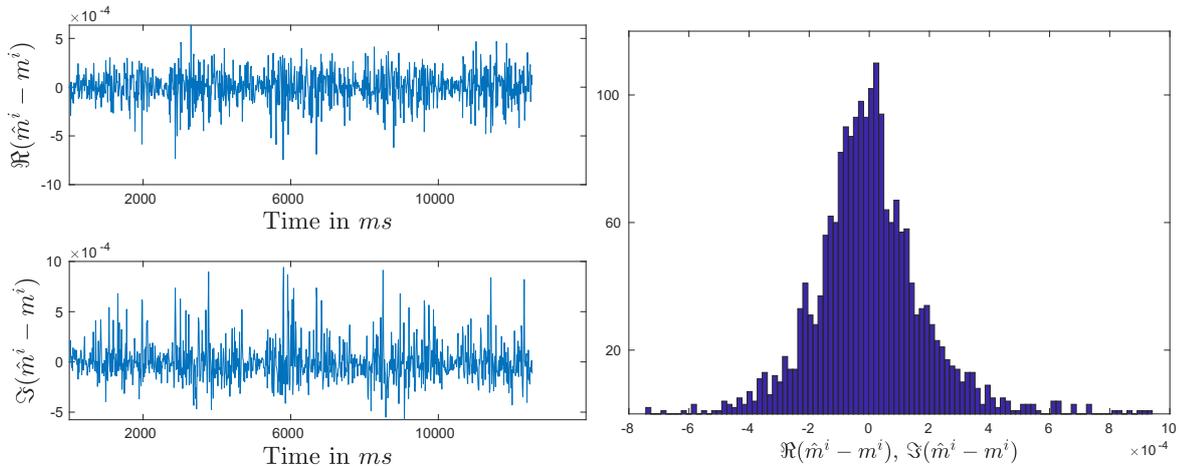


Figure 4.2: Results from phantom data. Left: residuals of real (top) and imaginary (bottom) part of the (unnormalized) magnetization in one voxel obtained by the dictionary matching method for the phantom data. Right: histogram of all residuals on the left. Figure from Metzner et al. (2021).

4.3.1 Phantom and In Vivo Measurements

Figure 4.2 shows the (estimated) residuals $\hat{m}^i - m^i$ for a typical voxel together with their histogram. The residuals seem to approximately follow a Gaussian distribution which supports the choice of the statistical model used for the Bayesian uncertainty quantification.

In Figure 4.3 a comparison of calculated uncertainties and the observed variation of estimated values for T_1 and T_2 within each tube is presented. More precisely, the figure shows mean values of the estimates obtained by the dictionary matching method for T_1 and T_2 within the nine different tubes. Additionally, both the standard deviations of the estimates within the tubes (blue errorbars) and the square root of within-tubes means of squared uncertainties (red errorbars) are displayed for the different tubes. The calculated mean uncertainties reflect the within-tube variability of the estimates well, yet they are slightly larger than the observed variability of the estimates for all nine tubes.

The range of the dictionary has to be chosen with great care since it has an influence on both: the dictionary matching solution and the Bayesian uncertainties. To analyze this impact we evaluated the same phantom data with a larger and a smaller dictionary in T_2 . When taking a larger dictionary in T_2 (but not in T_1) with $T_2 = (20, 22, \dots, 500) \text{ ms}$, the within-tube standard deviations of the estimates differed from those obtained for the original dictionary by at most

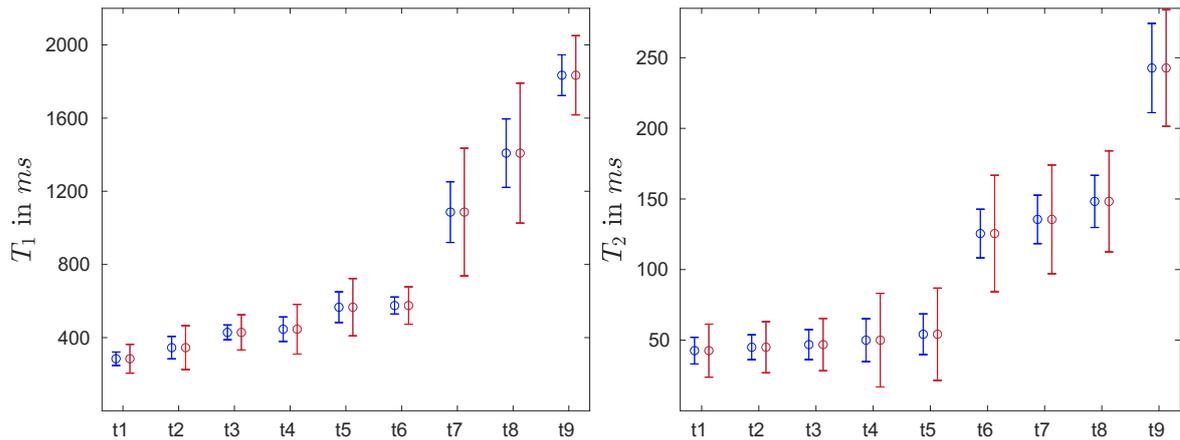


Figure 4.3: Mean values for the estimates obtained by the dictionary matching method within the tubes t_1, \dots, t_9 for T_1 (left) and T_2 (right), together with their standard deviation (blue errorbars) and mean uncertainty (red errorbars). Note that the errorbars are in ascending order in terms of the mean values for the estimates, respectively for T_1 and T_2 . Figure from Metzner et al. (2021).

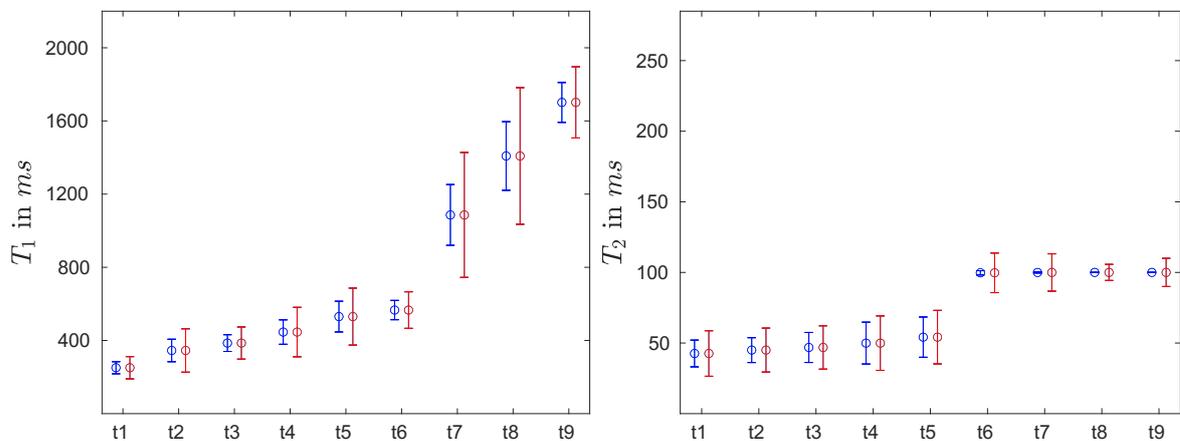


Figure 4.4: Results for a dictionary with a smaller range: mean values for the estimates obtained by the dictionary matching method within the tubes t_1, \dots, t_9 for T_1 (left) and T_2 (right), together with their standard deviation (blue errorbars) and mean uncertainty (red errorbars). Note that the errorbars are in ascending order in terms of the mean values for the estimates, respectively for T_1 and T_2 . Figure from Metzner et al. (2021).

10%. The same holds for the mean uncertainties, except for tube nine in which the mean uncertainty is 50% higher. Results for a smaller dictionary in T_2 with $T_2 = (20, 22, \dots, 100)$ ms are shown in Figure 4.4. Since some tubes have an actual value that is larger than the maximum value for T_2 in the (restricted) dictionary, the corresponding estimates equal the maximum value of T_2 in the dictionary. Note that the resulting uncertainties are significantly larger than the standard deviations of the estimates within these tubes which indicates that the true T_2 values are beyond the upper bound.

In order to analyze the impact of the amount of data, the same quantities as in Figure 4.3 are plotted in Figure 4.5, but for data where 5 radial lines in k -space are acquired per image. This was accomplished by a multi-shot MRF approach as presented by Flassbeck et al. (2019). The amount of data is hence increased by a factor of 5. The same scale is used as in Figure 4.3 which shows that the mean uncertainties and the standard deviation of the estimates in each tube are smaller now (roughly by a factor $\sqrt{5}$). Note that the ratios of mean uncertainties and standard deviations of the estimates remain nearly unchanged.

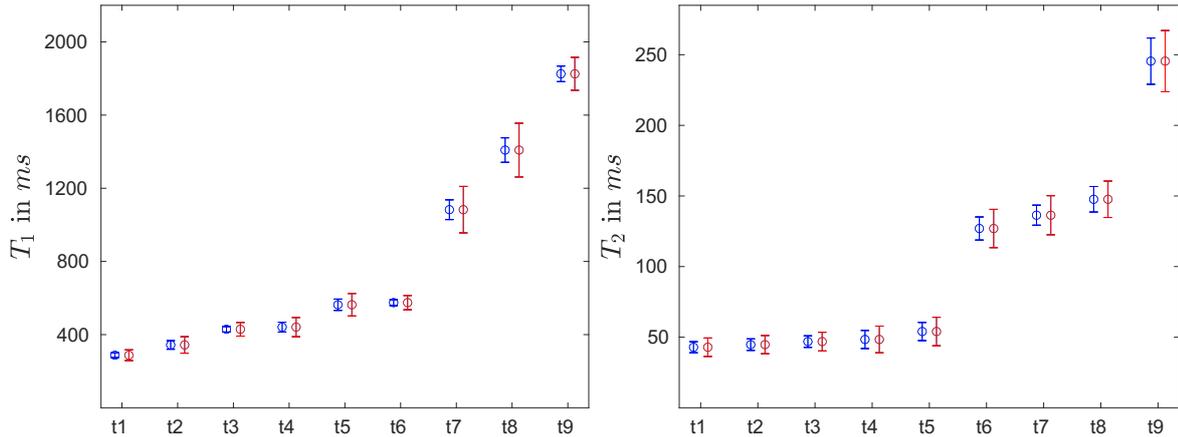


Figure 4.5: Results for 5 radial lines in k -space: mean values for the estimates obtained by the dictionary matching method within the tubes t_1, \dots, t_9 for T_1 (left) and T_2 (right), together with their standard deviation (blue errorbars) and mean uncertainty (red errorbars). Note that the errorbars are in ascending order in terms of the mean values for the estimates, respectively for T_1 and T_2 . Figure from Metzner et al. (2021).

Results from the reference measurements and results obtained by the dictionary matching method are compared in Figure 4.6 in terms of the uncertainties calculated for the latter. The uncertainties calculated for the dictionary matching solution for T_1 appear to describe well their deviations to the reference measurements. The same is true for the T_2 estimation but only for the smaller values of T_2 , while for larger values of T_2 a slight bias in the estimates produced by the dictionary matching method can be observed. Nonetheless, the calculated uncertainties for the larger T_2 values still characterize the size of the errors well in most cases.

In Figure 4.7 estimates of the dictionary matching as well as their uncertainty map are shown. Additionally, marginal distributions for T_1 and T_2 in two exemplary voxels of different tubes are presented. The dotted red lines in the images of the marginal distributions illustrate the bounds for T_1 and T_2 . Especially for large values of T_2 some of the distributions are truncated at these bounds, i.e. the bounds are truly informative. The derived marginal distributions can be used to calculate various probability statements, for example, to calculate the probability that T_1 is smaller than a certain limit. The probabilities for $T_1 < 1600$ ms of the two voxels in Figure 4.7 indicated by the arrows are 0.54 and almost 1, respectively.

The dictionary matching solution of the cardiac MRF in vivo data for T_1 and T_2 as well as their uncertainties are shown in Figure 4.8. An interesting fact can be seen: in some areas the uncertainty increases towards the centre of the body. This is reasonable since the coil sensitivity is lower there.

4.3.2 Simulated Data

In Figure 4.9, similar to Figure 4.3, mean uncertainties and standard deviations within each tube are compared for T_1 and T_2 . The uncertainties reflect the different within-tube variabilities of the estimates well, similarly to the phantom data above. Again, the calculated uncertainties appear to be slightly conservative.

The uncertainty of estimates with higher T_1 and T_2 values is also higher. However, the relative uncertainty (i.e. the uncertainty divided by the estimate) is roughly constant. When simulating further MRF data for which the same simulation setup was used, changing only

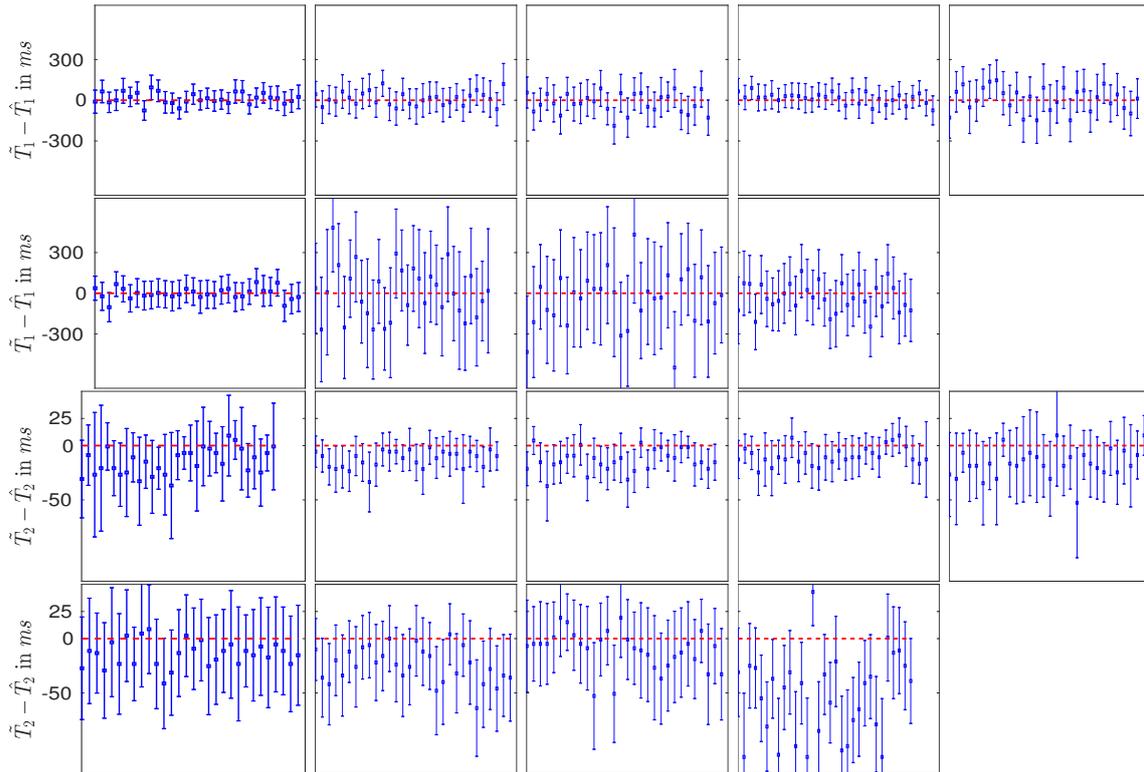


Figure 4.6: Differences between reference measurement (\tilde{T}_1, \tilde{T}_2) and results of dictionary matching method (\hat{T}_1, \hat{T}_2) in the 9 tubes of the phantom. The ordering from left to right is according to increasing $\tilde{T}_1 = (295, 339, 427, 435, 562, 587, 1088, 1432, 1834) ms$ and $\tilde{T}_2 = (31, 32, 33, 33, 37, 115, 120, 123, 191) ms$ values, respectively. Errorbars show the calculated uncertainty for the dictionary matching method. For clarity, results are shown only for a representative subset of voxels for each of the tubes. Figure from Metzner et al. (2021).

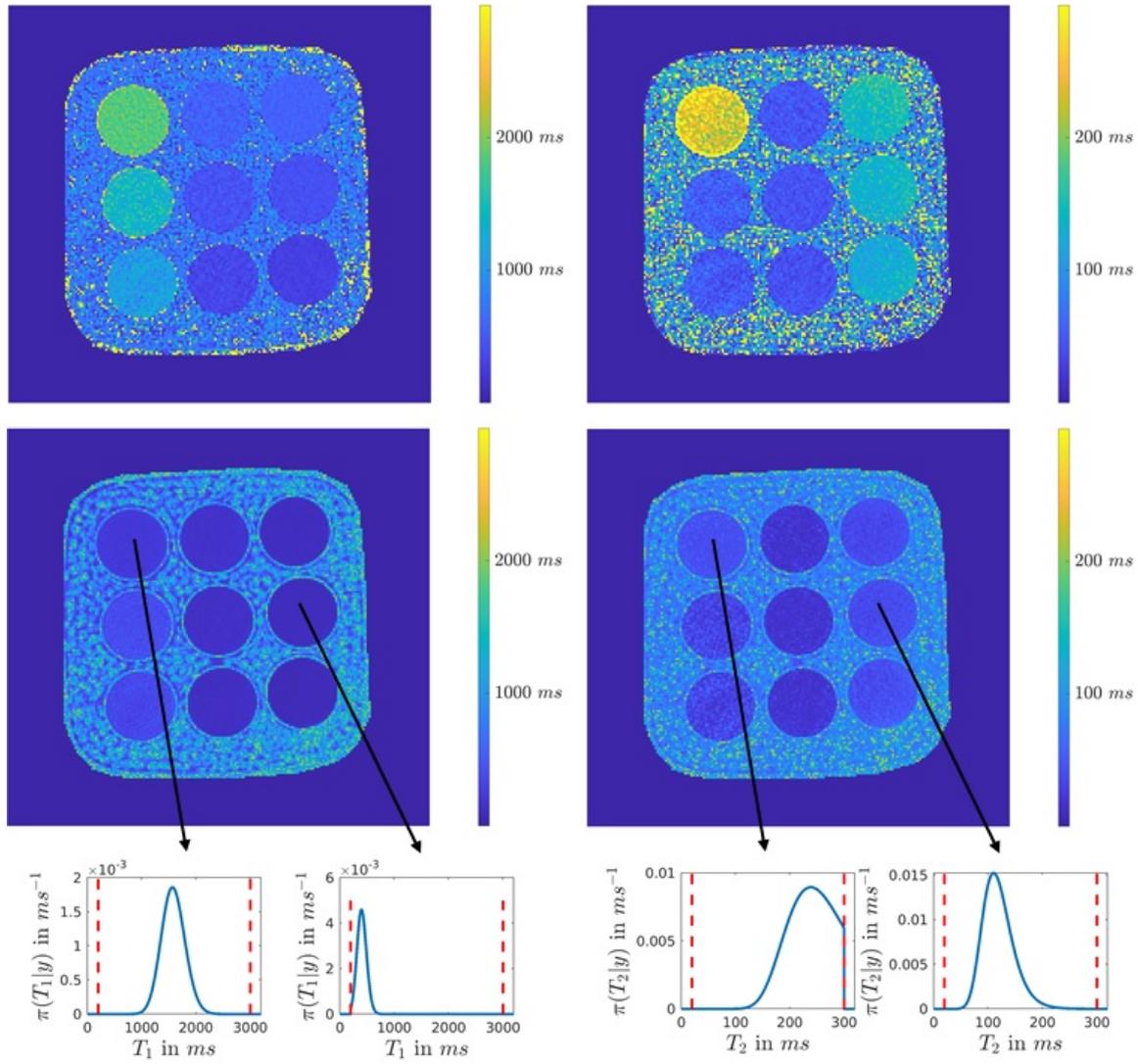


Figure 4.7: Top: Estimates obtained by the dictionary matching method for T_1 (left) and T_2 (right). Middle: uncertainties for the estimates of T_1 (left) and T_2 (right). Bottom: Examples of marginal distributions of T_1 and T_2 from two different voxels. The red dotted lines correspond to the bounds assumed for T_1 and T_2 . Figure from Metzner et al. (2021).

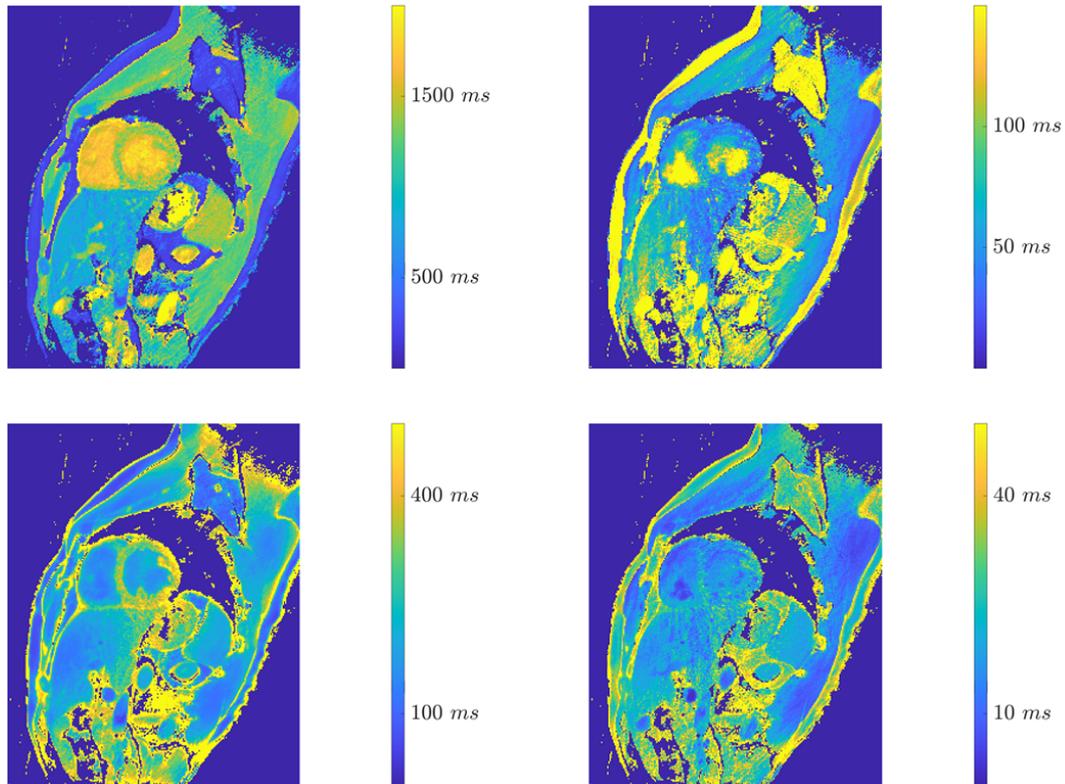


Figure 4.8: In vivo results. Top: Estimates obtained by the dictionary matching method for T_1 (left) and T_2 (right). Bottom: Uncertainties for the estimates of T_1 (left) and T_2 (right). Figure from Metzner et al. (2021).

repetition times (TR was set to 11 ms) or flip angle (the maximum of the magnitude of the sinusoidal curve was set to 60 degree), the relative uncertainties were again similar in both cases. The mean uncertainties in all nine tubes differed by a maximum of 10% when using the different repetition times, and by 20% when applying the different flip angle.

Figure 4.10 compares the estimates obtained by the dictionary matching method with the known ground truth in terms of the uncertainties calculated for the former for one particular tube. The results for this tube are representative for the other ones. Errors in the estimates obtained by the dictionary matching method are well described by the calculated uncertainties for both T_1 and T_2 .

4.4 Discussion

The residuals in Figure 4.2 obtained for the real and the imaginary part of the reconstructed magnetizations from the phantom data in a single voxel appear to approximately follow a Gaussian distribution. This supports the choice of the statistical model for the Bayesian uncertainty quantification. Note that the residuals shown in Figure 4.2 are typical and look similar in other voxels and other tubes. In Figure 4.1 two single reconstructed magnetization images are shown. Due to the severe aliasing artifacts it is not even possible to recognize the tubes. When looking at the sum of all these individual reconstructions, the tubes are now visible and the aliasing artifacts appear to be canceled out. This supports the assumption of the dictionary-based MRF method that aliasing artifacts act like noise over time.

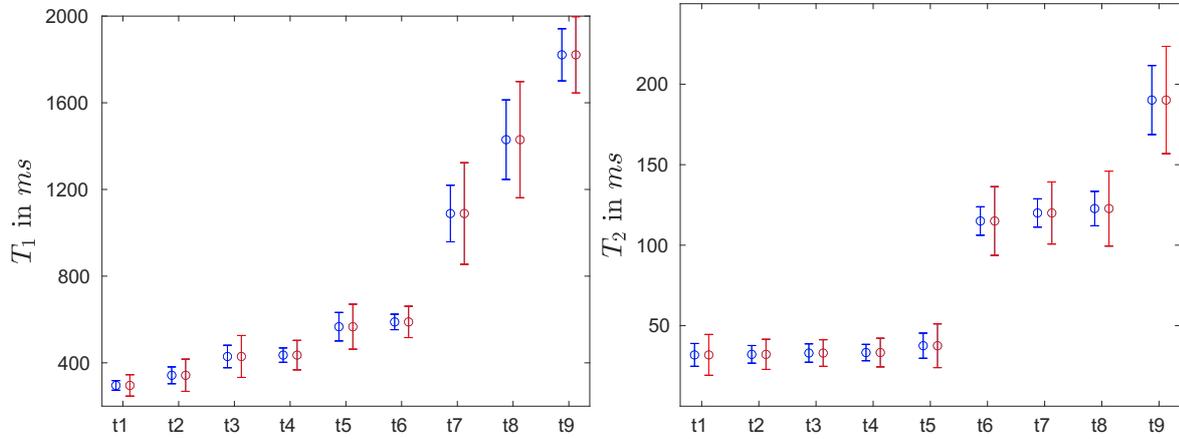


Figure 4.9: Simulation results: Mean values for the estimates obtained by the dictionary matching method within tubes t_1, \dots, t_9 for T_1 (left) and T_2 (right), together with their standard deviation (blue errorbars) and the mean uncertainty (red errorbars). Figure from Metzner et al. (2021).

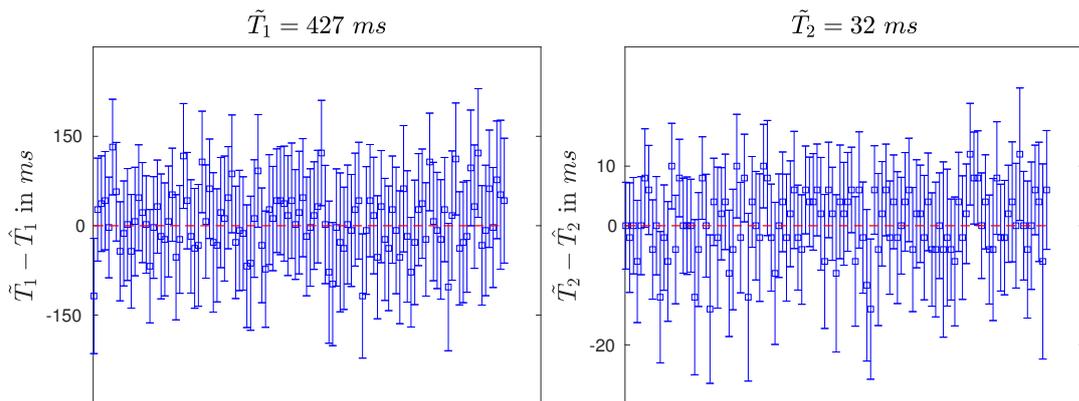


Figure 4.10: Simulation results: differences of the simulated ground truth \tilde{T}_1, \tilde{T}_2 and the estimates \hat{T}_1, \hat{T}_2 obtained by the dictionary matching method for one tube. Errorbars show calculated uncertainties. For clarity, results are shown only for a representative subset of voxels from the selected tube. Figure from Metzner et al. (2021).

The simulated MRF data were created by adding Gaussian noise of realistic size (the same size as for the phantom data) in k -space. When simulating data in the same way but without this additional measurement noise, results for the estimates of T_1 and T_2 essentially remain the same. This implies that the (unavoidable) aliasing errors due to the imperfect inverse Fourier transformation in the course of the observed magnetizations at each voxel are dominant for the calculated uncertainties. We have shown in Theorem 4.1.1 that the dictionary matching method is equivalent to fitting the modeled magnetizations to the observed ones. We want to highlight that the dictionary-based MRF method as well as our proposed uncertainty quantification are therefore based on the very same assumptions and can only be recommended if the (aliasing) errors appear like random noise, ideally following a Gaussian distribution (cf. Corollary 4.1.2). Whether this assumption is fulfilled is dependent on the MRF sequence design. Hence, when the sequence has been designed properly such that the aliasing errors can be modeled as noise, one can be confident that the assumptions of the proposed Bayesian uncertainty quantification are met as well.

The uncertainty u defined in (4.18) has been taken to describe the *a posteriori* belief about the size of the errors. We compared the within-tube standard deviations of the estimates with the mean uncertainties and observed that they are slightly smaller for both the real phantom MRF data (Figure 4.3) and the simulated MRF data (Figure 4.9). If the within-tube standard deviations can be taken as a reference for the size of the errors, then this size is slightly overestimated by the proposed Bayesian inference. One possible source of error might be the presence of temporal and/or spatial correlations of the residuals. Although we have shown that the assumption of a Gaussian distribution for the errors in the sequence of magnetization is in line with the assumption underlying the dictionary matching method, the suggested statistical model will generally provide an approximation to the true distribution of errors only. The choice of this simple statistical model therefore represents a limitation of the proposed approach. Future work could address more sophisticated statistical models which, e.g., can account for dependencies of the residuals. Nonetheless, we want to highlight that one benefit of our simple model is that calculations can be carried out by numerical quadrature using only the dictionary and the observed data in an efficient way. This is hardly possible when more sophisticated statistical models, containing additional unknowns, are employed and might therefore be impracticable. Another possible source of error in the proposed approach is the calculation of the Bayesian uncertainties themselves. The computations of integrals such as (4.18) are carried out on the discrete dictionary and might be prone to numerical approximation errors. For the results shown in this chapter this error source is not relevant since the outcomes remain essentially unchanged when doubling the resolution of the dictionary. However, the resolution of the dictionary plays an important role in the uncertainty calculation (as well as in the dictionary matching method itself) and can be a potential source of error. It has to be chosen with great care and dictionaries with low resolution should be avoided.

While for the simulated data errors in the dictionary-based estimates are well characterized by the calculated uncertainties (Figure 4.10), this does not hold for all results obtained for the phantom MRF data (Figure 4.6). In the latter case a significant bias can be observed for the estimates for the larger T_2 values. Nonetheless, the size of single errors of the estimates are still largely captured by the calculated uncertainties. Since the reference measurements

can be trusted for all T_2 values, the results of the dictionary matching method appear to have some bias for large T_2 values. Note that the uncertainty in the reference measurement is small compared to the observed discrepancies to the dictionary matching method. The reason for the bias in the dictionary matching method probably is some error in the underlying physical model used to create the dictionary. The proposed Bayesian uncertainty quantification assumes that the employed physical model is correct and therefore underrates calculated uncertainties in cases when a significant model error occurs. In our case, we do not include inhomogeneities of the B_1 field to our physical model. The radiofrequency excitation is assumed to be perfect (which would only be achieved with an infinite sinc pulse) which is a simplification and leads to errors in T_2 .

The fact that the unavoidable (large) aliasing errors which are treated as Gaussian noise are the dominant uncertainty source for the dictionary matching method makes this method - as well as the proposed uncertainty quantification - robust. As long as other error sources like measurement errors, incorrect k -space positioning, or the above mentioned errors in the dynamic model of magnetization, are small compared to the aliasing errors, both the dictionary matching method as well as the proposed uncertainty quantification can be expected to produce reasonable results.

As described in Appendix B we used an isochromat simulation with 201 isochromats per voxel as physical model. The Extended Phase Graph (EPG) formalism (Weigel (2015)) is another possibility to describe the magnetization for which no further parameters such as the number of isochromats are needed. We compared the results of both models and observed that the mean values of the estimates of all nine tubes have a deviation less than 1%. The same holds for within-tube standard deviations and for mean uncertainties.

In addition to the resolution of the dictionary, the results of the estimates as well as of the uncertainties also depend on its considered range. We examined this influence by creating two dictionaries, one with a smaller, and one with a larger range for T_2 . When taking the dictionary with the larger range of values for the T_2 relaxation time, the mean uncertainties across the tubes behave similarly compared to the standard deviations of the estimates. When taking the dictionary with a too small range for the T_2 relaxation times (i.e. some tubes have an actual value that is larger than the employed upper bound for T_2), biased estimates result and calculated uncertainties can no longer be expected to provide a reasonable characterization of the errors in the estimates. As expected, the range of the dictionary hence plays an important role both for the results achieved by the dictionary matching method and the proposed uncertainty quantification. It should therefore be chosen with great care.

In Zhao et al. (2018) the Cramér–Rao bound (CRB) was used to characterize the accuracy of estimates obtained in an MRF experiment. It was also utilized for optimizing the experimental design (in Zhao et al. (2018) the design parameters are chosen as repetition times and flip angle). Similarly to that, our proposed Bayesian uncertainty quantification also depends on the chosen protocol, and it could likewise be used for optimization of the experimental design. Future work could address this important feature and compare resulting protocols with those achieved by CRB minimization.

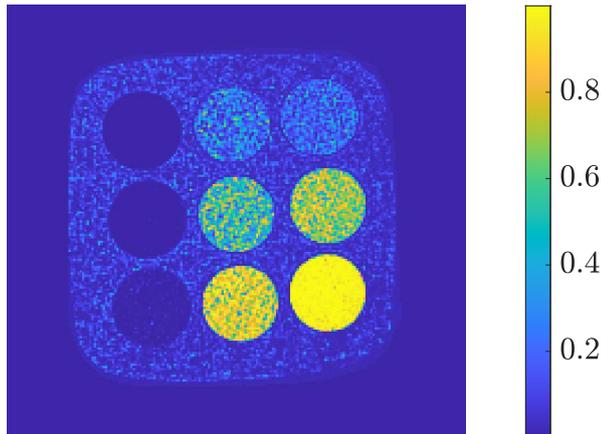


Figure 4.11: Probability map for the hypothesis $T_1 < 500 \text{ ms}$. Approximately true values for T_1 as obtained from the reference measurement are from left to right and top to bottom: $\tilde{T}_1 = (1834, 562, 587, 1432, 427, 435, 1088, 339, 295) \text{ ms}$. Figure from Metzner et al. (2021).

It was also possible to calculate uncertainties with our method for in vivo data. The results for the estimates of the dictionary matching method as well as their uncertainties are reasonable and show the applicability to complicated in vivo data.

One important advantage of Bayesian inference is that not only uncertainties as defined in (4.18) but a whole probability distribution is obtained for every voxel and every parameter. Basically, a complete uncertainty quantification can be made utilizing these distributions. In some cases, a single summary derived from it may be sufficient. In other cases, e.g., when the distributions deviate significantly from a Gaussian distribution (cf. the truncated posteriors in Figure 4.7 for example), the whole distribution should be considered. The posterior distribution can be used to calculate probabilities about a hypothesis such as $T_1 < \bar{T}_1$ for a specific \bar{T}_1 . A probability map of the hypothesis $T_1 < 500 \text{ ms}$ is shown in Figure 4.11 for the phantom data. It can be seen that T_1 is almost certainly smaller than 500 ms in all voxels of the tube on the bottom right, while the opposite is true for, e.g., the tube on the bottom left. In other tubes a clear decision cannot be made, which is represented through probabilities for this hypothesis which are neither close to 1 nor to 0. This sort of information might be useful in clinical follow-up measurements when trying to monitor possible changes in the properties of the tissue.

We decided to apply numerical quadrature techniques relying on the discrete, pre-computed dictionary. It is also possible to use other methods for Bayesian inferences which usually require extensive numerical calculations such as Markov Chain Monte Carlo methods (MCMC) (Robert and Casella (2013)). Note that results achieved by MCMC methods need some form of convergence checks. An advantage of the proposed approach is that its results are deterministic (in contrast to MCMC) and can be calculated efficiently. As input only the data (i.e. sequence of magnetization at every voxel) and the dictionary are required, which provides a convenient interface as these are the same inputs already needed for the dictionary matching method.

4.5 Summary

The dictionary matching method restricts itself to the calculation of estimates for the tissue-related parameters. The additional computation of reliable uncertainties for these estimates could be beneficial for allowing a meaningful assessment of single MRF results.

We developed a Bayesian uncertainty quantification for the estimates gained from the dictionary matching solution using a noninformative prior combined with informative priors for the relaxation times. Results are obtained through the analytical derivation of the marginal posterior of the relaxation times and can be carried out fast due to the usage of numerical quadrature techniques based on the pre-computed dictionary. Propriety of this posterior has been ensured. We showed that the proposed uncertainty quantification can be viewed as being based on the very same assumption as the dictionary matching method itself.

On the one hand, the proposed Bayesian uncertainty quantification has been tested on MRF data of a suitably designed phantom. The values of the relaxation times can be assumed to be constant within each tube. A consistency check was carried out by comparing the standard deviation within each tube to their mean uncertainty. The results are also compared to additional reference measurements. Similar outcomes are obtained for simulated data which altogether show that the proposed approach yields a reasonable uncertainty quantification. On the other hand, the proposed uncertainty quantification has also been applied to an in vivo measurement.

In addition to calculating uncertainties, the Bayesian approach also offers the possibility to derive probability statements which can increase confidence in conclusions drawn from MRF results. We believe that the calculation of uncertainties for the dictionary-based MRF estimates can be beneficial in applications for the assessment of the estimates.

5

Large-Scale Bayesian Modeling

The dictionary matching method requires pseudo-inverse Fourier transformations applied to measured Fourier data. Due to the undersampling in k -space, severe aliasing artifacts in the reconstructed images arise. The dictionary matching treats these artifacts as noise (Ma et al. (2013)). Nonetheless, modeling the data in k -space directly can avoid these artifacts and produce more accurate estimates. However, k -space modeling results in a large-scale problem since all parameters have to be treated in parallel. For instance, applying a maximum likelihood estimation results in a large-scale optimization problem in which the number of variables is in the order of 10^5 .

In this chapter our contribution is a *Bayesian inference for a large-scale k -space MRF model with different prior distributions for the tissue-related parameters*. A general class of regression problems in which the parameters model the spatial distribution of some property is introduced. MRF is one example but other applications might be possible. Different prior distributions are considered and the resulting properties of the posteriors explored such as propriety or existence of moments. The impact of a noninformative, a Gaussian Markov Random Field and a partition prior is compared in terms of the accuracy in the estimation of the sought parameters. Maximum a posteriori (MAP) estimates are determined for which a trust-region algorithm using a second-order approximation is presented. In order to solve the large-scale optimization an approximate calculation of the Hessian is provided.

We show that posterior distributions do not exist when an improper constant prior is assumed for the tissue parameters, and improper GMRF priors have to be augmented appropriately to reach propriety of the posterior. The accuracy of the estimates increases significantly when incorporating substantial prior knowledge that could be gained from a segmentation.

After introducing a general statistical model for regression tasks and three different prior distributions in Section 5.2, properties of the resulting posteriors are explored in Section 5.3. This is followed by a description of the calculation methods. Applicability and results are shown for simulated MRF data in Section 5.4.

Most of the work presented in this chapter is published in Metzner et al. (2019).

5.1 Related Work

k -space modeling of the MRF data has previously been shown to deliver improved estimates over the conventional dictionary matching approach. Davies et al. (2014) introduced a compressed sensing framework and proposed an iterative projected Landweber algorithm for such an estimation. By proving the restricted isometry property convergence to the sparse solution was shown in the noiseless case. Zhao et al. (2016) and Wübbeler and Elster (2017) proposed a maximum likelihood estimation in conjunction with a nonlinear regression task as another alternative. Sophisticated numerical procedures are applied for the calculation of the estimates. Wübbeler and Elster (2017) also derived a sparse approximation to the Hessian and used it to characterize the uncertainty associated with the obtained estimates.

From a statistical point of view the analysis of MRF data represents a large-scale nonlinear regression task. As the availability of large-scale data sets increases, solving these problems gain much interest. Applications range from the analysis of genome-wide association studies (Smith and Fahrmeir (2007)), logistic regression of text categorization (Aseervatham et al. (2011)) or Bayesian model averaging for functional magnetic resonance imaging (Smith and Fahrmeir (2007); Lee et al. (2014)). One key focus in this context is variable selection (e.g. Guan and Stephens (2011); Wu et al. (2014); Jiang et al. (2012)). Another direction of research considers particularly structured problems such as additive models (Meier et al. (2009)) or sparse regression models (Huang et al. (2008)). In MRF, the number of unknowns is in the order of 10^5 or larger which leads to a challenging regression task. Furthermore, the underlying physical model is nonlinear.

Bayesian inference for large-scale problems has already been carried out in several applications, see, e.g., Genkin et al. (2007); Smith and Fahrmeir (2007); Mumcuoglu et al. (1996); Seeger and Nickisch (2011). Standard calculations such as Markov Chain Monte Carlo fail due to the small step sizes required when the dimensionality scales up (Beskos et al. (2009)). Alternatives might be using sequential Monte Carlo (Beskos et al. (2014)), Hamiltonian Monte Carlo (Hoffman and Gelman (2014)), or Variational Bayes (Braun and McAuliffe (2010)). However, an alternative is to apply approximate methods such as the Laplace approximation (cf., e.g., Kass et al. (1990)) together with a MAP estimation (cf. Chapter 3). The Integrated Nested Laplace Approximation (INLA) (Rue et al. (2009)) approach is a tool for approximate Bayesian inference for latent Gaussian models. This class includes many statistical models and is widely used in current research (e.g. Ugarte et al. (2014); Santermans et al. (2016); Crewe and McCracken (2015)). Unfortunately it is not applicable to the problem of MRF since the data are not conditionally independent of the parameters. McGivney et al. (2018) already carried out Bayesian methods to MRF in the context of the partial volume effect.

Bayesian inference includes available prior knowledge into the analysis, allows probability statements to be made after the data have been observed, and it lends itself naturally to a complete uncertainty characterization of the results in terms of the posterior distribution. Applications such as MRF where the parameters model the spatial distribution of some property can benefit from using prior knowledge about that spatial distribution. For example, when spatial smoothness can be assumed, Gaussian Markov random field priors (Rue and Held (2005)) are well suited to employ such prior knowledge in a Bayesian inference. GMRF priors

have been applied successfully in many applications such as image classification (Nishii and Eguchi (2006)), image segmentation (Schmidt et al. (2013)), and magnetic resonance imaging (Schmid et al. (2006)).

5.2 Model

5.2.1 Statistical Model and Assumptions

The following class of homoscedastic Gaussian sampling models is considered in which the $n \times 1$ data vector Y is modeled as

$$Y \sim N(X(\theta_2)\theta_1, \sigma^2 \mathbf{I}_n). \quad (5.1)$$

$X(\theta_2)$ is a $n \times p$ matrix depending on the $p \times 1$ vector θ_2 for $n > 2p$, θ_1 is $p \times 1$ vector and \mathbf{I}_n denotes the identity matrix of dimension n . In general, $X(\theta_2)$ depends nonlinearly on θ_2 . Note that this is a general class of regression problems and the following results can not only be used in MRF, where the matrix $X(\theta_2)$ is nondiagonal, but also for other applications. Note also that the class of linear problems is also included in (5.1).

We make the following assumptions (which hold in the case of MRF, cf. Appendix A): First, the matrix $X(\theta_2)$ is bounded from above, i.e.

$$\|X(\theta_2)\| < c \quad (5.2)$$

for $c > 0$ and all θ_2 . In (5.2) $\|\cdot\|$ represents the Frobenius norm. Second,

$$\int_{\Theta_2} d\theta_2 = \infty, \quad (5.3)$$

which, for example, is satisfied when the domain Θ_2 of θ_2 equals \mathbb{R}^p or $(\mathbb{R}^+)^p$.

The goal is the inference of the parameter vectors θ_1 and θ_2 , given the observed data y . Different types of priors are considered, see below. The likelihood function is given by

$$l(\theta_1, \theta_2, \sigma; y) \propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \chi^2(\theta_1, \theta_2)\right), \quad (5.4)$$

for

$$\chi^2(\theta_1, \theta_2) := \|y - X(\theta_2)\theta_1\|^2. \quad (5.5)$$

In (5.5) $\|\cdot\|$ denotes the 2-norm. In the following it should be clear from the context if $\|\cdot\|$ stands for the 2-norm or the Frobenius norm.

5.2.2 Priors

Different priors are considered for a Bayesian inference and properties of the resulting posterior distributions are explored. The different types of parameters are modeled as being a priori

independent, i.e. all considered priors have the form

$$\pi(\theta_1, \theta_2, \sigma) = \pi(\theta_1)\pi(\theta_2)\pi(\sigma). \quad (5.6)$$

When θ_1 and θ_2 model spatial distributions of some property, similar priors might be considered for them. To ease notation, θ will denote either θ_1 or θ_2 throughout this subchapter.

The nuisance parameter σ acts like a scale parameter in the likelihood (5.4). We assume that no specific prior knowledge is available about it, and its prior is taken as

$$\pi(\sigma) = 1/\sigma \quad (5.7)$$

for $\sigma > 0$ and zero otherwise. This is a common choice as a noninformative prior for the variance parameter, cf. Kass and Wasserman (1996).

In the following three different classes of priors are considered for θ : a noninformative, a Gaussian Markov Random Field and a partition prior.

The first prior taken for θ is

$$\pi(\theta) \propto 1. \quad (5.8)$$

This is a noninformative prior and might be used when no prior knowledge about θ is available.

As second prior a Gaussian Markov Random Field prior (Rue and Held (2005)) is considered modeling spatial smoothness of the parameter θ (cf. Chapter 3)). In the case of MRF, θ represents the entries of all voxels for one unknown parameter (e.g. T_1). It is highly likely that neighboring voxels have similar parameters and this can be expressed via a GMRF prior:

$$\pi(\theta|\lambda) \propto \lambda^{-\frac{p-1}{2}} \exp\left(-\frac{1}{2\lambda}\theta^T K \theta\right), \quad (5.9)$$

where the neighboring matrix K of size $p \times p$ is given by

$$K_{rs} = \begin{cases} -1, & r \sim s \\ n_r, & r = s \\ 0, & \text{otherwise} \end{cases}. \quad (5.10)$$

The equivalence relation \sim indicates that two voxels are neighboring voxels. The number of neighboring voxels of voxel r is denoted by n_r and λ is a hyperparameter.

The neighboring matrix K is symmetric, positive semi-definite with a rank of $p-1$ according to Theorem 3.2.3.

It holds (cf. Example 3.2.2)

$$(\theta)_i | (\theta)_{-i} \sim N\left(\frac{1}{n_i} \sum_{k \sim i} (\theta)_k, \frac{\lambda}{n_i}\right).$$

For that reason the prior acts like a smoothness prior. The intensity of the smoothing can be handled via the hyperparameter λ . The smaller the λ , the more smoothing is enforced.

The last considered prior will be called a partition prior. The idea is that parameters belonging to the same segment are likely to have similar parameters. For this approach a

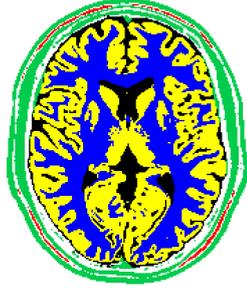


Figure 5.1: Illustration of partition prior gained from a segmentation of a slice in the human brain. Red segment: fat, blue segment: white matter, green segment: muscle and skin, yellow segment: gray matter, black segment: cerebrospinal fluid. Figure from Metzner et al. (2019).

segmentation of the image into q segments is assumed to be known beforehand. Note that single segments need not be connected, i.e. that no true spatial correlation for pixels within the same segment needs to exist. After that each parameter is assigned to a single partition. For every partition, prior knowledge about expected mean and variance values is assumed to be available. The motivation of this prior comes from MR images. In Figure 5.1 a segmentation which can be achieved from a fast, previous T_1 -weighted image (Brown et al. (2014)) for a slice of a human brain is shown. Values of, e.g., the T_1 relaxation times can be expected to be alike in each segment. Note that voxels of a partition do not need to be connected.

Let θ^i denote the parameters of θ corresponding to the i th segment of the partition. There are in total q different θ^i of length q_i . The hyperparameters are $\mu = (\mu_1, \dots, \mu_q)^T$ and $\lambda = (\lambda_1, \dots, \lambda_q)^T$. The partition prior is given as

$$\theta^i | \mu_i, \lambda_i \sim N(\mu_i \mathbf{1}_{q_i}, \lambda_i \mathbf{I}_{q_i}), \quad i = 1, \dots, q, \quad (5.11)$$

where $\mathbf{1}_{q_i} = (1, \dots, 1)^T \in \mathbb{R}^{q_i}$. $\theta^i | \mu_i, \lambda_i$ and $\theta^j | \mu_j, \lambda_j$ are assumed to be independent for $i \neq j$.

In the following theorem, a relationship between the partition prior and a GMRF prior is given when taking a particular noninformative prior for the mean μ .

Theorem 5.2.1 (Metzner et al. (2019)). *Let $\mu_i | \psi \sim N(0, \psi)$ for the location parameter of the partition prior (5.11). For $\psi \rightarrow \infty$, the resulting marginal prior reaches an intrinsic GMRF prior.*

Proof. For the marginal prior, we obtain

$$\begin{aligned} \pi(\theta^i | \lambda_i, \psi) &= \int \pi(\theta^i, \mu_i | \lambda_i, \psi) d\mu_i = \int \pi(\theta^i | \mu_i, \lambda_i) \pi(\mu_i | \psi) d\mu_i \\ &\propto \int \exp\left(-\frac{1}{2\lambda_i} [\theta^i - \mu_i \mathbf{1}_{q_i}]^T [\theta^i - \mu_i \mathbf{1}_{q_i}] - \frac{1}{2\psi} \mu_i^2\right) d\mu_i \\ &\propto \exp\left(-\frac{1}{2\lambda_i} (\theta^i)^T \left[I - \frac{\psi}{\lambda_i + n_i \psi} \mathbf{1}_{q_i} \mathbf{1}_{q_i}^T\right] \theta^i\right), \end{aligned}$$

and therefore reaches a particular GMRF prior with positive definite precision matrix

$$Q_i = \frac{1}{\lambda_i} \left(I - \frac{\psi}{\lambda_i + q_i \psi} \mathbf{1}_{q_i} \mathbf{1}_{q_i}^T \right).$$

For $\psi \rightarrow \infty$, Q_i approaches

$$\tilde{Q}_i = \frac{1}{q_i \lambda_i} \left(q_i \mathbf{I}_{q_i} - \mathbf{1}_{q_i} \mathbf{1}_{q_i}^T \right),$$

which is rank-deficient. The GMRF prior then becomes an intrinsic GMRF prior as in (5.9). \square

5.3 Properties of Posteriors and Numerical Calculations

We will now consider properties of the resulting posterior distributions. The posterior is given via Bayes' theorem (together with the conditional independence assumption (5.6)) as

$$\pi(\theta_1, \theta_2, \sigma, \omega_1, \omega_2 | y) \propto \pi(\theta_1 | \omega_1) \pi(\theta_2 | \omega_2) \pi(\sigma) l(\theta_1, \theta_2, \sigma; y), \quad (5.12)$$

where $l(\theta_1, \theta_2, \sigma; y)$ denotes the likelihood function (5.4), $\pi(\sigma)$ is the improper prior (5.7) and $\pi(\theta_i | \omega_i)$ stands for one of the three priors (5.8), (5.9) or (5.11) introduced above with its corresponding hyperparameters ω_i , $i = 1, 2$. We will focus on the propriety of the posterior and the existence of moments.

5.3.1 Properties of Posteriors

The selection of the noninformative prior for θ_2 does not yield a proper posterior.

Theorem 5.3.1 (Metzner et al. (2019)). *Consider model (5.1), together with assumptions (5.2) and (5.3), and a prior of the form $\pi(\theta_1, \theta_2, \sigma) \propto \pi(\theta_1)1/\sigma$. Then, the resulting posterior (5.12) does not exist.*

Proof. Since $\|X(\theta_2)\|$ is bounded (cf. assumption (5.2)), $\chi^2(\theta_1, \theta_2)$ in (5.5) is (for fixed θ_1) bounded from above. The likelihood $l(\theta_1, \theta_2, \sigma; y)$ in (5.4) is then bounded from below, given σ and θ_1 . It immediately follows from assumption (5.3) that

$$\int l(\theta_1, \theta_2, \sigma; y) \pi(\theta_1 | \omega_1) 1/\sigma d\theta_2 = \infty$$

holds. \square

Using a truncated constant prior for θ_2 leads to a proper posterior according to the following Theorem.

Theorem 5.3.2 (Metzner et al. (2019)). *Consider model (5.1), together with assumption (5.2). Additionally, assume $\min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) > 0$, and consider a prior of the form $\pi(\theta_1, \theta_2, \sigma) = \pi(\theta_1) \pi(\theta_2) \pi(\sigma)$. Assume further that the noninformative prior (5.8) is taken for θ_1, θ_2 , truncated w.r.t. θ_1, θ_2 , i.e. $\pi(\theta_1, \theta_2, \sigma) \propto 1/\sigma$ for $(\theta_1, \theta_2) \in \Theta = [(\underline{\theta}_1, \underline{\theta}_2), (\bar{\theta}_1, \bar{\theta}_2)]$, $-\infty < \underline{\theta}_i \leq \bar{\theta}_i < \infty$, $i = 1, 2$ and $\sigma > 0$, and zero otherwise. Then, the posterior (5.12) exists.*

Proof.

$$\begin{aligned} \int l(\theta_1, \theta_2, \sigma; y) \pi(\theta_1, \theta_2, \sigma) d\sigma d\theta_1 d\theta_2 &\propto \int \frac{1}{(\chi^2(\theta_1, \theta_2))^{n/2}} d\theta_1 d\theta_2 \\ &\leq \left(\min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) \right)^{-n/2} \text{vol}(\Theta) < \infty. \end{aligned} \quad \square$$

It is reasonable to assume prior knowledge about the parameters in the sense of (conservative) bounds for their domain. This allows for a Bayesian inference under an otherwise noninformative prior. However, results such as posterior moments will (strongly) depend on these bounds since the posterior does not exist without them. Hence, the bounds are truly informative and employed bounds therefore need to be justified. Note that when applying approximate methods such as a Laplace approximation a dependence on these bounds might never be noticed while the true posterior results depend strongly on them.

The choice of a GMRF prior without further augmentation does not lead to a proper posterior according to

Theorem 5.3.3 (Metzner et al. (2019)). *Consider model (5.1), together with assumption (5.2), $\Theta_2 = \mathbb{R}^p$, and a prior of the form $\pi(\theta_1, \theta_2, \sigma) = \pi(\theta_1)\pi(\theta_2)\pi(\sigma)$ with $\pi(\sigma) = 1/\sigma$. When the GMRF prior $\pi(\theta_2|\lambda_2)$ from (5.9) is taken for θ_2 , the posterior (5.12) does not exist.*

Proof. The structure matrix in (5.10) is rank-deficient with $K(1, \dots, 1)^T = 0$. Assume that a GMRF prior is taken for θ_2 . Consider the integral

$$\int l(\theta_1, \theta_2, \sigma; y) \pi(\theta_2|\lambda_2) \pi(\theta_1) \pi(\sigma) d\theta_2, \quad (5.13)$$

and apply a transformation of θ_2 according to $\theta_2 = \zeta_2 1 + A\zeta_{-2}$, where $\zeta_2 \in \mathbb{R}$ and $\zeta_{-2} \in \mathbb{R}^{p-1}$, and where the constant $p \times (p-1)$ matrix A is chosen such that $A^T 1 = 0$ and $\text{rank}(A) = p-1$ hold. According to assumption (5.2), $\|X(\theta_2)\|$ is bounded from above for all ζ_2 . This implies that the likelihood $l(\theta_1, \theta_2, \sigma; y)$ in (5.4) is bounded from below in dependence on ζ_2 . Since $K1 = 0$, $\pi(\theta_2|\lambda_2)$ does not depend on ζ_2 . It follows that the integral (5.13) does not exist. \square

The above proof suggests that the reason for the impropriety is the rank-deficient matrix K . Since K has rank $p-1$, we need to compensate for this one dimension. The next theorem shows that with the help of a proper prior on the mean of θ_2 , propriety can be achieved. Note, that only one further prior for the mean of θ_2 is needed.

Theorem 5.3.4 (Metzner et al. (2019)). *Consider model (5.1), together with assumption (5.2) and let $\Theta_1 = \Theta_2 = \mathbb{R}^p$. Assume $\min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) > 0$ and suppose that the eigenvalues of $X(\theta_2)^T X(\theta_2)$ are bounded from below for all θ_2 by a constant greater than zero. Consider a prior of the form $\pi(\theta_1, \theta_2, \sigma) = \pi(\theta_1)\pi(\theta_2)\pi(\sigma)$ with $\pi(\sigma) = 1/\sigma$. Assume a GMRF prior is taken for θ_1 and for θ_2 , and that these GMRF priors are augmented by (i) an inverse gamma hyperprior on the λ_l , i.e. $\lambda_l \sim IG(\alpha_l, \beta_l)$, $l = 1, 2$, and (ii) a proper prior for the mean of θ_2 . Then, the posterior (5.12) exists if $\alpha_l > 0$ holds for $l = 1, 2$. Furthermore, for $l = 1, 2$ the marginal posterior $\pi(\theta_l|y, \alpha, \beta)$ has moments up to order i if in addition $2\alpha_l > i$ also holds.*

Proof. The posterior distribution is given by

$$\begin{aligned}
& \pi(\theta_1, \theta_2 | y, \alpha, \dots, \beta) \\
& \propto \int l(\theta_1, \theta_2, \sigma; y) \pi(\sigma) \pi \left(p^{-1} \sum_i (\theta_2)_i \right) \prod_{l=1}^2 \pi(\theta_l | \lambda_l) \pi(\lambda_l | \alpha_l, \beta_l) d\sigma d\lambda_l \\
& \propto \pi \left(p^{-1} \sum_i (\theta_2)_i \right) \int l(\theta_1, \theta_2, \sigma; y) \pi(\sigma) d\sigma \\
& \times \prod_{l=1}^2 \int \lambda_l^{-p/2+1/2-\alpha_l-1} \exp \left(-\frac{1}{2\lambda_l} (\theta_l^T K \theta_l + 2\beta_l) \right) d\lambda_l \\
& \propto \frac{1}{(\chi^2(\theta_1, \theta_2))^{n/2}} \prod_{l=1}^2 \frac{1}{(\theta_l^T K \theta_l + 2\beta_l)^{p/2+\alpha_l-1/2}} \pi \left(p^{-1} \sum_i (\theta_2)_i \right),
\end{aligned}$$

where $\pi(p^{-1} \sum_i (\theta_2)_i)$ denotes some proper prior on the mean of θ_2 .

By assumption, it is $\min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) =: \chi_{min}^2 > 0$ and the eigenvalues of $X^T X$ are bounded from below, i.e. $\lambda(X(\theta_2)^T X(\theta_2)) \geq \lambda_{min} > 0$ for all θ_2 and all eigenvalues $\lambda(X(\theta_2)^T X(\theta_2))$ of $X(\theta_2)^T X(\theta_2)$, except for null sets.

Let $\hat{\theta}_1(\theta_2) := (X^T X)^{-1} X^T y$ be the minimum of $\chi^2(\theta_1, \theta_2) = \|y - X\theta_1\|^2$ with respect to θ_1 for $X := X(\theta_2)$.

It follows

$$\begin{aligned}
& \chi^2(\theta_1, \theta_2) \\
& = \|y - X\theta_1\|^2 = \frac{1}{2} \left(y^T y - 2y^T X\theta_1 + \theta_1^T X^T X\theta_1 \right) \\
& = \|y - X(X^T X)^{-1} X^T y\|^2 + (\theta_1 - (X^T X)^{-1} X^T y)^T X^T X (\theta_1 - (X^T X)^{-1} X^T y) \\
& = \|y - X\hat{\theta}_1\|^2 + (\theta_1 - \hat{\theta}_1)^T X^T X (\theta_1 - \hat{\theta}_1) \\
& \geq \chi_{min}^2 + \left((\theta_1 - \hat{\theta}_1)^T X^T X (\theta_1 - \hat{\theta}_1) \right).
\end{aligned}$$

Let $\tilde{\theta}_1 := \theta_1 - \hat{\theta}_1$ and apply the transformation according to $\tilde{\theta}_1 = \zeta_1 \mathbf{1} + A\zeta_{-1}$, where $\zeta_1 \in \mathbb{R}$ and $\zeta_{-1} \in \mathbb{R}^{p-1}$, and where the constant $p \times (p-1)$ matrix A is chosen such that $A^T \mathbf{1} = 0$

and $\text{rank}(A) = p - 1$ holds. Using $K1 = 0$ leads to

$$\begin{aligned}
 & \int \frac{1}{(\chi^2(\theta_1, \theta_2))^{n/2}} \frac{1}{(\theta_1^T K \theta_1 + 2\beta_1)^{p/2+\alpha_1-1/2}} d\theta_1 \\
 & \leq \int \frac{1}{(\chi_{\min}^2 + \lambda_{\min} \|\theta_1 - \hat{\theta}_1\|^2)^{n/2}} \frac{1}{(\theta_1^T K \theta_1 + 2\beta_1)^{p/2+\alpha_1-1/2}} d\theta_1 \\
 & = \int \frac{1}{(\chi_{\min}^2 + \lambda_{\min} \|\tilde{\theta}_1\|^2)^{n/2}} \frac{1}{(2\beta_1 + (\tilde{\theta}_1 + \hat{\theta}_1)^T K (\tilde{\theta}_1 + \hat{\theta}_1))^{p/2+\alpha_1-1/2}} d\tilde{\theta}_1 \\
 & = \int \frac{1}{(\chi_{\min}^2 + \lambda_{\min} \|\zeta_1 1 + A\zeta_{-1}\|^2)^{n/2}} \\
 & \quad \times \frac{1}{(2\beta_1 + (A\zeta_{-1} + \hat{\theta}_1)^T K (A\zeta_{-1} + \hat{\theta}_1))^{p/2+\alpha_1-1/2}} d\zeta_1 d\zeta_{-1} \\
 & = \int \frac{1}{(\chi_{\min}^2 + \lambda_{\min} (\|\zeta_1 1\|^2 + \|A\zeta_{-1}\|^2))^{n/2}} \\
 & \quad \times \frac{1}{(2\beta_1 + (A\zeta_{-1} + \hat{\theta}_1)^T K (A\zeta_{-1} + \hat{\theta}_1))^{p/2+\alpha_1-1/2}} d\zeta_1 d\zeta_{-1} \\
 & \leq \int \frac{1}{(\chi_{\min}^2 + \lambda_{\min} \|\zeta_1 1\|^2)^{n/2}} d\zeta_1 \\
 & \quad \int \frac{1}{(2\beta_1 + (A\zeta_{-1} + \hat{\theta}_1)^T K (A\zeta_{-1} + \hat{\theta}_1))^{p/2+\alpha_1-1/2}} d\zeta_{-1} \\
 & \leq c_1 c_2 =: c
 \end{aligned}$$

for $n > 1$, $\alpha_1 > 0$ and all θ_2 .

Let us now consider the integral

$$\int \frac{1}{(\theta_2^T K \theta_2 + 2\beta_l)^{p/2+\alpha_l-1/2}} \pi(p^{-1} \sum_i (\theta_2)_i) d\theta_2. \quad (5.14)$$

By again applying the transformation $\theta_2 = \zeta_2 1 + A\zeta_{-2}$, and by using $p^{-1} \sum_i (\theta_2)_i = \zeta_2$, this integral exists if $\alpha_l > 0$. Hence, the posterior exists if $(n > 1$ and) $\alpha_l > 0$ for $l = 1, 2$.

The i -th moments of the marginal posterior $\pi(\theta_l | y, \alpha, \beta)$ exist if

$$\int \pi(\theta_l | y, \alpha, \beta) (\theta_l)_{p_1} \dots (\theta_l)_{p_i} d\theta_l < \infty \quad (5.15)$$

holds for all $p_1, \dots, p_i \in \{1, \dots, p\}$. Similar considerations as above show that (5.15) holds when $2\alpha_l > i$. \square

The improper GMRF prior with no further augmentation together with the improper prior for σ lead to an improper posterior distribution due to the rank-deficiency of K . This can be solved by adding only one further prior for the mean of θ_2 .

The partition prior is proper since it is a simple normal distribution. Together with the improper prior for σ a proper posterior is achieved and the moments exist according to the next theorem.

Theorem 5.3.5 (Metzner et al. (2019)). *Consider model (5.1), together with assumption (5.2) and let $\Theta_1 = \Theta_2 = \mathbb{R}^p$. Assume $\min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) > 0$ and consider a prior of the form $\pi(\theta_1, \theta_2, \sigma) = \pi(\theta_1)\pi(\theta_2)\pi(\sigma)$ with $\pi(\sigma) = 1/\sigma$. Assume the partition priors $\theta_l^i | \mu_{li}, \lambda_{li} \sim N(\mu_i 1_{q_i}, \lambda_i I_{q_i})$ are taken for θ_1 and for θ_2 , together with the hyperpriors $\lambda_{li} \sim IG(\tilde{\alpha}_{li}, \tilde{\beta}_{li})$*

5. Large-Scale Bayesian Modeling

relevant for the i -th partition of θ_l , $l = 1, 2$. Then, the posterior (5.12) exists if $\tilde{\alpha}_i > 0$ holds for all $l = 1, 2$ and $i = 1, \dots, q_l$. Furthermore, all moments up to order k of the marginal posterior $\pi(\theta_l|y, \mu, \tilde{\alpha}, \tilde{\beta})$ exist if $2\tilde{\alpha}_i > k$ for all $i = 1, \dots, q_l$.

Proof. As posterior distribution, we get

$$\begin{aligned} & \pi(\theta_1, \theta_2|y, \mu, \tilde{\alpha}, \tilde{\beta}) \\ & \propto \int l(\theta_1, \theta_2, \sigma; y) \prod_{l=1}^2 \prod_{i=1}^q \pi(\theta_{li}|\mu_{li}, \lambda_{li}) \pi(\lambda_{li}|\tilde{\alpha}_{li}, \tilde{\beta}_{li}) \pi(\sigma) d\sigma d\lambda \\ & \propto \frac{1}{(\chi^2(\theta_1, \theta_2))^{n/2}} \prod_{l=1}^2 \prod_{i=1}^q \int \lambda_{li}^{-\tilde{\alpha}_{li}-1-q_{li}/2} \\ & \quad \times \exp\left(-\frac{1}{2\lambda_{li}}(2\tilde{\beta}_{li} + [\theta_{li} - \mu_{li}1]^T[\theta_{li} - \mu_{li}1])\right) d\lambda_{li} \\ & \propto \frac{1}{(\chi^2(\theta_1, \theta_2))^{n/2}} \frac{1}{\prod_{l=1}^2 \prod_{i=1}^q (2\tilde{\beta}_{li} + [\theta_{li} - \mu_{li}1]^T[\theta_{li} - \mu_{li}1])^{\tilde{\alpha}_{li}+q_{li}/2}}, \end{aligned}$$

which leads for $\chi_{min}^2 := \min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) > 0$ to

$$\begin{aligned} & \int \pi(\theta|y, \mu, \tilde{\alpha}, \tilde{\beta}) d\theta_1 d\theta_2 \\ & \leq \frac{1}{(\chi_{min}^2)^{n/2}} \int \frac{1}{\prod_{l=1}^2 \prod_{i=1}^q (2\tilde{\beta}_{li} + [\theta_{li} - \mu_{li}1]^T[\theta_{li} - \mu_{li}1])^{\tilde{\alpha}_{li}+q_{li}/2}} d\theta_1 d\theta_2 \\ & = \frac{1}{(\chi_{min}^2)^{n/2}} \prod_{l=1}^2 \prod_{i=1}^q \int \frac{1}{(2\tilde{\beta}_{li} + [\theta_{li} - \mu_{li}1]^T[\theta_{li} - \mu_{li}1])^{\tilde{\alpha}_{li}+q_{li}/2}} d\theta_{li}. \end{aligned}$$

It follows that the posterior exists if $\tilde{\alpha}_i > 0$ holds for all l and i . Furthermore, moments up to order k of the marginal posterior $\pi(\theta_l|y, \mu, \tilde{\alpha}, \tilde{\beta})$ exist if $2\tilde{\alpha}_i > k$ holds for $i = 1, \dots, q$. \square

This is an interesting result since the GMRF and the partition prior have a close relationship according to Theorem 5.2.1. The equivalence between these two priors holds when a normal distribution with a variance converging to ∞ is assumed for the mean of the partition prior. This leads to an uninformative prior which is not the case for the partition prior. For the propriety of the posterior when using a GMRF prior one further prior has to be introduced whereas this is not necessary for the case of the partition prior.

It is also possible to combine the GMRF and partition prior.

Theorem 5.3.6 (Metzner et al. (2019)). *Consider model (5.1), together with assumption (5.2) and let $\Theta_1 = \Theta_2 = \mathbb{R}^p$. Assume $\min_{\theta_1, \theta_2} \chi^2(\theta_1, \theta_2) > 0$ and suppose that the eigenvalues of $X(\theta_2)^T X(\theta_2)$ are bounded from below for all θ_2 by a constant greater than zero. Consider a prior of the form $\pi(\theta_1, \theta_2, \sigma) = \pi(\theta_1)\pi(\theta_2)\pi(\sigma)$ with $\pi(\sigma) = 1/\sigma$. Assume that for each $\theta_l, l = 1, 2$, either the GMRF prior from (5.9) is taken, or the partition prior according to (5.11). Assume further that GMRF priors are augmented by (i) an inverse gamma hyperprior on the λ_l , i.e. $\lambda_l \sim IG(\alpha_l, \beta_l)$ where $\alpha_l > 0$ and for $l \neq 1$ also by (ii) a proper prior for the mean of θ_l . For the partition priors $\pi(\theta_{li}|\mu_{li}, \lambda_{li})$, the μ_{li} are assumed to be known, and for the λ_{li} , the hyperpriors $\lambda_{li} \sim IG(\tilde{\alpha}_{li}, \tilde{\beta}_{li})$ are taken where $\tilde{\alpha}_{li} > 0$. Then, the posterior exists. For $l = 1, 2$ the marginal posterior $\pi(\theta_l|y, \alpha, \beta, \tilde{\alpha}, \tilde{\beta})$ has moments up to order i if $2\alpha_l > i$ when a*

GMRF prior augmented by a prior for the mean has been taken for θ_l . When a partition prior has been taken for θ_l , then the marginal posterior $\pi(\theta_l|y, \alpha, \beta, \tilde{\alpha}, \tilde{\beta})$ has moments up to order k if $2\tilde{\alpha}_{li} > k$ for $i = 1, \dots, q_l$.

Proof. This statement easily follows from the proofs of Theorems 5.3.4 and 5.3.5. \square

5.3.2 Numerical Calculations

Bayesian inference for high-dimensional problems can be very challenging. In order to calculate moments of the posterior distributions, high-dimensional integrals have to be calculated. When analytical expressions are not available numerical procedures have to be applied. A common approach uses MCMC methods. In Chapter 3 the fundamental idea of MCMC is presented. In high dimensions it is challenging to guarantee convergence of the samples. We tested the Hamiltonian Monte Carlo algorithm using the No-U-Turn sampler (Hoffman and Gelman (2014)) for the application of MRF. In the paper by Hoffman and Gelman (2014) good results for a multivariate normal distribution up to dimension 250 are shown. The HMC sampler employs the gradient of the posterior distribution which is in the case of MRF a very time consuming task. Additionally, we carried out tests for multivariate t -distributions with a block diagonal scale matrix of increasing dimensions (up to dimension 10^5). We determined an underestimation of the variance leading us to the conclusion that HMC will not deliver reasonable results in these dimensions.

When direct sampling from the posterior is unfeasible one can resort to approximate calculation techniques such as Laplace approximations (Kass et al. (1990)) as described in Chapter 3. Analytical marginalization w.r.t. hyperparameters is helpful in this context and can be achieved for the GMRF and the partition prior using appropriate hyperpriors.

The marginal posterior is then maximized numerically to get the MAP estimate of the parameters. In the case of MRF, this is a challenging optimization task due to the high dimensionality. We employed a trust-region algorithm (Nocedal and Wright (2006); Fletcher (2013)) based on a second-order approximation of the objective function as in Wübbeler and Elster (2017). Since calculating the exact Hessian is not feasible in the case of MRF, an approximate Hessian is used and its derivation is explained below. A pseudo-code of our trust-region method is given in Figure 5.2. Note that the algorithm is suitable for the whole class of regression models (5.1), not only in the MRF case. In this algorithm θ^k denotes the current iterate of θ and the operator ϕ maps the parameters of θ^k onto a feasible region $\Theta = [\underline{\theta}, \bar{\theta}]$ by

$$(\phi[\theta^k])_i \begin{cases} \theta_i^k, & \theta_i^k \leq \bar{\theta}_i^k \\ \bar{\theta}_i^k, & \theta_i^k \geq \bar{\theta}_i^k \\ \theta_i^k, & \text{otherwise} \end{cases} .$$

Here, θ_i^k denotes one parameter in one voxel, i.e. $\theta^k = (\theta_1^k, \dots, \theta_p^k)^T$. The trust-region method establishes a second-order model around the current iterate θ^k . This is then minimized within the active trust region. In the case of MRF, the Hessian is just approximated, therefore our trust-region algorithm also calculates a gradient descent step and then either takes the Newton or the gradient step. Next, the minimizer is evaluated with the objective function.

5. Large-Scale Bayesian Modeling

In dependence on the ratio of the predicted and realized gain in the objective function, the iterate θ^k and the trust region radius are replaced. To ensure numerical stability in the calculation of the Newton step, a symmetric, positive-definite matrix Λ is added onto the Hessian. Convergence of the iterates is ensured through the convergence of the trust-region algorithm (Toint (1997); Deng et al. (1993)).

The values of the trust-region algorithm are chosen as follows: $\Lambda = \text{Id}_p$, $\Delta = [10^{-3}, 100]$, $k_{max} = 1000$, $\mu = 0.5$, $\eta = 0.7$, $\gamma_0 = 0.5$, $\gamma_1 = 0.7$, $\gamma_2 = 1.2$, $\beta_1 = 0.1$. The feasible region of Θ is given by large, reasonable bounds for the different parameter types: $\rho \in [0.1, 2]$, $T_1 \in [100, 10000]$, $T_2 \in [10, 1000]$ and $\delta f \in [-300, 300]$.

```

1: procedure TRUST-REGION ALGORITHM
2:   Given:  $\theta^0$ ,  $\Lambda \succ 0$ ,  $\Delta > 0$ ,  $k_{max}$ ,  $\mu \in (0, 1)$ ,  $\eta \in (\mu, 1)$ ,  $0 < \gamma_0 \leq \gamma_1 < 1 \leq \gamma_2$ ,  $\beta_1 \in (0, 1)$ 
3:   for  $k = 0 : k_{max}$  do
4:     Calculate  $\chi^2(\theta^k)$ ,  $g^k = \nabla_{\theta} \chi^2(\theta^k)$  and approximate Hessian  $H^k$ 
5:     Define  $A^k := H^k + \Lambda$  and  $h^k(\theta^k + s) := \chi^2(\theta^k) + s^T g^k + \frac{1}{2} s^T (H^k + \Lambda) s$ 
6:     Calculate trial step  $s$ :
7:     Calculate  $t_1 = \arg \min_{\psi} h^k(\phi[\theta^k - \psi g^k])$  such that  $\|\phi[\theta^k - \psi g^k] - \theta^k\| \leq \Delta$  and
8:      $t_2 = \arg \min_{\psi} h^k(\phi[\theta^k - \psi(A^k)^{-1} g^k])$  such that  $\|\phi[\theta^k - \psi(A^k)^{-1} g^k] - \theta^k\| \leq \Delta$ 
9:     if  $(\chi^2(\theta^k) - h^k(\phi[\theta^k - t_2(A^k)^{-1} g^k])) \geq \beta_1(\chi^2(\theta^k) - h^k(\phi[\theta^k - t_1 g^k]))$  then
10:        $s = \phi[\theta^k - t_2(A^k)^{-1} g^k] - \theta^k$ 
11:     else
12:        $s = \phi[\theta^k - t_1 g^k] - \theta^k$ 
13:     end if
14:     Calculate  $\rho_k = (\chi^2(\theta^k) - \chi^2(\theta^k + s)) / (\chi^2(\theta^k) - h^k(\theta^k + s))$ 
15:     if  $\rho_k > \mu$  then
16:        $\theta^{k+1} = \theta^k + s$ 
17:       if  $\rho_k > \eta$  then
18:         Select  $\Delta \in [\Delta, \gamma_2 \Delta]$ 
19:       else
20:         Select  $\Delta \in [\gamma_1 \Delta, \Delta]$ 
21:       end if
22:     else
23:        $\theta^{k+1} = \theta^k$  and select  $\Delta \in [\gamma_0 \Delta, \gamma_1 \Delta]$ 
24:     end if
25:   end for
26:   return  $\theta^{k_{max}}$ 
27: end procedure

```

Figure 5.2: Implementation of the trust-region algorithm used to calculate the MAP.

The choice of the starting values is crucial for a nonconvex problem as in the MRF setting. After the optimization has terminated, the marginal posterior can then be approximated by

$$\theta|y, \mu \sim N(\hat{\theta}, V), \quad (5.16)$$

where $\hat{\theta}$ denote the MAP estimates from the marginal posterior, V the inverse of the Hessian of the negative log marginal posterior at $\hat{\theta}$ and μ the additional hyperparameters. With the help of (5.16) approximated variances and credible regions can be calculated. Again, in high dimensional cases it can be difficult to actually calculate V . It may be possible when $X(\theta)$

has a special structure but in the case of MRF it was not feasible. Alternatively, we used an approximate calculation of V using a sparse block-diagonal approximation of the Hessian of χ^2 derived by Wübbeler and Elster (2017).

The gradient $g = (\partial\chi^2/\partial\theta^T)^T$ of the function χ^2 (defined in (5.20)) is obtained as

$$g = -\operatorname{Re} \left(\sum_{l=1}^L J_l^H \mathcal{F}_l^H (y_l - \mathcal{F}_l m_l(\theta)) \right), \quad (5.17)$$

where $J_l = \partial m_l / \partial \theta^T$ is a sparse, $N \times 4N$ Jacobian of the magnetization model m_l (N denotes the number of voxels per image). An approximate Hessian $H = (\partial^2 \chi^2 / \partial \theta \partial \theta^T)$ is then given by

$$H \approx \operatorname{Re} \left(\sum_{l=1}^L J_l^H \mathcal{F}_l^H \mathcal{F}_l J_l \right), \quad (5.18)$$

where a Gauss Newton approximation has been applied. This $4N \times 4N$ Hessian is furthermore approximated by a Block diagonal matrix where only the $N \times 4 \times 4$ blocks on the diagonal of H are taken.

5.4 Application to Magnetic Resonance Fingerprinting

The statistical model and the priors presented above describe a general class of regression problems and will now be applied to simulated, realistic MRF data. A truncated noninformative as well as two informative priors are analyzed. In the case of the informative priors, a combination of the GMRF and the partition prior is chosen for the different types of parameters. The MAPs of the posterior distributions are calculated and by applying a Laplace approximation uncertainties via the Hessian can be obtained. The goal is to show the practicability of a k -space modeling statistical approach in the case of MRF as well as to demonstrate the impact of using informative prior knowledge.

In MRF the conditions (5.2) and (5.3) are fulfilled due to the structure and range of the parameters (cf. Appendix A). It is also true that the number of observations exceed the number of unknown parameters. In this simulation the number of observations is in the order of 1×10^6 while the number of parameters is in the order of 1×10^5 .

5.4.1 Physical Model

In Figure 5.3 four parameters of a simulated, realistic slice of a human brain taken from BrainWeb (2018) are presented. The tissue-related parameters are the T_1 and T_2 relaxation times as well as the proton density ρ . A fourth parameter has additionally to be taken into account due to the choice of the acquisition sequence, namely the smooth off-resonance frequency δf . This is a parameter depending on the MR device. A slice is modeled such that it consists out of N different voxels. Each voxel is attached with a set of ρ , T_1 , T_2 and δf . In this case study N will be equal to $256 \times 256 \approx 6.5 \times 10^4$ and the total number of unknown parameters is therefore approximately 2.6×10^5 . Altogether, these four parameter maps are taken to be the ground truth in the simulations.

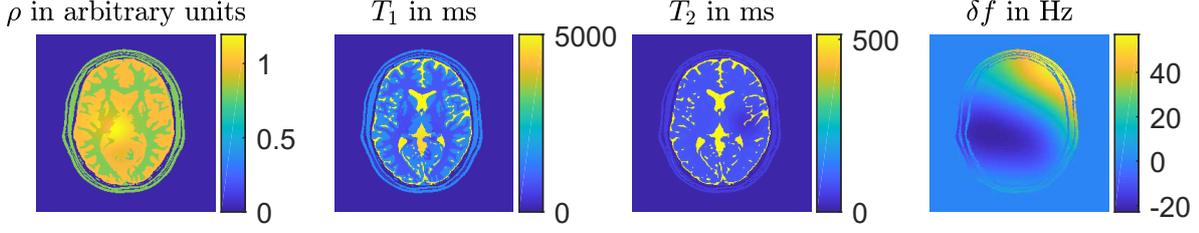


Figure 5.3: Values of tissue related parameters from BrainWeb (2018) in a slice of a human brain, together with values for δf , used for the simulation setting. Figure from Metzner et al. (2019).

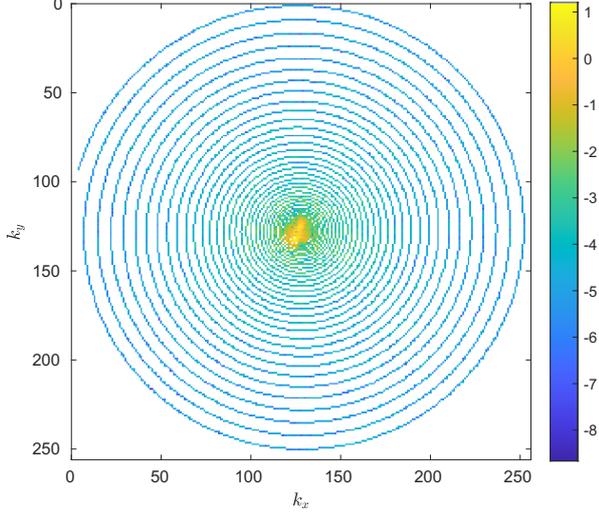


Figure 5.4: Undersampled data in Fourier domain on a logarithmic scale. Figure from Metzner et al. (2019).

Let the $N \times 1$ vector m_l denote the complex-valued magnetization in the N voxels after applying the l th high frequency pulse for $l = 1, \dots, L$. For this study a so-called balanced steady-state free precession (bSSFP) sequence is assumed for the scanning process. The magnetization can be described via a three-dimensional linear discrete time system relying on the Bloch equations (Nishimura (1996)). For more details regarding the exact calculation of the magnetization m_l we refer to Appendix A. Let y_l denote the $n_l \times 1$ complex-valued vector of measurements in k -space. This data can then be modeled according to

$$y_l = \mathcal{F}_l m_l(\theta), \quad l = 1, \dots, L, \quad (5.19)$$

where θ represents the $4N$ parameters ρ , T_1 , T_2 and δf and \mathcal{F}_l stands for a nonuniform fast Fourier operator (Greengard and Lee (2004)). The Fourier operator samples nonuniformly along a trajectory identified via the k -space data points. Here, a spiral readout is taken and Figure 5.4 is showing an example of the absolute values of the Fourier data across the trajectory on a logarithmic scale for a specific readout.

In the following, θ_i , $i = 1, \dots, 4$ will stand for the parameters ρ , T_1 , T_2 and δf respectively, and $\theta = (\theta_1^T, \theta_2^T, \theta_3^T, \theta_4^T)^T$. In the case of MRF, ρ is a linear factor of the magnetization model and therefore plays the role of θ_1 in (5.1) while the rest of the parameters enter like θ_2 in (5.1).

Furthermore, the function χ^2 in (5.5) will be replaced by

$$\chi^2(\theta) = \sum_{l=1}^L \left\{ \|\operatorname{Re}(y_l - \mathcal{F}_l m_l(\theta))\|^2 + \|\operatorname{Im}(y_l - \mathcal{F}_l m_l(\theta))\|^2 \right\}. \quad (5.20)$$

5.4.2 Simulation Setting

The MRF data were simulated using the parameters of the slice of the human brain taken from BrainWeb (2018) (Figure 5.3). The values from BrainWeb (2018) were constant for each tissue type. In order to be more realistic the values were changed by adding smooth Gaussian functions. The off-resonance frequency δf was simulated as a superposition of randomly chosen 2D Gaussian peaks, i.e. δf varies spatially in a smooth way. A variable density spiral with an undersampling factor of 8 is taken and the repetition times were chosen to be 60 ms. The flip angle follows the magnitude of a sinusoidal curve with a maximum of 60° and a period of 50 acquisitions. A total of $L = 50$ scans were obtained. Using a bSSFP sequence, the physical model can be described via a three-dimensional linear discrete time system that relies on the Bloch equations. The complex-valued magnetization $m_l(\theta)$ of these parameters can therefore be calculated for $l = 1, \dots, 50$. These magnetizations were then transformed into the Fourier domain by applying a nonuniform fast Fourier transformation (Greengard and Lee (2004)). Gaussian noise is added onto the complex and imaginary part of the Fourier data following the statistical model with variance 10^{-6} . This scheme is closely related to the one used by Wübbeler and Elster (2017).

The trust-region algorithm was used to calculate the MAP of the marginal posteriors. Ten starting values (the same values were taken for all calculations) for the optimization were used and calculated as follows. The true values of the tissue parameters were distorted by a multivariate random deviate with zero mean and covariance three times the (approximate) inverse negative Hessian of the log posterior of the noninformative prior at the ground truth. The results of the optimizations (RMSE, coverages, interval length of coverages) for the different starting values are robust, i.e., they only have a maximum deviation of 3%.

5.4.3 Employed Priors and Marginal Posteriors

Two priors $\pi(\theta, \sigma) = \pi(\theta)\pi(\sigma)$ with $\pi(\sigma) = 1/\sigma$ and $\pi(\theta) = \pi(\theta_1)\pi(\theta_2)\pi(\theta_3)\pi(\theta_4)$ were considered.

First, a truncated noninformative prior is used. For each θ_i , $i = 1, \dots, 4$ the noninformative prior $\theta_i \propto 1$ is taken, augmented by reasonable, large bounds for the domain $\Theta = [\underline{\theta}, \bar{\theta}]$ of the parameters. Here, $\underline{\theta}, \bar{\theta}$ denote the bounds of the domain. The resulting marginal posterior (σ is integrated out) is then given as

$$\pi(\theta|y) \propto \frac{1}{(\chi^2(\theta))^{n/2}} 1_{\Theta}(\theta). \quad (5.21)$$

The number n stands for the number of all observations. Note, that one single complex-valued data point in the Fourier domain counts as two observations.

Second, a combination of a partition and a GMRF prior is used. For the tissue-related parameters θ_i , $i = 1, \dots, 3$ (i.e. ρ , T_1 and T_2) a partition prior is taken. For this prior a

segmentation of the slice of the human brain is assumed to be available beforehand. The mean values in each partition and for every parameter are taken as the true mean values. These $q_l \times 1$ mean vectors are denoted by $\mu_l, l = 1, \dots, 3$. The variance parameters of the individuals partitions are modeled with inverse gamma distributions depending on the shape and scale parameters α_{li} and β_{li} for $i = 1, \dots, q_l$. That means $\lambda_{li} | \alpha_{li}, \beta_{li} \sim IG(\alpha_{li}, \beta_{li})$. For the device parameter $\theta_4 = \delta f$ a GMRF prior is taken together with a normal distribution for the mean of the δf_i in order to ensure propriety. The parameters of the normal distributions are μ_4 and σ_4 while for the variance parameter of the GMRF prior an inverse gamma distribution is taken with $\lambda_4 | \alpha_4, \beta_4 \sim IG(\alpha_4, \beta_4)$. The marginal posterior (σ and $\lambda_i, i = 1, \dots, 4$ are integrated out) is given as (Metzner et al. (2019))

$$\begin{aligned} & \pi(\theta | y, \mu, \alpha, \beta, \alpha_4, \beta_4, \mu_4, \sigma_4) \\ & \propto \frac{1}{(\chi^2(\theta))^{n/2}} \frac{1}{\prod_{l=1}^3 \prod_{i=1}^{q_l} (2\beta_{li} + (\theta_{li} - \mu_{li}1)^T (\theta_{li} - \mu_{li}1))^{\alpha_{li} + q_{li}/2}} \\ & \times \frac{1}{(\theta_4^T K \theta_4 + 2\beta_4)^{N/2 + \alpha_4 - 1/2}} \exp\left(-\frac{1}{2\sigma_4^2} \left(\frac{1}{N} \sum_i (\theta_4)_i - \mu_4\right)^2\right) \frac{1}{\sigma_4}. \end{aligned} \quad (5.22)$$

A second order neighborhood is assumed for the structure matrix K , i.e. one voxel has at least 3 and at most 8 neighbors. Two different informative priors were considered in order to observe the influence of variance parameters of the priors. Therefore the shape parameters of all inverse gamma distributions were either fixed to be 100 (later called informative informative-II) or 10^{-3} (later called informative informative-I). The scale factors were calculated in dependence on these shape factors as follows: for the partition prior the scale factors of all partitions for one parameter were computed in the way that that the mode of the inverse gamma distributions (the mode depends on scale and shape factor) is equal to the maximum variance of all partitions of that parameter. For ρ the maximum standard deviation of all segments equals 0.04, for T_1 145.60 *ms* and for T_2 9.90 *ms*. In a similar way it is calculated for the GMRF prior. In that case, the scale factor is chosen such that the mode of the inverse gamma distribution is equal to the variance of δf , which equals 0.37^2 Hz^2 .

The posterior distributions (5.21) and (5.22) are proper and their first and second moments exist according to Theorem 5.3.2 and 5.3.6.

5.4.4 Results and Discussion

A slice of the human brain is modeled to have 256×256 voxels. In order to decrease the number of unknowns, only voxels containing tissue are considered. This is reasonable since only those voxels contribute to the measured MR signal. The information for creating a suitable mask could be gained by calculating the sum of all individual magnetization images or by a fast MRI pre-scan which can then also be used for the segmentation information. In this simulation, the total number of unknowns is about $4 \times 2.5 \times 10^4$.

In Figure 5.5 the MAP for the unknown parameters and their deviations from the ground truth are shown when using the truncated noninformative prior. The same is done in Figure 5.6 for the results of the optimization with the informative prior (with shape factor 100). Additionally the width of the 95% credible intervals are shown in a third row. They are

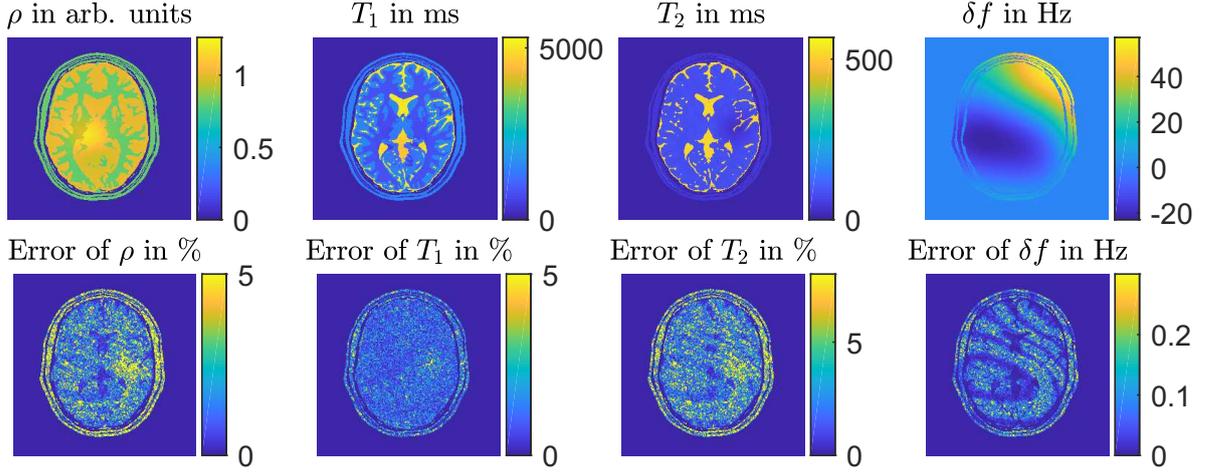


Figure 5.5: Reconstructed images and reconstruction errors for the noninformative prior augmented by large bounds. Figure from Metzner et al. (2019).

Table 5.1: Proportions over roughly 2.5×10^4 voxels of 95% credible intervals covering the true values employed for simulating the data. Noninformative relates to the posterior obtained for the noninformative prior augmented by large bounds, informative-I to that for the informative prior with a shape of factor 10^{-3} and informative-II to that for the informative prior with a shape factor of 100. Table from Metzner et al. (2019).

	ρ	T_1	T_2	δf
noninformative	80.69%	78.68%	81.86%	81.53%
informative-I	96.73%	91.42%	85.89%	93.91%
informative-II	97.83%	92.58%	97.94%	94.58%

calculated as 2 times 1.96 times the standard deviation of the parameter (calculated via the approximate Hessian presented in (5.3.2)). The width of the credible intervals summarize the posterior uncertainty about the estimation and appear to characterize the observed errors well.

The root-mean-square error (RMSE) of every of the four parameter types is displayed in Figure 5.7 for the results of the different posterior distributions. Note that the influence of each error on the RMSE is proportional to the size of the squared error. Larger errors therefore have a disproportionately large impact on the RMSE.

In Table 5.1 the percentage of voxels for which the 95% credible intervals cover the true values (the simulated ground truth) are shown.

The average width in percentage of the 95% credible intervals are summarized in Table 5.2.

Table 5.2: Average width over roughly 2.5×10^4 voxels of relative 95% credible intervals, i.e. 95% credible intervals divided by the true value of the tissue parameters, for ρ, T_1, T_2 as well as average width of 95% credible intervals for δf . Noninformative relates to the posterior obtained for the noninformative prior augmented by large bounds, informative-I to that for the informative prior with a shape factor of 10^{-3} and informative-II to that for the informative prior with a shape factor of 100. Table from Metzner et al. (2019).

	ρ	T_1	T_2	δf in Hz
noninformative	8.29%	3.76%	11.82%	0.24
informative-I	4.18%	2.56%	6.68%	0.19
informative-II	4.85%	3.34%	7.58%	0.20

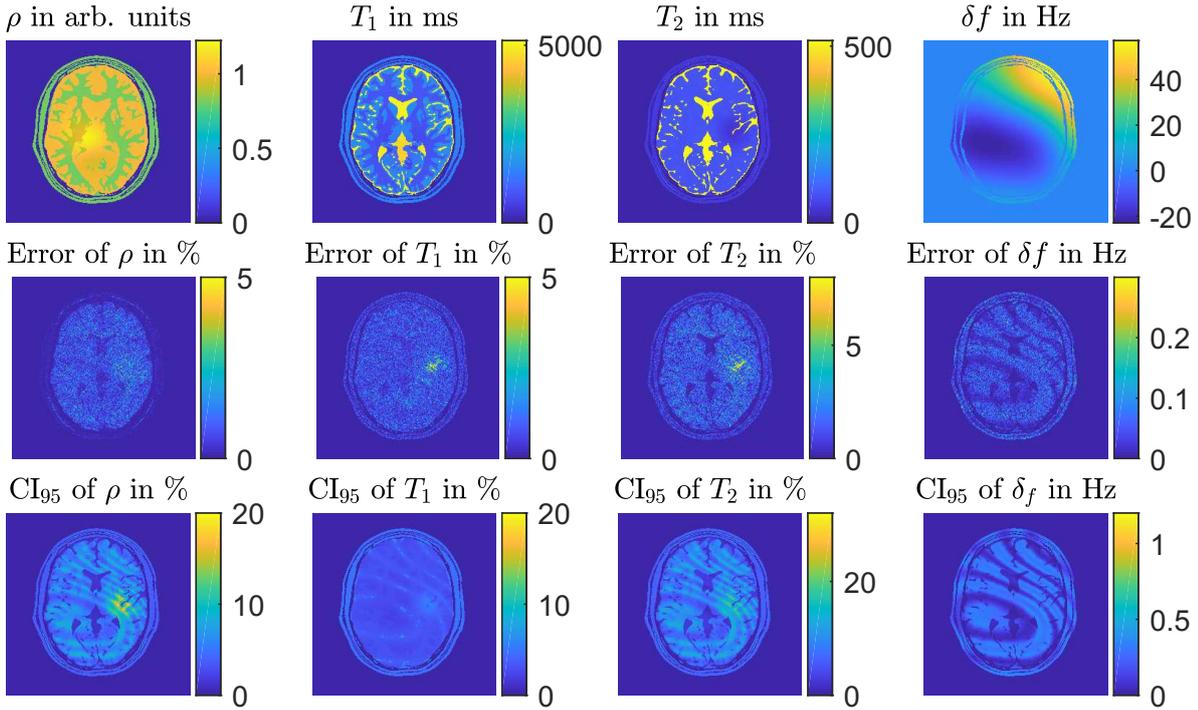


Figure 5.6: First two rows: reconstructed images and reconstruction errors for the informative prior with shape factor 100. Last row: width of 95% credible intervals denoted by CI_{95} . Figure from Metzner et al. (2019).

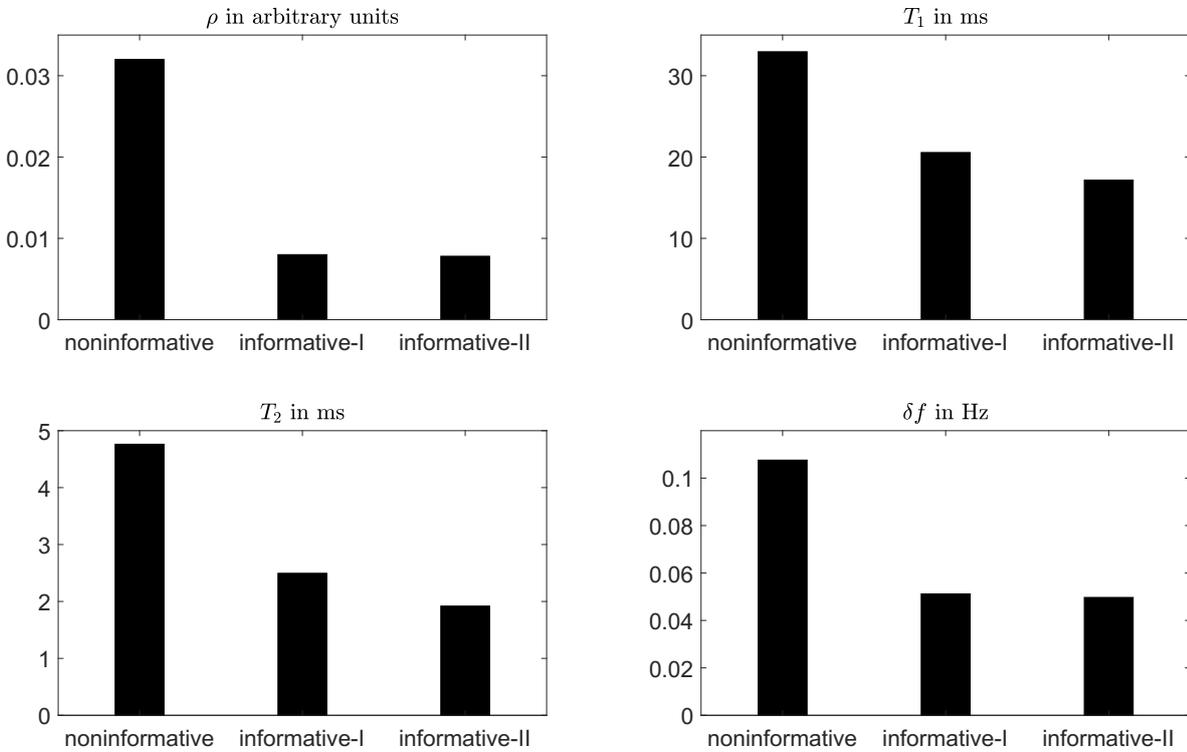


Figure 5.7: Root-mean-square errors of the reconstructed images for the different choices of priors. Noninformative relates to the posterior obtained for the noninformative prior augmented by large bounds, informative-I to that for the informative prior with a shape factor of 10^{-3} and informative-II to that for the informative prior with a shape factor of 100. Figure from Metzner et al. (2019).

The obtained images for the parameters ρ , T_1 , T_2 and δf in Figure 5.5 and 5.6 are the MAP estimates of the respective posterior distributions. The quality is reasonable and it can be seen that incorporating prior knowledge leads to a significant reduction of the observed errors. This is also shown in the root-mean-square errors. Additional information about different partitions (for ρ , T_1 and T_2) and smoothness (for ρ) can lead to better results in terms of deviation from the ground truth and overall root-mean-square errors. The results appear to be quite robust regarding the choice of the shape parameter for the inverse gamma distribution. The overall root-mean-square errors do not change significantly when using either a shape factor of 100 or 10^{-3} . The number of voxels where the 95% credible interval cover the true value (cf. Table 5.1) can be seen as encouraging, especially since about 10^5 variables have been inferred simultaneously in a nonlinear problem. The proportion of these voxels increases, the average width of the 95% credible intervals decreases, when using the informative prior.

The dependence of results for the truncated noninformative prior on the selected additional bounds on the parameters was not investigated since the nonexistence of the posterior for a purely noninformative prior remains hidden when using a Laplace approximation.

We examined the possible reduction in acquisition time which can be achieved when incorporating prior knowledge. It turned out that the root-mean-square errors of the results for the truncated noninformative prior were approximately equal to the results for the informative prior (with shape factor 100) when only the first 20% of the data were used.

We also computed results for a significantly larger noise level σ^2 of the Fourier data. Not only deteriorates the quality of the MAP estimates, but also the number of voxels for which the credible intervals cover the true values decreases. This effect is probably caused by the nonlinearity of the physical model for MRF and by the limitations of a Laplace approximation. Such an approximation describes the behavior of the posterior well in a vicinity around the MAP, but not necessarily in regions far apart.

5.5 Summary

We developed a general Bayesian framework based on a homoscedastic Gaussian sampling model for a class of nonlinear, large-scale regression problems in which the parameters model the spatial distribution of some property, and where the regression function satisfies certain assumptions. MRF is one example for such a class but there could also be other possible applications. Basic advantage of a Bayesian inference is that prior knowledge can be included into the analysis. Furthermore, a Bayesian analysis lends itself in a natural way to an uncertainty characterization, and probability statements can be made conditional on the data.

The assignment of a prior is a crucial part of any Bayesian inference and different choices have been explored. We have shown that when a constant prior on the physical parameters is assigned, together with the standard noninformative prior for the variance, then the posterior does not exist. Another result of this paper is that Gaussian Markov random field priors, augmented by a proper prior on the mean of the parameter, do yield a proper posterior. These priors model spatial smoothness which can be assumed for the physical parameters. The possibility to include further prior knowledge about the physical parameters gained from their segmentation has been expressed in terms of a partition prior. The resulting posterior of the

partition prior is proper and the existence of its moments has been explored. Furthermore, a close relationship between a partition and intrinsic GMRF priors has been shown which may provide guidance in the choice of a prior for Bayesian spatial modeling in regression problems of the considered type.

MRF is one example of the presented class of regression models. An approximate Bayesian inference based on a Laplace approximation at the MAP estimate has been presented for simulated, realistic MRF data. The results demonstrate that a Bayesian inference of high dimensional MRF data is practically feasible despite of the huge dimensionality. Furthermore, significant benefit is gained when prior knowledge is incorporated. Although the presented Bayesian inference is approximate only, the results are quite encouraging. The quality of the reconstructed images is high, and for the credible intervals reasonable coverage proportions were observed. Limitations and practical aspects of the Bayesian k -space modeling for MRF data are discussed in Chapter 6.

6

Comparison of Dictionary Matching and Bayesian Modeling

The MRF Bayesian k -space modeling approach presented in Chapter 5 directly models the undersampled data in the Fourier domain by a physical model. It therefore avoids applying a pseudo-inverse Fourier transformation which leads to aliasing artifacts in the magnetization space. The dictionary matching method on the other hand performs a template matching process of the approximately reconstructed magnetization images. Although the k -space modeling captures the data without any loss (in contrast to the dictionary matching approach), the model in the Fourier domain is dependent on all the parameters in parallel. As a consequence, determination of the maximum a posteriori (MAP) estimate leads to a high-dimensional and non-convex optimization task which is time consuming and challenging to solve. The dictionary matching method only depends on the desired parameters of one voxel and the matching of the voxels can thus be performed independently for each voxel. This is relatively easy and fast to compute. From a practical point of view it is important to know whether it is worth to optimize the complicated Bayesian k -space model or if it is sufficient to compute the dictionary matching solution. Additionally, analyzing the sensitivity to model errors could give valuable insights about the robustness of the two approaches.

In this chapter our contribution is a *comparison of the dictionary matching and the Bayesian k -space modeling approach* in terms of the accuracy of the estimates dependent on the level of noise, the MRF sampling scheme and computational complexity as well as in terms of their sensitivity to errors in the physical model.

The whole chapter is based on simulations so that we can compare the results with their ground truth.

The first aspect deals with the impact of the level of noise in the data. A comparison of the dictionary matching solution and the MAP estimate of the Bayesian k -space modeling approach for the two MR sequences taken in Chapter 4 and 5 is presented. The signal-to-noise ratio will play an important role, whether the MAP estimation of the Bayesian k -space modeling approach leads to a significant improvement.

The second aspect discusses the impact of the MRF sampling scheme. Three different sampling schemes, each with approximately the same amount of data per image, will be compared in terms of the solutions obtained by the dictionary matching and the Bayesian k -space modeling approach.

The third aspect is concerned with the impact of errors in the physical model. An error in the flip angle pattern is simulated and the qualities of the results from the dictionary matching and the Bayesian k -space modeling approach are assessed. At last a comparison of computational aspects of both approaches is presented.

6.1 Impact of Noise

The signal-to-noise ratio (SNR) plays an important role in MRI (Macovski (1996); Redpath (1998)) as a fundamental measure of image quality. We will define the SNR as the ratio of the signal power to the noise power of the k -space data. There are many sources that affect the SNR. These sources can be divided into two basic classes (Nishimura (1996)): the physical and instrumental parameters such as the B_0 field or the receiver coil geometry and the imaging sequence parameters, such as the total scan time or the spatial resolution. The quality of the results from acquired k -space data heavily depend on the level of noise in the data. In general, a high SNR is desired which means that the noise should be as low as possible. In the following, we will discuss the impact of the SNR to the results from the dictionary matching and from the Bayesian k -space modeling.

6.1.1 Data

Three different data sets will be considered. The first data set was created according to Chapter 5. A slice of the human brain was taken from BrainWeb (2018) as the ground truth for the parameter maps and the parameter values were changed by adding a smooth Gaussian function in order to be more realistic. A variable density spiral with an undersampling factor of 8 was simulated and the repetition times were chosen to be 60 *ms*. The flip angle follows the magnitude of a sinusoidal curve with a maximum of 60 degree and a period of 50 acquisitions. A total of $L = 50$ scans were obtained. A bSSFP sequence was used and the complex-valued magnetization $m_l(\theta)$ of these parameters can therefore be calculated for $l = 1, \dots, 50$ (cf. Appendix A). Note that throughout this thesis the reconstructed magnetization was not normalized. These magnetizations were then transformed into the Fourier domain by applying a nonuniform fast Fourier transformation (Greengard and Lee (2004)). Gaussian noise was added to the complex and imaginary part of the Fourier data with a standard deviation of 10^{-3} (the maximum value of the real and the imaginary part of the simulated Fourier data were roughly 5). This scheme is closely related to the one used by Wübbeler and Elster (2017). We will refer to this as the bSSFP data set.

A second data set was created in the same way, the only difference is the level of noise added to the Fourier data. For this data set, Gaussian noise was added to the complex and imaginary part of the Fourier data with a standard deviation of 1.97×10^{-2} . The noise is therefore higher than before, which means that the SNR is lower. We will refer to this as the bSSFP data set with lower SNR.

The third data set was generated according to Chapter 4. A phantom with nine tubes was created as the ground truth for the parameter maps (cf. Chapter 2). Each tube had its individual T_1 and T_2 value and the values across the tubes were constant. Repetition times were chosen to be 12.54 ms , the inversion time as 42 ms , echo times as 7 ms , and the flip angle follows the magnitude of a sinusoidal curve with maximum 70 degree. A total number of 1000 radial lines were acquired and for the parameter estimation one image per radial line was reconstructed using the nonuniform fast Fourier transformation. The four complex coil sensitivity maps were chosen accordingly to the ones acquired from the phantom scan in Chapter 4. A FISP sequence was used and the complex-valued magnetization $m_l(\theta)$ of these parameters could therefore be calculated for $l = 1, \dots, 1000$ (cf. Appendix B). These magnetizations were then transformed into the Fourier domain by applying a nonuniform fast Fourier transformation. Gaussian noise with signal-to-noise ratio similar to the real scanner data from Chapter 4 was added to the complex and imaginary part of the Fourier data with a standard deviation of 3.43×10^{-6} (the maximum value of the real and the imaginary part of the simulated Fourier data were of the order 10^{-3}). We will refer to this as the FISP data set.

Note that the FISP data set and the bSSFP data set with lower SNR have roughly the same SNR whereas the bSSFP data set has a higher SNR (actually, the bSSFP data set with lower SNR was created in such a way that the SNR matches the FISP data). Note further that it is generally realistic for the FISP sequence to have a higher SNR (Jiang et al. (2015)). We created the bSSFP data set with lower SNR in order to examine the influence of the SNR.

In general, there are several ways to define the SNR in the field of MR. In this chapter, the SNR is determined as the signal-to-noise ratio in k -space. For qualitative imaging, the SNR is typically defined in the magnetization space as the signal-to-noise ratio of the reconstructed magnetizations of a specific tissue. If the k -space data were fully sampled on a Cartesian grid, then the noise in the magnetization space could be directly computed from the noise in the k -space. Since we used undersampling patterns, this is not applicable here. As an approximation we created a setting in the following way so that similar techniques were applied as in standard qualitative imaging. The same sequence parameters and the same ground truth were used to calculate the magnetization courses for every voxel. These courses were averaged over the temporal domain such that only one magnetization image is obtained as a result. For the radial sampling used for the FISP data set, the Nyquist sampling rate is almost fulfilled if $\sqrt{N}\pi/2 = 384\pi/2$ radial lines were acquired (Feng et al. (2014)). We therefore simulated 604 radial lines and calculated the k -space data using the inverse nonuniform fast Fourier transformation. Subsequently, the respective noise was added to the k -space data 100 times leading to 100 different k -space data sets. From that, 100 magnetization images could be reconstructed using the nonuniform fast Fourier transformation. The 2D SNR map could then be calculated as the ratio of the mean and the standard deviation of the absolute magnetization values of each voxel. For the FISP data set an average SNR of 32 was gained in the tube corresponding to the values $T_1 = 562\text{ ms}$ and $T_2 = 33\text{ ms}$. In a region of white matter an average SNR of 135 was obtained for the bSSFP data set respectively 8 for the bSSFP data set with lower SNR.

The dictionaries used for calculating the dictionary matching solutions from the bSSFP data sets were calculated according to the physical model given in Appendix A for $T_1 =$

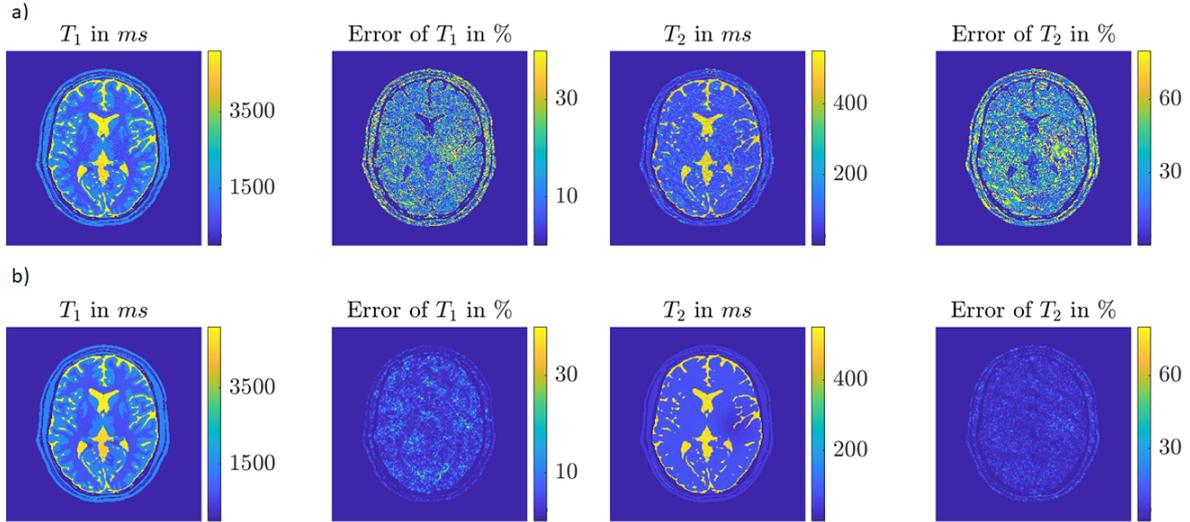


Figure 6.1: Results for the bSSFP data set. a) Dictionary matching solution with error maps for T_1 and T_2 . b) MAP estimate of the Bayesian k -space modeling approach with error maps for T_1 and T_2 .

(100, 120, \dots , 5020) ms , $T_2 = (50, 60, \dots, 520) ms$ and $\delta f = (-100, -98, \dots, 100) Hz$. The dictionary for the FISP data set was calculated for $T_1 = (200, 205, \dots, 3000) ms$ and $T_2 = (20, 22, \dots, 300) ms$ with the physical model given in Appendix B. For all three data sets, a density compensation function was calculated by using the Voronoi method (Rasche et al. (1999)).

The optimization of the Bayesian k -space modeling approach was carried out similarly as in Chapter 5 for the noninformative prior. The MAP estimate of the posterior distribution (cf. Chapter 5) was calculated by employing a trust-region method which makes use of an approximate Hessian. A total number of 1000 iterations were computed and the starting values were chosen as the solution obtained by the dictionary matching method.

6.1.2 Results

Results from the dictionary matching and the Bayesian k -space modeling approach for the different data sets are presented subsequently. In Figure 6.1 the parameter maps together with the error maps are shown for the bSSFP data set. The error is significantly smaller for the MAP estimate of the Bayesian k -space modeling approach than for the dictionary matching solution. This is also reflected by the overall root-mean-square error (RMSE). The RMSE for the dictionary matching solution is 379.8 ms for T_1 and 68.1 ms for T_2 , for the MAP estimate of the Bayesian k -space modeling approach it is 119.9 ms and 12.7 ms respectively.

In Figure 6.2 results for the FISP data set are shown. The error is similar for both solutions. The RMSE is 148.5 ms for T_1 and 45.8 ms for T_2 for the dictionary matching solution, whereas it is 117.8 ms for T_1 and 43.5 ms for T_2 for the MAP estimate of the Bayesian k -space modeling approach, respectively.

Exemplary and typical residuals of the corresponding k -space data for the dictionary matching solutions as well as for the MAP estimates of the Bayesian k -space modeling approach are shown in Figure 6.3. Let y denote the simulated data and θ denote here either

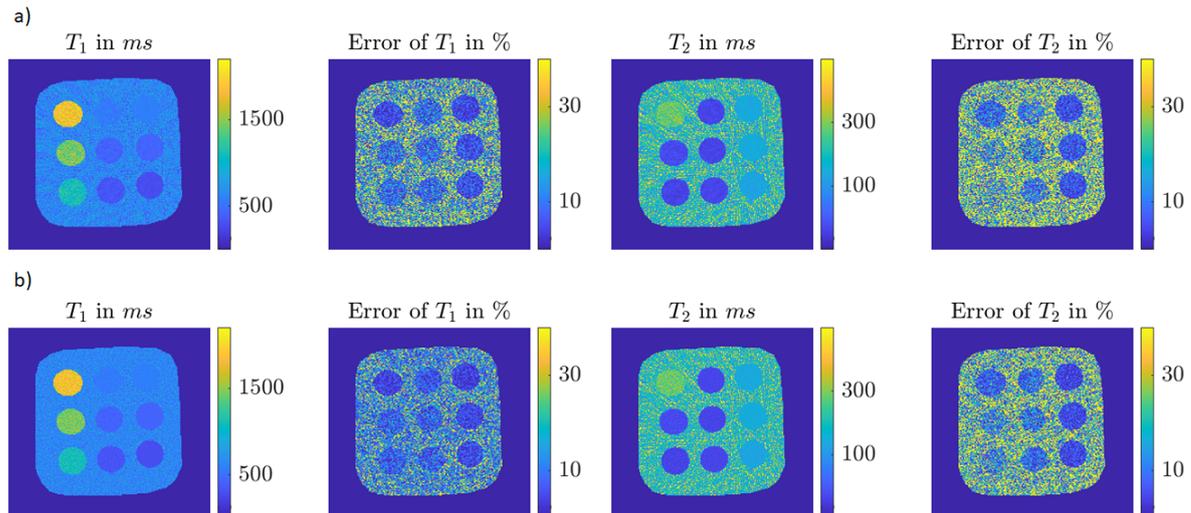


Figure 6.2: Results for the FISP data set. a) Dictionary matching solution and corresponding error maps for T_1 and T_2 . b) MAP estimate of the Bayesian k -space modeling approach and corresponding error maps for T_1 and T_2 .

the dictionary matching solution or the MAP estimate of the Bayesian k -space modeling approach. The magnetization can then be calculated according to the physical model $m_l(\theta)$ for $l = 1, \dots, L$, where L denotes the total number of images (cf. Appendix A for the bSSFP sequences and Appendix B for the FISP sequence). The fitted data $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_L)$ for the dictionary matching solution $\tilde{\theta}$ in k -space can be calculated according to:

$$\tilde{y}_l = \mathcal{F}_l(m_l(\tilde{\theta})), \quad l = 1, \dots, L. \quad (6.1)$$

Note that the coil sensitivity maps for the FISP data set have to be taken additionally into account as explained in Chapter 4. The residuals in Figure 6.3 for the bSSFP data set decrease by roughly a factor of 10 when comparing the dictionary matching solution and the MAP estimate of the Bayesian k -space modeling approach. For the FISP data set the residuals have a similar range. They differ a little around the k -space center (around the 192th k -space point) where the dictionary matching solution has a peak in contrast to the MAP estimate of the Bayesian k -space modeling approach. A homoscedastic Gaussian sampling model is assumed for the Bayesian k -space modeling approach. The residuals in Figure 6.3 for the MAP estimate of the Bayesian k -space modeling approach appear to be homoscedastic and they approximately follow a Gaussian distribution. The size of these residuals fits to the simulated Gaussian standard deviation in k -space. This is not the case for the residuals of the dictionary matching solution.

Results in terms of parameter and error maps for the bSSFP data set with lower SNR are shown in Figure 6.4. The range of the errors is similar for both solutions. The RMSE is 445.5 ms for T_1 and 88.3 ms for T_2 for the dictionary matching solution and 571.3 ms and 86.5 ms respectively for the MAP estimate of the Bayesian k -space modeling approach.

In Figure 6.5 exemplary and typical residuals for the bSSFP data set with lower SNR are shown for the two solutions. The size of these residuals is similar but the residuals are smaller around the k -space center, i.e. the first k -space points, for the MAP estimate of the Bayesian

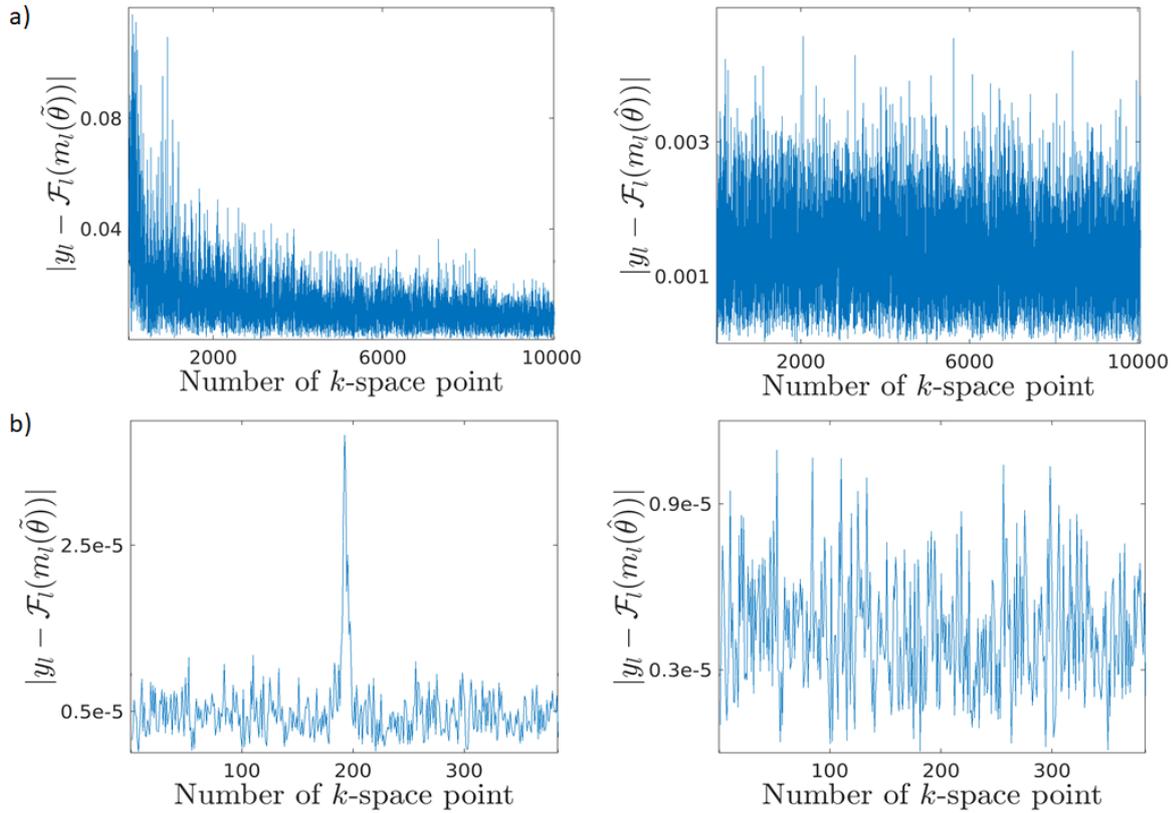


Figure 6.3: Example of residuals in k -space for one trajectory (of image l) in k -space. The absolute difference of the simulated data y_l and the fitted data of either the dictionary matching solution $\hat{\theta}$ or the MAP estimate of the Bayesian k -space modeling approach $\hat{\theta}$ is plotted. a) Results for the bSSFP data set. Left: Residuals for the dictionary matching solution $\hat{\theta}$ in k -space. Right: Residuals for the MAP estimate of the Bayesian k -space modeling approach $\hat{\theta}$ in k -space. b) The same as in a) but for the FISP data set.

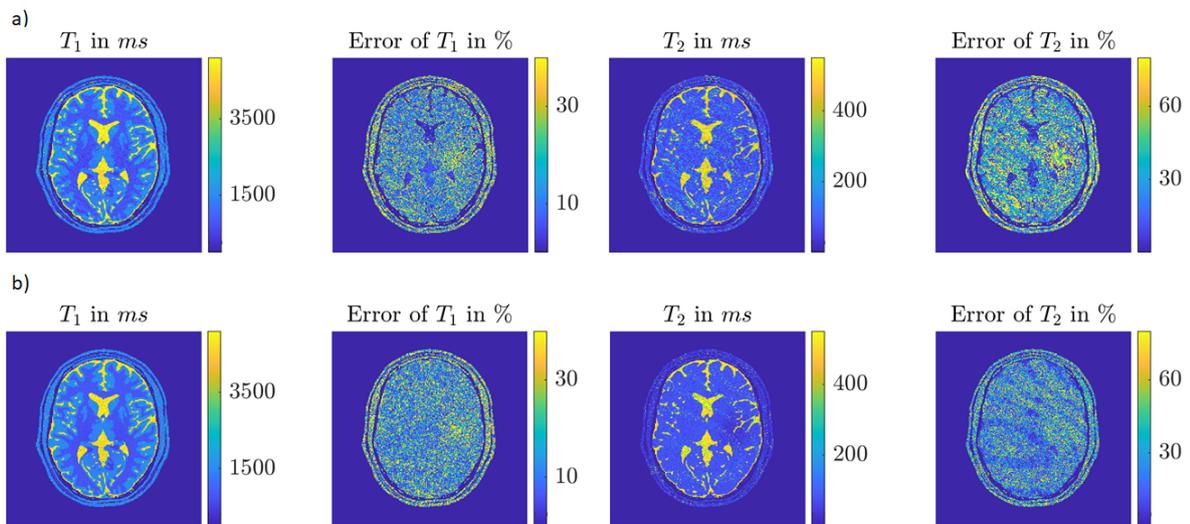


Figure 6.4: Results for the bSSFP data set with lower SNR. a) Dictionary matching solution with error maps for T_1 and T_2 . b) MAP estimate of the Bayesian k -space modeling approach with error maps for T_1 and T_2 .

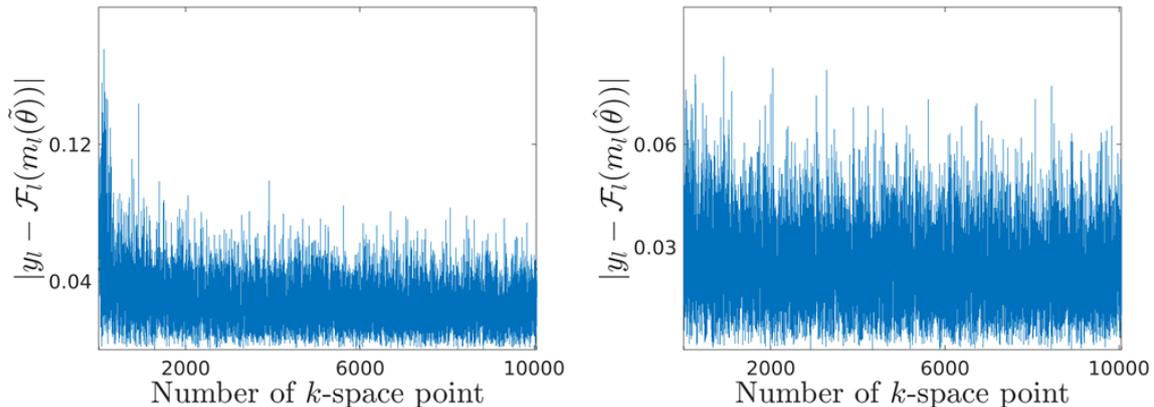


Figure 6.5: Results as in Figure 6.3 for the bSSFP data set with lower SNR. Left: Residuals for the dictionary matching solution $\tilde{\theta}$ in k -space. Right: Residuals for the MAP estimate of the Bayesian k -space modeling approach $\hat{\theta}$ in k -space.

k -space modeling approach. The model assumption about a Gaussian noise in k -space seems to be fulfilled for the MAP estimate of the Bayesian k -space modeling approach. The residuals of the dictionary matching solution also reflect the size of the standard deviation of the simulated Gaussian noise, especially off the k -space center, in contrast to the bSSFP data set above.

6.1.3 Discussion

The results above imply that the SNR is a critical factor for practical aspects of the Bayesian k -space modeling approach. The bSSFP data set with lower SNR and the FISP data set have a similar SNR whereas the bSSFP data set has a higher SNR. The MAP estimate of the Bayesian k -space modeling approach leads to a significant improvement over the dictionary matching solution for the bSSFP data set with higher SNR. This is not the case for the FISP data set. The results for the bSSFP data set with lower SNR are similar to the ones from the FISP data set. This implies that the MAP estimation of the Bayesian k -space modeling approach only improves the dictionary matching solution significantly if the aliasing errors arising from the imperfect inverse Fourier transformation dominate the error of the dictionary matching solution, i.e. when the SNR is high enough.

The residuals support these findings. For the bSSFP data set with higher SNR the residuals could be improved by roughly a factor of 10 while the residuals for the other two data sets are of similar size. Only for points around the k -space center (for a radial acquisition these points are in the middle of the radial line and for a spiral sampling these points are at the beginning, c.f. Figure 6.6) an improvement with the Bayesian k -space modeling approach can be achieved. This supports the explanation that for low SNR the noise dominates the errors in the pseudo-inverse Fourier rather than aliasing errors.

Provided that the physical model contains no errors, the assumed sampling model is correct, and the employed prior is consistent with the ground truth, the Bayesian k -space modeling approach is expected to be superior to the dictionary matching method. However, calculating the MAP can be very challenging due to the high-dimensionality and the nonlinearity of the physical model in practice. Nevertheless, we have shown that there are relevant cases in which the MAP estimate of the Bayesian k -space modeling improves the dictionary matching solution

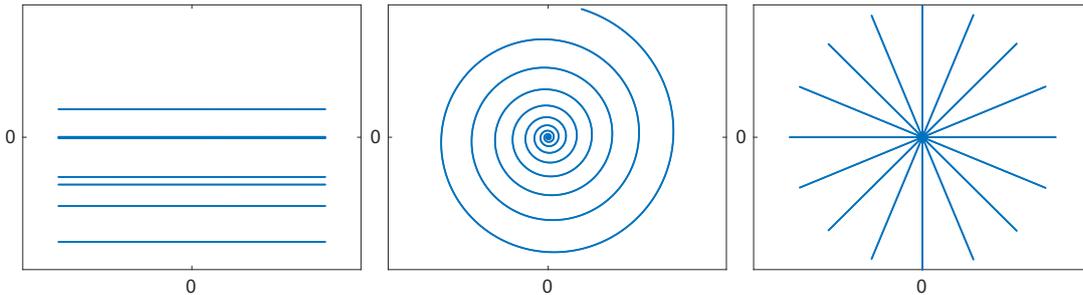


Figure 6.6: Different sampling schemes. Left: Cartesian sampling, middle: spiral sampling, right: radial sampling.

significantly. On the other hand, for the FISP data set the MAP estimate of the Bayesian k -space modeling approach does not yield significantly better results. For this scenario, the dictionary matching solution already gives a good estimate of the parameters. From a practical point of view it is therefore essential to decide if expensive MAP estimation of the Bayesian k -space modeling approach is worth to calculate or if the dictionary matching solution already yields a reasonable result. One indicator could be to look at the residuals of the dictionary matching solution in k -space. If the residuals look Gaussian and more importantly have roughly the same standard deviation than the noise in k -space then it might be unnecessary to also compute the MAP estimate of the Bayesian k -space modeling approach. If the standard deviation of the residuals is much higher than the standard deviation of the noise in k -space then the MAP estimate of the Bayesian k -space modeling approach might significantly improve the dictionary matching solution.

6.2 Impact of Sampling Scheme

After discussing the influence of the SNR on the quality of the results from the dictionary matching and the Bayesian k -space modeling approach, we will now consider the impact of different sampling schemes. In Figure 6.6 a Cartesian, a spiral and a radial trajectory for the k -space sampling are shown. The number of points for each trajectory is chosen along the lines of Doneva et al. (2017). We will compare the results of the dictionary matching and the Bayesian k -space modeling approach when using these three sampling schemes. For comparability roughly the same number of points is used. The SNR of the different data sets is chosen to match with the level of noise in real scanner data from the phantom (cf. Chapter 4).

6.2.1 Data

For the Cartesian read-out, eight Cartesian lines with 384 points each were simulated for every image. Three out of eight lines were chosen as the three lines through and around the k -space center. The other five ones were chosen randomly. Similarly, eight radial lines were simulated in every image with 384 points each. In real measurements this can be accomplished by a multi-shot MRF approach as presented by Flassbeck et al. (2019). The angle between two adjoined lines equals $360/8$ degree and the eight radial lines were altered by the golden angle between two adjoined images. A variable density spiral with an undersampling factor of 52 was simulated for the spiral readout leading to an acquisition of 3183 data points per image. For

Cartesian and radial sampling on the other hand 3072 data points per image were simulated. Figure 6.6 shows the first trajectory of each simulated sampling scheme.

The same FISP sequence with the same parameters was used for every one of the three sampling schemes. A total amount of 1000 images were simulated. Repetition times were chosen to be 12.54 *ms*, the inversion time as 42 *ms*, echo times as 7 *ms*, and the flip angle follows the magnitude of a sinusoidal curve with maximum 70 degree. This is basically the same sequence as used in Chapter 4 only with a different amount of data per image. The four complex coil sensitivity maps were chosen accordingly to the ones acquired from the phantom scan in Chapter 4.

A phantom with nine tubes was chosen as the ground truth for the T_1 and T_2 values. Each tube had its individual T_1 and T_2 value and the values across the tubes were constant (cf. Chapter 2). The magnetization was then calculated according to the physical model given in Appendix B.

The modeled magnetization was transformed into the Fourier domain using a nonuniform fast Fourier transformation (Greengard and Lee (2004)), and Gaussian noise with similar signal-to-noise ratio as the real scanner data from Chapter 4 was added with a standard deviation of 3.43×10^{-6} (the maximum value of the real and the imaginary part of the simulated Fourier data were of the order 10^{-3}). Note that the same noise was added for all three sampling schemes.

The dictionaries used for calculating the dictionary matching solutions were calculated according to the physical model given in Appendix B for $T_1 = (200, 205, \dots, 3000)$ *ms* and $T_2 = (20, 22, \dots, 300)$ *ms*. The density compensation function was calculated by using the Voronoi method (Rasche et al. (1999)) for the radial and spiral sampling.

The MAP estimation of the Bayesian k -space modeling approach was carried out similarly as in Chapter 5 for the noninformative prior. A trust-region method was employed using an approximate Hessian. A total number of 1000 iterations were computed for each optimization and the starting values were chosen as the corresponding dictionary matching solutions.

6.2.2 Results

The results from the dictionary matching approach for T_1 and T_2 are shown in Figure 6.7. The parameter maps are plotted as well as the error maps for all three sampling schemes. It is not possible to recognize the tubes in the dictionary matching solution for the Cartesian sampling whereas this is possible for the results of radial and spiral sampling. The highest errors occur for the Cartesian sampling and the lowest errors for the radial sampling.

The MAP estimates of the Bayesian k -space modeling approach for T_1 and T_2 are shown in Figure 6.8. Again, the parameter maps are plotted as well as the error maps for all three sampling schemes. All MAPs and error maps appear to be quite similar.

The regions of interest are the nine tubes. Due to the simulations, the values in each tube are constant. In Table 6.1 the RMSEs of the dictionary matching solutions of each tube for T_1 and T_2 are shown for the three sampling schemes. The highest errors occur for the Cartesian sampling and the lowest errors for the radial sampling. Note that the RMSE is a measure for the quality of the fit that is scale-dependent and sensitive to outliers.

6. Comparison of Dictionary Matching and Bayesian Modeling

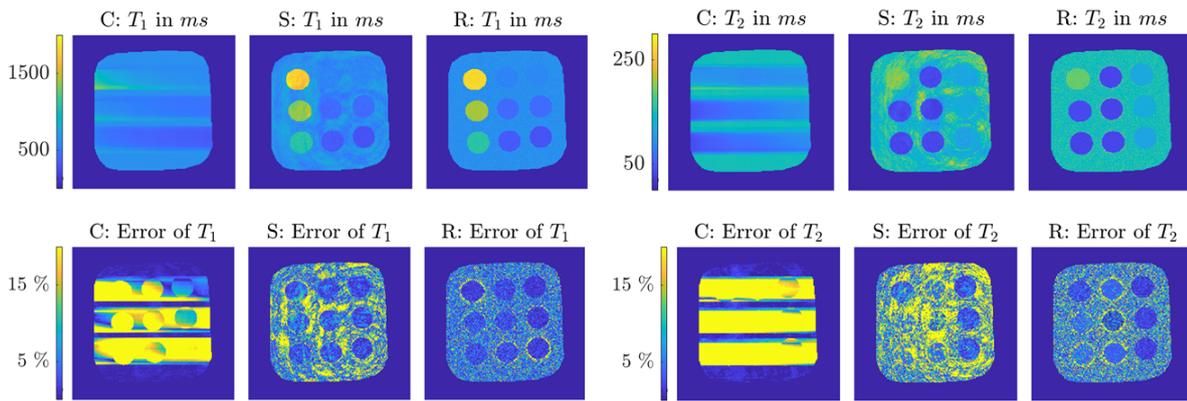


Figure 6.7: Dictionary matching solutions and error maps of T_1 and T_2 for the three different sampling schemes (C stands for Cartesian, S for spiral and R for radial readout).

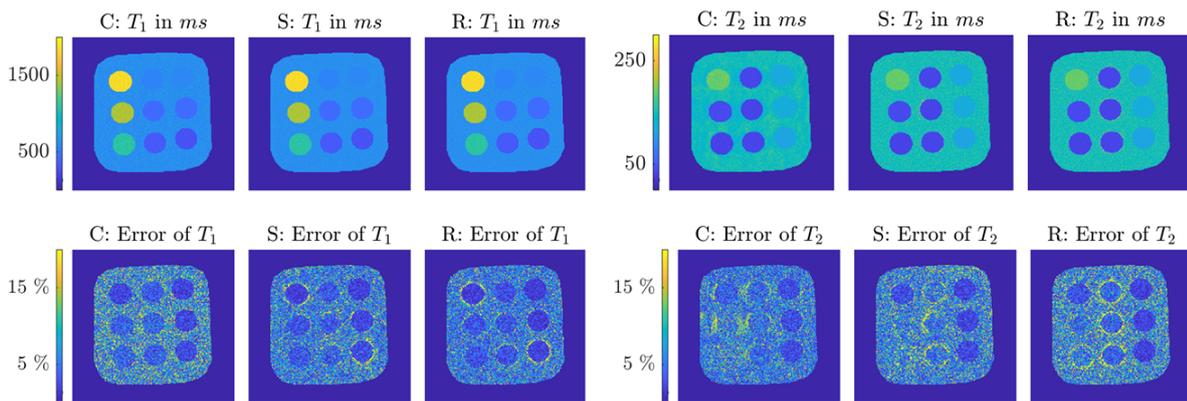


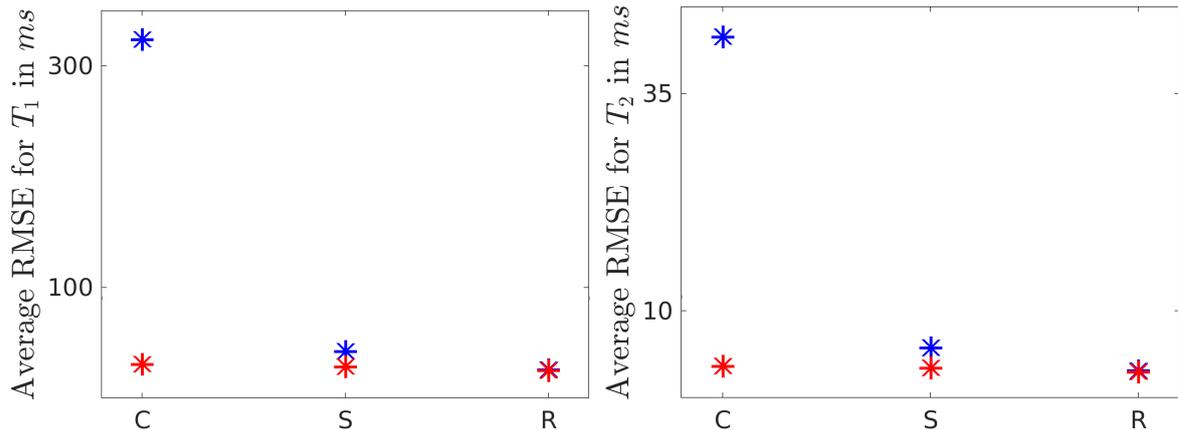
Figure 6.8: MAP estimates of the Bayesian k -space modeling approach and error maps of T_1 and T_2 for the three different sampling schemes (C stands for Cartesian, S for spiral and R for radial readout).

Table 6.1: Root-mean-square errors in ms of the dictionary matching solutions of T_1 and T_2 for the three different sampling schemes in the nine tubes t_1, \dots, t_9 .

	t1	t2	t3	t4	t5	t6	t7	t8	t9
Cart.: RMSE for T_1 in ms	989.7	130.4	41.6	815.5	113.4	55.1	590.5	100.9	78.6
Spiral: RMSE for T_1 in ms	114.8	22.9	19.1	101.8	22.3	18.5	54.1	15.1	11.0
Radial: RMSE for T_1 in ms	50.0	17.7	11.2	63.9	16.3	9.8	42.8	12.8	7.5
Cart.: RMSE for T_2 in ms	83.3	62.6	20.7	30.8	46.7	29.2	33.9	40.0	26.5
Spiral: RMSE for T_2 in ms	13.0	4.0	5.0	5.2	4.6	5.4	5.6	3.8	5.3
Radial: RMSE for T_2 in ms	7.4	2.1	3.0	3.1	2.0	3.3	2.8	1.9	2.8

Table 6.2: Root-mean-square errors in ms of the MAP estimates of the Bayesian k -space modeling approach of T_1 and T_2 for the three different sampling schemes in the 9 tubes t1,...,t9.

	t1	t2	t3	t4	t5	t6	t7	t8	t9
Cart.: RMSE for T_1 in ms	49.8	23.8	15.2	71.3	21.9	12.6	56.9	15.7	9.3
Spiral: RMSE for T_1 in ms	49.5	24.8	14.1	55.0	21.9	12.8	47.1	18.7	10.9
Radial: RMSE for T_1 in ms	46.8	19.0	12.1	56.8	1.28	10.1	42.1	13.3	7.9
Cart.: RMSE for T_2 in ms	8.8	1.9	4.1	3.1	2.1	4.3	2.9	1.8	3.7
Spiral: RMSE for T_2 in ms	7.0	2.6	3.3	3.2	2.2	3.5	3.2	2.5	3.4
Radial: RMSE for T_2 in ms	7.3	1.9	2.9	2.6	1.8	3.2	2.7	1.8	2.7

**Figure 6.9:** The average RMSE over the nine tubes of the dictionary matching solutions (blue asterisks) and the MAP estimates of the Bayesian k -space modeling approach (red asterisks) for the different sampling schemes and for T_1 and T_2 (C stands for Cartesian, S for spiral and R for radial readout).

Additionally, the RMSE is shown for the MAP estimates of the Bayesian k -space modeling approach in Table 6.2. The errors are similar for the different sampling schemes, however, the best results are achieved with the radial sampling.

A summary of the results of Table 6.1 and Table 6.2 is displayed in Figure 6.9.

Figure 6.10 shows residual plots of the k -space data of the dictionary matching solutions and the MAP estimates of the Bayesian k -space modeling approach for the different sampling schemes. All plots belonging to the dictionary matching solutions show a peak for data in the k -space center. Looking at the scale of the image, the residuals for the dictionary matching solution of the Cartesian sampling are extremely poor whereas they are acceptable for the radial sampling. All residuals corresponding to the MAP estimates of the Bayesian k -space modeling approach are reasonable in terms of their amplitude as well as they seem to look Gaussian which would satisfy the model assumptions. The standard deviation of the residuals in k -space shown in Figure 6.10 for the Cartesian sampling is 2×10^{-3} for the dictionary matching solution and 2.2×10^{-6} for the MAP estimate of the Bayesian k -space modeling approach. Recall that the standard deviation of the noise in k -space was 3.43×10^{-6} .

6.2.3 Discussion

Three different sampling schemes with roughly the same amount of data for each image are used to simulate three different MRF data sets. For each of the three data sets, the dictionary

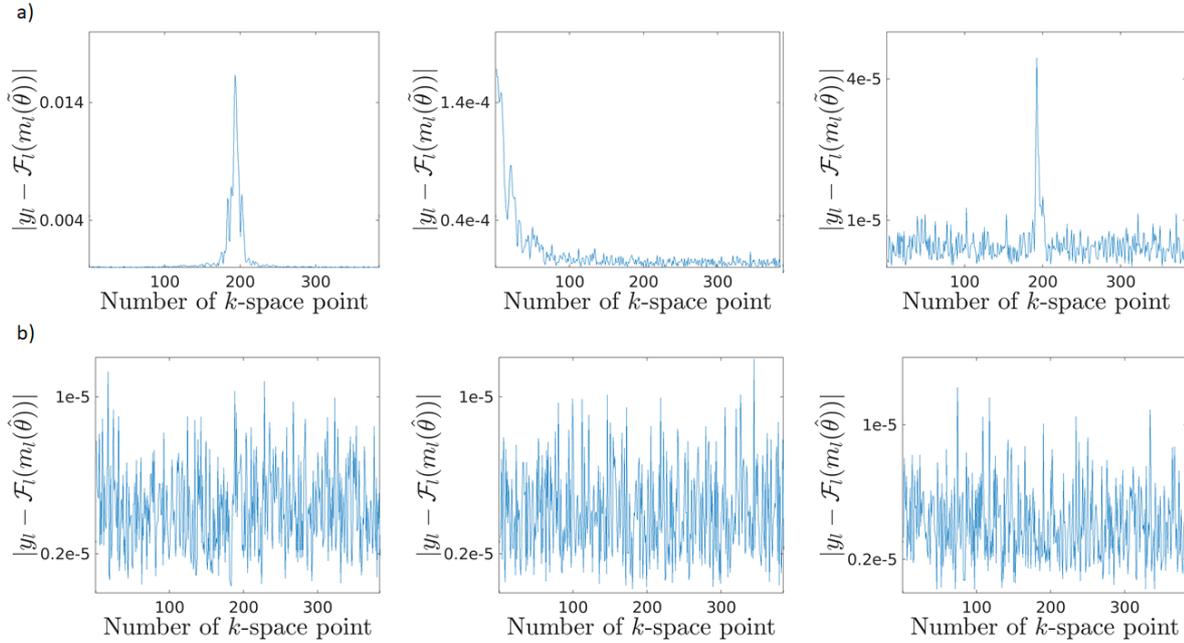


Figure 6.10: Example of residuals in k -space for one trajectory (of image l) in k -space for the different sampling schemes. a) Residuals of the dictionary matching solution $\hat{\theta}$ for Cartesian (left), spiral (middle) and radial (right) sampling. b) Residuals of the MAP estimate of the Bayesian k -space modeling approach $\hat{\theta}$ for Cartesian (left), spiral (middle) and radial (right) sampling.

matching and the MAP estimate of the Bayesian k -space modeling approach are computed. The resulting T_1 and T_2 maps are shown in Figure 6.7 and 6.8.

The dictionary matching solution for the Cartesian sampling is poor. It is not even possible to recognize the different tubes. The deviation to the ground truth is extremely large. When using the spiral sampling, the dictionary matching solution provides reasonable results. There is a pattern in the background of the tubes but the tubes themselves are estimated to a sufficient degree. The best solution is achieved with the radial sampling. The tubes can be identified and the estimated values are excellent.

When looking at the RMSE of the tubes for the different dictionary matching solutions (Table 6.1), the same results can be detected. The error for the Cartesian sampling is very high, whereas it is better for the spiral and the radial sampling. The results for the radial sampling are better by a factor of 1.2 to 2.3 for T_1 and 1.6 to 2.3 for T_2 compared to the spiral sampling solution. The median of the absolute values of the residuals which is a more robust measure delivers similar results.

The MAP estimates of the Bayesian k -space modeling approach (Figure 6.8) show similar results for all sampling schemes. The starting values of the employed optimizations were chosen to be the corresponding dictionary matching solutions. Even in the Cartesian case, where a very poor dictionary matching solution was computed, the optimization of the Bayesian k -space modeling approach was able to reach to a satisfying solution. This is an excellent result as it shows the versatility of the Bayesian k -space modeling approach. In this particular case it was possible to identify the T_1 and T_2 maps with the sophisticated model in k -space but not with the simple dictionary matching approach.

Residual plots of the dictionary matching solutions for the different sampling schemes support the statements above. The solution of the Cartesian sampling yields the worst residual

plot whereas the radial sampling yields the best. The residuals of the MAP estimates of the Bayesian k -space modeling approach yield similar results for all sampling schemes. The residuals appear to be uncorrelated and homoscedastic and they approximately follow a Gaussian distribution.

The MAP estimates of the Bayesian k -space modeling approach for the spiral and radial sampling provide good results, as it can also be seen from the RMSE (Table 6.2). The RMSE in the tubes are very similar for the different sampling schemes, but again, the radial sampling delivers predominantly the best results.

Since the computation of the MAP estimate of the Bayesian k -space modeling approach is highly challenging it would be great to have an indicator to whether to calculate it or just use the dictionary matching solution. Similar to the previous subchapter, it turned out that by looking at the residuals in k -space and comparing their standard deviation with the standard deviation of the noise in k -space important information can be gained. If the residuals look Gaussian and more importantly have roughly the same standard deviation as the noise in k -space, it might then be unnecessary to compute the MAP estimate of the Bayesian k -space modeling approach. In this case it might be enough to only calculate the dictionary matching solution.

The results might lead to an important feature of the Bayesian k -space modeling approach. When using (MRF-) optimized sequences, the dictionary matching method will provide satisfying results. However, for non-optimized sequences such as the Cartesian sampling pattern above for which the dictionary matching yields poor estimates, the optimization of the Bayesian k -space modeling approach was able to significantly improve the solution. In Rieger et al. (2017) and Koolstra et al. (2019) Cartesian sampling patterns are employed for MRF leading to reasonable results. Nonetheless, the possibility to use non-optimized sequences and still compute excellent results might be a practically relevant advantage of the Bayesian k -space modeling approach. This is especially important since Cartesian sampling patterns are highly relevant in clinical practice.

6.3 Impact of Errors in the Physical Model

We will now compare the results from the dictionary matching and the Bayesian k -space modeling approach in terms of their behavior to errors in the physical model. In general, the concept of MRF only works if a physical model for the parameters is available. This physical model is an approximation of the underlying physics and will always be imperfect. In the following, a key error source in quantitative MRI will be evaluated, namely errors of the flip angle (Hurley et al. (2012)). The radiofrequency pulse interacts with human tissues and therefore the flip angle adjusted at the scanner is not equal to the actual flip angle. The errors can also arise from slice profile imperfections and local changes in the B_1 field and are already addressed for MRF (e.g. Ma et al. (2017)). We will analyze the impact of such errors in the flip angle. For that, data with flip angle $r\alpha$ for $r \leq 1$ were simulated and then evaluated by the dictionary matching and the Bayesian modeling approach assuming the flip angle equals α (and not $r\alpha$). The SNR for the different data sets is chosen to match with the level of noise in real scanner data from the phantom (cf. Chapter 4).

6.3.1 Data

Eleven different FISP data sets were created according to Chapter 4. For each data set, a total number of 1000 images were simulated and for the parameter estimation one image per radial line was reconstructed using the nonuniform fast Fourier transformation. Repetition times were chosen to be 12.54 *ms*, the inversion time as 42 *ms*, echo times as 7 *ms*, and the flip angle follows the magnitude of a sinusoidal curve with maximum 70 degree. For the simulation $r\alpha$ was used as flip angle for $r = (0.8, 0.82, \dots, 1)$. The four complex coil sensitivity maps were chosen accordingly to the ones acquired from the phantom scan in Chapter 4.

A phantom with nine tubes was chosen as the ground truth for the T_1 and T_2 values. Each tube had its individual T_1 and T_2 value and the values across the tubes were constant (cf. Chapter 2). The magnetization was then calculated according to the physical model given in Appendix B.

The modeled magnetization was transformed into the Fourier domain using a nonuniform fast Fourier transformation, and Gaussian noise with similar signal-to-noise ratio as the real scanner data from Chapter 4 was added with a standard deviation of 3.43×10^{-6} (the maximum value of the real and the imaginary part of the simulated Fourier data were of the order 10^{-3}). Note, that the same noise was added for all data sets.

The dictionaries used for calculating the dictionary matching solutions were calculated according to the physical model given in Appendix B for $T_1 = (200, 205, \dots, 3000)$ *ms* and $T_2 = (20, 22, \dots, 300)$ *ms*. The density compensation function was calculated by using the Voronoi method (Rasche et al. (1999)).

The optimization of the Bayesian k -space modeling approach was carried out similarly as in Chapter 5 for the noninformative prior. A trust-region method was employed using an approximate Hessian. A total number of 1000 iterations were computed for each optimization and the starting values were chosen as the corresponding dictionary matching solutions.

6.3.2 Results

Eleven different dictionary matching solutions and MAPs of the Bayesian k -space modeling approach were computed for the factors $r = (0.8, 0.82, \dots, 1)$.

The dictionary matching solutions and the MAP estimates of the Bayesian k -space modeling approach for $r = 0.8$ are shown in Figure 6.11 together with their error maps. The errors of T_1 and T_2 are in a similar range for both approaches, yet the errors of the dictionary matching method seem to be slightly smaller.

In Figure 6.12 the average root-mean-square errors of the nine tubes for T_1 and T_2 are shown in dependence on r . For T_1 the factor r has almost no impact on the overall RMSE of the dictionary matching solutions. The RMSE of the MAP estimates of the Bayesian k -space modeling approach are better than of the dictionary matching solutions in all cases except for $r = 0.8$ and $r = 0.82$, i.e. when a relatively high error in the physical model is present. For the MAP estimates of the Bayesian k -space modeling approach however, there is a strong influence on r . The smaller r the worse the RMSE.

For T_2 , the factor r has an impact on both types of solutions. The RMSE gets better for a larger factor, i.e. the closer the flip angle used for evaluation is to the actual simulated flip angle the better the RMSE gets. The average RMSE for the MAP estimates of the Bayesian

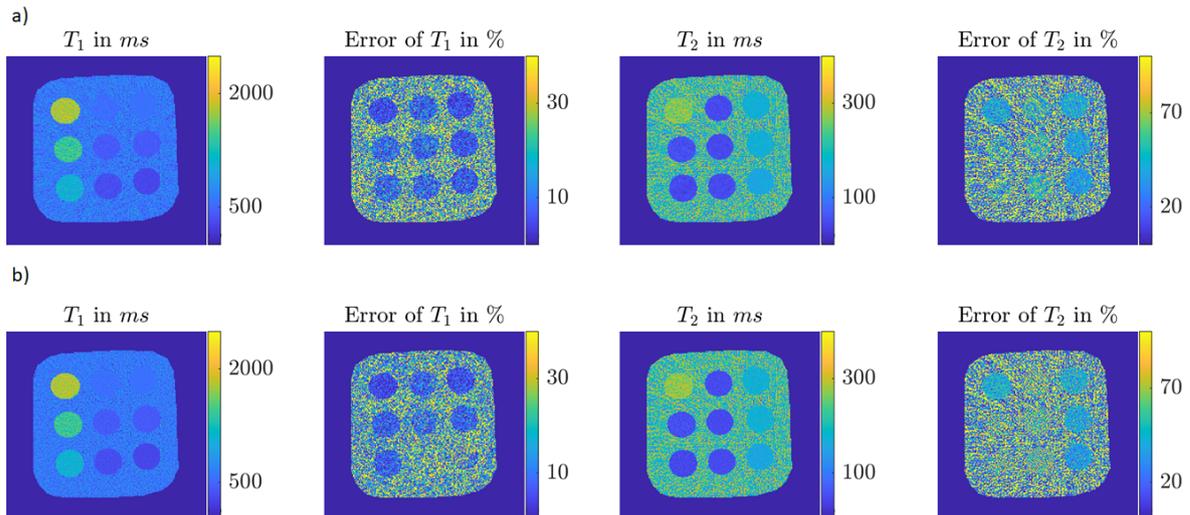


Figure 6.11: Results for data which were simulated with 0.8 times the flip angle α while the dictionary matching and the Bayesian modeling solution were evaluated with α . a) Dictionary matching solution with error maps for T_1 and T_2 . b) MAP estimate of the Bayesian k -space modeling approach with error maps for T_1 and T_2 .

k -space modeling approach and the dictionary matching solutions are similar for the different values of r .

6.3.3 Discussion

The impact of errors in the flip angle pattern has been analyzed by simulating several data sets with several flip angles $r\alpha$ for $r \leq 1$ but evaluated by the dictionary matching and the Bayesian k -space modeling approach with flip angle α .

For T_1 the average RMSE of the MAP estimates of the Bayesian k -space modeling approach is higher for larger errors in the flip angle pattern than for smaller errors. I.e., the Bayesian k -space modeling method is very sensitive to errors in the flip angle for T_1 whereas the dictionary matching produces similar results for all factors. The MAP estimate of the Bayesian k -space modeling approach has a lower RMSE than the dictionary matching solution in almost all cases.

The average RMSE for the two approaches for T_2 behave similarly. The smaller the errors in the flip angle pattern, the smaller the RMSE. I.e. both solutions are sensitive to flip angle errors for the T_2 estimation and similar results for both approaches are achieved in terms of their average RMSE.

Altogether, this leads to the conclusion that the dictionary matching approach is less sensitive to errors in the flip angle pattern for the T_1 estimation than the Bayesian k -space modeling approach. In that case, the Bayesian k -space modeling approach produces high errors in contrast to the dictionary matching. However, if the correct model is used the Bayesian k -space modeling approach leads to better results.

The observed behavior can be expected, since dictionary matching should be insensitive to model errors that have less impact than aliasing errors, while Bayesian k -space modeling in connection with outliers from the Gaussian model is necessarily sensitive. The flip angle is

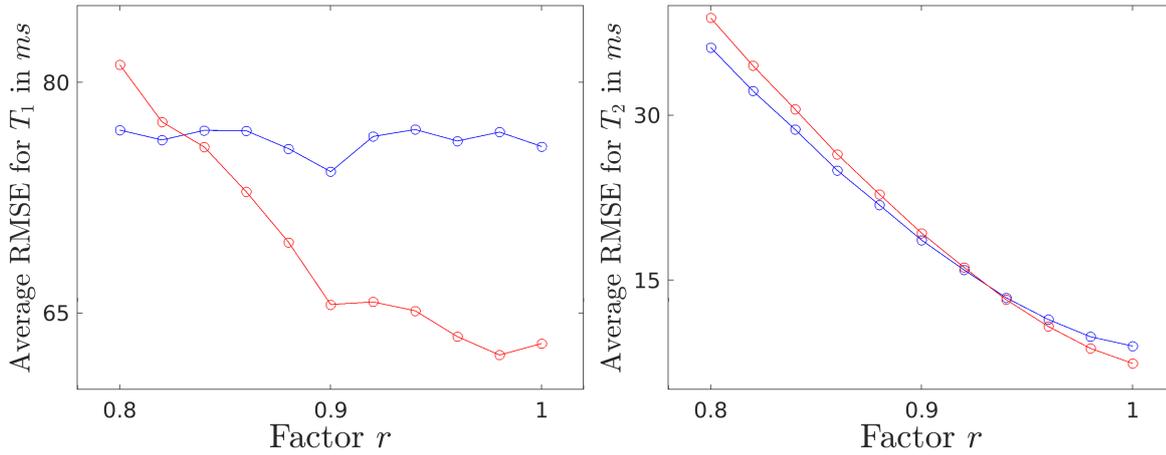


Figure 6.12: The average RMSE of the nine tubes of the dictionary matching solutions (blue line) and the MAP estimates of the Bayesian k -space modeling approach (red line) for T_1 and T_2 in dependence on the factor r which was used for simulating data with r times the flip angle α . Note that the evaluation of the dictionary matching solution and the MAP estimate of the Bayesian k -space modeling approach was carried out with α .

just one instant of model errors, similar behavior could be expected for other model errors and should be considered in future research.

6.4 Computational Aspects

The computation of the dictionary matching solution is composed of two parts. On the one hand a dictionary which is based on a physical model has to be calculated (cf. Chapter 2). The computation time of the dictionary is dependent on the number of parameters and the complexity of the physical model. Using the bSSFP model with a range in the parameters of $T_1 = (100, 120, \dots, 5020) \text{ ms}$, $T_2 = (50, 60, \dots, 520) \text{ ms}$ and $\delta f = (-100, -98, \dots, 100) \text{ Hz}$ it took 6 seconds to compute the dictionary. Using the FISP model (which is in general computationally more advanced since 201 different spins were simulated per voxel) with a range in the parameters of $T_1 = (200, 205, \dots, 3000) \text{ ms}$ and $T_2 = (20, 22, \dots, 300) \text{ ms}$ it took 587 seconds to compute the dictionary. Note that in the FISP scenario coil sensitivity maps were used which also extend the computation time. On the other hand, the magnetization curves in the image space will be compared to each entry of the dictionary. This matching process took 200 seconds in the bSSFP case and 227 seconds in the FISP case. Note that using finer grids would increase dictionary calculation drastically.

The MAP estimate of the Bayesian k -space modeling approach is calculated via a trust region algorithm (cf. Chapter 5) using 1000 iterations. It is a challenging task and several parameters have to be set upfront (cf. Chapter 5). The choice of the starting value is crucial and convergence aspects have to be taken into account. In the bSSFP case each iteration of the algorithm took around 4 seconds, the whole optimization took around 4.6×10^3 seconds. For the computationally more advanced FISP case each iteration took around 300 seconds and the total optimization took around 3.1×10^5 seconds. If the dictionary matching solution is taken as the starting value, the computation times for that have to be additionally incorporated.

Hence, the MAP estimation of the Bayesian k -space modeling approach is computationally more challenging than the dictionary matching approach. It is therefore necessary to weigh up the benefits of calculating the MAP, ideally prior to the actual computation.

6.5 Summary

We discussed several practical aspects concerning the quality of the dictionary matching and the MAP estimate of the Bayesian k -space modeling approach throughout this chapter. In a first part the influence of the SNR is explored. A bSSFP and a FISP data set were created. It was demonstrated that the MAP estimate of the Bayesian k -space modeling approach improved the dictionary matching solution only for the bSSFP data set. The SNR was higher in these data than in the FISP data set. When simulating another bSSFP data set and only changing the noise in a way that the SNR is similar to the FISP data set, the MAP estimate of the Bayesian k -space modeling approach also does not improve the dictionary matching solution. This suggests that the MAP estimate of the Bayesian k -space modeling approach enhances the dictionary matching solution only if the SNR is high enough.

In the second part of this chapter the influence of the sampling scheme is discussed. Three different data sets with the same parameters except for the sampling schemes were simulated. Cartesian, spiral and radial trajectories were chosen with nearly the same amount of data points in each image. The solutions of the dictionary matching and the MAP estimates of the Bayesian k -space modeling approach are computed for the three data sets. The dictionary matching solution for this particular Cartesian sampling scheme was extremely poor, whereas it was reasonable for the spiral and radial sampling schemes. The MAP estimates of the Bayesian k -space modeling approach were all very similar. It is interesting that the optimization of the Bayesian k -space modeling approach converges to a satisfying solution for the poor Cartesian dictionary matching solution. This clearly shows the potential advantage of the Bayesian k -space modeling approach. Nonetheless, for the two other sampling schemes the optimization does not yield a large improvement.

Subsequently, the impact of errors in the physical model is analyzed. A model describing the underlying physics is required for the dictionary matching as well as for the Bayesian k -space modeling approach. We simulated errors in the flip angle pattern and explored the effect for both approaches. It turned out that the dictionary matching approach is less sensitive to flip angle errors for the T_1 estimation. The robustness of a method to model violations is highly relevant in practice and should be more examined in the future in the context of MRF.

All in all several aspects concerning the benefit of the (computational demanding) Bayesian k -space modeling approach compared to the (easy-to-compute) dictionary matching method are presented in this chapter. It was shown that the SNR as well as the sampling scheme both have an influence on this benefit.

Under the assumptions that the physical model contains no errors, the assumed sampling model is correct, and the employed prior is consistent with the ground truth, the Bayesian k -space modeling approach is expected to be superior to the dictionary matching method. However, difficulties of the optimization task can lead to a worse performance. Due to the high computational times of the Bayesian k -space modeling, the dictionary matching would be

preferred in practice. Future research could therefore examine experimental design aspects on creating conditions under which the dictionary matching method works as good as the Bayesian k -space modeling approach. Some sequences, e.g., might produce a lower SNR than others and might hence be favored if only the dictionary matching solution should be computed. Another angle whether to also compute the MAP estimate of the Bayesian k -space modeling approach might be to look at the residuals in k -space of the dictionary matching solution. This turned out to give valuable insights about the quality of the result throughout this chapter. If the residuals in k -space look Gaussian and more importantly have roughly the same standard deviation as the noise in k -space then it might be unnecessary to also compute the MAP estimate of the Bayesian k -space modeling approach. If the standard deviation of the residuals is much higher than the standard deviation of the noise in k -space, then the MAP estimate of the Bayesian k -space modeling approach might significantly be better than the dictionary matching solution. Note that an estimation of the standard deviation of the noise in k -space is needed for that.

7

Conclusion and Outlook

The goal of this thesis is to apply Bayesian statistics to further enhance the data analysis of Magnetic Resonance Fingerprinting, which is a novel technique for quantitative magnetic resonance imaging. The widely used dictionary-based MRF approach delivers estimates for the tissue-related parameters simultaneously and within short acquisition time. Although this method has shown to work well in practice, it usually restricts itself to the calculation of estimates. Uncertainties that reliably characterize the accuracy of the tissue-related parameter estimates are essential to assess the significance of observed differences in single MRF results. No uncertainty quantification for the dictionary-based MRF estimates has been achieved so far. Additionally, the original MRF approach applies an imperfect Fourier transformation which leads to aliasing artifacts in the approximately reconstructed magnetization images. These aliasing errors could be avoided if the data is modeled in the Fourier space, the so-called k -space, directly. These two issues, uncertainty quantification and Fourier space modeling, were dealt with in this thesis using instruments of Bayesian statistics. In particular, three objectives were defined in the beginning of this work that are addressed subsequently.

In Chapter 4, a Bayesian uncertainty quantification for MRF is presented. A reliable uncertainty quantification for the dictionary-based MRF approach is novel and is developed as the first objective of this thesis. For that, a statistical analysis of the dictionary-based MRF approach is conducted. It turned out that the original MRF approach is equivalent to a maximum likelihood estimation of a statistical model which assumes a Gaussian distribution for the aliasing errors. The same statistical model is assumed for the Bayesian uncertainty quantification together with prior distributions for the tissue-related parameters. The prior knowledge combines a standard noninformative prior for the proton density and the variance with an informative prior for the relaxation times. The latter utilizes as prior knowledge the ranges chosen for the relaxation times when calculating the dictionary. A Bayesian inference is carried out leading to an analytical expression for the joint posterior distribution of the relaxation times. The existence of the posterior distribution could be ensured. Numerical integration utilizing the pre-computed dictionary yields whole probability distributions for the relaxation times in every voxel in a short time. Uncertainties and probability statements can then be calculated accordingly. The Bayesian uncertainty quantification was validated

on simulated data as well as on phantom and in vivo measurements. The proposed Bayesian uncertainty quantification and the dictionary-based MRF method are based on the very same assumption, namely that the aliasing errors can be treated as noise. This assumption and the hence simple statistical model provide an approximation to the true distribution of errors. More sophisticated statistical models which, e.g., can account for dependencies of the reconstructed and modeled magnetization and which could then, in turn, have consequences for the dictionary-based MRF estimation could be considered in the future. However, numerical calculations become challenging for more sophisticated statistical models which would contain additional unknowns. Furthermore, the proposed uncertainty quantification depends on the chosen protocol such as the flip angles and the repetition times and could thus be used for protocol optimization. Future work could address this feature and might be able to help design optimal sequences.

In Chapter 5, a general statistical model which assumes Gaussian distributed errors for a class of nonlinear, large-scale regression problems is considered together with three different prior distributions for the sought parameter. These are a noninformative, a Gauss Markov Random Field (GMRF) and lastly a so-called partition prior. In order to guarantee propriety of the resulting posterior distributions as well as the existence of moments, additional precautions have to be made. E.g., the GMRF prior has to be augmented by a proper prior on the mean of the parameter. It was also possible to show a close relationship between a partition prior and an intrinsic GMRF prior. The considered class of problems includes MRF with a data model in the Fourier domain. Previous work already applied a maximum likelihood estimation of a particular statistical model in k -space but no Bayesian inference was carried out so far. The second objective of this thesis is a Bayesian inference for a large-scale k -space MRF model with different prior distributions for the tissue-related parameters. In contrast to the dictionary-based MRF method, the Fourier domain model is dependent on all tissue-related parameters in parallel. The high-dimensionality of the MRF parameters made common techniques for Bayesian computation such as Markov Chain Monte Carlo methods infeasible. We decided to calculate the maximum a posteriori estimate leading to a high-dimensional and non-convex optimization problem. The applicability to simulated MRF data was shown and the influence of the choice of the prior distribution was evaluated. Using the partition prior instead of the augmented noninformative prior for the relaxation times and the proton density improved the estimates significantly. It was also possible to calculate credible intervals with reasonable coverage probabilities under repeated sampling. Further research could address the choice of other prior distributions for the parameters. Another possibility of improvement is the calculation time. Although the Bayesian k -space modeling approach captures the MR data without any loss, the resulting optimization problem is time consuming and computationally challenging to solve, especially for complicated models. Accelerating the optimization is therefore necessary before using it in clinical practice. A further line of research could be to avoid the calculation of a maximum a posteriori estimate and to develop a Markov Chain Monte Carlo sampling procedure to approximate the high-dimensional posterior distribution itself. Sampling in high dimensions is an active field of research in the statistical community and the Bayesian k -space modeling approach could benefit from its work.

Two approaches for MRF were presented in this thesis. The original, dictionary-based method was shown to be equivalent to a maximum likelihood estimation of a statistical model which assumes a Gaussian distribution for the errors including the aliasing artifacts. This approach is easy to implement and can be carried out fast. However, due to the imperfect Fourier transformation aliasing artifacts in the magnetization courses are introduced. A direct modeling of the MR data in the Fourier domain would avoid these artifacts. The proposed Bayesian k -space modeling approach establishes a statistical model in the Fourier domain which assumes Gaussian distributed errors in k -space. This approach should be superior to the dictionary-based MRF approach under the following assumptions: the physical model contains no errors, the assumed sampling model is correct, and the employed prior is consistent with the ground truth. However, the high-dimensional optimization task is computationally challenging to solve. Additionally, dependent on the complexity of the physical model, the calculation of the maximum a posteriori estimate can be very time consuming. In our third and last objective that is discussed in Chapter 6, we compare the dictionary-based MRF and Bayesian k -space modeling approach in terms of the accuracy of the estimates dependent on the level of noise, the MRF sampling scheme and computational complexity as well as in terms of their sensitivity to errors in the physical model. Several interesting insights could be shown. It turned out that only if the SNR is high enough and therefore the aliasing errors dominate the error of the dictionary-based MRF estimate, the Bayesian k -space modeling approach improves the dictionary-based MRF approach. Since some sequences produce a lower SNR than others, they might hence be favored if only the easy-to-compute dictionary-based MRF estimate should be computed. The analysis of the impact of the sampling scheme has shown that the Bayesian k -space modeling approach but not the dictionary-based MRF approach succeeded in a case where a non-optimized MRF sequence was used. This aspect shows the versatility of the direct modeling in the Fourier domain. Future research should determine if the same effect occurs for other non-optimized sequences. Since optimizing a sequence in a way such that the dictionary-based MRF approach delivers good results is challenging, this feature might be practically relevant and hence points to scenarios where the Bayesian k -space modeling approach could be used. An important indicator whether the maximum a posteriori estimate of the Bayesian k -space modeling approach should be computed is to look at the residuals of the dictionary-based MRF estimate in the Fourier domain. A comparison of the standard deviation of these residuals to the standard deviation of the Fourier noise hints at the quality of the estimate. Throughout this chapter only simulated data were considered and future research should verify these findings for real data.

MRF is a recent approach in quantitative MRI. The additional uncertainty quantification developed in this thesis gives valuable insight about assessing the significance of single results and of multiple, time-delayed results from the same subject. The fast acquisition time of an MRF sequence allows for an application in clinical practice where an uncertainty for each estimate could help in practical decisions about the nature of a tissue. Future applications to patient data should show practical relevant examples where the uncertainties could improve diagnostics. The Bayesian k -space modeling approach has the potential to improve the dictionary-based estimates significantly and could be used for, e.g., non-optimized MRF sequences. The possibility to include prior knowledge about spatial dependencies of the

7. Conclusion and Outlook

parameters additionally improves the estimates and makes it robust against challenges due to the high-dimensionality of the problem. Validating the Bayesian k -space modeling approach in terms of in vivo and, in particular, patient data, should be another topic of future research. The general framework of the large-scale Bayesian modeling also allows for an application to other high-dimensional regression problems in which the parameters model some spatial dependency.



Physical Model of a bSSFP Sequence

In Chapter 5 a bSSFP sequence for the data recording process is used. The dynamic magnetization of an arbitrary MRI sequence can be described by a set of three coupled differential equations, the so-called Bloch equations (see, e.g., Nishimura (1996) and Chapter 2). Let the three-dimensional vector $M_l^i = (M_{x,l}^i, M_{y,l}^i, M_{z,l}^i)^T$ define the magnetization in the i -th voxel after the l -th high-frequency pulse. When using a bSSFP sequence the magnetization model can be expressed in terms of a three-dimensional linear discrete time system (Hargreaves et al. (2001))

$$M_{l+1}^i = P_l^i M_l^i + Q_l^i, \quad (\text{A.1})$$

where P_l^i is a 3×3 matrix and Q_l^i a 3×1 vector with initial magnetization $M_0^i = (0, 0, -\rho_i)^T$ in each voxel. The matrix P_l^i and the vector Q_l^i depend nonlinearly on the tissue-related parameters $\rho_i, T1_i, T2_i$ and δf_i and the (known) parameters of the radiofrequency pulse FA_l, TR and TE . Choosing the echo time as $TE = TR/2$ in every readout step, and by defining $\gamma = \pi TR$ and $\tau = TR/2$, the elements of the matrix P_l^i are given by

$$(P_l^i)_{1,1} = \exp(-2\tau/T2_i) \left(\cos^2(\gamma \delta f_i) - \cos(FA_l) \sin^2(\gamma \delta f_i) \right), \quad (\text{A.2})$$

$$(P_l^i)_{1,2} = \exp(-2\tau/T2_i) \cos(\gamma \delta f_i) \sin(\gamma \delta f_i) (\cos(FA_l) + 1), \quad (\text{A.3})$$

$$(P_l^i)_{1,3} = \exp(-\tau/T1_i - \tau/T2_i) \sin(FA_l) \sin(\gamma \delta f_i), \quad (\text{A.4})$$

$$(P_l^i)_{2,1} = -(P_l^i)_{1,2} \quad (\text{A.5})$$

$$(P_l^i)_{2,2} = \exp(-2\tau/T2_i) \left(\cos^2(\gamma \delta f_i) \cos(FA_l) - \sin^2(\gamma \delta f_i) \right), \quad (\text{A.6})$$

$$(P_l^i)_{2,3} = \exp(-\tau/T1_i - \tau/T2_i) \sin(FA_l) \cos(\gamma \delta f_i), \quad (\text{A.7})$$

$$(P_l^i)_{3,1} = (P_l^i)_{1,3}, \quad (\text{A.8})$$

$$(P_l^i)_{3,2} = -(P_l^i)_{2,3}, \quad (\text{A.9})$$

$$(P_l^i)_{3,3} = \exp(-2\tau/T1_i) \cos(FA_l), \quad (\text{A.10})$$

and the elements of vector Q_l^i by

$$(Q_l^i)_1 = \rho_i (1 - \exp(-\tau/T1_i)) \exp(-\tau/T2_i) \sin(\gamma \delta f_i) \sin(FA_l), \quad (\text{A.11})$$

$$(Q_l^i)_2 = \rho_i (1 - \exp(-\tau/T1_i)) \exp(-\tau/T2_i) \cos(\gamma \delta f_i) \sin(FA_l), \quad (\text{A.12})$$

$$(Q_l^i)_3 = \rho_i (1 - \exp(-\tau/T1_i)) \exp(-\tau/T1_i) \cos(FA_l) - \exp(-\tau/T1_i) + 1. \quad (\text{A.13})$$

Only the complex-valued transverse magnetization

$$M_l^i = M_{x,l}^i + iM_{y,l}^i \quad (\text{A.14})$$

of the i -th voxel contributes to the MRI signal (see, e.g., Nishimura (1996)). The resulting magnetization in all voxels after the l -th pulse is then given by

$$m_l(\theta) = (M_l^1, \dots, M_l^N)^T.$$

Looking at the single entries of the matrix P_l^i and the vector Q_l^i , it is obvious that for arbitrary T_1 and T_2 , $\|m_l(\theta)\|$ and thus $\|\chi^2(\theta)\|$ from (5.20) are bounded from above. Condition (5.2) is therefore fulfilled.

B

Physical Model of a FISP Sequence

In Chapter 4 a FISP sequence is used for recording the Fourier data. The physical model underlying the magnetization process can be described by a three dimensional linear system of equation, for more details we refer to Hargreaves et al. (2001). Unlike the bSSFP sequence, a dephasing gradient is applied in the middle of each repetition time TR in the FISP sequence. The signal is therefore an average of the spins in one voxel which we will be simulated by taking Nf different spins in one voxel. The following Matlab[®] code is used to explain the implemented sequence. The flip angles will be denoted by $flip$, echo times by TE and the inversion time by TI .

```
function Msig = FISP_sequence(T1,T2,TE,TI,TR,flip ,Nf)
flip = pi*flip/180; % degree to radians
Nex = length(TR);
phi = linspace(-0.5,0.5,801)*8*pi;
Msig = zeros(1,Nex);
M0 = [zeros(2,Nf);-ones(1,Nf)];
[Ati,Bti] = freeprecess(TI,T1,T2);
M = Ati*M0 + Bti;
on = ones(1,Nf);
for n = 1:Nex
    [Ate,Bte] = freeprecess(TE(n),T1,T2);
    [Atr,Btr] = freeprecess(TR(n)-TE(n),T1,T2);
    A = Ate * xrot(flip(n));
    B = Bte;
    M = A*M+B*on;
    Msig(n) = mean( squeeze(M(1,:)+1i*M(2,:)) );
    M = Atr*M+Btr*on;
    for k = 1:Nf
        M(:,k) = zrot(phi(k))*M(:,k);
    end
end
```

B. Physical Model of a FISP Sequence

end

end

```
function [Afp, Bfp]=freeprecess (T, T1, T2)
```

```
E1 = exp(-T/T1);
```

```
E2 = exp(-T/T2);
```

```
Afp = [E2 0 0; 0 E2 0; 0 0 E1];
```

```
Bfp = [0 0 1-E1]';
```

end

```
function Rx=xrot(phi)
```

```
Rx = [1 0 0; 0 cos(phi) sin(phi); 0 -sin(phi) cos(phi)];
```

end

```
function Rz=zrot(phi)
```

```
Rz = [cos(phi) sin(phi) 0; -sin(phi) cos(phi) 0; 0 0 1];
```

end



Matlab Code for the Bayesian Uncertainty Quantification

In Chapter 4 a Bayesian uncertainty quantification for the dictionary matching method is presented. The following code was used to calculate uncertainties and marginal posterior distributions.

```
function [U_T1,U_T2,pi_T1_save,pi_T2_save] = Bayesian_Uncertainty_Quant_MRF(theta_MRF,mxy_hat,T1_dict,T2_dict,dict,mask)

%% INPUT:

%% theta_MRF ... the solution of the dictionary matching method of size
%% 2 x Nx x Ny where Nx and Ny denote the number of voxels in x and y
%% direction respectively, theta_MRF(1,::) should be the T1 estimate and
%% theta_MRF(2,::) should be the T2 estimate

%% mxy_hat ... observed sequence of magnetiations of size Nex x Nx x Ny
%% where Nex denotes the sequence length

%% T1_dict ... vector of unique T1 values in the dictionary, e.g.
%% [200:5:3000];

%% T2_dict ... vector of unique T2 values in the dictionary, e.g.
%% [20:2:300];

%% dict ... dictionary of size Nex x length(combvec(T1_dict,T2_dict))

%% !!! Caution: the dictionary has to be created according to
%% dict = physical_mrf_model(combvec(T1_dict,T2_dict)); !!!
```

C. Matlab Code for the Bayesian Uncertainty Quantification

```
%% mask ... a mask of the voxels of size Nx x Ny, an entry 1 represents a
%% voxel for which the uncertainty should be calculated, else the entry
%% should be 0

%% OUTPUT:

%% U_T1 and U_T2 ... Uncertainty maps for T1 and T2 of size Nx x Ny

%% pi_T1_save and pi_T2_save ... marginal distributions for T1 and T2 of
%% size length(T1_dict) x length(find(mask)) and
%% length(T2_dict) x length(find(mask)) respectively

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

[Nex,Nx,Ny] = size(mxy_hat);
if nargin < 6
    mask = ones(Nx,Ny);
end
indmask = find(mask);
Var_T1 = zeros(1,length(indmask));
Var_T2 = zeros(1,length(indmask));
pi_T1_save = zeros(length(T1_dict),length(indmask));
pi_T2_save = zeros(length(T2_dict),length(indmask));
mr = real(dict);
mi = imag(dict);
detMtM = (dot(mi,mi)+dot(mr,mr)).*(dot(mi,mi)+dot(mr,mr));
d = dot(mi,mi) + dot(mr,mr);

parfor i = 1:length(indmask)
    m2 = mxy_hat(:,indmask(i));
    yr_rep = repmat(real(m2),1,size(mr,2));
    yi_rep = repmat(imag(m2),1,size(mr,2));
    muhat_r = (dot(mr,yr_rep) + dot(mi,yi_rep))./d;
    muhat_i = (dot(-mi,yr_rep) + dot(mr,yi_rep))./d;

    chi_r = yr_rep-(mr.*muhat_r-mi.*muhat_i);
    chi_i = yi_rep-(mi.*muhat_r+mr.*muhat_i);
    Gridlogpi_T1T2 = (1-Nex).*log(dot(chi_r,chi_r)+dot(chi_i,chi_i)) - ...
        1/2.*log(detMtM);

    GridReslogpi_T1T2 = reshape(Gridlogpi_T1T2,length(T1_dict),...
        length(T2_dict));
```

```

GridReslogpi_T1T2 = GridReslogpi_T1T2';

pi_T1T2_small = GridReslogpi_T1T2 - max(GridReslogpi_T1T2(:));
pi_T1T2 = exp(pi_T1T2_small);
pi_T1 = trapz(T2_dict, pi_T1T2, 1);
cT1 = trapz(T1_dict, pi_T1);
pi_norm_T1 = pi_T1/cT1;
pi_T2 = trapz(T1_dict, pi_T1T2, 2);
cT2 = trapz(T2_dict, pi_T2);
pi_norm_T2 = pi_T2/cT2;

pi_T1_save(:, i) = pi_norm_T1;
pi_T2_save(:, i) = pi_norm_T2;

Var_T1(i) = trapz(T1_dict, (T1_dict - ...
    theta_MRF(1, indmask(i)).^2.*pi_norm_T1);
Var_T2(i) = trapz(T2_dict, (T2_dict - ...
    theta_MRF(2, indmask(i)).^2.*pi_norm_T2');
end

U_T1 = zeros(Nx, Ny);
U_T2 = zeros(Nx, Ny);
U_T1(indmask) = sqrt(Var_T1);
U_T2(indmask) = sqrt(Var_T2);

end

```


References

- Abragam, A. (1983). *Principles of Nuclear Magnetism*. Oxford University Press.
- Aseervatham, S., Antoniadis, A., Gaussier, É., Bulet, M., and Denneulin, Y. (2011). A sparse version of the ridge logistic regression for large-scale text categorization. *Pattern Recognition Letters*, 32(2):101–106.
- Assländer, J., Glaser, S. J., and Hennig, J. (2017). Pseudo steady-state free precession for mr-fingerprinting. *Magnetic Resonance in Medicine*, 77(3):1151–1161.
- Awate, S. P. and Whitaker, R. T. (2007). Feature-preserving mri denoising: a nonparametric empirical Bayes approach. *IEEE Transactions on Medical Imaging*, 26(9):1242–1255.
- Badve, C., Yu, A., Dastmalchian, S., Rogers, M., Ma, D., Jiang, Y., Margevicius, S., Pahwa, S., Lu, Z., Schluchter, M., et al. (2017). Mr fingerprinting of adult brain tumors: initial experience. *American Journal of Neuroradiology*, 38(3):492–499.
- Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.
- Bernardo, J. M. and Smith, A. F. (2009). *Bayesian theory*, volume 405. John Wiley & Sons.
- Beskos, A., Crisan, D., and Jasra, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *The Annals of Applied Probability*, 24(4):1396–1445.
- Beskos, A., Roberts, G., and Stuart, A. (2009). Optimal scalings for local Metropolis-Hastings chains on nonproduct targets in high dimensions. *The Annals of Applied Probability*, pages 863–898.
- Betancourt, M. and Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. *Current Trends in Bayesian Methodology with Applications*, 79(30):2–4.
- Bipin Mehta, B., Coppo, S., Frances McGivney, D., Ian Hamilton, J., Chen, Y., Jiang, Y., Ma, D., Seiberlich, N., Gulani, V., and Alan Griswold, M. (2019). Magnetic resonance fingerprinting: a technical review. *Magnetic Resonance in Medicine*, 81(1):25–46.
- BrainWeb (2018). Simulated brain database. <http://brainweb.bic.mni.mcgill.ca/brainweb/>. Accessed: 2018-04-12.
- Braun, M. and McAuliffe, J. (2010). Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*, 105(489):324–335.
- Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R. (2014). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley & Sons.
- Buonincontri, G. and Sawiak, S. J. (2016). Mr fingerprinting with simultaneous b1 estimation. *Magnetic Resonance in Medicine*, 76(4):1127–1135.
- Cauley, S. F., Setsompop, K., Ma, D., Jiang, Y., Ye, H., Adalsteinsson, E., Griswold, M. A., and Wald, L. L. (2015). Fast group matching for mr fingerprinting reconstruction. *Magnetic Resonance in Medicine*, 74(2):523–528.

REFERENCES

- Cawley, P. J., Maki, J. H., and Otto, C. M. (2009). Cardiovascular magnetic resonance imaging for valvular heart disease: technique and validation. *Circulation*, 119(3):468–478.
- Chen, Y., Jiang, Y., Pahwa, S., Ma, D., Lu, L., Twieg, M. D., Wright, K. L., Seiberlich, N., Griswold, M. A., and Gulani, V. (2016). Mr fingerprinting for rapid quantitative abdominal imaging. *Radiology*, 279(1):278–286.
- Christen, T., Pannetier, N., Ni, W. W., Qiu, D., Moseley, M. E., Schuff, N., and Zaharchuk, G. (2014). Mr vascular fingerprinting: A new approach to compute cerebral blood volume, mean vessel radius, and oxygenation maps in the human brain. *Neuroimage*, 89:262–270.
- Clare, S. and Jezzard, P. (2001). Rapid t1 mapping using multislice echo planar imaging. *Magnetic Resonance in Medicine*, 45(4):630–634.
- Cohen, O., Zhu, B., and Rosen, M. S. (2018). Mr fingerprinting deep reconstruction network (drone). *Magnetic Resonance in Medicine*, 80(3):885–894.
- Cooley, J. W., Lewis, P. A., and Welch, P. D. (1969). The fast Fourier transform and its applications. *IEEE Transactions on Education*, 12(1):27–34.
- Crewe, T. and McCracken, J. (2015). Long-term trends in the number of monarch butterflies (lepidoptera: Nymphalidae) counted on fall migration at long point, ontario, canada (1995–2014). *Annals of the Entomological Society of America*, 108(5):707–717.
- Cuocolo, R., Cipullo, M. B., Stanzione, A., Ugga, L., Romeo, V., Radice, L., Brunetti, A., and Imbriaco, M. (2019). Machine learning applications in prostate cancer magnetic resonance imaging. *European Radiology Experimental*, 3(1):1–8.
- Davies, M., Puy, G., Vanderghenst, P., and Wiaux, Y. (2014). A compressed sensing framework for magnetic resonance fingerprinting. *SIAM Journal on Imaging Sciences*, 7(4):2623–2656.
- Davis, P. J. and Rabinowitz, P. (2007). *Methods of numerical integration*. Courier Corporation.
- Deng, N., Xiao, Y., and Zhou, F. (1993). Nonmonotonic trust region algorithm. *Journal of Optimization Theory and Applications*, 76(2):259–285.
- Deoni, S. C., Peters, T. M., and Rutt, B. K. (2005). High-resolution t1 and t2 mapping of the brain in a clinically acceptable time with despot1 and despot2. *Magnetic Resonance in Medicine*, 53(1):237–241.
- Deshmane, A., McGivney, D. F., Ma, D., Jiang, Y., Badve, C., Gulani, V., Seiberlich, N., and Griswold, M. A. (2019). Partial volume mapping using magnetic resonance fingerprinting. *NMR in Biomedicine*, 32(5):e4082.
- Doneva, M., Amthor, T., Koken, P., Sommer, K., and Börnert, P. (2017). Matrix completion-based reconstruction for undersampled magnetic resonance fingerprinting data. *Magnetic Resonance Imaging*, 41:41–52.
- Dongarra, J. and Sullivan, F. (2000). Guest editors’ introduction: The top 10 algorithms. *Computing in Science & Engineering*, 2(1):22–23.
- Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.
- Fang, Z., Chen, Y., Liu, M., Xiang, L., Zhang, Q., Wang, Q., Lin, W., and Shen, D. (2019). Deep learning for fast and spatially constrained tissue quantification from highly accelerated data in magnetic resonance fingerprinting. *IEEE Transactions on Medical Imaging*, 38(10):2364–2374.
- Feng, L., Grimm, R., Block, K. T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D. K., and Otazo, R. (2014). Golden-angle radial sparse parallel mri: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric mri. *Magnetic Resonance in Medicine*, 72(3):707–717.

- Flassbeck, S., Schmidt, S., Bachert, P., Ladd, M. E., and Schmitter, S. (2019). Flow mr fingerprinting. *Magnetic Resonance in Medicine*, 81(4):2536–2550.
- Fletcher, R. (2013). *Practical methods of optimization*. John Wiley & Sons.
- Fox, C. W. and Roberts, S. J. (2012). A tutorial on variational Bayesian inference. *Artificial Intelligence Review*, 38(2):85–95.
- Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC Press.
- Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, pages 733–760.
- Genkin, A., Lewis, D. D., and Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, 49(3):291–304.
- Greengard, L. and Lee, J.-Y. (2004). Accelerating the nonuniform fast Fourier transform. *SIAM Review*, 46(3):443–454.
- Guan, Y. and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics*, pages 1780–1815.
- Gutmann, M. U. and Hyvärinen, A. (2012). Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *The Journal of Machine Learning Research*, 13(1):307–361.
- Hamilton, J. I., Jiang, Y., Chen, Y., Ma, D., Lo, W.-C., Griswold, M., and Seiberlich, N. (2017). Mr fingerprinting for rapid quantification of myocardial t1, t2, and proton spin density. *Magnetic Resonance in Medicine*, 77(4):1446–1458.
- Hargreaves, B. A., Vasanawala, S. S., Pauly, J. M., and Nishimura, D. G. (2001). Characterization and reduction of the transient response in steady-state mr imaging. *Magnetic Resonance in Medicine*, 46(1):149–158.
- Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.
- Huang, J., Horowitz, J. L., and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics*, pages 587–613.
- Hurley, S. A., Yarnykh, V. L., Johnson, K. M., Field, A. S., Alexander, A. L., and Samsonov, A. A. (2012). Simultaneous variable flip angle–actual flip angle imaging method for improved accuracy and precision of three-dimensional t1 and b1 measurements. *Magnetic Resonance in Medicine*, 68(1):54–64.
- Jaubert, O., Cruz, G., Bustin, A., Schneider, T., Koken, P., Doneva, M., Rueckert, D., Botnar, R., and Prieto, C. (2020). Free-running cardiac magnetic resonance fingerprinting: Joint t1/t2 map and cine imaging. *Magnetic Resonance Imaging*, 68:173–182.
- Jiang, X., Jiang, J., and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, pages 1479–1506.
- Jiang, Y., Ma, D., Seiberlich, N., Gulani, V., and Griswold, M. A. (2015). Mr fingerprinting using fast imaging with steady state precession (fisp) with spiral readout. *Magnetic Resonance in Medicine*, 74(6):1621–1631.
- Kara, D., Fan, M., Hamilton, J., Griswold, M., Seiberlich, N., and Brown, R. (2019). Parameter map error due to normal noise and aliasing artifacts in mr fingerprinting. *Magnetic Resonance in Medicine*, 81(5):3108–3123.
- Kass, R. E., Tierney, L., and Kadane, J. B. (1990). The validity of posterior expansions based on Laplace’s method. *Bayesian and Likelihood Methods in Statistics and Econometrics*, 7:473–488.

REFERENCES

- Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370.
- Koolstra, K., Beenakker, J.-W. M., Koken, P., Webb, A., and Börnert, P. (2019). Cartesian mr fingerprinting in the eye at 7t using compressed sensing and matrix completion-based reconstructions. *Magnetic Resonance in Medicine*, 81(4):2551–2565.
- Körzdörfer, G., Jiang, Y., Speier, P., Pang, J., Ma, D., Pfeuffer, J., Hensel, B., Gulani, V., Griswold, M., and Nittka, M. (2019). Magnetic resonance field fingerprinting. *Magnetic Resonance in Medicine*, 81(4):2347–2359.
- Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191.
- Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. S. (2014). Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis*, 9(3):699–732.
- Leiner, T., Rueckert, D., Suinesiaputra, A., Baeßler, B., Nezafat, R., Išgum, I., and Young, A. A. (2019). Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *Journal of Cardiovascular Magnetic Resonance*, 21(1):1–14.
- Lemasson, B., Pannetier, N., Coquery, N., Boisserand, L. S., Collomb, N., Schuff, N., Moseley, M., Zaharchuk, G., Barbier, E., and Christen, T. (2016). Mr vascular fingerprinting in stroke and brain tumors models. *Scientific reports*, 6(1):1–11.
- Liao, C., Bilgic, B., Manhard, M. K., Zhao, B., Cao, X., Zhong, J., Wald, L. L., and Setsompop, K. (2017). 3d mr fingerprinting with accelerated stack-of-spirals and hybrid sliding-window and grappa reconstruction. *Neuroimage*, 162:13–22.
- Liao, C., Wang, K., Cao, X., Li, Y., Wu, D., Ye, H., Ding, Q., He, H., and Zhong, J. (2018). Detection of lesions in mesial temporal lobe epilepsy by using mr fingerprinting. *Radiology*, 288(3):804–812.
- Long, Q., Scavino, M., Tempone, R., and Wang, S. (2013). Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39.
- Ma, D., Coppo, S., Chen, Y., McGivney, D. F., Jiang, Y., Pahwa, S., Gulani, V., and Griswold, M. A. (2017). Slice profile and b1 corrections in 2d magnetic resonance fingerprinting. *Magnetic Resonance in Medicine*, 78(5):1781–1789.
- Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., and Griswold, M. A. (2013). Magnetic resonance fingerprinting. *Nature*, 495(7440):187–192.
- Ma, D., Jiang, Y., Chen, Y., McGivney, D., Mehta, B., Gulani, V., and Griswold, M. (2018). Fast 3d magnetic resonance fingerprinting for a whole-brain coverage. *Magnetic Resonance in Medicine*, 79(4):2190–2197.
- Macovski, A. (1996). Noise in mri. *Magnetic Resonance in Medicine*, 36(3):494–497.
- Marroquín, J. L., Vemuri, B. C., Botello, S., Calderon, E., and Fernandez-Bouzas, A. (2002). An accurate and efficient Bayesian method for automatic segmentation of brain mri. *IEEE Transactions on Medical Imaging*, 21(8):934–945.
- Mazor, G., Weizman, L., Tal, A., and Eldar, Y. C. (2018). Low-rank magnetic resonance fingerprinting. *Medical Physics*, 45(9):4066–4084.
- McGivney, D., Deshmane, A., Jiang, Y., Ma, D., Badve, C., Sloan, A., Gulani, V., and Griswold, M. (2018). Bayesian estimation of multicomponent relaxation parameters in magnetic resonance fingerprinting. *Magnetic Resonance in Medicine*, 80(1):159–170.

- McGivney, D. F., Pierre, E., Ma, D., Jiang, Y., Saybasili, H., Gulani, V., and Griswold, M. A. (2014). Svd compression for magnetic resonance fingerprinting in the time domain. *IEEE Transactions on Medical Imaging*, 33(12):2311–2322.
- Meier, L., Van de Geer, S., and Bühlmann, P. (2009). High-dimensional additive modeling. *The Annals of Statistics*, 37(6B):3779–3821.
- Menezes, G. L., Knuttel, F. M., Stehouwer, B. L., Pijnappel, R. M., and van den Bosch, M. A. (2014). Magnetic resonance imaging in breast cancer: a literature review and future perspectives. *World Journal of Clinical Oncology*, 5(2):61.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American statistical association*, 44(247):335–341.
- Metzner, S. and Wübbeler, G. (2021). Software-working group 8.42. <https://www.ptb.de/cms/en/ptb/fachabteilungen/abt8/fb-84/ag-842/software.html>. Accessed: 2021-03-01.
- Metzner, S., Wübbeler, G., and Elster, C. (2019). Approximate large-scale Bayesian spatial modeling with application to quantitative magnetic resonance imaging. *ASTA Advances in Statistical Analysis*, 103:333–355.
- Metzner, S., Wübbeler, G., Flassbeck, S., Gatefait, C., Kolbitsch, C., and Elster, C. (2021). Bayesian uncertainty quantification for magnetic resonance fingerprinting. *Physics in Medicine & Biology*, 66(7):075006.
- Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6):47–60.
- Müller-Gronbach, T., Novak, E., and Ritter, K. (2012). *Monte Carlo-Algorithmen*. Springer-Verlag.
- Mumcuoglu, E., Leahy, R., and Cherry, S. (1996). Bayesian reconstruction of pet images : methodology and performance analysis. *Physics in Medicine & Biology*, 41(9):1777–1807.
- Nishii, R. and Eguchi, S. (2006). Image classification based on Markov random field models with Jeffreys divergence. *Journal of Multivariate Analysis*, 97:1997–2008.
- Nishimura, D. (1996). *The Principles of Magnetic Resonance Imaging*. Stanford University.
- Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.
- Oakley, J. E. and O’Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769.
- Panda, A., Mehta, B. B., Coppo, S., Jiang, Y., Ma, D., Seiberlich, N., Griswold, M. A., and Gulani, V. (2017). Magnetic resonance fingerprinting—an overview. *Current Opinion in Biomedical Engineering*, 3:56–66.
- Polders, D. L., Leemans, A., Luijten, P. R., and Hoogduin, H. (2012). Uncertainty estimations for quantitative in vivo mri t1 mapping. *Journal of Magnetic Resonance*, 224:53–60.
- Rasche, V., Proksa, R., Sinkus, R., Bornert, P., and Eggers, H. (1999). Resampling of data between arbitrary grids using convolution interpolation. *IEEE Transactions on Medical Imaging*, 18(5):385–392.
- Redpath, T. W. (1998). Signal-to-noise ratio in mri. *The British Journal of Radiology*, 71(847):704–707.
- Rieger, B., Zimmer, F., Zapp, J., Weingärtner, S., and Schad, L. R. (2017). Magnetic resonance fingerprinting using echo-planar imaging: Joint quantification of t1 and relaxation times. *Magnetic Resonance in Medicine*, 78(5):1724–1733.
- Riihimäki, J., Vehtari, A., et al. (2014). Laplace approximation for logistic Gaussian process density estimation and regression. *Bayesian Analysis*, 9(2):425–448.

REFERENCES

- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.
- Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.
- Roberts, G. O., Rosenthal, J. S., et al. (2004). General state space Markov chains and mcmc algorithms. *Probability Surveys*, 1:20–71.
- Roemer, P. B., Edelstein, W. A., Hayes, C. E., Souza, S. P., and Mueller, O. M. (1990). The nmr phased array. *Magnetic Resonance in Medicine*, 16(2):192–225.
- Roos, M., Martins, T. G., Held, L., Rue, H., et al. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis*, 10(2):321–349.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.
- Rue, H., Martino, S., and Chopin, N. (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.
- Salvatore, C., Cerasa, A., Battista, P., Gilardi, M. C., Quattrone, A., and Castiglioni, I. (2015). Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer’s disease: a machine learning approach. *Frontiers in Neuroscience*, 9:307.
- Santermans, E., Robesyn, E., Ganyani, T., Sudre, B., Faes, C., Quinten, C., Van Bortel, W., Haber, T., Kovac, T., Van Reeth, F., et al. (2016). Spatiotemporal evolution of ebola virus disease at sub-national level during the 2014 west africa epidemic: model scrutiny and data meagreness. *PLOS One*, 11(1):e0147172.
- Sarty, G. E. (2003). The relationship between the Nyquist criterion and the point spread function. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering: An Educational Journal*, 17(1):17–24.
- Scheffler, K. and Lehnhardt, S. (2003). Principles and applications of balanced ssfp techniques. *European Radiology*, 13(11):2409–2418.
- Schmid, V. J., Whitcher, B., Padhani, A. R., Taylor, N. J., and Yang, G.-Z. (2006). Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 25(12):1627–1636.
- Schmidt, P., Schmid, V. J., Gaser, C., Buck, D., Bühlren, S., Förschler, A., and Mühlau, M. (2013). Fully Bayesian inference for structural mri: Application to segmentation and statistical analysis of t2-hypointensities. *PLOS One*, 8(7).
- Seeger, M. W. and Nickisch, H. (2011). Large scale Bayesian inference and experimental design for sparse linear models. *SIAM Journal on Imaging Sciences*, 4(1):166–199.
- Smith, M. and Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 102(478):417–431.
- Sommer, K., Amthor, T., Doneva, M., Koken, P., Meineke, J., and Börnert, P. (2017). Towards predicting the encoding capability of mr fingerprinting sequences. *Magnetic Resonance Imaging*, 41:7–14.
- Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate Bayesian computation. *PLOS Computational Biology*, 9(1):e1002803.
- Tang, S., Fernandez-Granda, C., Lannuzel, S., Bernstein, B., Lattanzi, R., Cloos, M., Knoll, F., and Assländer, J. (2018). Multicompartment magnetic resonance fingerprinting. *Inverse problems*, 34(9):094005.
- Tofts, P. (2005). *Quantitative MRI of the brain: measuring changes caused by disease*. John Wiley & Sons.
- Toint, P. L. (1997). Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints. *Mathematical Programming*, 77(3):69–94.

- Tokdar, S. T. and Kass, R. E. (2010). Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60.
- Ugarte, M. D., Adin, A., Goicoa, T., and Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research*, 23(6):507–530.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.
- Virtue, P., Stella, X. Y., and Lustig, M. (2017). Better than real: Complex-valued neural nets for mri fingerprinting. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3953–3957. IEEE.
- Vlaardingerbroek, M. T. and Boer, J. A. (2013). *Magnetic Resonance Imaging: Theory and Practice*. Springer.
- Wei, Q., Dobigeon, N., and Tournier, J.-Y. (2015). Bayesian fusion of multi-band images. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1117–1127.
- Weigel, M. (2015). Extended phase graphs: dephasing, rf pulses, and echoes-pure and simple. *Journal of Magnetic Resonance Imaging*, 41(2):266–295.
- Weishaupt, D., Köchli, V. D., and Marincek, B. (2008). *How does MRI work? An Introduction to the Physics and Function of Magnetic Resonance Imaging*. Springer.
- Wu, S., Xue, H., Wu, Y., and Wu, H. (2014). Variable selection for sparse high-dimensional nonlinear regression models by combining nonnegative garrote and sure independence screening. *Statistica Sinica*, 24(3):1365–1387.
- Wübbeler, G. and Elster, C. (2017). A large-scale optimization method using a sparse approximation of the hessian for magnetic resonance fingerprinting. *SIAM Journal on Imaging Sciences*, 10(3):979–1004.
- Xia, Y., Ji, Z., and Zhang, Y. (2016). Brain mri image segmentation based on learning local variational Gaussian mixture models. *Neurocomputing*, 204:189–197.
- Yang, M., Ma, D., Jiang, Y., Hamilton, J., Seiberlich, N., Griswold, M. A., and McGivney, D. (2018). Low rank approximation methods for mr fingerprinting with large scale dictionaries. *Magnetic Resonance in Medicine*, 79(4):2392–2400.
- Yu, A. C., Badve, C., Ponsky, L. E., Pahwa, S., Dastmalchian, S., Rogers, M., Jiang, Y., Margevicius, S., Schluchter, M., Tabayoyong, W., et al. (2017). Development of a combined mr fingerprinting and diffusion examination for prostate cancer. *Radiology*, 283(3):729–738.
- Zhao, B., Haldar, J. P., Liao, C., Ma, D., Jiang, Y., Griswold, M. A., Setsompop, K., and Wald, L. L. (2018). Optimal experiment design for magnetic resonance fingerprinting: Cramer-Rao bound meets spin dynamics. *IEEE Transactions on Medical Imaging*, 38(3):844–861.
- Zhao, B., Setsompop, K., Ye, H., Cauley, S. F., and Wald, L. L. (2016). Maximum likelihood reconstruction for magnetic resonance fingerprinting. *IEEE Transactions on Medical Imaging*, 35(8):1812–1823.