Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN FISICA - XXXII CICLO

Settore Concorsuale: 02/D1

Settore Scientifico Disciplinare: FIS/07

Advances in the Role of Quantitative NMR in Medicine: Deep Learning applied to MR Fingerprinting and Trabecular Bone Volume Fraction Estimation through Single-Sided NMR

Coordinatore di Dottorato:

Prof.ssa Silvia Arcelli

Presentata da:

Marco Barbieri

Supervisore:

Prof. Gastone Castellani

Co-supervisore:

Dott.ssa Claudia Testa

ESAME FINALE

Anno 2020

Abstract

Nuclear Magnetic Resonance (NMR) has been a powerful and widespread tool since its birth thanks to its flexibility in assessing properties of physical systems without being invasive and without using ionizing radiations. Although applications of NMR for medical purposes have rapidly developed since the introduction of MR imaging (MRI), most of the clinical protocols retrieve qualitative information about biological tissues. Being able to retrieve also quantitative information with NMR may be beneficial to identify biomarkers for understanding and describing the pathophysiology of complex diseases in many tissues. However, established quantitative MRI (qMRI) methods require long scan times that not only can represent more exposure to image artifacts and more discomfort for the patient, but they also increase the costs of MRI protocols. To improve the clinical feasibility of quantitative NMR, one can focus on optimizing qMRI protocols to increase data acquisition efficiency, i.e. minimizing the acquisition times and maximising the number of retrieved information. Alternatively, one can focus on the application of low-cost, portable and low maintenance NMR devices in the medical field, such as single-sided devices. This Ph.D thesis presents studies that aim to advance the role of quantitative NMR in medicine using the two directions stated above.

The first part of the thesis proposes a deep learning approach based on deep Fully Connected Networks, here simply referred to Neural Networks (NN), for pixel-wise MR parameter prediction task in Magnetic Resonance Fingerprinting (MRF), a fast multiparametric qMRI methodology, as a solution to overcome the curse of dimensionality affecting the gold standard dictionary approach. The key feature was the ability to effectively train the NN with only simulated data.

With this aim, it was assessed, by means of simulations, the benefits of the NN approach in overcoming the curse of dimensionality in respect of the dictionary approach. Preliminary robustness to undersampling artifacts was also assessed. Results showed that NN performance was at least as good as the dictionary-based approach in reconstructing MR parameter. The difference in performance increased with the number of estimated parameters, because the dictionary method suffers from the coarse resolution of the MR parameter space sampling.

Then the results were translated in vivo, and the goal was to modify the training pipeline

so that the NN would be accurate and robust to artifacts affecting real MRF signals, such as those arising from undersampling. Using an ablation approach, the NN was trained with different pipelines, using only simulated MRF data, and their performance were tested against the T_1 and T_2 maps generated with the gold standard dictionary approach. Results allowed to heuristically determine the optimal training strategy to train a NN able to predict T_1 and T_2 maps that are in strong agreement with those obtained with the dictionary approach.

The second part of the Ph.D thesis proposes a methodology to assess the trabecular bone-volume-to-total-volume (BV/TV) ratio, an important morphological parameter, using single-side NMR by means of NMR relaxometry measurements. Nowadays there are not well established methodologies to assess trabecular BV/TV that are suitable for wide screening campaigns of the population at risk of bone fractures related to diseases such as osteoporosis. Although we recently proposed a technique for BV/TV estimation based on a CPMG pulse sequence using single-sided NMR, the methodology had two main limitations. First, the presence of soft tissues other than trabecular bone that can be intercepted by the sensitive volume of the single-sided NMR could affect the estimation of the BV/TV parameter. Second, the methodology requires to use a signal acquired from a reference sample (bulk bone marrow) to estimate the BV/TV ratio. The aim of the second part of this Ph.D thesis was to overcome these limitations.

The first problem was addressed by weighting the signal acquisition by molecular diffusion. Experiments based on an *ad-hoc* designed Diffusion Weighted $T_1 - T_2$ correlation pulse sequence demonstrated the feasibility of the filtering. A 1D measurement was then established using a diffusion prepared CPMG acquisition. This modification resulted in a correct estimation of trabecular BV/TV in a sample where also muscle tissue was present within the sensitive volume of the single-sided device.

The problem of the need of using the signal coming from a reference signal for BV/TV estimation was then addressed. The limitation was overcome by exploiting the information about the T_2 relaxation times of the trabecular bone that comes from the CPMG decay analysis. Such analysis allows to separate the intra trabecular signal (mainly due to water bound to collagen) and the signal coming from the inter trabecular spaces (mainly due to bone marrow). A set of trabecular bone samples, cored from pig shoulders, were analysed with NMR and with micro computed tomography (micro-CT), the gold standard technique, to assess ground truth bone morphological parameters. Starting from the computation of intra and inter trabecular signal contributions, BV/TV and bone-surface-to-total-volume (BS/TV) ratios were estimated using theoretical models, which require an esteem of the intra trabecular porosity and of the mean trabecular radius (trabeculae are modelled as cylinders). Comparison between NMR and micro-CT showed high level of agreement for

BV/TV and high correlation with moderate agreement with BS/TV.

In conclusion, although the works reported in this Ph.D thesis requires further steps to be considered for clinical application, they may be considered as deliverables in the direction forward the advance of the role of quantitative NMR in medicine.

To my family

Acknowledgements

As I am used to say, there are very little things that one can accomplish alone, and they are even not guaranteed for the entire course of life. A not comprehensive list of activities one can possibly accomplish alone includes: breath, swallow, walking, taste and hearing. Although they are very important, they might not be the most exciting activities one can imagine. Well, a Ph.D programme will never be appended to this list. I cannot imagine how this journey would have been without the support of the people that walked me through the Ph.D, and those who, even if not directly related to my scientific research, were with me during this journey. This short chapter (really, too short!) is for all of you.

First, I want to thank my Ph.D supervisor, Prof. Gastone Castellani, that took the risk of accepting to guide me in this path three years ago. I really want to thank Dr. Claudia Testa, who guided me through the field of MRI to build everything that is included in this manuscript. I want to thank her for the uncountable talks about research directions and future perspectives. I've always found a person willing to hear, and ready to share her experience with the aim of allowing me to make decisions with a gaze at the whole picture. I also want to thanks Prof. Daniel Reomndini for all the guidance he provided me in relation to machine learning and deep learning tools, and for the productive collaboration in the field of data analysis.

Second, I thank Prof. Paola Fantazzini who was my teacher during my first NMR class, and my thesis supervisor for master of science degree. To her I owe my enthusiasm for everything is NMR related. She continued to share with me all her knowledge and expertise in the field of NMR relaxometry. A consistent part of this manuscript could have not been included without the effort and the guidance that she continued to provide to our group, despite she would be technically retired.

I want to thank my family, Manuela and Giorgio (mom and dad) to have been always supportive. I can just imagine all the doubts about my decision in pursuing a Ph.D, considering the job insecurity this word has always been meaning in Italy. Despite it, I have been always encouraged to follow what I though was right, and taking risk for that. I thank my twin brother Alberto, who also is pursuing a Ph.D, but in Philosophy. We shared good

and bad moments of this path, and it is wonderful to finish this jouorney almost together. Whatever the future will reserve to us, I know we will be there for each others.

I also want to thank all the member of the NMR group of Bologna. A giant thanks goes to Dr. Leonardo Brizi, a friend and colleague. I say this without rhetoric, your help was invaluable. Not only for the time you dedicated to me during my first times, but for all the conversations we had about every aspect of life. This really prevented this long path from being alienating. I thank Prof. Villiam Bortolotti, especially because every so and so he called me in my desk (He has always been the only one calling on my desk phone!). Every time I heard that ring, I would not even ask hello, I would just say his name. I thank Remi Kogon, the French proponent of the group, for your will to help every time I need a help.

I also want to thank Dr. Enrico Giampieri for always pointing out tools that without him I would have never thought about. A special thanks also goes to my friend Dr. Francesco Solera for the very appreciated help he provided me when I was doing my first steps into the field of deep learning.

I want to thank Prof. Brian Hargreaves for supervising me during my visiting period at Stanfrod (CA, USA). First of all, I want to thank him because he accepted me as visiting student, even if he barley knew me at all. I thank him for all the guidance he provided to me during those months, and for the care he demonstrated. In relation to my abroad period, I want to thank all the group members of the body MRI group of Stanford. You all made my experience full of colours and enjoyable. I want to thank you name by name. Thanks to: Prof. Garry Gold, Dr. Valentina Mazzoli, Dr. Akshay Chaudari, Dr. Feliks Kogan, Elka Rubin, Lauren Watkins, Philip Lee, Dr. Chaterin Morgan, Dr. Daehyun Yoon, Arjun Desai, Steffi Perkins, Alex Toews, Dr. Marianne Black, Yuxin Hu, Dr. Jianmin Yuan and Kate Young.

Finally, I thank all my friends. Albi (also my brother), Dario, Alle, Gaia, Menni that together form the "legendary" *Los Vacancieros*. My indoor football team mates: Cappe, Paul, Save, Manfre, Simo, Lory, Mirko, Edo and Nico.
My *old* university mates from Modena: Anna, Giulia, Bacca, Alle and Bibi. Bibi and Alle, I like how even if we are apart we've managed to enrich our friendships during these years. I want also to make a special thanks to Massimo, Soncio, Marianna, Laura, Dino, Pinni, Anna and Marco. A last special thanks to Pizza, a Ph.D student (on his way to be done!) and a friend that has shared with me the joy and the efforts of doing a Ph.D.

Contents

Introduction

Nuclear Magnetic Resonance (NMR) has been a powerful and widespread tool since the first works of Bloch and Purcell in 1946, thanks to its flexibility in assessing properties of physical systems without being invasive and without using ionizing radiations. Although every nucleus characterized by a non-zero total quantum spin number can be probed with NMR, ^{1}H nuclei are the common target because of their relatively high NMR sensitivity and natural abundance. Given a system of nuclei, a variety of properties can be probed with NMR, but the most fundamental are the longitudinal relaxation time T_1 and the transverse relaxation time T_2, which depend on the chemical and physical environments that ^{1}H nuclei perceive. For example, ^{1}H nuclei in water molecule confined within small compartments have T_1 and T_2 relaxation times that differ from relaxation times of bulk water. This dependence is exploited in medicine to investigate tissue properties such as myelin water fraction in the brain [101] and collagen bound water fraction in cortical bone [94]. More generally, the dependence of NMR parameters on the environment that the nuclei perceive, gives wide versatility to NMR, which has been used for diverse applications ranging from biology to chemistry and medicine.

Medicine has been one of the main field of application of NMR because of its potential to probe a variety of tissue properties without being invasive. Although usefulness of NMR in medicine applications was evident since the early years after its birth [35, 96], medical applications have rapidly developed since the invention of Magnetic Resonance Imaging (MRI), which allows to spatially resolve the NMR signal [71]. The high interest in NMR as imaging technique for medical applications lies on the fact that, differently from X-ray imaging techniques, NMR does not use ionizing radiations and pixels intensity is not restricted to a single physical property of the tissue, but it can depend on a multiplicity of parameters. Through the years, the diagnostic potential of MRI has been widely validated and it has been applied to many body districts. Despite its development, clinical application of MRI protocols mainly return qualitative information about tissues. MR images can be acquired with different pulse sequences, and they show anatomical features that are more or less enhanced by the type of NMR parameter that weights the signal evolution (T_1 or T_2 weighted,

for example). Although qualitative MRI has been shown to have high clinical impact on diagnosis, interest in quantifying values of NMR tissue parameters in each pixel has increased. The quantification may bring many benefits through a deeper biological knowledge of the diseases, so that diagnosis of such diseases and responses to potential treatments can proceed in a more satisfactory way. Being able to provide quantitative images can be helpful for assessing disease progression longitudinally and for cross-sectional studies. For example, the use of T_2 maps has been demonstrated to be promising to help in understanding many diseases, such as osteoarthritis in knees [86], osteoporosis in bones [94] and multiple sclerosis in the brain [77].

However, quantitative MRI (qMRI) methods typically provide information on a single parameter at a time, they require significant scan time, and are often highly sensitive to system imperfections. In the context of MRI, long scan times not only can represent more exposure to image artifacts (such as those caused by motion during acquisition) and an increased discomfort for the patient, but they also increase the costs of MRI protocols. Whole body MRI scanners use superconducting electromagnet to generate the polarizing magnetic field, which requires liquid nitrogen for the cooling. Hence, MRI scanners are high cost and maintenance. Overall, these characteristics limit the clinical application of quantitative MRI for medical purposes.

To make quantitative NMR clinically feasible, one can focus on optimizing qMRI protocols in such a way to increase data acquisition efficiency, i.e. minimizing the acquisition times and maximising the number of retrieved information. MRI research community has been widely working on this field for many body districts ranging from brain [36] to muscoloskeleton [64]. Among them, Magnetic Resonance Fingeriprinting (MRF) was firstly proposed in 2013 [75] as fast qMRI methodology to retrieve multiparametric maps in one-shot measurement exploiting variable sampling pattern and a template matching processing step based on a precomputed dictionary containing simulated theoretical signals. Interest in MRF has rapidly increased among the research community due to its fastness and promising applicability in a clinical setting, which, however, is not established yet.

Another way to bring quantitative NMR into clinical practise is to focus on the application of low-cost, portable and low maintenance NMR devices in the medical field. Single-sided NMR scanners, for example, use permanent magnet to generate the polarising field (that is a low-intensity field compared to whole-body MRI scanners) and they detect the NMR signal from a sensitive volume that is outside the magnet. Their helpfulness for assessing porous media, both in laboratory and in-situ, has been widely demonstrated in literature. Interest in using such devices for medical purpose is slowly increasing, and their use has been recently shown to be promising for characterizing breast [2] and bone [23, 24] tissues. However, more

steps are needed to consider clinical application of such devices.

This Ph.D thesis presents studies regarding NMR methodologies that aims to help the application of quantitative NMR to clinical medicine. This manuscript is a summary of part of the work that I carried out in the NMR group of the Department of Physics and Astronomy at the University of Bologna (Italy) during my Ph.D program. In relation to this program, I also spent seven months in the Body MRI group of the department of Radiology of the Stanford University (CA, USA), supervised by Prof. Brian Hargreaves. The work is organized in two parts composed of three chapters each, with a conclusive chapter that summarizes the main findings of the presented studies and future directions.

Part I: Circumventing the Curse of Dimensionality in Magnetic Resonance Fingerprinting through a Deep Learning Approach

A deep learning approach based on deep Fully Connected Networks for pixel-wise MR parameter prediction task in MRF is presented. The part is divided into three chapters.

Chapter 1 presents the research issues that guided the developing of the project. The curse of dimensionality that affects the standard dictionary approach is presented as limitation for the scalability of the MRF methodology when the number of NMR parameters to be retrieved increases. A background section is also presented to describe fundamental concepts of MRF and the related state of the art.

Chapter 2 presents a simulation study in which the Deep Learning application is set up. Simulated experiments are conducted to show the advantages in using the deep learning approach against the standard dictionary approach for parametric map reconstructions when the number of parameters to be retrieved increases. Different MRF pulse sequences are tested, and robustness to artifacts caused by undersampling the k-space is also discussed.

Chapter 3 focuses on the application of the deep learning approach set up in the previous chapter, to real acquired data. The main goal is to investigate which training pipeline, in terms of data augmentation and preprocessing steps, yields the best results in terms of robustness to artifacts caused by the undersampling. The real MRF data were acquired at the Stanford Radiological Sciences Laboratory (RSL), part of the Department of Radiology of the Stanford University (CA, USA), where I spent seven months during my Ph.D programme.

Part II: Evaluating the Bone Volume Fraction of Trabecualr Bone using NMR relaxometry through single-sided scanners

The main aim is to propose a technique to assess the bone volume fraction, an important bone morphological parameter, of trabecular bone using single-side NMR by means of NMR

relaxometry measurements. Although experiments presented are performed in laboratory, the long-term aim is to propose a low-cost NMR methodology that could be used to wide spread campaing to assess fracture risk associated with Osteoporosis disease. The part is dived into three chapters.

Chapter 4 introduces the problem at hand highlighting the benefits of assessing microstructural bone parameters for correct prediction of fracture risk in the Osteroporosis disease. A section is dedicated to describing the characteristics of the single-sided devices used to perform the experiments reported in the following chapters, i.e. the NMR-MOUSE and the NMR-MOLE. The last section of the chapter describes the first work of single-sided NMR applied to bone volume fraction estimation of trabecular bones [23]. The section highlights the main limitations of the proposed technique, and this parts introduces the following chapters which investiga how to overcome them.

Chapter 5 addresses the problem of detecting the signal from the trabecular bone marrow and suppressing the signals from other tissues (muscle and cartilage), which may be present within the sensitive volume of a single-sided NMR scanner. The goal is achieved by signal filtering exploiting molecular diffusion. Part of the material of this chapter has been recently published [7].

Chapter 6 investigates the feasibility of a new methodology to assess the micro-structural properties of trabecular bone using single-sided NMR scanners without the need of a marrow reference sample. The original methodology presented in [23] requires the comparison between the acquired signal from the trabecular bone and the signal coming from a reference sample of bone marrow. Since signal is not spatially resolved using a reference sample for in vivo appreciation is not ideal. Comparison between the bone volume fractions of a set of pig trabecular bone samples obtained using the new proposed NMR methodology and the bone volume fractions obtained using the gold standard micro computed tomography is presented to validate the technique. The work has been carried out in collaboration with the Rizzoli Orthopedic Institute (Bologna, Italy) where micro-CT acquisitions were performed. The NMR measurements with the NMR-MOUSE were performed at the Department of Physics and Astronomy of the University of Bologna (Italy).

Part I

Circumventing the Curse of Dimensionality in Magnetic Resonance Fingerprinting through a Deep Learning Approach

Chapter 1

Introduction

1.1 Outline of the problem

MRF is a fast qMRI methodology able to obtain multi-parametric maps in a one-shot measurement [75]; many applications of the technique have been investigated since its birth, ranging from brain imaging [73] to body MRI [29], [30].

The key concept of MRF is to apply a train of radio-frequency (RF) pulses whose flip angle (FA) and repetition time (TR) vary according to a pattern designed to differentiate the signal response for different tissues as much as possible, so that for each voxel a so-called fingerprint of the tissue is acquired. MRF usually employs a non-Cartesian k-space sampling scheme, using variable density spirals [59, 75] or radial waveforms [4, 32], to acquire data rapidly. The original methodology uses a nearest neighbour approach to match the experimental fingerprint with a precomputed dictionary of simulated signal evolutions using the dot product as similarity measure. The method has been demonstrated to produce accurate MR parameter maps even in presence of severe artifacts caused by the k-space undersampling acquisition [25, 59, 75]. This is because the dot product takes advantage of the incoherence between the artifacts caused by the undersampling and the MR signal. An exploding dictionary size, matching accuracy, robustness to undersampling artifacts are some of the main challenges to overcome before MRF can be accepted in a clinical setting. The more parameters are encoded into the pulse sequence simulations (T_1, T_2, B_0, T_2^*, B_1^+, etc...), the bigger the size of the precomputed dictionary has to be. Confounding factors such as B_1^+ field inhomogeneities are known to be a source of artefacts in MRI [106], and they have to be considered in MRF to improve parameter estimation [25, 76]. However, big dictionaries are hard to handle because they are costly both in memory usage efficiency, occupying up to 150 GB, and in computational time for the matching procedure [76]. Moreover, a priori generation of the entire high-resolution dictionary can result in high computational costs when, for

example, simulating the full 3D voxel profiles to perform single voxel proton spectroscopy with MRF [68]. Hence, dictionary-based MRF suffers from the curse of dimensionality [45], which limits the scalability of the MRF methodology. The problems arising from handling big dictionaries have driven the MRF community toward methods aiming to compress the size of the dictionaries by both applying Single Value Decomposition (SVD) to generate low-rank approximation of the fingerprinting signals, and reducing the entries of the dictionaries by using a polynomial fitting [81,82,110]. The improvements that one can achieve with these methods still depend on the size of the problem under investigation [13].

Another strategy to overcome the limitations of dictionary-based template matching is to take advantage of a Machine Learning algorithm that, after a supervised training procedure, can predict the MR parameters given the experimental fingerprint as input. The Neural Network (NN) algorithm is one possibility, as NNs have been demonstrated to be universal function approximators given enough training data and model complexity [56]. A sufficiently large NN in principle can learn the Inverse Transfer Function (ITF) that maps the acquired MRF signal into the MR parameter space. Moreover, once the NN is trained, the prediction operation is computationally more efficient with respect to template matching, because no exhaustive search across a dictionary has to be performed. More importantly, under optimal training conditions, it is able to predict MR parameters of unknown signals, and thereby limits quantization artifacts that can arise from a dictionary approach in which template matching approximates MR parameters to those present in the dictionary. Depending on the task and the architecture, NN can require thousands of training data items, which in the case of MRI can be both expensive and time-consuming. However, for application to MRF, the NN model can be trained using simulated data. Simulating input data is the original way in which dictionaries were produced in MRF, and the reliability of such dictionaries has been widely demonstrated [75]. Some studies have investigated the feasibility of applying Neural Networks to MRF, in conjunctions with numerical simulations [8, 53, 104], with phantoms, or with in-vivo acquisitions [6, 9, 33, 39, 46]. Among possible deep learning architectures, fully-connected neural networks (FCNs) have features that make them particularly suited to circumventing the curse of dimensionality in MRF, while maintaining most of the desirable features of the original approach. Convolutional neural networks (CNNs), for example, were designed to take advantage of local connection and correlations [72], and they require spatial structure to be present in the training data set. Hence, real acquired data are usually used instead of simulated data in MRF applications [6, 39]. However, if the amount of training data is not large enough it can generate overfitting, especially when a large number of parameters are to be retrieved. In contrast, FCNs used to perform pixel-wise processing without exploring neighbor pixels can easily be trained with simulated data. Indeed, because

the network does not take advantage of spatial information one can easily create a training data set with simulated MRF signals using a simple MR simulator. A method combining pixel-wise processing with a training procedure based on simulated data has two desirable features: the approach is inherently structure independent, since no spatial connections are taken into account; the lack of training data need never be a problem. For these reasons, this work focuses on the application of FCNs to MRF. From now on the NN acronym is used to indicate FCN, with the exception of section 1.3.2, where NN generally refers to the neural network algorithm. Despite the advantages associated with NN approaches, a critical point is to train the NN model in such a way to be both accurate and robust to artifacts affecting real MRF measurements, such as those resulting from undersampling the k-space. The training procedure is critical for this aim, because it helps to perform model regularization [48] during training.

Although the feasibility of NN approaches has been demonstrated in previous works [33, 53] these methods are still in their infancy. In particular, the capacity of NNs to circumvent the curse of dimensionality in MRF has not been carefully studied since the number of parameters to be retrieved by the NNs has always been limited to the minimum number of parameters required based on the MRF pulse sequence used. For example, in [33], where an IR Fast Imaging with Steady state Precession (IR-FISP) [59] sequence was used, only T_1 and T_2 were retrieved, and in [53], where a IR balanced Steady State Precession (IR-bSSFP) [75] sequence was used, only T_1, T_2 and B_0 were retrieved. It is important to study the scalability performance of NN approaches to assess their actual benefits over the dictionary approach. Moreover, little attention has been given to which pipeline should be used, in terms of training set design and data augmentation strategies during training, to allow the NN model to learn an accurate and robust ITF for a given MRF pulse sequence. For example, Virtue et al [104] randomly sampled the MR parameter space, whereas Hoppe et al [53] and Cohen et al [33] used a grid sampling. Moreover, different strategies have been applied to take noise into account, such as in [104] and [53], in which no noise was added in the training procedure, while in [33] white Gaussian noise with zero mean and 1% standard deviation was added during the training phase.

The goal of the first part of this Ph.D thesis is to set up a deep learning approach for MRF based on (1) a FCN architecture (2) using simulated data for training, and (3) comparing its performance with the standard dictionary approach in terms of capability of overcoming the curse of dimensinoality and robustness to undersampling artifacts.

This part is divided in two chapters. The primary aim of the first chapter is to set up the NN approach and investigate how its performance scales with the number of parameters to be retrieved in comparison with the standard dictionary approach. Secondly, the chapter aims

to preliminary investigate training strategies to allow a NN to learn an accurate and robust ITF for a given MRF pulse sequence. Given the nature of the objectives, the experiments reported were conducted with simulated data only.

In the second chapter the NN approach is tested on real acquired data and the training procedure is modified accordingly to promote robustness to undersampling artifacts.

1.2 Background: introduction to Magnetic Resonance Fingerprinting

1.2.1 General overview of the MRF framework

MRF is a quantitative MRI methodology that aims to retrieve multiple parameter maps from a one-shot measurement. As every quantitative MRI methodology, it is divided into two main steps: the MRI acquisition scheme, which involves pulse sequence design and k-space sampling, and the reconstruction algorithm, aiming to map the acquired signal $S(t)$ to the MR parameter space (T_1, T_2, B_0, B_1^+, etc...). The latter step is mathematically well described by the concept of the Inverse Transfer Function (ITF), and the goal of the algorithm is to perform an inverse mapping to map the acquired signal to the MR parameter space.

Considering the main differences between conventional qMRI and MRF is helpful to understand the MRF methodology. In Fig. 1.1 a schematization of conventional qMRI (left box) and qMRI based on MRF (right box) is presented. Let us consider the general problem of producing a MR parameter map from an MRI acquisition. Let us not consider the k-space sampling scheme and let us focus on the application of a pulse sequence to voxels containing different tissues.

Let us say $\mathbf{x}$ is the acquired MR signal, with $\mathbf{x} \in \mathcal{C}^n$, where n is the size of the acquired signal, and let us say $\mathbf{p} \in \mathcal{R}^m$ to be a m-ple in the MR parameter space, where m is the number of parameters such as T_1, T_2, B_0 ect.

Applying a pulse sequence to a system characterized by MR parameters, $\mathbf{p}$ generates a function $g(\mathbf{p})$ that maps $\mathbf{p}$ into the output signal $\mathbf{x}$. Quantitative mapping applies g^{-1} to the acquired signal $\mathbf{x}$ to obtain $\mathbf{p}$.

Inverse mapping is strongly related to pulse sequence design. In conventional quantitative mapping pulse sequences have usually been designed so that the evolution of the magnetization is described by an analytic function of the MR parameters, that is an analytic form of $g(\mathbf{p})$. Examples are the Inversion Recovery (IR) sequence for T_1 mapping, and Carr-Purcell-Meiboom-Gill (CPMG) sequence for T_2 mapping, in which a mono-exponential model or a multi-exponential model are used to describe g, and for the multi-exponential case an Inverse

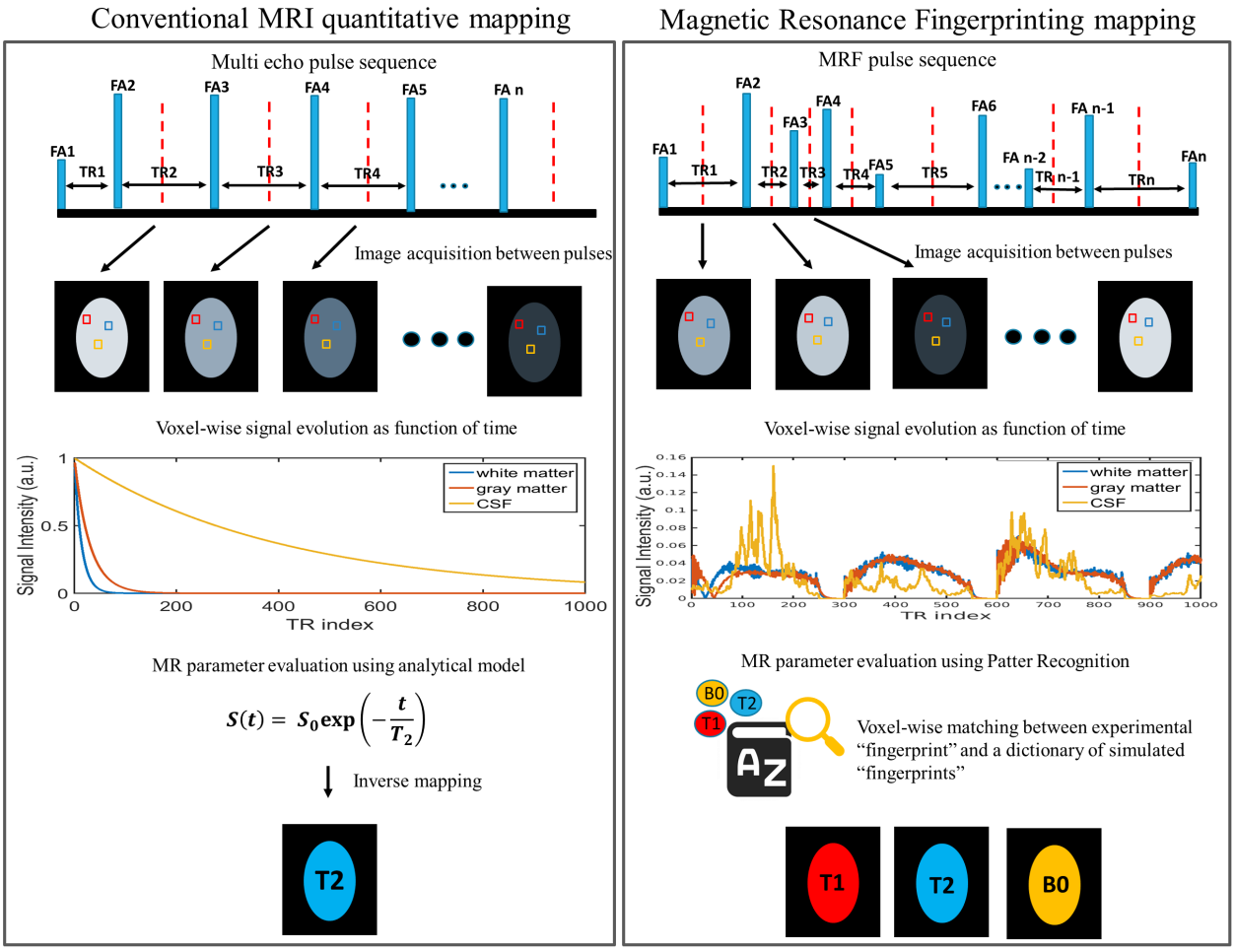


Figure 1.1: Schematization of the conventional approach (left box) to MR parameter mapping and the Magnetic Resonance fingerpinting approach (right box).

Laplace Transform is commonly used to obtain T_1 and T_2 distributions from the acquired signals [17, 20]. For example, left box in Fig. 1.1 schematize the framework for T_2 mapping using a multi-echo sequence (i.e. a CPMG sequence). With this way of performing parameter mapping only one parameter at a time can be acquired, so that if one wants to obtain T_1, T_2 and B_0 maps three pulse sequences have to be run. Moreover, fluctuations due to noise heavily impact on the accuracy of the fitting procedure, which requires to repeat measurements and averaging them to damp the noise. However, between repeated measurements, the equilibrium magnetization has to be restored, so that at least a minimum time is required before repeating the measurement. All of these are critical drawbacks in the context of MRI, that limits the clinical feasiblity of quantiative MRI.

In MRF (see right box in Fig. 1.1), the pulse sequence is designed to make the magnetization evolution sensitive to multiple parameters simultaneously using variable FA and TR patterns, providing a fingerprint-like signal evolution for a particular tissue. This makes hard to model the transfer function $g(\mathbf{p})$ (and consequently $g^{-1}(\mathbf{p})$) with analytic functions. In such a scenario, template matching with a precomputed dictionary can be used to overcome the lack of an analytic model for the ITF. The output of the transfer function can be simulated for a m-ple of MR parameters $\mathbf{p}_i$ using the Bloch equations, to obtain an estimated $\hat{g}(\mathbf{p}_i)$. Hence, for a given MRF pulse sequence, a dictionary is built by sampling a set of MR parameters $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_l\}$, where $\mathbf{p} \in \mathcal{R}^m$. Then, the simulations of the Transfer Function related to that pulse sequence are stored into a dictionary $\mathcal{D} = \{\hat{g}(\mathbf{p}_1), \hat{g}(\mathbf{p}_2), ..., \hat{g}(\mathbf{p}_l)\}$, where l is the length of the dictionary and $\hat{g}(\mathbf{p}_i) \in \mathcal{C}^n$. When real data are acquired from a voxel, the signal is described by $g(\mathbf{p}_{GT})$ where $\mathbf{p}_{GT}$ indicates the ground truth MR parameter vector characterizing the voxel tissue. The dictionary-based approach maps $g(\mathbf{p}_{GT})$ into $\mathcal{D}$ by computing the nearest neighbor between $g(\mathbf{p}_{GT})$ and each $\hat{g}(\mathbf{p}_i)$ present in $\mathcal{D}$ according to a similarity measure, which is usually the dot product. The MRF approach to quantitative mapping substantially improves limitations of conventional mapping. With one acquisition it is possible to obtain information about multiple parameters. Since MR parameters are acquired simultaneously, maps are already registered to each others and all of them have the same resolution. In conventional mapping, because different pulse sequences have different limits and acquisition times it is very common that even when multiple parameter maps are acquired they have different resolution and ideally they have to be co-registered.

1.2.2 MRF pulse sequences and k-space sampling

The MRF framework does not impose strict constrains on designing pulse sequences, as long the primary goal of differentiate tissues as much as possible is mantained. The design of the sequence determines the acquisition efficiency, the properties that can be estimated and

their accuracy, and the corresponding clinical application. Two main types of sequences have been developed for MRF, they are the IR-bSSFP [75], and the IR-FISP [59]. The first sequence is characterized by high SNR and it is highly sensitive to T_1, T_2 and off-resonances caused by B_0 field inhomogeneites, but it suffers from banding artifacts when a wide range of off-resonance spins are present in the Field of View (FOV). The IR-FISP sequence has been proposed to limit the sensitivity to off-resonances to eliminate banding artifacts adding an unbalanced gradient moment at the end of each TR. An unbalanced gradient produces a phase twist within each voxel, thereby it retains signal coherence and it reduces the sequence sensitivity to off-resonance. However, this comes at cost of lower SNR compared to the IR-bSSFP sequence, and it is not sensitive to B_0. All the sequences are usually inversion recovery-prepared, with this meaning that an inversion pulse (i.e. flip angle of π) is added at the beginning of the sequence to increase T_1 sensitivity. Generally FAs vary from 10 to 60 degrees, and TRs are in the order of 9-13 ms.

Since the MRF framework requires many images to be acquired in a single-shot measurement, with TR in the order of ten milliseconds, the image readout needs to be short, in the order of few ms. With this aim, the k-space is severely undersampled. Reconstructing the image from such undersampled k-space produces an image heavily corrupted by undersampling artifacts, where the type of the artifacts is characterized by the undersampling scheme used. Most MRF applications use a variable-density spiral trajectory to sample the k-space [59,75], but radial and Cartesian trajectory have been used as well [4, 32]. To ensure that all the k-space is eventually sampled during the acquisition, the k-space trajectory used for the first TR is then rotated in the following TR and so for the next TRs, so that during time all the k-space frequencies are acquired. In Fig. 1.2 examples of spiral k-space sampling trajectories are reported to illustrate how the k-space is sampled during time, how the reconstructed images look like and tissue signals appear.

1.2.3 Dictionary generation and pattern recognition algorithm

The MRF dictionary contains simulated signals generated using a pre-defined set of combinations of MR parameters. The MR parameters that can be included in the dictionary clearly depends on the sensitivity of the MRF pulse sequence used for acquisition. The range of MR parameters should cover the broadest possible range of physiologically relevant parameters. Generally, the dictionary has to be computed only once for each pulse sequence, and it can be applied to every acquisition that uses the same MRF sequence. This is not true for some cardiac applications to MRF, where the dictionary has to be dynamically adapted to the subject [52].

As last step of the MRF framework, each acquired fingerprint is matched to the dictio-

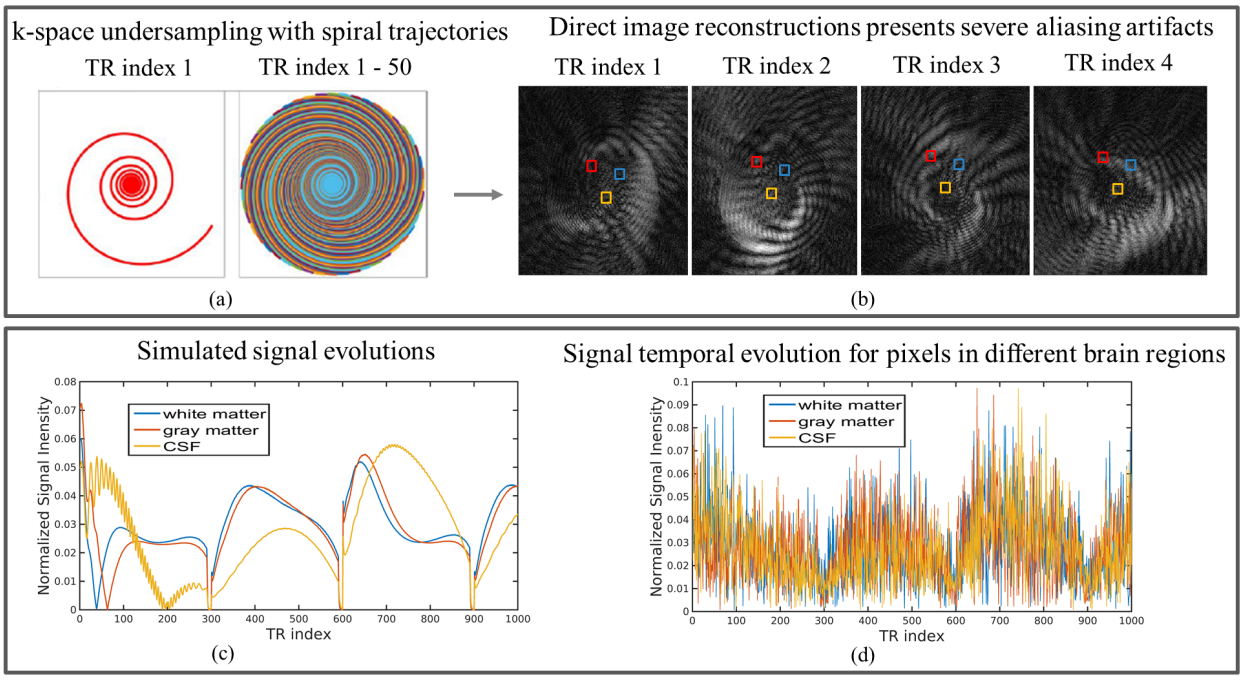


Figure 1.2: Summary of k-space sampling paradigm in MRF and his effect on the signal. (a) Example of spiral trajectories: during readout only one spiral interleave is acquired, and successive TRs have the same spiral interleave but rotated with respect to the previous TR. (b) Example of reconstructed images from k-space undersampled data. (c) Simulated signals for WM, GM and CSF tissues. (d) Real acquired signals for WM, GM and CSF tissues are heavily affected by undersampling.

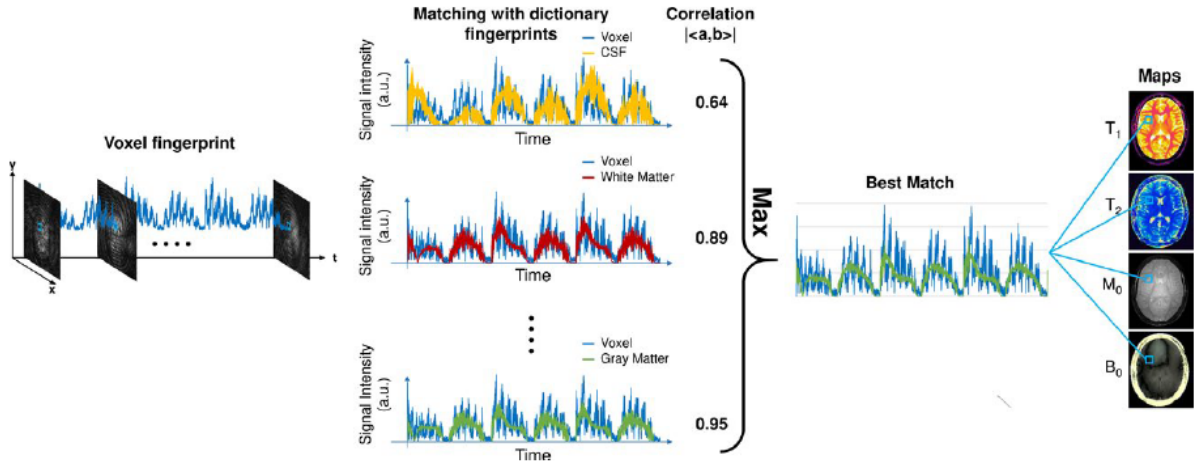


Figure 1.3: Original pattern recognition pipeline proposed in [75], based on template matching using the scalar product as metric of similarity. The Figure is taken from [13]

.

nary using a pattern recognition algorithm. The standard and most validated approach uses a nearest neighbor algorithm using the scalar product as metric of similarity. The scalar product has been demonstrated to be both accurate and very robust to undersampling artifacts, providing that those artifacts are spatio-temporally incoherent and incoherent with the fingerprint time course [99]. The algorithm, which is schematized in Fig. 1.3, evaluates the complex scalar product between the normalized voxel fingerprint and each dictionary entry. The dicionary entry that presents the maximum absolute value is selected as best match and the corresponding MR parameter set is assigned to that voxel.

1.3 Background: the curse of dimensionality in MRF

The main drawback of the dictionary based approach is that the number of dictionary entries scales rapidly with the number of MR parameters encoded into the dictionary, according to the curse of dimensionality [10]. Adding a new parameter to be retrieved with MRF without affecting the resolution of the dictionary leads to an exponential growth of the number of entries to be inserted into the dictionary.

Since both computational and memory usage limitations have to be taken into account, the number of entries needs to be kept under control, especially for clinical application where the parameter maps have to be generated rapidly. Simply reducing the number of entries increases the sparsity of the dictionary, which in turn can produce high biases in nearest neighbor algorithms [45], especially when they are applied to high dimensional data spaces, such as in MRF where usually signals are represented by 1000-dimensional vectors. Finding ways to speed-up or keep feasible the matching procedure in clinically acceptable times is therefore very important. Several methods have been proposed to overcome these limitations and they can be grouped in non-deep learning-based methods and deep learning-based methods.

1.3.1 Dictionary k-SVD Low-rank Approximation and Group Matching

The first algorithm for solving the problem related to a big dictionary was proposed by McGivney et al [82]. The method consists in performing a low-rank approximation of the dictionary using the SVD along the time dimension. Singular vectors are ranked according to their energy ratio, and only the first k of them are kept to create a k-dimensional subspace. The dictionary is then projected into this subspace and also the real acquired fingerprints so that the matching is performed in a lower-dimensional space, which speed up the process.

A formal description of k-SVD low-rank approximation is here given. Let us give a matrix of simulated MRF signals $\mathbf{D} \in \mathcal{R}^{n \times t}$, where n is the number of tissue entries and t is the number of time points. The SVD decomposition of such a matrix is $\mathbf{D} = \mathbf{USV}^\top$, where $\mathbf{U}$ is an $n \times n$ matrix and $\mathbf{V}$ is a $t \times t$ matrix, and their columns correspond to the left and right singular vectors of $\mathbf{D}$, respectively. $\mathbf{S}$ is a diagonal matrix containing the singular values. $\mathbf{P} = \mathbf{DV}$ is an $n \times t$ matrix that is the projection of $\mathbf{D}$ in the base of the right singular vectors. Then, only the first k columns of $\mathbf{P}$ are taken. In the work of McGivney et al values of k for MR fingerprints with 1000 time points were 25 for the IR-FISP sequence (3.4 speed-up factor) and 200 for IR-bSSFP sequence (4.8 spped-up factor).

Yang et al [110] proposed randomized SVD to bypass the need of computing the whole dictionary in advance, which can be itself a challenge when the number of MR parameters included in the dictionary increases. On top of this they proposed to take advantage of the smoothness of the MR parameter dimension in the randomized SVD space to further reduce the size and computation time of the dictionary. The tissue-property dimension of the randomized SVD space of a coarse dictionary is fitted to a polynomial hypersurface, which is then used to estimate the MR parameters at finer resolution.

Cauley et al [28] proposed a fast group matching approach to speed-up computational time for the matching procedure. Dictionary entries are grouped based on correlation, and a representative fingerprint signal for each group is calculated based on the mean of all the atoms within the group. Once the dictionary has been separated into groups the correlation between the voxel fingerprint and each group's representative fingerprint is firstly computed. Secondly, groups are pruned, based on relative or absolute correlation threshold and SVD low-rank approximation-based matching is applied on the atoms from the remaining groups to identify the voxel MR parameters. The speed-up factor found in the original work was of 70.

For all of these methods there is a trade-off between accuracy compared to the standard nearest-neighbor approach and speed-up factos. Moroever, speed-up factors are dependent on the MRF data (i.e. IR-FISP or IR-bSSFP data), dictonary size and reconstruction parameters (i.e. k in k-SVD and numbers of groups in group matching approach). Hence, they can vary across different applications.

1.3.2 Deep Learning approaches

NN approaches have been proposed to overcome limitations of MRF related to dictionary size. The rationale is that given a training data set $\mathcal{T} = \{\hat{g}(\mathbf{p}_1), \hat{g}(\mathbf{p}_2), ..., \hat{g}(\mathbf{p}_h)\}$, built by Bloch equation simulations, an NN approximation of the inverse transfer function $g^{-1}(S(t))$, defined $\hat{g}^{-1}(S(t))$, can be learned [56]. Once trained, an MRF voxel signal $S(t)$ is the input to

the NN model, which applies $\hat{g}^{-1}(S(t))$ to retrieve $\hat{\mathbf{p}}$, an estimate of $\mathbf{p}_{GT}$. Being a regression procedure, the model can correctly predict unknown examples, depending on how well it has been trained, avoiding the issues of the template-matching approach. Moreover, a strong mathematical rationale exists behind these applications. For example Hornik demonstrated that standard multilayer feedforward networks are capable of approximating any measurable function to any desired degree of accuracy [56].

Several works have investigated the application of neural networks to MR Fingerprinting, using different type of architectures ranging from CNNs to recurrent neural networks (RNNs).

Virtue et al. [104] train both a real and a complex-valued NN to estimate T_1, T_2 and B_0 off-resonances from a 500 time point fingerprint given as input, produced by an IR-bSSFP sequence. The work was carried out only with simulated data. The main focus of this paper was to investigate whether a complex-valued NN is better than a real-valued net. The MR parameter space was sampled using uniform random sampling. No noise was added during training, while white Gaussian noise with SNR = 40 dB was added to test performance.

Hoppe et al. [53] trained a CNN with 3 hidden layers to predict T_1 and T_2 parameters given a 3000-time point fingerprint as input, produced by an IR-FISP sequence. The NN was trained with simulated data. The MR parameter space was sampled using grid sampling. No noise was added, during either the training or the testing phases.

Cohen et al. [33] trained a neural network composed of 2 fully-connected hidden layers to predict T_1 and T_2 parameters. The pulse sequences used in this work were an optimized EPI-MRF sequence [34], composed of 25 pulses, and an IR-FISP sequence with a sliding-window reconstruction [27] to reduce undersampling artifacts. The NN was trained with simulated data, and the MR parameter space was sampled with uniform grid spacing. White Gaussian noise with zero mean and 1% standard deviation was added during the training procedure. The work presents tests both on simulated and on real data.

Balsiger et al. [6] trained a CNN to predict T_1, T_2 and PD using an optimized IR-FISP sequence given a 4D spatio-temporal patch as input. In particular, the authors wanted to take advantage of the neighboring pixels during reconstruction. They designed the NN to process 5×5 patches in order to predict the MR parameter of the fingerprint in the center of the patch. The training phase was carried out using real data: six brains were scanned with MRF as well with standard protocols to obtain ground truth maps. A leave-one-out cross-validation procedure was used to train and test performance.

Fang et al. [39] designed a two-step deep learning model which included: first, a feature extraction module using a FCN to perform dimensionality reduction on the time domain; second, a spatially-constrained quantification module using a CNN to directly estimate T_1 and T_2 MR parameter maps. Training was carried out using acquired in-vivo brain data,

while T_1 and T_2 maps obtained using the standard dictionary approach were used as ground truth maps. The experiment was conducted using the $IR-FISP$ pulse sequence

Chapter 2

Setting the Deep Learning Application by Means of Simulated Experiments

2.1 Introduction

In this chapter the NN approach is firstly set up and tested by means of simulations. The ability of NNs to generalize in relation to the method used to sample parameter space, either uniform random sampling or grid sampling, a factor of known importance [100], [90] is investigated. Noise robustness is investigated by using different strategies to add white Gaussian noise to MRF training signals. To show the generalization ability of the NN approach and how it scales with the number of predicted MR parameters, in comparison with the dictionary approach, the NN has been applied to different MRF pulse sequences. They are: the IR-FISP, its variant for B_1^+ estimation [25], the IR-bSSFP and a modification of it to simultaneously estimate T_1, T_2, B_0 off-resonances and B_1^+ field inhomogeneities. A numerical brain phantom was used to assess the performance of the NN and the standard dictionary approaches in reconstructing MR parameter maps. Moreover, robustness to undersampling was studied by using different degrees of acceleration factor using radial sampling.

2.2 Methods

2.2.1 MRF pulse sequences and simulations

The simulations were performed using four MRF pulse sequences: the MRF IR-FISP pulse sequence as described in [59] that encodes T_1 and T_2, and its variant, IR-FISP B_1, which accounts for B_1 inhomogeneity by adding abrupt flip angle changes as described in [25]; the MRF IR-bSSFP as described in the original MRF article [75], which encodes T_1, T_2 and

B_0 off resonance, and a modification of it, IR-bSSFP B_1, to take B_1 field inhomogeneities into account. The latter sequence adds abrupt changes to the end of the IR-bSSFP sequence, following the same criterion used for the IR-FISP B_1 sequence. Flip angle and time repetition patterns are depicted in Fig. 2.1.

All sequence simulations were carried out using MATLAB (MathWorks). The Bloch equations were used for IR-bSSFP-type sequences, while the Extended Phase Graph (EPG) algorithm [61] was used for IR-FISP-type sequence simulations.

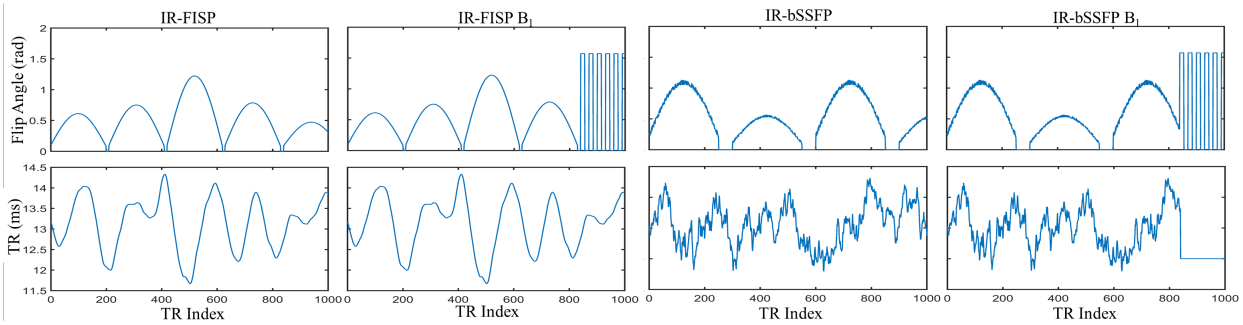


Figure 2.1: MRF pulse sequences used to generate the synthetic signals.

2.2.2 Noise

The noise affecting the MRF signal was modeled as white Gaussian noise, whose variance is expressed in terms of Signal to Noise Ratio (SNR). SNR has been defined in two ways as reported in equations 2.1 and 2.2. In Equation 2.1, P_{signal} and P_{noise} represent the average power of the MRF signal and the noise respectively, and $E(signal)$ indicates the expectation value of the MRF signal. Equation 2.2 is used when noise is added to a brain numerical phantom. In this case, following [112] A_{signal} indicates the average signal intensity within a region of white matter in the first image of the MRF time series, while σ_{noise} represents the standard deviation of the noise.

Equation 2.2 expresses SNR in decibels (dB), but in this work all SNRs are expressed in linear scale by applying the proper conversion. The definition of SNR used during the different experiments of this work has been indicated explicitly in the corresponding sections.

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{E(signal)}{\sigma^2_{noise}} \tag{2.1}$$

$$SNR = 20 \times \log_{10}\left(\frac{A_{signal}}{\sigma_{noise}}\right) \tag{2.2}$$

Only white Gaussian noise was used as a noise model in the experiments reported in this

chapter, since it is the most common noise model for NMR simulations, and it is widely used in MRF literature [31], [112], [97], [33]. As the simulated MRF signal is a complex signal, identically distributed Gaussian noise was added both to the real and imaginary parts. This indeed simulates two-channel acquisition, and preserves the Rician distribution when the magnitude of the signal is considered [50].

2.2.3 Fully Connected Neural Network models

The Deep Neural Network application was developed using the Python package Keras with the TensorFlow [1] back-end.

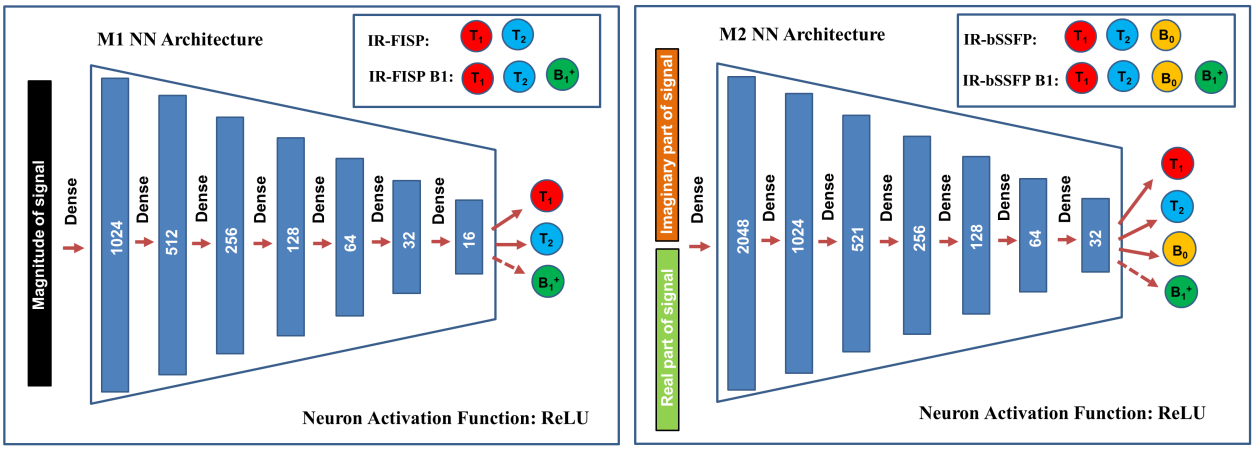


Figure 2.2: NN architectures for model M1 (left box) and model M2 (right box). The two models are fully connected NNs with the Rectified Linear Unit (ReLU) used as a neuron activation function for hidden layers, while a linear activation function has been used for the output layer. Blue boxes represents layers and the numbers inside indicate the number of neurons. Model M1 has been used for IR-FISP and IR-FISP B_1 pulse sequences, whereas model M2 has been used for IR-bSSFP and IR-bSSFP B_1 sequences.

Two NN models have been defined: one architecture to process IR-FISP data, and another architecture to process IR-bSSFP data. Both of them are feedforward nets with 9 fully connected layers, and the Rectified Linear Unit (ReLU) has been used as the activation function for the neurons in the first 8 layers, while a linear activation function has been chosen for the output layer. The NN architectures are reported in Fig.2.2. M1 takes the magnitude of the complex MR signal produced by an IR-FISP or IR-FISP B_1 sequence as input, while M2, which is used to estimate B_0, takes as input the concatenated real and imaginary parts of the complex MR signal produced by the IR-bSSFP or IR-bSSFP B_1 sequences. The output layer size matches the number of MR parameters to be estimated. These are: T_1 and T_2 for IR-FISP; T_1, T_2 and B_1^+ for IR-FISP B_1; T_1, T_2 and B_0 for IR-bSSFP and T_1, T_2, B_0 and B_1^+ for IR-bSSFP B_1, where B_0 indicates stationary field off-resonances,

and B_1^+ indicates B_1 excitation field inhomogeneities expressed in terms of a correction ratio, for which 1 represents the nominal flip angle.

Preliminary experiments, not presented in the current work, in which different network sizes, batch sizes and learning rates were tested, have guided the architectures selected. However, it is useful to highlight some of the motivations that have driven the design of the NN model architectures described above. A bottleneck shape was selected for two main reasons. The first is to avoid an exploding number of model parameters: many layers lead to overfitting, since fully connected layers with large number of neurons rapidly increase the number of model parameters. The second is that a bottleneck shape forces the network to encode more meaningful representations the deeper is the layer considered [72]. Hence, fixed the number of hidden layers to seven, the size of the first hidden layer was selected to have around the same number of entries as the expected input, which is a 1D vector of length 1000 for model M1, and in each successive layer the number of neurons was halved. Further trials were performed by doubling all the number of neurons in every layer and by halving the number of neurons per layer. We saw that increasing the NN size did not significantly increase the performance of the NN model, whereas halving the size of the NN decreased its performance. The same criterion was used for model M2. Hence, the architecture used for IR-bSSFPs sequences works well even for IR-FISPs sequences, whereas the contrary does not hold. However, because computational time decreases as the size of the network decreases, the smallest architecture has been selected for IR-FISPs sequences.

The training procedure was supervised, using the Mean Squared Error (MSE) between NN estimated parameters and ground truth parameters as loss function, and the Adam algorithm [63] was used for model weight optimization. In particular, 500 epochs with 1000 gradient steps for each epoch were used, with a fixed batch size of 500. Initial learning rates of 3×10^{-5}, for model M1, and 1×10^{-4}, for model M2, were used. The application was run on a cluster with 16 dual-core CPUs.

2.2.4 Training strategies: training sets, test sets and data augmentation

A very important step in building an NN application is the training procedure. While the selection of the NN architecture sets the model complexity, the goal of the training procedure is to perform model regularization [48] to promote the learning of the target function while preventing overfitting on unseen data. Training data set creation [100], [90] and data augmentation through added noise [103] are important forms of model regularization especially when data can be simulated, as in the case of MRF. The sampling and data

Model	Sequence	Set Label	T_1 (ms)	T_2 (ms)	B_0 (Hz)	B_1^+	Count
M1	IR-FISP	R1	[10 ÷ 4000]	[1 ÷ 3000]	0	1	75 555
		G1	[10 ÷ 4000]*	[1 ÷ 3000]*	0	1	
M1	IR-FISP B_1	R2	[10 ÷ 4000]	[1 ÷ 3000]	0	[0.5 ÷ 1.5]	164 475
		G2	[10 ÷ 4000]**	[1 ÷ 3000]**	0	[0.5 ÷ 1.5]**	
M2	IR-bSSFP	R3	[10 ÷ 4000]	[1 ÷ 3000]	[-400 ÷ 400]	1	175 041
		G3	[10 ÷ 4000]***	[1 ÷ 3000]]***	[-400 ÷ 400]***	1	
M2	IR-bSSFP B_1	R4	[10 ÷ 4000]	[1 ÷ 3000]	[-400 ÷ 400]	[0.5 ÷ 1.5]	396 550
		G4	[10 ÷ 4000]****	[1 ÷ 3000]****	[-400 ÷ 400]****	[0.5 ÷ 1.5]****	

Table 2.1

Summary of the training sets used to train the NN models. For G labeled sets, the parameter spaces were sampled using the following grids, imposing the constraint $T_2 < T_1$. *) T_1 and T_2 were incremented with steps of 10 ms; **) T_1 = [10:10:800, 850:50:1000, 1100:100:2000, 2500:500:4000] ms, T_2 = [1, 10:10:300, 350:50:1000, 600:100:1000, 1500:500:3000] ms and B_1^+ = [0.5:0.02:1.5]; ***) T_1 = [10, 20:20:500, 600:50:1000, 1100:100:2000, 2500:500:4000] ms, T_2 = [1, 10:10:500, 550:50:1000, 600:100:1000, 1500:500:3000] ms and B_0 = [-400:50:150, -100:10:-70, -60:2:60, 70:10:100, 150:50:400] Hz; ****) T_1 = [10, 20:20:300, 350:50:500, 600:100:1000, 1250:250:2000, 2500:500:4000], T_2 = [1, 10:10:300, 350:50:1000, 600:100:1000, 1500:500:3000] ms. B_0 = [-400:50:150, -100:10:-70, -60:2:60, 70:10:100, 150:50:400] Hz and B_1^+ = [0.5:0.1:1.5].

augmentation strategies are described below.

Parameter space sampling, random uniform and grid sampling

To assess how well the network models learn the ITF depending on the training set distribution, two training sets of the same size were generated for each pulse sequence: a random uniform set, in which parameter space is sampled using a random uniform distribution; a gridded set, in which parameter space is sampled with a variable mesh grid.

Table 2.1 summarizes the training set characteristics, where label R refers to random sampling, whereas label G refers to grid sampling.

Data augmentation and preprocessing

Three data augmentation strategies were tested, by training the NN models with different noise adding procedures:

- *W/O Noise*: using no data augmentation during training, which means that only noise-free examples are fed to the NN models, as in reference [104];

- *Fixed variance*: feeding the NN with noisy inputs affected by noise with 1% standard deviation, equal to a variance of 10^{-4}, as in reference [33]. This means that the network during the training sees just one fixed SNR given a set of MR parameters;

- *Variable variance*: feeding the NN with inputs affected by noise with different variances, expressed in terms of SNRs. For each training batch, a vector of SNRs was generated by randomly sampling the SNR values in the range 2 to 100, so that the batch contains data with different SNRs, and the variances of the noise added to the batch examples are estimated using equation 2.1.

The data augmentation step was done on-line. Hence, given a training set, it was unnecessary to store any new data either in hard or in flash memories to carry out the three data augmentation strategies. After the data augmentation step, each input was then normalized to unit norm.

For each model, the performance in predicting an MR parameter was estimated in terms of Mean Absolute Percentage Error (MAPE) and Mean Percentage Error (MPE) of the predicted parameter, evaluated on a test set composed by 30000 fingerprints generated using the same pulse sequence used for the training procedure, and sampling parameter space with a random uniform distribution. To test noise robustness, the prediction procedure described was repeated giving each model data with different SNRs = [3, 5, 10, 20, 30, 40, 50, 60, 70 ,80, 90, 100]. Moreover, to assess the variance of the MAPE and MPE as a function of the SNR, for each noise level the operation was repeated 10 times. The MAPE and MPE standard deviations were considered as a measure of variability. The MAPE was computed following equation (2.3), where $\widehat{p_k}$ is the estimated parameter value, while p_k is the ground truth parameter value and N is the number of examples considered. The MPE was computed according to equation 2.2.

Other error measurements used in this chapter are the Absolute Error (AE) and the Root Mean Squared Error (RMSE) defined in equations 2.4, 2.5 and 2.6 respectively. Lin's concordance coefficient (ρ_c) [74] has been also used as quantitative measure of agreement.

$$MAPE(\%) = \sum_{k=1}^{N} |\frac{\widehat{p_k} - p_k}{p_k}| \times 100 \tag{2.3}$$

$$MPE(\%) = \sum_{k=1}^{N} \frac{\widehat{p_k} - p_k}{p_k} \times 100 \tag{2.4}$$

$$AE = \widehat{p_k} - p_k \tag{2.5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\widehat{p_k} - p_k)^2} \tag{2.6}$$

2.2.5 Numerical Brain Phantom Simulations

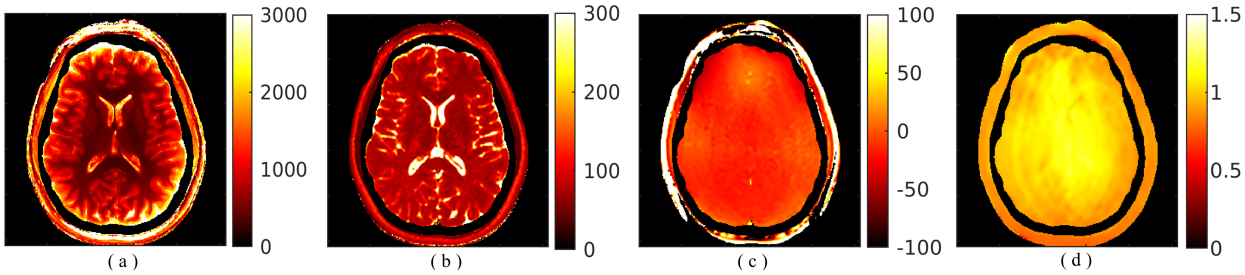


Figure 2.3: Quantitative brain maps computed from the Multi-Modal MRI Reproducibility Resource: (a) T_1 (ms), where T_1 values from 3000 ms to 4000 ms are shown as equal to 3000 ms; (b) T_2 (ms), where T_2 values from 300 ms to 2500 ms are shown as equal to 300 ms; (c) B_0 (Hz), where B_0 values less than |500| Hz are shown as equal to |100| Hz ; (d) B_1^+.

Realistic T_1, T_2, B_0 and B_1^+ maps where obtained by processing real acquisitions downloaded from the Multi-Modal MRI Reproducibility Resource repository. Landman et al [70] performed MRI acquisitions with standard quantitative protocols at 3 T: variable flip angle (VFA) imaging for T_1 mapping; double echo time imaging for T_2 mapping; two sequential 2D gradient echo with different echo times for B_0 mapping and Actual Flip-Angle Imaging (AFI) for B_1^+ mapping. The in-vivo brain images acquired with these protocols were then processed with MATLAB to obtain the quantitative maps. In particular, the qMRLab software [26] was used to process the VFA applying B_1^+ correction, since B_1^+ is known to be a confounding factor for the estimation of T_1 in VFA imaging [22]. Once obtained, these quantitative maps, reported in Fig. 2.3, were used as ground truth to simulate the MRF acquisition at the single pixel level, using the four pulse sequences described in section 2.2.1. Complex white Gaussian noise was added to simulated data using the SNR defined in equation 2.2.

It is worth pointing out that all MR parameters have been encoded in the MRF simulations. Hence, the B_1^+ map was taken into account for all four sequences; in the IR-FISP and IR-FISP B_1, B_0 was not be considered because it was not encoded by the sequence, while in the IR-bSSFP sequences B_0 is encoded and thus it has been considered in the simulations.

For each pulse sequence, the parameter maps were reconstructed by processing MRF data with both the trained NN models and the usual dictionary method based on the dot product. For this latter algorithm, the data sets G1, G2, G3 and G4 were used as dictionaries for the corresponding pulse sequences. Moreover, four dictionaries with around 400000 entries each

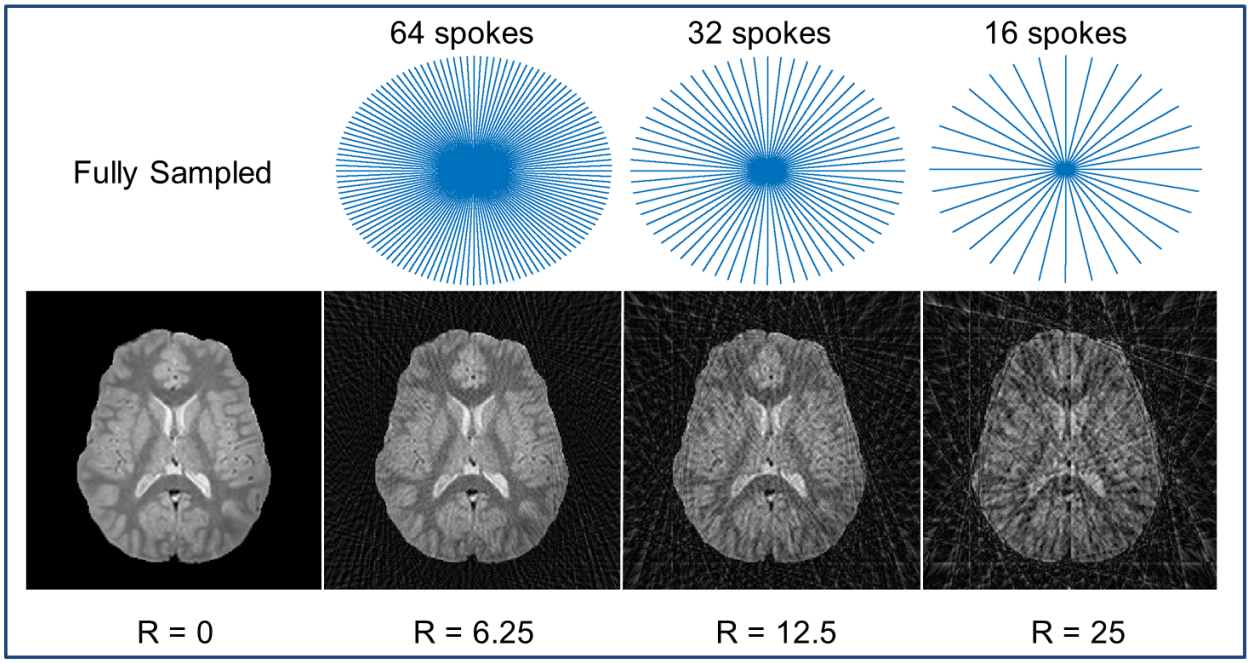
have been created to test the performance of the dictionary approach with dictionaries with higher resolution than the data set G1, G2, G3 and G4. MAPE and RMSE have been used as global measures of reconstruction quality, whereas the AE is used as a local measure. To have a complete overview of the performance, all three errors must be assessed. MAPE gives information about the mean relative error of parameter estimation, which has the advantage of giving an immediate sense of global performance, but it can be misleading when very small values, such as B_0 off-resonances, are taken into account because small absolute errors give high relative errors. In such cases, a more reliable global measure is the RMSE, which expresses a global error in the same unit of measurement as the parameter considered. It is also interesting to check a measure of the local error such as the AE, because it can be used to assess whether the error has a spatially uncorrelated distribution in the image or whether it shows spatial structure.

MRF undersampled acquisitions have been simulated using radial k-space sampling with different degrees of acceleration factor R that, for an acquired matrix of size $M \times M$, is defined according to equation 2.7:

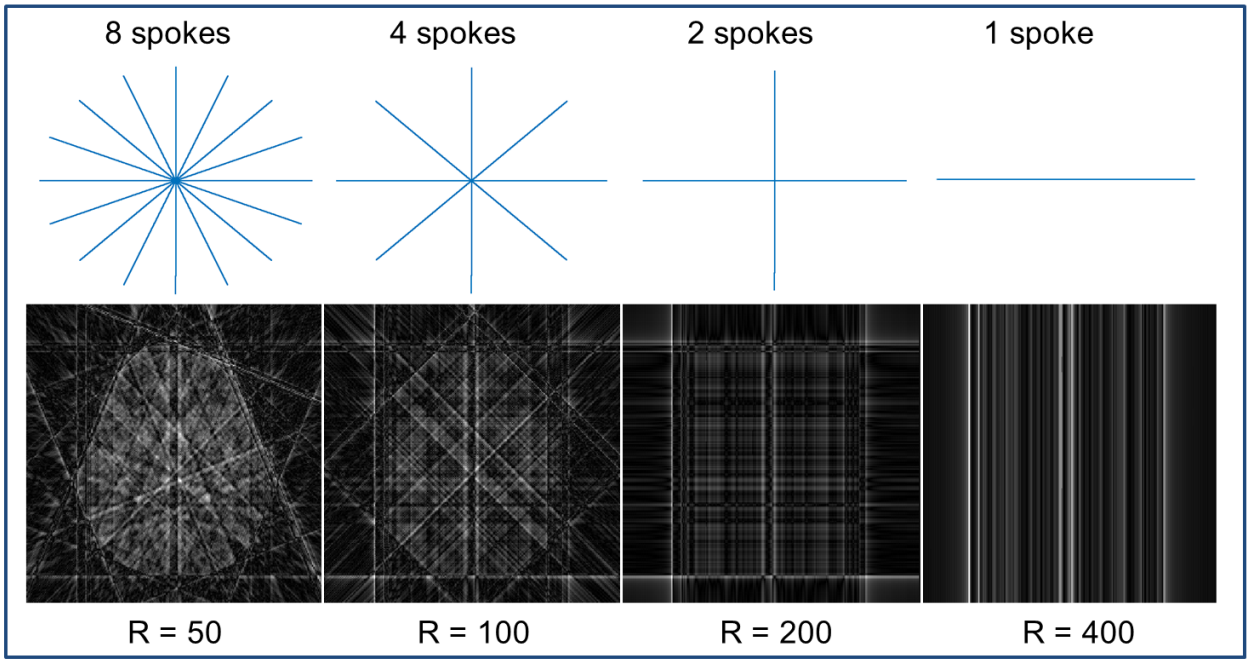
$$R = \frac{M \times \pi}{2N} \tag{2.7}$$

where N is the number of radial spokes used to sample k-space. The numerical brain phantom has a matrix size of 256 x 256, hence in this case M = 256. The performance in reconstructing the GT parameter maps using the NN and the dictionary approaches was evaluated for R = [0, 6.25, 12.5, 25, 50, 100, 200, 400]. Between different TRs, the radial trajectory is rotated with a golden angle increment [109]. Since the aim of the experiment is to test robustness against artifacts introduced only by the undersampling, no white Gaussian noise was added during the simulations. The IR-FISP B_1 and the IR-bSSFP sequences were used for this experiment. It is worth pointing out that in this case for the IR-bSSFP sequence B_1^+ has not been taken into account during simulations to avoid errors introduced by not estimating B_1^+. Moreover, high-resolution dictionaries (of around 400000 entries) were used to avoid errors introduced by dictionary quantization. Part of the Matlab code used for these simulations was taken from the freely available MRF reconstruction tool (`https://bitbucket.org/asslaender/nyu_mrf_recon`), described in more detail in [4]. For each TR, the fully sampled MRF image is retrospectively undersampled and then reconstructed using Jeffrey Fessler's NUFFT implementation with min-max interpolation [43].

In Fig. 2.4 examples of image reconstruction are reported for the different acceleration factors used. An example of the noise introduced in the MRF signal only due to undersampling is reported in Fig 2.5.



(a)



(b)

Figure 2.4: Imaging frame # 1 (i.e. corresponding to the first TR) for the simulated acquisition of the numerical brain phantom using the MRF IR-FISP B_1 sequence for different acceleration factors. Aliasing artifacts increase when R increases.

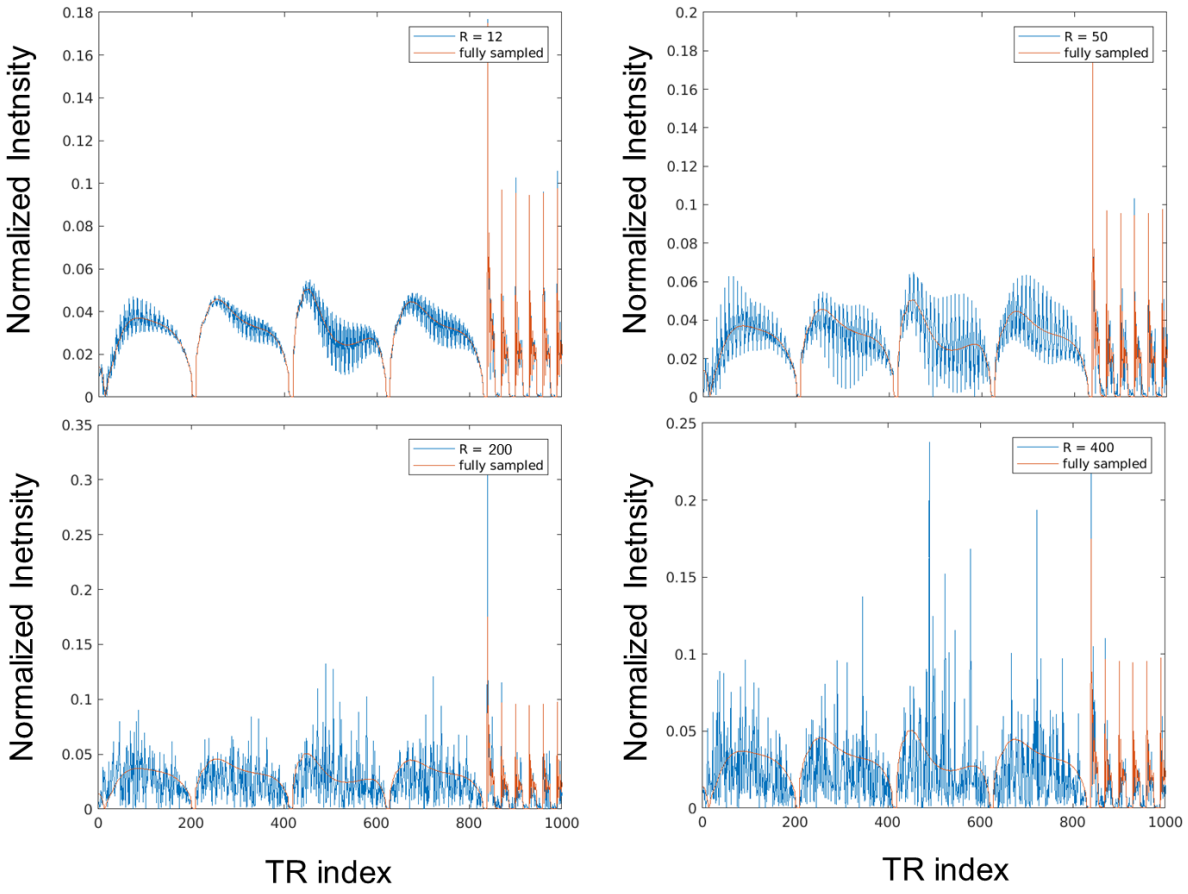


Figure 2.5: Example of a MRF IR-FISP B_1 signal for a pixel in the white matter region of the numerical brain phantom for different degrees of undersampling.

2.3 Results and Discussion

2.3.1 Training data distribution: Grid Vs Random sampling

The results of the training step of the NN models, fed with noiseless training examples and sampled using random and grid sampling, are summarized in Fig. 2.6, where training and test losses are plotted as a function of the training epoch.

For model M1, which processes IR-FISP-type data, there is a strong overlap of the training and test losses as clearly visible in Fig. 2.6 (a) and (b), indicating a good overall training phase, with no overfitting using either grid or random sampling. In contrast, for model M2, which processes IR-bSSFP-type data, the training losses using grid sampling are about one order of magnitude lower than the random counterparts, but test losses do not improve as shown in Fig. 2.6 (c) and (d), indicating that NN is overfitting the training data set. Hence, model M2 trained with grid-spaced sampling overfits the training set.

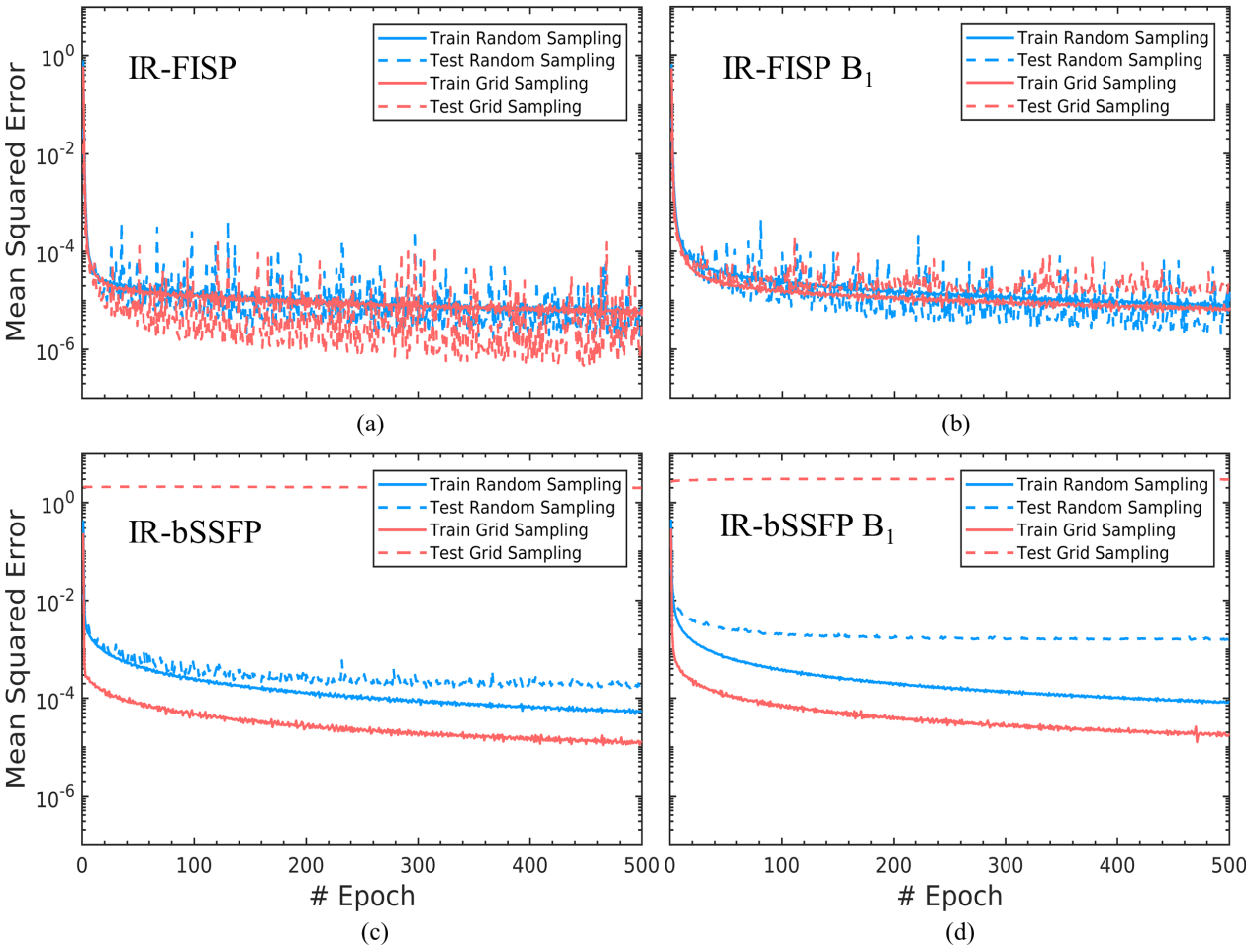


Figure 2.6: Comparison of the training and test loss functions for the four NN models trained with random sampled and grid sampled training data sets.

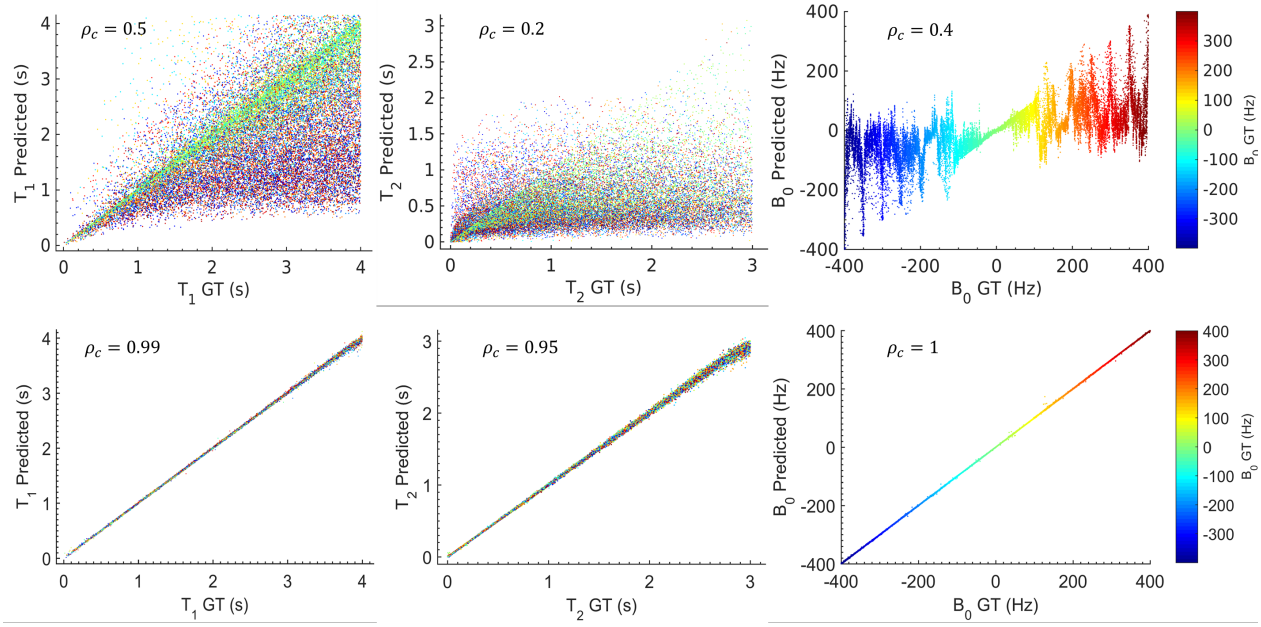


Figure 2.7: Predicted MR parameters against Ground Truth (GT) parameters using NN model M2 trained with IR-bSSFP MRF examples taken from sets G3 (grid sampling, top row) and R3 (random sampling, bottom row). Lin's concordance coefficients are also reported as quantitative measures of agreement. Note that the reported predictions refer to test data without added noise.

In Fig. 2.7 an example of the difference in performance between grid and random sampling for IR-bSSFP sequence is shown. Lin's concordance coefficient between predicted and GT values is reported as a quantitative measure of agreement. When the NN model is trained with a grid sampled training set, the NN does not generalize well for parameters that were not considered during training (ρ_c always less than or equal to 0.5). However, the B_0 parameter is well estimated when $B_0 = [-60 \div 60]$ Hz (sampling step was 2 Hz). Moreover, T_1 generally shows good agreement when B_0 is restricted to this range (green color labeled in the top left plot of Fig. 2.7). This is a hint that the overfitting problem with the grid is present when the grid does not have high enough sampling resolution. In contrast, when the same NN model is trained with a random sampled training set, the NN does generalize well (ρ_c equals or is greater than 0.95 for all the MR parameters). Considering what is shown above, it is possible to give an estimate of the number of examples needed to efficiently train model M2 using grid sampling without the overfitting problem. Using 10 ms resolution both for T_1 and T_2 axes, and 2 Hz resolution for B_0 off resonance axis, and using the same MR parameter ranges considered in the main manuscript; the resulting number of training examples would be 30 222 000, which would require 240 GB of memory for storing (considering signals of length 1000). The number is evaluated considering the constraint $T_1 \geqslant T_2$.

Summarizing, although grid sampling is a reasonable way to built dictionaries, it is not well suited to train NN for IR-bSSFP type sequences, since this method introduces a strong bias due to its regularity, and this leads the NN to overfit if the grid resolution is not high enough. This result is in accordance with the NN literature, where it has been shown that grid sampling normally requires many more training examples than random sampling to reach comparable performance [90] [97].

2.3.2 Noise robustness: learning less to learn better

The training and test losses of the NN models trained with different data augmentation strategies are reported in Fig. 2.8. It is worth noticing that for each model, the test loss, where the MSE is used as loss function, is checked at the end of each epoch by presenting examples never seen but with the same level of noise addition as that used in the training step. The results shown in Fig. 2.8 demonstrate that training is more effective when the *W/O noise* data augmentation strategy is used, since both training and test losses reach the lowest values at the end of the training. The worst condition for both training and test losses, at the end of the 500 training epochs, is when the *Variable variance* strategy is used.

Moreover, looking at the performance gap between training and test losses, model M1 does not show any differences regardless of the data augmentation strategy used, while model M2 always has a gap, indicating overfitting. Although a difference between training and test losses is expected, in this case we are testing the NNs with data coming from the same distribution, even if the precise representations were not seen during the training phase, so one might expect training and test losses to decrease in a similar fashion. Behavior like the one shown here means that model M2 is overfitting the training sets.

Models M1 and M2 handle two different MRF pulse sequences. Moreover, model M1 is trained with the magnitude of the MRF signals, while model M2 takes the concatenated real and imaginary parts of the MRF signal as input. To assess if the gap between training and test losses can be due to the latter difference, models M1 and M2 were also trained using the other type of input, i.e. concatenated real and imaginary part of IR-FISP signals for model M1, and the magnitude of IR-bSSFP signals for model M2. The results, reported in Fig. 2.9, excluded this hypothesis. Hence this gap is likely due to the size of R3 and R4 training sets. A straightforward solution is to increase the training set size. To test the goodness of this reasoning, a data set of 1000000 MRF examples was created using the IR-bSSFP B_1 sequence (using uniform random sampling). The 97% of it was used as a training set, and the 3% as a test set for checking test loss during the training phase, which was performed with the *Variable variance* data augmentation strategy. In Figure 2.10 the comparison between the training and test losses evaluated during the training of NN model M2 with 396 550 and 970

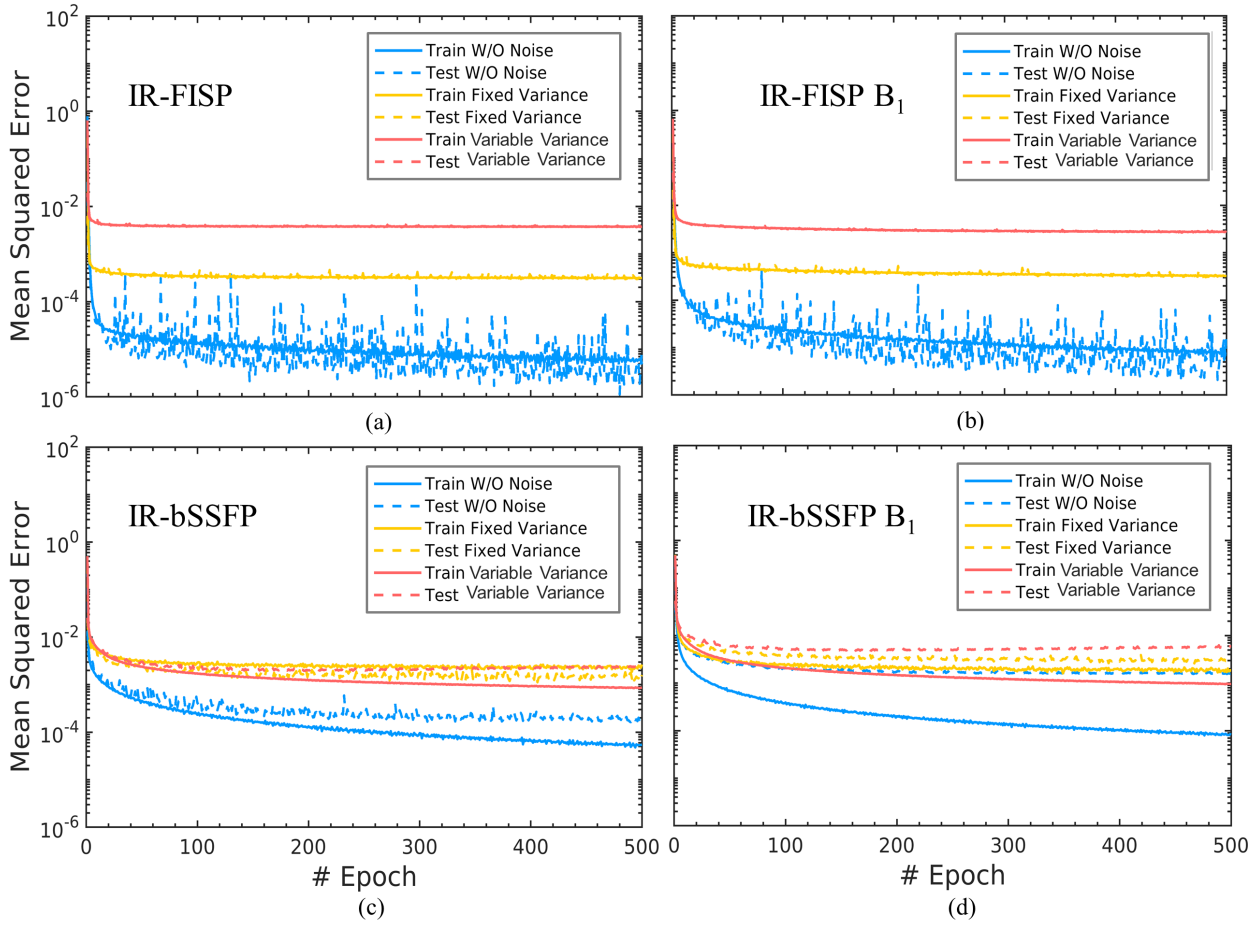


Figure 2.8: Training and test loss functions for the NN models trained with different data augmentation strategies: without noise, *Fixed variance* and *Variable variance*.

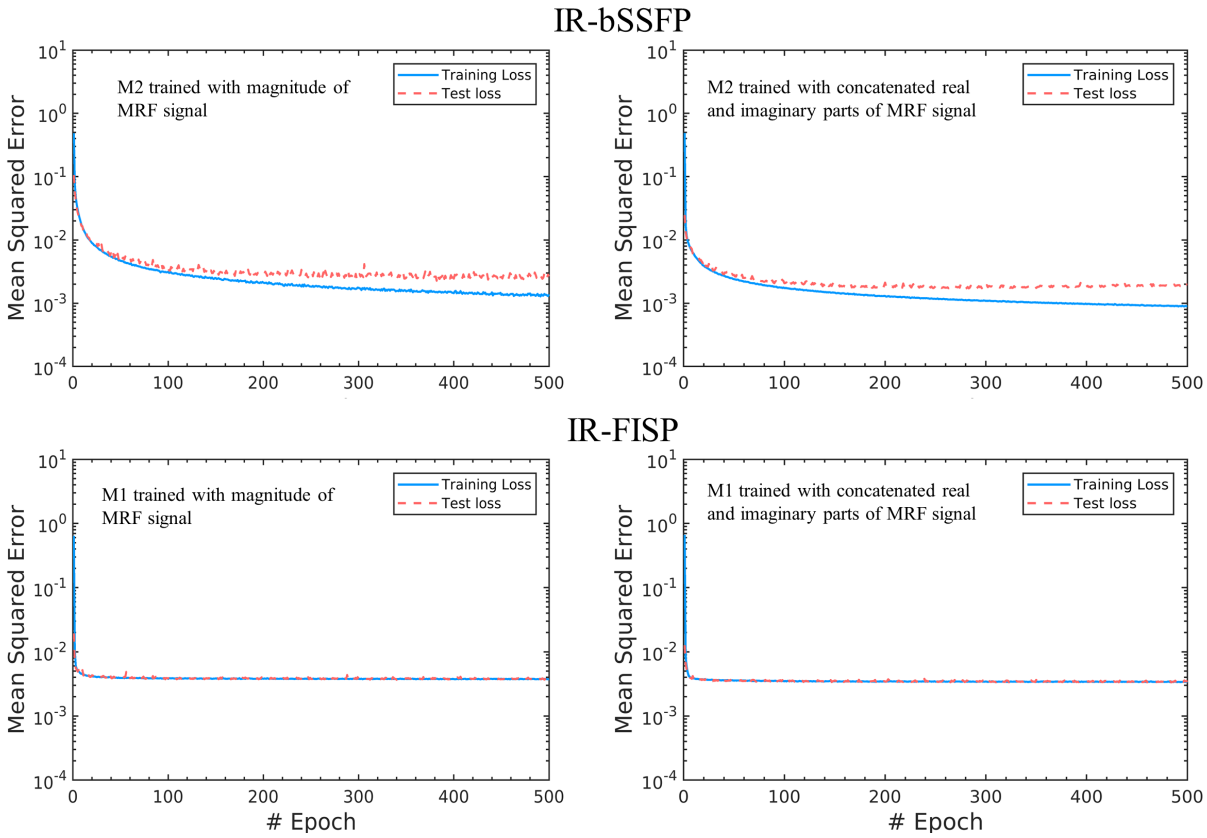


Figure 2.9: Training and test losses for models M1 and M2 trained with the magnitude of the MRF signals (left column) and concatenating real and imaginary parts of MRF signals (right column).

000 examples is reported. Although the training loss worsens when model M2 is trained with more examples, the gap between training and test loss is considerably reduced and the test loss reaches a lower value than the case in which the model is trained with fewer examples. Since the test loss is an estimator of the generalization performance of the NN model, it is possible to observe that the generalization of the model M2 is increased when more training examples are added.

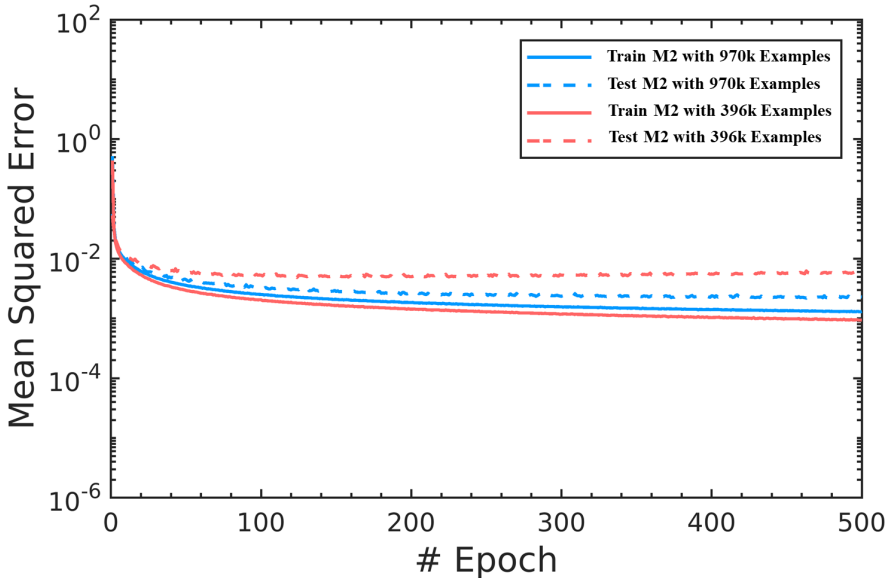


Figure 2.10: Training and test loss functions for the NN model M2 trained with 396 550 training examples and 970 000 training examples. Although the training loss worsens when model M2 is trained with more examples, the gap between training and test loss is considerably reduced and the test loss reaches a lower value than the case in which the model is trained with fewer examples. Since the test loss is an estimator of the generalization performance of the NN model, it is possible to observe that the generalization of M2 model is increased when more training examples are added.

The MAPE of parameters, evaluated on test sets for different data augmentation strategies as a function of SNR, are reported in Fig. 2.11. The error bar indicates two standard deviations of the MAPE, which comes from 10 repetitions of the prediction procedure on noisy data. Although Fig. 2.8 shows *Variable variance* to be the worst performing strategy during training, it is the most robust to noise among the three. Looking at Fig. 2.11 one can appreciate the significant improvement in high noise level robustness using this latter strategy for T_2 parameter in particular. For both models, M1 and M2, T_2 MAPE never goes over 12% and decreases rapidly under 10%, while MAPE for the other parameters is always lower than 5%. The standard deviation values are of the order 10^{-2}, which renders them almost invisible in the plots, indicating that all models show low variance for experiment repetitions. The MPE is reported as a function of SNR in Fig. 2.12, showing that the *variable variance*

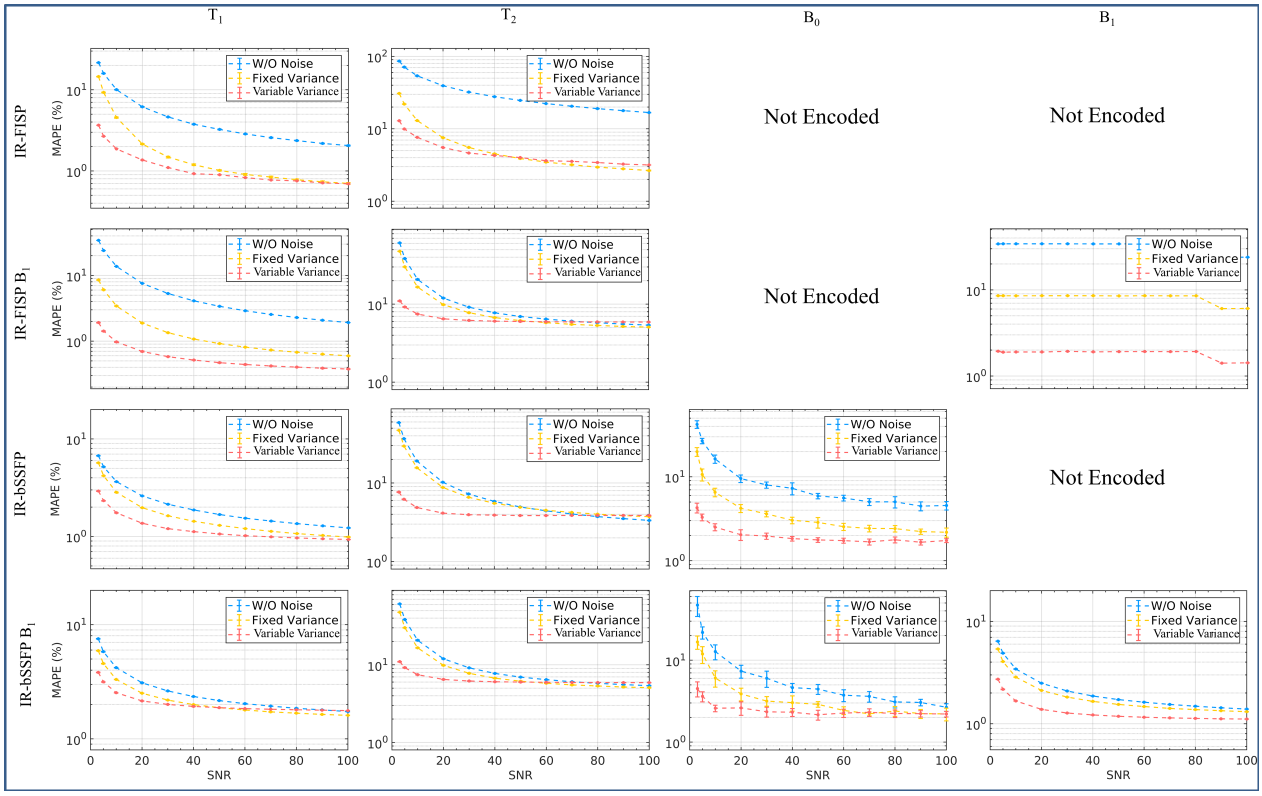


Figure 2.11: Comparison of parameter MAPEs evaluated on test sets for NN models trained with different data augmentation strategies.

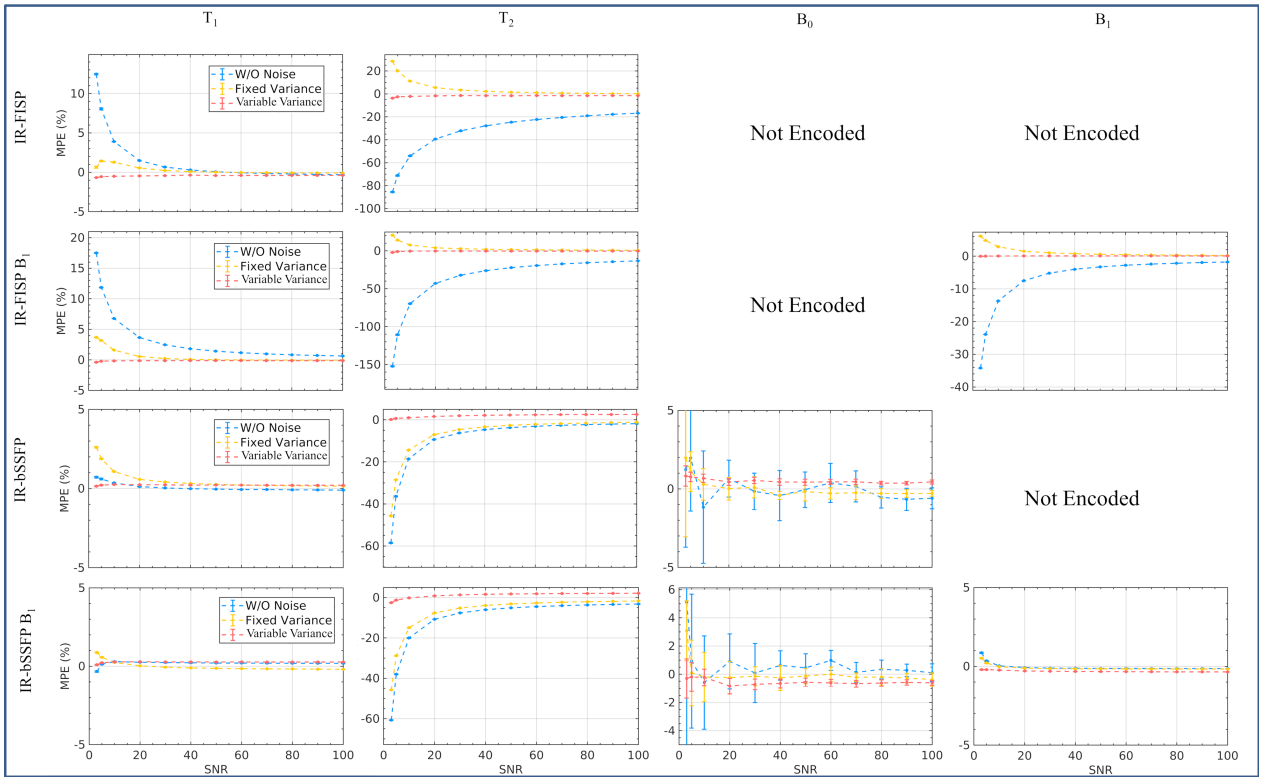


Figure 2.12: Comparison of the parameter mean percentage errors (MPEs) evaluated on test sets for the NN models trained with different data augmentation strategies.

strategy produces unbiased NNs models regardless of SNR (MPE around 0%), whereas the other strategies show increased MPE with a reduction of SNR. The explanation for this behavior is straightforward: with the *Variable variance* strategy, the NNs get to see more data characterized by heterogeneous levels of noise, improving generalization, while with the *Fixed variance* strategy the models eventually overfit for data with high SNRs, and, as expected, with the *W/O Noise* strategy the models perform worst in terms of noise robustness.

Although T_2 MAPEs are higher than those characterizing other MR parameters even when SNR is considerably higher, these are acceptable values given the wide range of estimated T_2 values (up to 3000 ms) and given that the TRs of the MRF pulse sequences considered are in the order of 10 ms. Considering the NMR transverse relaxation process it is expected that sequences with such TRs will have higher sensitivity when the T_2 values are in the order of the TRs employed. When T_2 relaxation times are orders of magnitude longer, we expect the pulse sequence to be less sensitive to T_2 variations, since there will be less transverse magnetization decay in each TR. A loss of sensitivity is also expected when T_2 relaxation times are much shorter than pulse TRs. In Figure 2.13 the results of simulations run using the IR-FISP sequence are summarized for three sets of examples with the following T_2 relaxation times: shorter, comparable and much longer than the TRs used in the pulse sequence (T_1 was kept fixed at 1.5 s and no noise was added in the simulations). The L_2 norms was computed among the simulated signals to quantitatively estimate the relative distances (each row represents the distance between one signal and all the other signals in the set) and the mean L2 norm as a global distance estimator was computed for each set of examples. As expected, when T_2 relaxation times are in the order of IR-FISP TRs (central column in Fig. 2.13), varying T_2 implies a significant variation of the MRF fingerprint, whereas when T_2 relaxation times are much shorter or longer than sequence TRs (left and right columns in Fig. 2.13, respectively), the MRF fingerprints vary much less with T_2 variations. This implies that, whatever method is used to perform the T_2 estimation, it will be more prone to estimation errors when T_2 relaxation times are orders of magnitude longer or shorter than TRs.

2.3.3 Interpolation and extrapolation capabilities of NN models

To address how well NN models M1 and M2 perform in interpolation and extrapolation tasks, the models are required to predict MR parameters for examples where the parameter ranges exceed those used during training. This experiment was conducted both for IR-FISP and IR-bSSFP sequences, using models M1 and M2, respectively. White gaussian noise, with variance such that each test signal has SNR = 100, was added before the prediction task.

In Fig. 2.14 the results for model M1, which processes IR-FISP data, are reported. Lin's concordance coefficient has been used to assess the goodness of agreement between

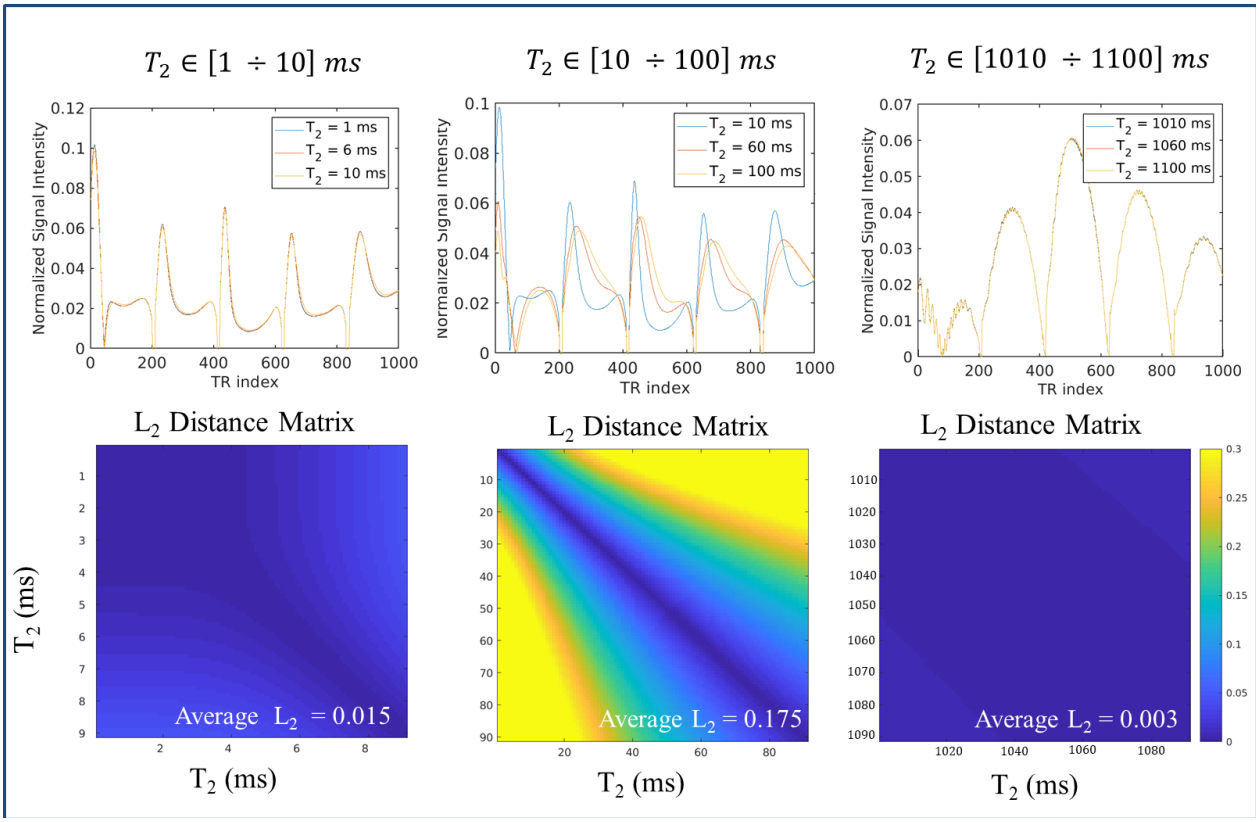


Figure 2.13: Simulated IR-FISP MR fingerprints with T_2 values belonging to three different classes: lower, comparable and much longer than TRs values used in the pulse sequence (T_1 was kept fixed to 1.5 s). Top row reports examples of simulated signal evolution for different T_2 values, while bottom row reports the L_2 distance matrix among signals.

ground truth and predicted parameters. The results show that very good agreement is reached when the test examples, never encountered during training, nevertheless have MR parameters within the training range. Indeed Lin's coefficient is 0.99 both for T_1 and T_2, whereas agreement decreases when the test signals have MR parameters partially outside the training ranges ($\rho_c = 0.55$ for T_1, and $\rho_c = 0.94$ for T_2), and show the worst agreement when all the MR parameters characterizing the test signals are outside the training ranges ($\rho_c = 0.3$ for T_1, and $\rho_c = 0.03$ for T_2). It is worth pointing out that Lin's coefficient for T_2 is significantly higher than that for T_1 when at least one parameter is inside the training ranges because there are many more test examples for which T_1 is outside the training range, while T_2 is inside the training range, than vice versa. Looking at Figure 2.15 the conclusions drawn for model M1 also hold for model M2, used to process IR-bSSFP data.

The results presented here show that both model M1 and M2 have good interpolation capacity, but poor extrapolation capacity. Thus, these NN models are reliable only if the MR parameters to be predicted have values within the training ranges.

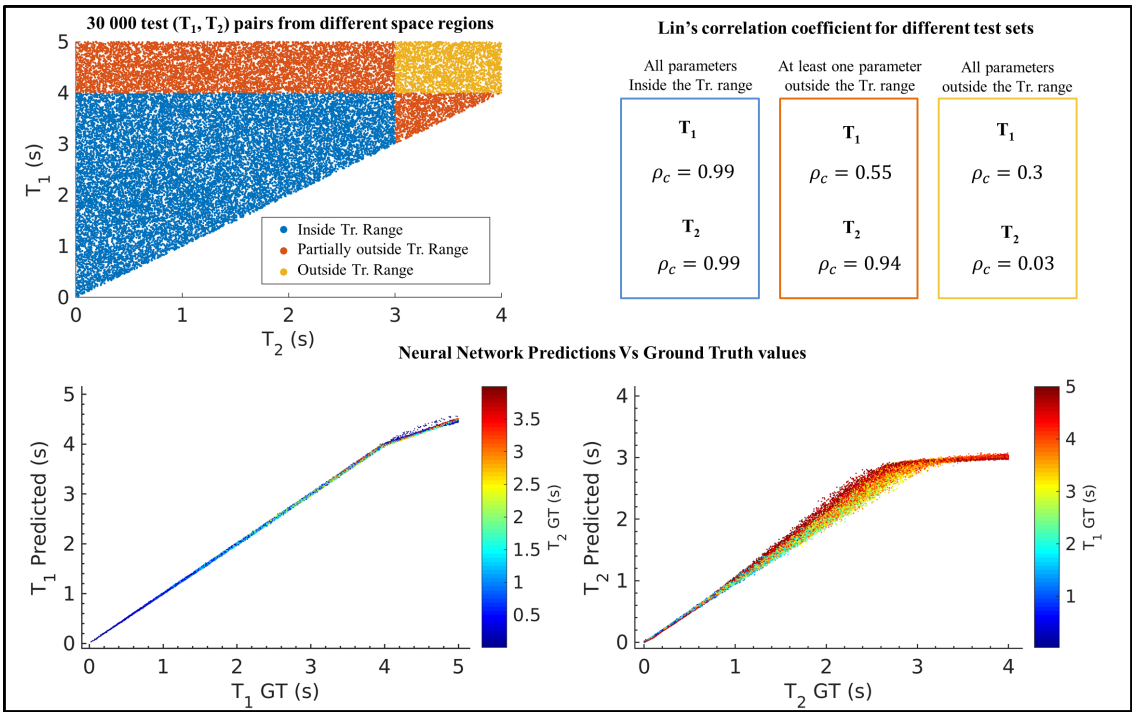


Figure 2.14: FISP interpolation and extrapolation capabilities.

2.3.4 Brain map reconstruction: Neural Networks vs Dictionaries

In Figure 2.16 and Fig. 2.17 the absolute error maps for each parameter, relative to NN reconstruction and dictionary matching respectively, are reported. They have been evaluated

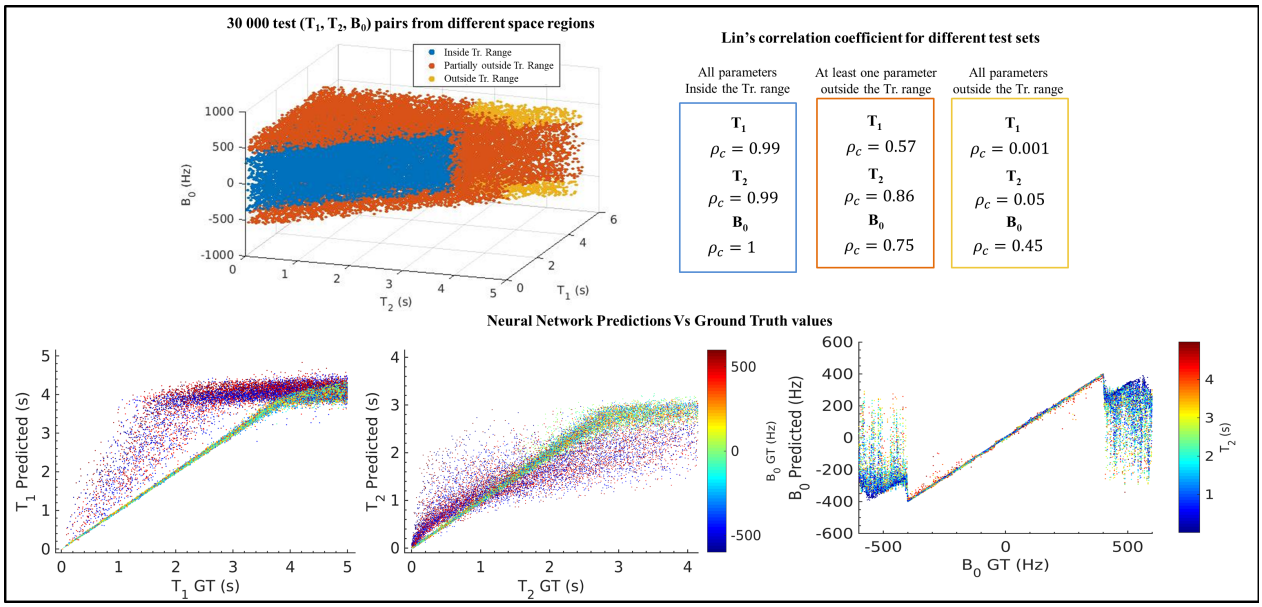


Figure 2.15: bSSFP interpolation and extrapolation capabilities.

by computing, pixel-wise and for each parameter map, the absolute error between the ground-truth parameter value and the predicted value. For each map, four global error estimators have been reported on the corners of the images: upper left corner: MAPE between the ground-truth pixel values and those reconstructed; upper right corner: MAPE between the ground-truth pixel values and those reconstructed, considering the brain without the scalp; lower left corner: RMSE between the ground-truth pixel values and those reconstructed; lower right corner: RMSE between the ground-truth pixel values and those reconstructed, considering the brain without the scalp. Although Fig. 2.16 and 2.17 report the reconstructions using noisy data with SNR = 5, in Fig. 2.18 the RMSE beetween reconstructed and ground-truth values as a function of SNR have been reported to show how the results change as a function of SNR.

Overall the NN approach performs better than the dictionary approach. This is particularly evident considering the IR-bSSFP B_1 sequence, where all the error estimators are lower than those characterizing the dictionary matching. This difference is even more evident if one looks at the error estimators evaluated without considering the scalp, where the B_0 off resonances go up to +/- 500 Hz, which are out of the range used for training the NN model M2, and, as expected, this affects parameters estimation accuracy[1]. Since most of the clinical information is in the brain region and the scalp region is of little interest, the global error estimators have also been computed without considering the scalp. In the latter case, the error estimators show better agreement with ground truth values even when the SNR is very

[1]See section 2.3.3 for extrapolation and interpolation capacities of NN model M2

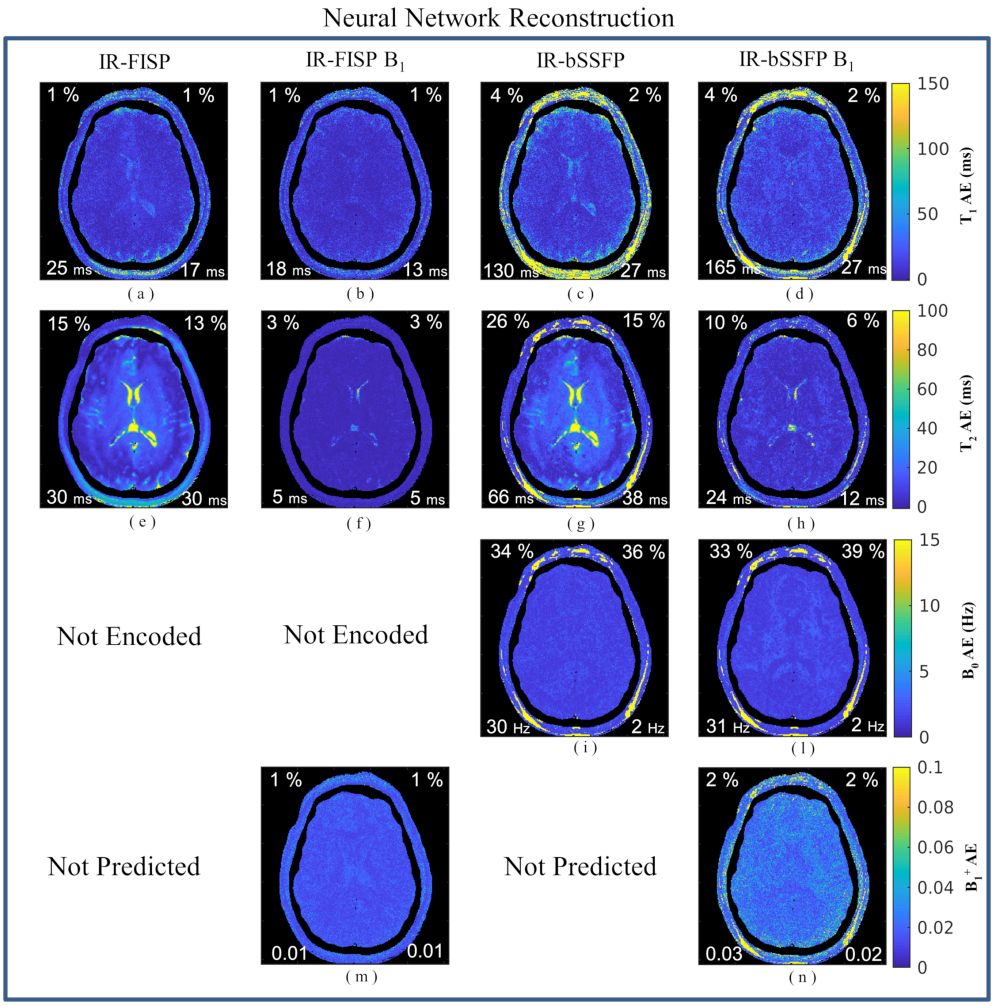


Figure 2.16: Absolute error maps between reconstructed and ground truth parameter maps using the NN approach. Global error estimators are reported on the four corners of each image: upper left corner: MAPE; upper right corner: MAPE evaluated without considering the scalp region; lower left corner: RMSE; lower right corner: RMSE evaluated without considering the scalp region.

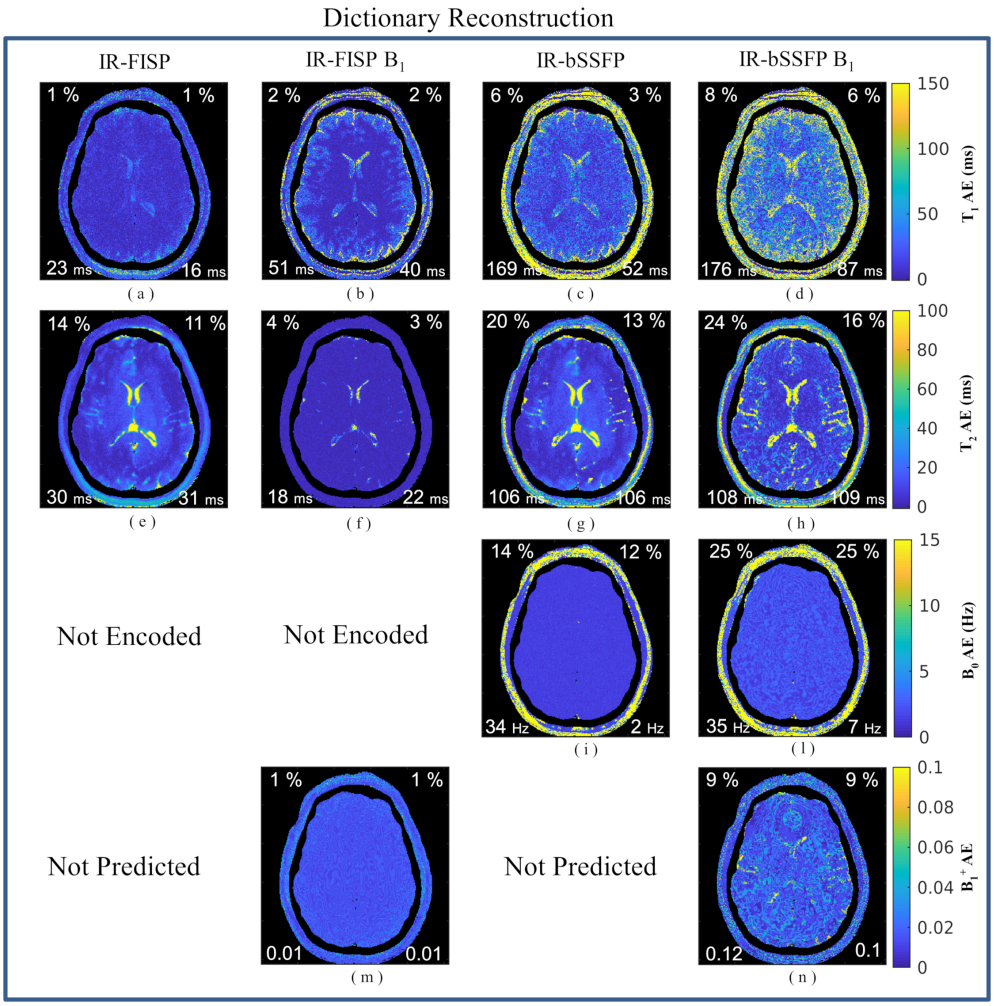


Figure 2.17: Absolute error maps between reconstructed and ground truth parameter maps using the Dictionary approach. Global error estimators are reported on the four corners of each image: upper left corner: MAPE; upper right corner: MAPE evaluated without considering the scalp region; lower left corner: RMSE; lower right corner: RMSE evaluated without considering the scalp region.

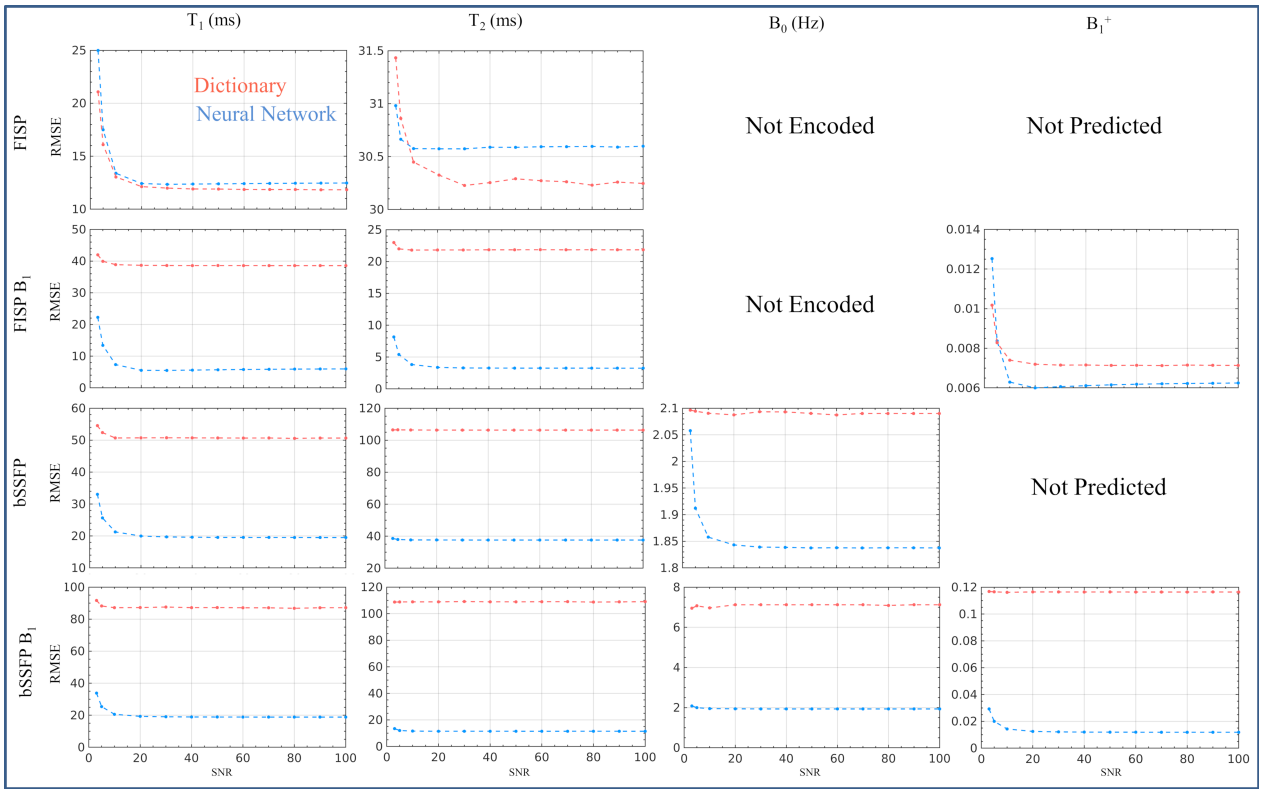


Figure 2.18: RMSE as a function of SNR for Brain reconstruction with NN and Dictionary.

low. It is worth pointing out that MAPE is a poor measure of quality for B_0 estimation, because small absolute errors made by the NN models generate high percentage errors, as highlighted by looking at the RMSEs, which are in the order of 2 Hz.

Considering IR-FISP, the two approaches have similar performance, probably because the number of the examples used to build the dictionary was high enough to ensure a proper MR space sampling. For IR-FISP B_1, on the other hand, the NN approach performs better in estimating T_1 and T_2. MAPEs and RMSEs evaluated for NN reconstruction are at least as good as the values calculated for dictionary reconstruction.

Analyzing Fig. 2.16 in greater detail, when B_1^+ field inhomogeneities are not taken into account, this leads to a drop in parameter estimation accuracy. T_2 is the most affected parameter (Fig. 2.16e, Fig. 2.16g). In particular, when B_1^+ is not taken into account during network training for IR-FISP sequences, T_2 MAPE increases from 2.5% to 13% (Fig. 2.16f and Fig. 2.16e respectively). The same drop in T_2 accuracy happens for IR-bSSFP sequences (Fig. 2.16g and Fig. 2.16h). In contrast, considering Fig. 2.17, one can conclude that taking B_1^+ field inhomogeneities into account does not improve parameter estimation for IR-bSSFP sequences: errors were higher when B_1^+ field inhomogeneities were predicted in the dictionary, which was the opposite of the expected result, and in contrast to what was found

using the NN approach. However, this behaviour is only due to the number of entries of the dictionary, being insufficient to guarantee an accurate sampling of the 4D MR parameter space. It is an example of the curse of dimensionality of nearest-neighbor algorithms. With the dictionary approach there is a trade-off between keeping the number of dictionary entries below a reasonable number and guaranteeing a good resolution. In favour of this claim, one can notice that not all dictionary reconstruction errors increase when B_1^+ is considered. The trade-off is still favorable in the case of the IR-FISP B_1 sequence (second column of Fig. 2.17). There is a slightly worsening in the T_1 errors, but a significant improvement in the T_2 errors when the B_1^+ is added to the dictionary.

To better test the hypothesis that the dictionary matching approach has lower performance than the NN approach mainly because of the lack of resolution in the dictionary, Fig. 2.19 reports the error maps of brain parameter maps reconstruction using 4 dictionaries where the number of entries has been kept fixed around 400000 entries, which produces higher resolution dictionaries. For example, for the IR-FISP sequence the resolution for T_2 is 2 ms in the range $[1 \div 500]$ ms, and 5 ms for T_2 in range $[505 \div 3000]$ ms, while for T_1 the resolution is 2 ms in the range $[1 \div 500]$ ms, 5 ms in the range $[505 \div 2000]$ ms and 10 ms in the range $[2010 \div 4000]$ ms. Comparing Fig. 2.16 and Fig. 2.19, one can appreciate that the dictionary approach reaches NN approach performance for IR-FISP, IR-FISP-B_1 and IR-bSSFP. With this result is evident that, as expected, the lack of resolution in the dictionaries generates high errors in MR map estimation. However, globally, increasing the resolution of the dictionary does not affect Lin's concordance coefficients, which are reported in Tables 2.2, 2.3, 2.4 and 2.5. With these higher resolution dictionaries the computational time to generate the maps is around 160s, which is about 100 times higher than the processing time for model M1 (for IR-FISP data), that is, 1.5 s, and 40 times higher than the processing time for model M2 (for IR-bSSFP data), 4 s. Hence, when using larger dictionaries one can appreciate that the performance of the dictionary approach improves in accordance with the degree of improvement in dictionary resolution, but at the cost of longer computational time.

Moreover, when using higher resolution dictionaries, taking into account B_1^+ inhomogeneities clearly improves T_2 estimation for IR-FISP B_1 sequence, as showed in second column of Fig. 2.19.

To process 35,000 pixels, the NN reconstruction takes about 1.5 s for IR-FISP-type sequences and 4 s for IR-bSSFP-type sequences. No significant change in processing time was observed when estimating 2 or 3 parameters with the IR-FISP-type sequences, or 3 or 4 parameters with the IR-bSSFP-type sequences. This is because the only difference in NN models for IR-FISP-type sequences, and NN for IR-bSSFP-type sequences, is in the output layers, and this just adds a small number of operations. On the contrary, with dictionary

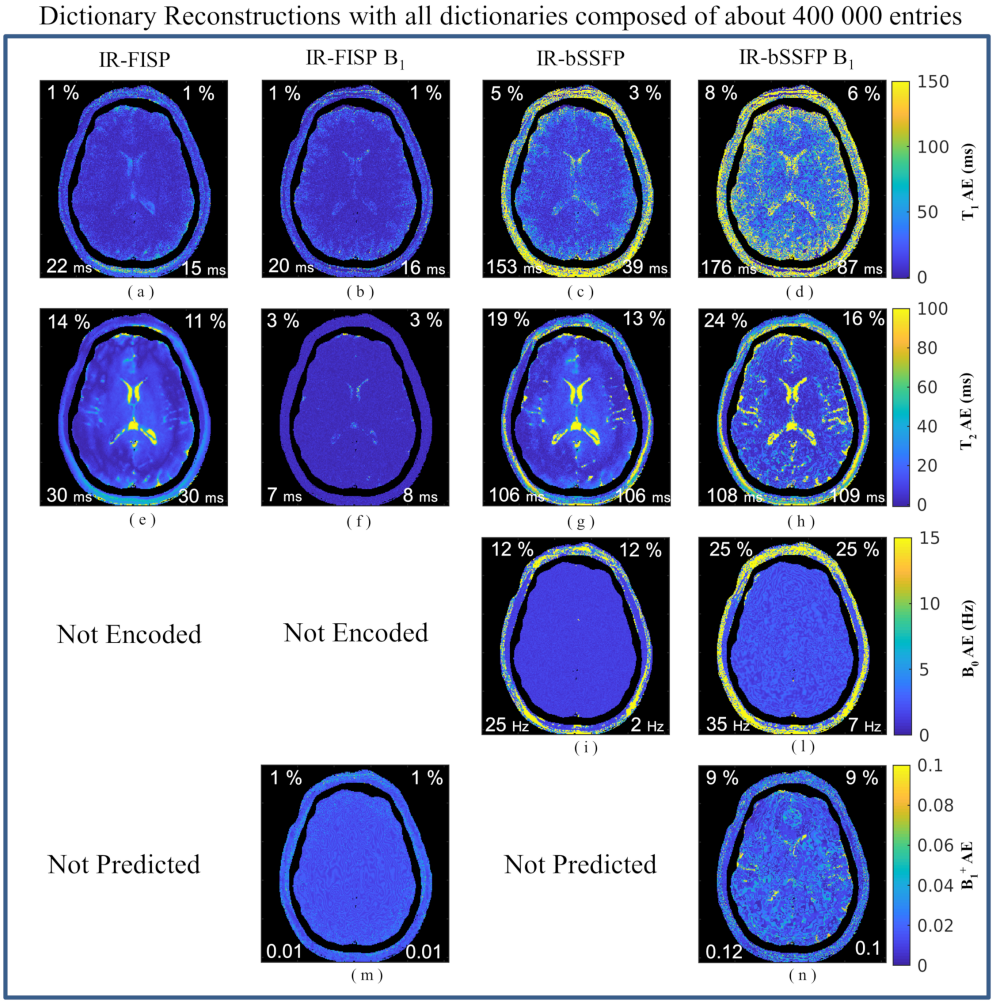


Figure 2.19: Absolute error maps between NN reconstructed and ground truth parameter maps with original B_1^+ field inhomogeneities map at SNR = 5. All dictionaries are composed of 400 000 entries. Global error estimators are reported on the four corners of each image: upper left corner reports the MAPE; upper right corner reports MAPE evaluated without considering the scalp region; lower left corner reports the RMSE while lower right corner reports RMSE evaluated without considering the scalp region.

Parameter	**NN (M1)**		**Dictionary (G1)**		**Dictionary 400k entries**	
	W scalp	W/O scalp	W scalp	W/O scalp	W scalp	W/O scalp
T_1	0.99	0.99	0.99	0.99	0.99	0.99
T_2	0.98	0.98	0.98	0.98	0.98	0.98

Table 2.2: Lin's concordance coefficient between GT MR parameters and those estimated using the NN and dictionary approaches for the numerical brain phatnom. MRF data were simulated using the IR-FISP sequence

Parameter	NN (M1)		Dictionary (G2)		Dictionary 400k entries	
	W scalp	W/O scalp	W scalp	W/O scalp	W scalp	W/O scalp
T_1	0.99	0.99	0.99	0.99	0.99	0.99
T_2	0.99	0.99	0.99	0.99	0.99	0.99
B_1^+	0.99	0.99	0.99	0.99	0.99	0.99

Table 2.3: Lin's concordance coefficient between GT MR parameters and those estimated using the NN and dictionary approaches for the numerical brain phatnom. MRF data were simulated using the IR-FISP B1 sequence

Parameter	NN (M2)		Dictionary (G3)		Dictionary 400k entries	
	W scalp	W/O scalp	W scalp	W/O scalp	W scalp	W/O scalp
T_1	0.99	0.99	0.98	0.99	0.99	0.99
T_2	0.90	0.99	0.66	0.67	0.67	0.67
B_0	0.97	0.99	0.95	0.98	0.98	0.99

Table 2.4: Lin's concordance coefficient between GT MR parameters and those estimated using the NN and dictionary approaches for the numerical brain phatnom. MRF data were simulated using the IR-bSSFP sequence

Parameter	NN (M2)		Dictionary 400k entries (G4)	
	W scalp	W/O scalp	W scalp	W/O scalp
T_1	0.98	0.99	0.98	0.99
T_2	0.99	0.99	0.66	0.68
B_0	0.96	0.98	0.95	0.82
B_1^+	0.95	0.97	0.60	0.52

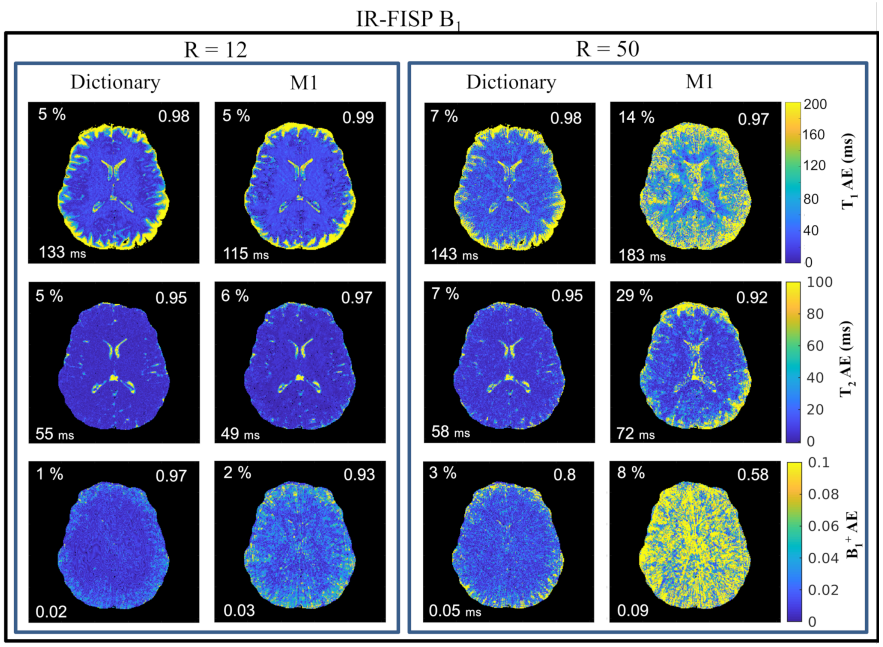
Table 2.5: Lin's concordance coefficient between GT MR parameters and those estimated using the NN and dictionary approaches for the numerical brain phatnom. MRF data were simulated using the IR-bSSFP B_1 sequence

matching the computational time strongly scales with the number of parameters, because the size of the dictionary increases. The computational time required to generate maps reported in Fig 2.16 are 30 s, 66 s, 71 s and 163 s for IR-FISP, IR-FISP B_1, IR-bSSFP and IR-bSSFP B_1 respectively. Overall the NN reconstruction is thus more efficient both in memory usage, since it does not need to store any dictionary in the RAM memory, and in computation time.

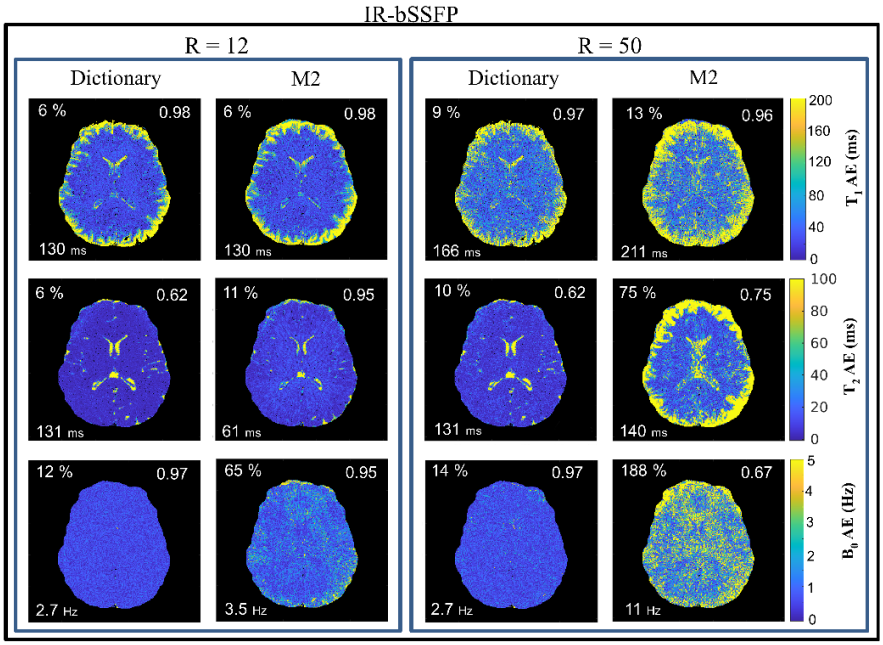
2.3.5 Robustness to undersampling artifacts

In Figure 2.20(a) and Fig. 2.20(b) the absolute error maps for each parameter, relative to NN reconstruction and dictionary matching respectively, are reported for two cases of acceleration factors (i.e. R = 12 and R = 50). In Tables 2.6 and 2.7 global estimators of error and agreement are reported for all the tested acceleration factors. The scalp region was not taken into account in the quantitative analysis. Since the performance of the methods in the scalp region was assessed in the previous section, the error analysis did not take into account this region.

Focusing on Fig. 2.20(a), with R = 12 (left box in the Figure) there is close agreement between GT and reconstructed parameter maps using both NN and dictionary approaches. For all the parameters ρ_c is greater or equal to 0.95 and 0.93 for the dictionary and NN approaches, respectively. On the contrary, when R = 50, the NN approach shows less robustness to undersampling artifacts than the dictionary approach. Although the global errors also increase for the dictionary approach, there is a better overall agreement between dictionary and GT than between NN and GT, especially for B_1^+ ($\rho_c = 0.8$ and $\rho_c = 0.58$ for the dictionary and the NN, respectively). The results summarized in Figure 2.20(b) support a similar conclusion: the dictionary approach is more robust to undersampling artifacts than the NN approach. It is worth commenting that although MAPE for T_2 is relatively low (MAPE = 6 %) using the dictionary approach, the agreement with T_2 GT values is relatively low ($\rho_c = 0.62$), even in the case of moderate undersampling (R=12). On the contrary, with R = 12, NN model M2 reaches a close agreement with GT T_2 values ($\rho_c = 0.95$), although MAPE is higher than the one for the dictionary (MAPE = 11 %). This is because the dictionary approach makes lower errors in the white matter region compared to the NN approach, but much higher errors in the CSF region, as hilighted by the AE maps. In particular, the dictionary approach tends to systematically underestimate CSF T_2 values, and such bias is highlighted in the low concordance coefficient. This finding is also confirmed when the data are fully sampled, as reported in the first row of Tab 2.7. Since in this case a high resolution dictionary has been used, there is an evidence of biases introduced by the dictionary approach that cannot be attributed to the dictionary quantization. However, such bias is not reported either for the other parameters or for IR-FISP based sequences.



(a)



(b)

Figure 2.20: Absolute error maps between reconstructed and ground truth parameter maps using dictionary and NN approaches for the IR-FISP B_1 (a) and the IR-bSSFP (b) sequences. Two different acceleration factors, R=12 (left box) and R=50 (right box) are reported. Global error estimators are reported at left corner of each image: upper left corner: MAPE; lower left corner: RMSE. In the upper right corner the Lin's concordance coefficient between estimated and GT parameters is reported.

R	M1 MAPE (%) [ρ_c]			Dictionary MAPE (%) [ρ_c]		
	T_1	T_2	B_1^+	T_1	T_2	B_1^+
0	0.6 [0.99]	2 (0.97]	0.5 [0.99]	0.8 [0.99]	3 [0.95]	0.5 [0.99]
6.25	5 [0.99]	5 [0.97]	1 [0.98]	5 [0.98]	5 [0.96]	0.8 [0.99]
12.5	5 [0.99]	6 [0.97]	2 [0.93]	5 [0.98]	5 [0.96]	1 [0.97]
25	8 [0.99]	13 [0.96]	5 [0.80]	6 [0.98]	7 [0.95]	2 [0.93]
50	14 [0.97]	29 [0.92]	8 [0.58]	7 [0.98]	10 [0.95]	3 [0.80]
100	20 [0.94]	61 [0.79]	10 [0.43]	7 [0.98]	12 [0.95]	4 [0.70]
200	32 [0.83]	120 [0.61]	12 [0.32]	8 [0.98]	20 [0.93]	6 [0.48]
400	59 [0.54]	195 [0.36]	14 [0.19]	12 [0.97]	34 [0.88]	10 [0.26]

Table 2.6
MAPEs and Lin's concordance coefficients between estimated and GT parameter maps for NN and dictionary approaches as a function of different acceleration factors (Rs) using the IR-FISP-B1 sequence

R	M2 MAPE (%) [ρ_c]			Dictionary MAPE (%) [ρ_c]		
	T_1	T_2	B_1^+	T_1	T_2	B_1^+
0	0.8 [0.99]	6 [0.95]	19 [0.96]	2 [0.99]	5 [0.60]	12 [0.97]
6.25	6 [0.99]	10 [0.95]	48 [0.95]	6 [0.98]	6 [0.62]	12 [0.97]
12.5	6 [0.98]	11 [0.95]	66 [0.95]	6 [0.98]	6 [0.62]	12 [0.97]
25	9 [0.98]	23 [0.92]	102 [0.89]	8 [0.98]	9 [0.62]	12 [0.97]
50	13 [0.96]	75 [0.74]	188 [0.67]	9 [0.98]	10 [0.62]	14 [0.97]
100	25 [0.92]	275 [0.30]	410 [0.34]	11 [0.97]	13 [0.62]	63 [0.85]
200	81 [0.68]	774 [0.10]	1513 [0.12]	15 [0.95]	20 [0.62]	220 [0.42]
400	179 [0.28]	1494 [0.01]	4052 [0.04]	22 [0.91]	38 [0.59]	714 [0.13]

Table 2.7
MAPEs and Lin's concordance coefficients between estimated and GT parameter maps for NN and dictionary approaches as a function of different acceleration factors (Rs) using the IR-bSSFP sequence

2.4 Conclusions

In this chapter, a deep neural network approach applied to MRF data is presented. The approach is tested on simulated data for prediction accuracy, noise robustness and tolerance to artifacts caused by undersampling. Some intrinsic limitations of dictionary-based approaches for MRF, and how NNs approaches can overcome them, have been discussed in detail. In particular, methods have been studied to improve parameter space sampling design and how to achieve better noise robustness during NN training are explored. To demonstrate the generalization of the proposed NN, four different MRF pulse sequences were considered: the IR-FISP sequence and its variant to account for B_1^+ field inhomogeneities, the original IR-bSSFP and a newly proposed variant to account for B_1^+ field inhomogeneities. The results demonstrate the low accuracy of grid sampling for training data sets, since this method

introduces a strong bias due to its regularity, and this leads the NN to data overfitting, as shown for IR-bSSFP type sequences. Random uniform sampling was found to perform better in spanning the parameter space for the IR-bSSFP type sequence, whereas no significant difference between grid and random sampling was found for the IR-FISP type sequence.

Three different data augmentation strategies were than tested with regard to robustness to white Gaussian noise. Feeding the network models with noisy data with different noise variances yielded the best results in terms of noise robustness, providing a MAPE lower than 15% (evaluated on a wide range of parameter values) even in case of very noisy data (SNR = 3). These results exceed those obtained with the data augmentation strategies used in [104] and [33], which were also tested.

Finally, a comparison between the NN and dictionary approaches was performed using a numerical brain phantom based on real standard quantitative MRI protocols. The results of this comparison show that when considering noisy data, but without undersampling, the NN performance is equal to or greater than the dictionary-based approach, depending on the resolution of the dictionary. However, an increase in the resolution of the dictionary comes at the cost of longer computational time, whereas the NN approach showed a constant computational time independent of the number of parameters to be retrieved. A neural network model learns an approximation of the ITF from a set of training examples. Once trained, processing is computationally fast and efficient, not requiring a large dictionary stored in memory. To increase parameter estimation accuracy, it is important to take into account as many scanner imperfections as possible, such as B_0 and B_1^+ inhomogeneities. With the classical dictionary-based approach to MRF, each added parameter results in an exponential increase in the number of dictionary entries, making it difficult to consider other relevant MR parameters, such as diffusivity. These results show the advantages of NN approaches in MR Fingerprinting parameters estimation. Thousands of training examples might be required to learn an accurate and robust ITF with NN, as shown for the IR-bSSFP and IR-bSSFP B_1 sequences. However, because the NN model is trained in batches, the data can be stored in the hard disk, and just few of them are loaded into the RAM in each step of the training. Moreover, because the training examples are synthetic data, one can even build the simulation step into the training pipeline, without the need to store any data even on the hard disk. The simulation step requires the associated code to be fast and efficient, otherwise it becomes a bottleneck in terms of time needed to train the network model. However, parallel coding is well suited to this kind of problem and can be used to make speed up computation.

The investigation of robustness to undersampling artifacts allowed us to characterize the limits for which the current form of the NN approach is less robust to undersampling artifacts than the dictionary approach. Hence, future efforts must focus on promoting this

form of robustness. To this end, different noise models could be investigated to assess the generalization performance of the NN models, as well as data preprocessing methods for noise reduction, such as SVD or the method recently proposed by Bo et al [16], in which high-quality time-series images are reconstructed from highly-aliased images. Preliminary results of this on-going study have recently been presented [9], in which an SVD pre-processing filter in the time domain was proposed to mitigate the effect of undersampling for in-vivo undersampled MRF FISP data. Other recent contributions have investigated the application of FCNNs to MRF data with in-vivo application [46].

The above results demonstrate than NN approaches show promise for application to MRF and may help the community working in MRF to build more trust in deep learning approaches, pushing fingerprinting pulse sequence design beyond the limitations affecting the dictionary approach due the curse of dimensionality, and allowing the consideration of other meaningful MR parameters, such as diffusivity.

Chapter 3

Optimal Training Strategy to Promote Robustness to Undersampling Artifacts: Applications to Real Acquired Data

3.1 Introduction

The aim of this chapter is to extend the results presented in chapter 2 by investigating how to promote robustness in the NN model during training to perform accurate pixel-wise MR parameter estimation. In particular, fixed the NN architecture, different pipelines are used to train the NN model, and the performances are evaluated by comparing the estimations obtained by the NN approach with those obtained using the standard dictionary matching approach. In terms of data augmentation operations, phase offsets multiplication and online noise adding are proposed to promote robustness to undersampling artifacts and phase offsets affecting real acquired data. Usually MRF algorithms use the magnitude of the signal, so that it is not necessary to consider the phase of it [59, 75]. However, this is possible because of the robustness of the dot product to undersampling artifacts. Using the magnitude allows one to get rid of the problem of considering the phase of the MRF signal, but at cost of loosing information that can be critical with the purpose of building a robust NN application. Hence, magnitude and complex form of the signal are both tested as input of the network to asses what kind of input yields the best performance. In chapter 2 it has been shown that without applying any de-noising filter, the NN was less robust to artifacts produced by the undersampling. In terms of preprocessing steps, k-SVD truncation is proposed as

de-nosing filter [3]. Moreover, three noise models are used to investigate the capability of the NN to build robustness to undersampling when different noise models are used, and they are: additive Gaussian noise, multiplicative Gaussian noise and a combination of additive and multiplicative Gaussian noises. All these operations are combined together to form different training pipelines. By using an ablation approach it has been possible to heuristically identify the optimal training pipeline which allows to train a NN for the estimation of MR parameter maps that are in strong agreement with those obtained using the dictionary approach.

3.2 Materials and Methods

3.2.1 MRF pulse sequences and simulations

Two IR-FISP MRF pulse sequences [59] were used in this work: one is composed of 400 pulses (L = 400), while the other is composed of 1000 pulses (L = 1000). The latter sequence is used to perform a fullysampled acquisition, whereas the first sequence is used to perform an undersampled acquisition. A shorter pulse sequence was used for the fullysampled scan to maintain a reasonable acquisition time. The flip angle (FA) and repetition time (TR) patterns for the two sequences have been reported in Fig. 3.1.

The MRF dictionaries and the data sets for NN training were produced performing numerical simulations of the above pulse sequences using the Extended Phase Graph (EPG) algorithm [61]. All the simulations were performed with MATLAB (MathWorks).

During the training phase of the NN model, the simulated signals are corrupted with noise to promote robust learning [103]. Although additive white Gaussian noise is the most common noise model for NMR simulations, and it is also widely used in the MRF literature [31], [112], [97], [33], Gaussian multiplicative noise has been recently proposed as noise model able to mimic the undersampling artifacts affecting signal time evolution in MRF [62]. Motivated by this work, noise-free simulated signals are modeled to be corrupted by a sum of multiplicative and additive Gaussian noises following equation 3.1:

$$S(t) = S_{clean}(t) + \alpha(t) + n(t) \tag{3.1}$$

where $n(t)$ represents Gaussian additive noise, whereas $\alpha(t)$ is proportional to the MRF signal, and the coefficients of proportionality ω are taken from a Gaussian distribution with zero mean and variance σ^2. Hence, the i^{th} component of the noise can be expressed as $(\alpha)_i = |(S)_i|(\omega)_i$.

Three different models of noise were used in this work, and they are: only Gaussian additive; only Gaussian multiplicative and the sum of Gaussian multiplicative and additive

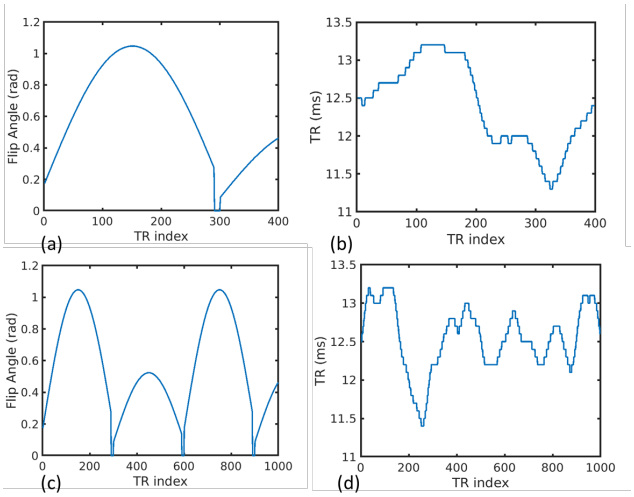


Figure 3.1: IR-FISP MRF pulse sequences used in this work. Top row depicts the FA (a) and TR (b) patterns used to perform the fullysampled acquisition. Bottom row depicts the FA (c) and TR (b) patterns used to perform the undersampled acquisition.

noises. Equations 3.2, 3.3 and 3.4 express the signal-to-noise-ratio (SNR) of the signal in presence of the different noise models, where $E[signal^2]$ is the power of the signal.

$$SNR_{additive} = \frac{E[signal^2]}{\sigma^2_{additive}} \tag{3.2}$$

$$SNR_{multiplicative} = \frac{1}{\sigma^2_{multiplicative}} \tag{3.3}$$

$$SNR_{(multi.+add.)} = \frac{\sigma^2_{multiplicative} E[signal^2] + \sigma^2_{additive}}{\sigma^2_{multiplicative}\sigma^2_{additive}} \tag{3.4}$$

Equation 3.3 was derived as follow:

$$SNR = \frac{E[s^2]}{E[n^2]} \tag{3.5}$$

where s is the signal and n is the noise. For the multiplicative noise considered in this work $n = |s| \times \omega$, where s and ω are independent so that equation 3.5 becomes:

$$SNR = \frac{E[s^2]}{(E[s^2] \times E[\omega^2])} = \frac{1}{\sigma^2_{omega}} \tag{3.6}$$

where σ^2_{omega} is labelled $\sigma^2_{multiplicative}$ in equation 2.2.

As example, in Figure 3.2 a simulated white matter fingerprint affected by the three type of noises is reported for SNR = 2.

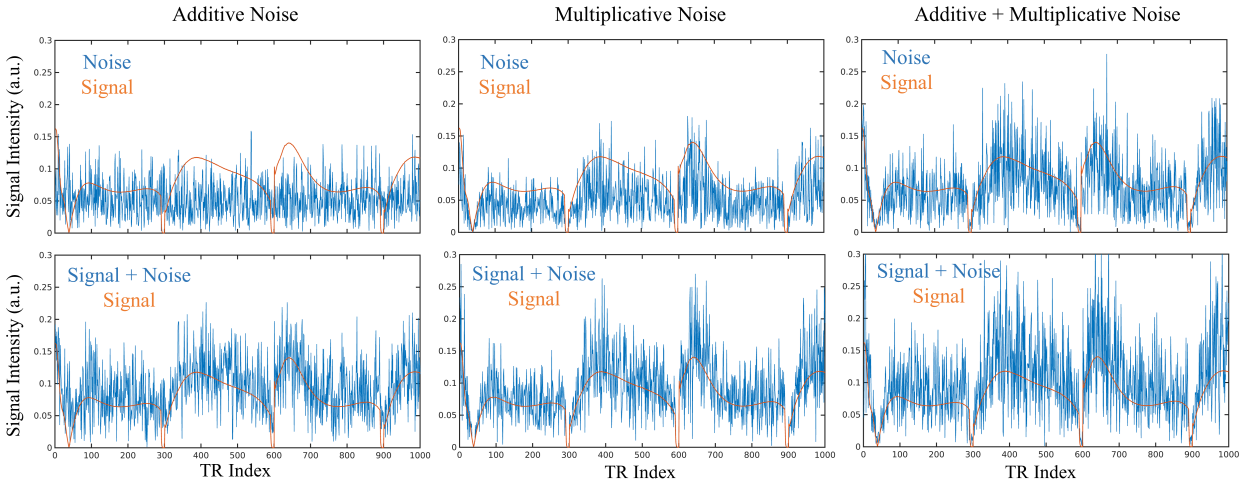


Figure 3.2: Examples of adding different type of noises to a clean white matter fingerprint. Plots in the top row report the clean white matter fingerprint with the simulated noise vector overlie. Plots in the bottom row report the clean white matter fingerprint with the fingerprint corrupted by the noise overlie.

3.2.2 Neural Network model and training pipelines

The NN model is composed by fully connected layers, where the ReLU is used as neuron activation function for the hidden layers, while a linear activation function is used for the output layer. Figure 3.3 depicts the NN model along with the procedure used to train it. The number of layers and neuron per layer were selected by an iterative procedure of tries and tests which are not reported in this work, but more details can be found in [8].

The network training pipeline is composed of 3 steps. First, a data augmentation step, where the input signals undergo a number of operations aiming to simulate artifacts that affect the real acquired signals. In this study, two possible data augmentation operations were considered.

The first operation is phase augmentation (PA), where a batch of simulated signals are multiplied by different phase offsets (θ), which are randomly sampled from the set [-π, π], to make the NN to see signals with different phases. This step aims to simulate the fact that in a real-scenario the application of the pulse sequence will not produce signals with phases that are the same encoded in the dictionary, but the white matter fingerprint, for example, will have phase offsets depending on the region in which the voxel is due to, for example, field inhomogeneities and receiver. The assumption here is that a phase offset is constant during the application of the pulse sequence. Also, this step is important only when the input of the network is the complex signal.

The second operation is noise adding with variable variance. Inputs signals inside a batch are corrupted by noise with different variances. The importance of using this noise adding

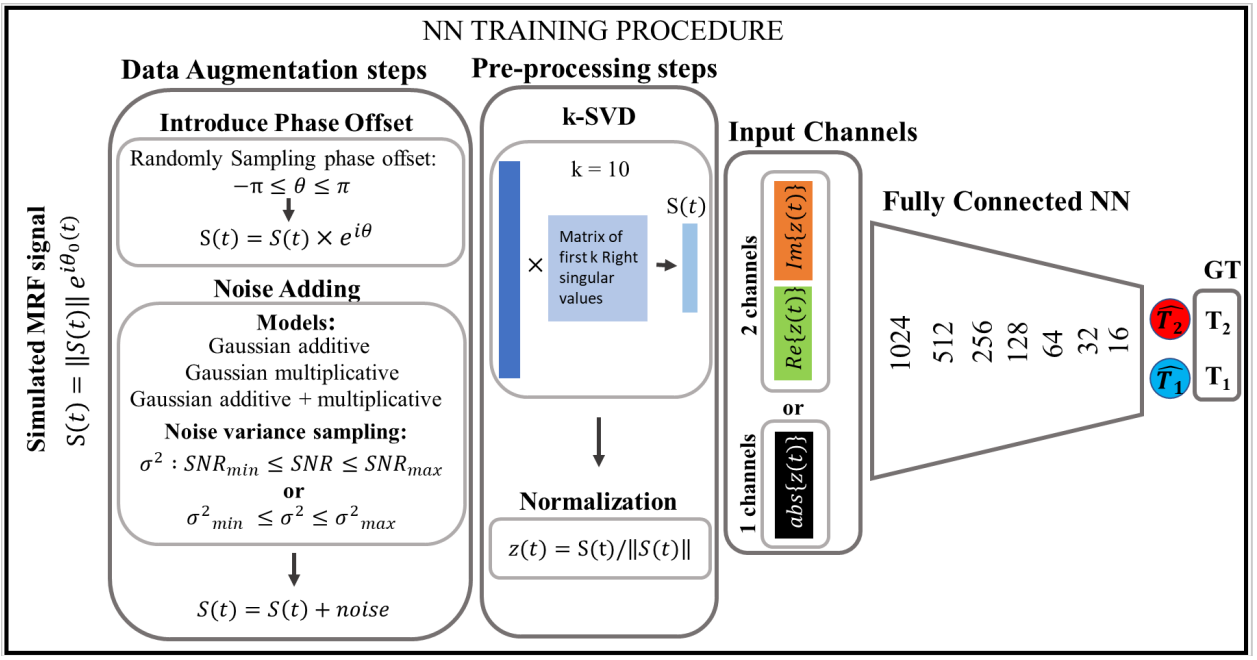


Figure 3.3: Schematization of the NN trainign procedure used in this work along with the representation of the fully connected NN model. During training each signal can undergoes data augmentation steps, in which the phase of the clean signal is multiplied by a random phase offset and complex noise (either Gaussian additive or Gaussian multiplicative) is added with random variance. Successively, the signal can undergo pre-processing steps, which are SVD truncation and normalization. Finally, the type of input to feed to the network is selected being either the magnitude of the signal or the complex signal.

strategy during training has been demonstrated in [8]. The noise model to be used can be one of the model described in section 3.2. The different variances can be generated by either directly sampling them from a uniform distribution in the range $[\sigma^2_{min}, \sigma^2_{max}]$, or by randomly sampling the SNR values in the range $[SNR_{min}, SNR_{max}]$. It is worth remarking the inverse relationship between variance and SNR. Hence, by uniformly sampling the variances the NN model during training will see more examples with low SNR compared to examples with high SNR.

If the two operations are performed together, the training batch contains MRF examples with a variety of MR parameters, that are corrupted by different phase offsets and noise levels.

After the data augmentation step, a preprocessing step is performed. In this step, the training examples undergo the same operations that will be applied during the prediction task on real acquired data. For this specific application, the preprocessing operations that can be performed are: k Singular Value Decomposition truncation (k-SVD) as proposed in [82] and signal normalization, where each input is normalized to have norm equals to 1. For k-SVD, let us give a matrix of simulated MRF signals $\mathbf{D} \in \mathcal{R}^{n \times t}$, where n is the number of tissue entries and t is the number of time points. The SVD decomposition of such a matrix is $\mathbf{D} = \mathbf{USV}^\top$, where $\mathbf{U}$ is an $n \times n$ matrix and $\mathbf{V}$ is a $t \times t$ matrix, and their columns correspond to the left and right singular vectors of $\mathbf{D}$, respectively. $\mathbf{S}$ is a diagonal matrix containing the singular values. $\mathbf{P} = \mathbf{DV}$ is an $n \times t$ matrix that is the projection of $\mathbf{D}$ in the base of the right singular vectors. Then, only the first k columns of $\mathbf{P}$ are taken. In this work k was set equal to 10, and the matrix $\mathbf{V}$ was computed in advance using the same training data set used for NN training. Setting the degree of the truncation to $k = 10$ produce an energy ratio of 0.994, meaning that basically all the meaningful information is maintained. The rationale for using k-SVD truncation for noise reduction is that by keeping just the first principal components of the signal, which are ranked according to the amount of variance that each component can explain, it is possible to significantly decrease the noise [3] since most of the variance is explained by the underlying signal pattern, which is modulated by the MR parameters. The noise, instead, explains less amount of the total variance and therefore it is projected onto the minor principal components.

In the last step, the examples are fed to the NN to perform the back-propagation algorithm for weights optimization. In this last step, one decides whether to feed the network with the magnitude of the signals, or to concatenate the real and imaginary parts of the MRF signals to have a complex input.

Two training data sets were generated (one per acquisition scheme) by simulating the MRF signals from 120 000 (T_1, T_2) pairs, which were sampled from an uniform distribution,

Label	Data Augmentation	Preprocessing	Input Type
P1	Additive Noise	Normalization	Complex
P2	Additive Noise	Normalization	Magnitude
P3	Additive Noise Phase Augmentation	Normalization	Complex
P4	Additive Noise	k-SVD Normalization	Magnitude
P5	Additive Noise Phase Augmentation	k-SVD Normalization	Complex

Table 3.1: Summary of the different pipelines utilized to train the NN model

Noise Model	Variance Sampling Method	Label
Additive	SNR $\in$ [2, 100]	P5
	$\sigma^2 \in$ [0.0002, 0.01]	P5.1
Multiplicative	$\sigma^2 \in$ [0.01, 2]	P5.2
Multiplicative + Additive	$\sigma^2_{add} \in$ [0.0002, 0.01] and $\sigma^2_{multi} \in$ [0.01, 2]	P5.3

Table 3.2: Variance sampling methods for the training pipeline P5.

where $T_1 \in [10, 4000]$ ms and $T_2 \in [1, 3000]$ ms.

In Table 3.1, the training pipelines used to train the NN model are reported. For pipeline P1, P2, P3 and P4 only Gaussian additive noise was used, and the variances during training were generated by randomly sampling the SNR values in the range [2, 100]. Whereas, for pipeline P5 all of the three noise models were tested, and Tab. 3.2 reports the ways in which the variances were sampled during the training procedure for the different noise models tested. The table also reports the labels that will be used in the following sections.

The training procedures were supervised, using the Mean Squared Error (MSE) between NN estimated parameters and ground truth parameters as loss function, and the Adam algorithm [63] was used as optimizer. In particular, 500 epochs with 1000 gradient steps for each epoch were used, with a fixed batch size of 500. Initial learning rate was 1×10^{-4}. A cluster with 16 dual cores CPUs was used to run the NN application, which was developed using the Python package Keras with the TensorFlow [1] back-end.

3.2.3 In-vivo experiments, quantitative map reconstructions and data analysis

The MRI acquisitions were performed using a 3.0 T GE SIGNA Premier scanner. Three volunteers were scanned under IRB approval using an 8-channel head coil (FOV 22 x 22 cm, resolution 1.2 mm, slice thickness 5 mm). Two images of the same volunteer's brain slice were acquired using two different acquisition procedures. One image was acquired using the MRF-FISP sequence with L = 1000 with one spiral interleave per TR (bandwidth = +/- 125 kHz and readout length of 3.5 ms), which results in undersampling the k-space. The other image was acquired using the sequence with L = 400, and it was a fullysampled MRF scan, which means 36 spiral interleaves were acquired. More specifically, this means that the standard undersampled MRF scan is repeated 36 times rotating the starting spiral angle to acquire missing k points and produce a fullysampled k space acquisition.

Quantitative T_1 and T_2 maps were reconstructed using a conventional dictionary approach using a fine grain dictionary. Two high-resolution dictionaries were created, according to the different MRF-FISP sequence, by sampling the (T_1, T_2) parameter space using the following grid: T_1 = [10:2:500, 505:5:1000, 1010:10:4000] ms and T_2 = [1, 2:2:500, 505:5:1000, 1010:10:3000] ms, and imposing the constrain $T_2 \leqslant T_1$. In addition, T_1 and T_2 maps were reconstructed using the NN models trained with the different pipelines and compared with those obtained using the dictionary approach. To evaluate the goodness of the NN reconstructions local and global error estimators were used. The pixel-wise absolute-percentage-error (APE) has been used as local error estimator (eq 3.7). Mean-absolute-percentage-error (MAPE) and root-mean-squared-error (RMSE), computed according equations 3.9 and 3.10, respectively, were used as global error measurements. Finally, Lin's concordance coefficient [74] was used as global measure of agreement between the NN and the dictionary approaches. The scalp region was not considered for the purpose of error analysis.

$$APE = |\frac{\widehat{p_k} - p_k}{p_k}| \times 100 \tag{3.7}$$

$$MPE(\%) = \sum_{k=1}^{N} \frac{\widehat{p_k} - p_k}{p_k} \times 100 \tag{3.8}$$

$$MAPE(\%) = \sum_{k=1}^{N} |\frac{\widehat{p_k} - p_k}{p_k}| \times 100 \tag{3.9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\widehat{p_k} - p_k\right)^2} \tag{3.10}$$

For both the fullysampled and undersampled acquisitions, noise characterizing the temporal signal evolution was studied by selecting a background region of interest (ROI). The pixel intensities across the temporal dimension were stored into a histogram with 100 bins. Kolmogorov-Smirnov test was used to quantitatively test the normality of the noise distribution, and quantile-quantile plot (QQ plot) was used to graphically asses the plausibility of this hypothesis. Moreover, the Fourier spectrum of the temporal noise was computed.

3.2.4 Experiment design

The main aim of this work is to investigate the key steps that have to be included during training to promote accuracy and robustness in a NN model trained with simulated data. Since many variables are considered, it is important to clearly state the rationales underlying the design of the conducted experiments.

As common practice in building NN applications data sets are divided in training, validation and test sets. The training set is composed of simulated data, while validation and test sets are composed of real MRF acquisitions, which are described in the above sections. The validation set is used to select the best training pipeline among those considered, and it is composed of the fully sampled and undersampled MRF acquisitions of volunteer #1. The test set is used to estimate the actual performance of the NN model trained with the selected pipeline. The undersampled acquisitions of volunteers #2 and #3 constitute the test set.

Given the considerably high number of training pipelines considered, it is worth pointing out that the way in which the validation set is used to discriminate the best training pipeline, follows an ablational study design. Both training pipeline features and complexity of the validation data are progressively added, so that only features in the pipeline that produce a significant improvement in the NN performance are promoted to the next level. Hence, the fully sampled acquisition is firstly used to assess the performance of the training pipelines P1, P2 and P3 with the main aim of assessing the role of the phase augmentation step when the NN handles complex data as input. These pipelines do not include the SVD de-noising preprocessing step and only Gaussian additive noise is used as model. Indeed, fullysampled data are characterized by high SNR, and the main contribution to temporal noise is expected to be the thermal noise that is well modeled by white additive Gaussian noise. The best pipelines are then promoted to the next step, where the undersampled acquisition is used. At this point data are affected from severe undersampling artifacts, and

additive Gaussian noise is not expected to be a good noise model. Hence, the performance of pipelines P3, P4 and P5 are assessed by comparing the NN reconstructions with those obtained using the dictionary approach. Pipeline P3 does not include the SVD de-noising filter, and P4 and P5 differ in the type of inputs used to feed the NN, which are magnitude and complex, respectively. With this experiment the main aims are to assess the role of the SVD de-noising filter, and to identify the best type of input to be used to feed the NN when dealing with undersampled data. The best training pipeline is then promoted to the last step in which different noise models are considered: only additive Gaussian noise; only multiplicative Gaussian noise and a combination of additive and multiplicative Gaussian noises. The three training pipelines are called P5.1, P5.2, P5.3. Moreover, the variable variance strategy is carried out by sampling the variance from a uniform distribution in the range $[\sigma^2_{min}, \sigma^2_{max}]$, with ranges expressed in Tab. 3.1, instead of sampling the SNR values from a uniform distribution. Since there is no guarantee that the dictionary approach provides the best estimation accuracy in presence of undersampling, selecting the best pipeline using the agreement with dictionary reconstructions of undersampled data as a metric would not be accurate. Indeed, in the absence of evident deviations between the T_1 and T_2 values estimated with the NN approach with those commonly reported in the literature for brains, deviations from the dictionary approach are not immediate evidence of worst accuracy. For this reason, the performance of both the NNs and dictionary reconstructions are evaluated by a comparison with T_1 and T_2 maps obtained with the dictionary matching approach performed on the fully sampled acquisition.

3.3 Results

3.3.1 Validation set: fullysampled acquisition

The noise histogram of the temporal evolution of the signal within a background ROI is reported in Fig. 3.4 (a) along with its Q-Q plot (Fig. 3.4 (b)). The Q-Q plot suggests that the noise histogram well fit a Gaussian distribution, and the K-S test quantitatively confirms the graphical suggestion: the test cannot reject the null hypothesis (histogram coming from a Gaussian distribution) with 95% of confidence. The Fourier spectrum of the background temporal noise is reported in Fig 3.4 (c), which does not clearly highlight specific frequencies contributing to the noise spectrum.

In Figure 3.5, T_1 and T_2 absolute percentage error maps between MR parameter maps evaluated using the NN and the dictionary approaches are reported for pipeline P1 (Fig. 3.5 (a) and (d)), P2 (Fig. 3.5 (b) and (e)) and P3 (Fig. 3.5 (c) and (f)). When the NN is

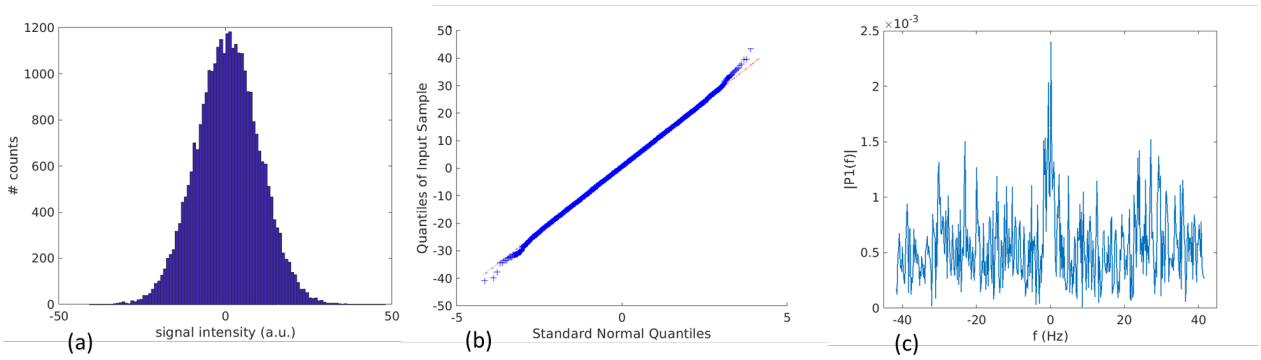


Figure 3.4: Analysis of a background ROI of the fullysampled acquisition. (a) histogram, (b) QQ plot and (c) Temporal Fourier spectrum.

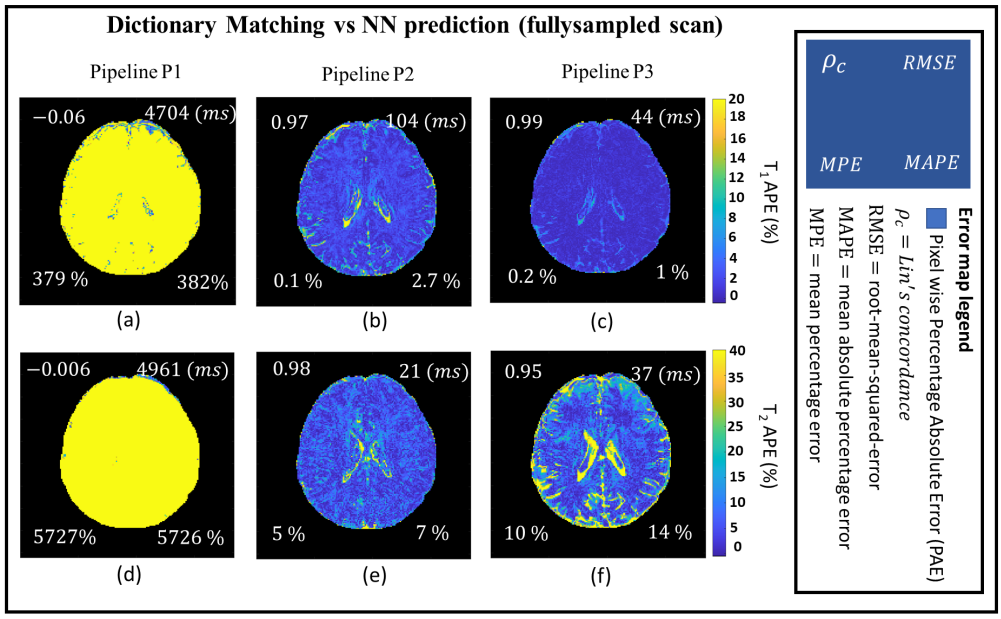


Figure 3.5: Performance estimation of the NN T_1 and T_2 predictions for the fullysampled acquisition using different training pipelines, where T_1 and T_2 maps obtained with the dictionary approach are used as GT. The noise model used during NN training was complex Gaussian additive.

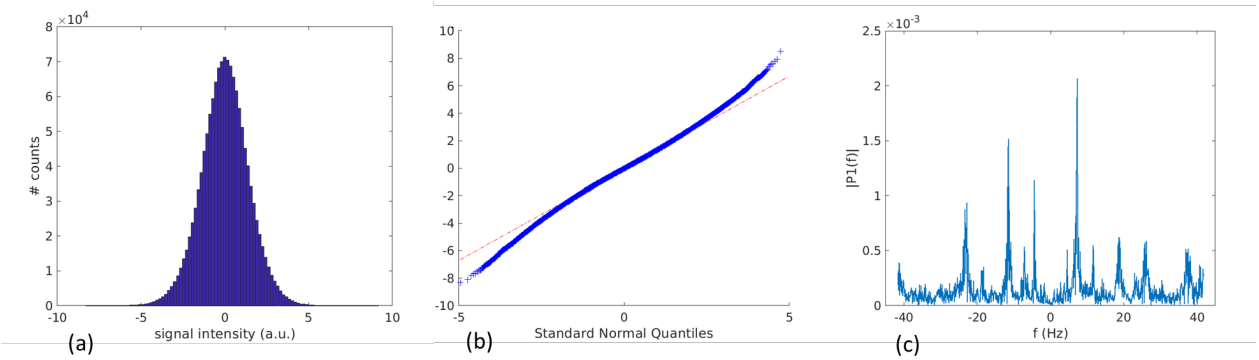


Figure 3.6: Analysis of a background ROI of the undersampled acquisition. (a) histogram; (b) QQ plot and (c) temporal Fourier spectrum.

trained using pipeline P1, both T_1 and T_2 maps show severe deviations from the GT maps (MAPE $>$ 300 % and RMSE $>$ 4000 ms), with high bias and no correlation with the GT values, as highlighted by the MPEs (grater than 300 %) and Lin's Coefficients, which are close to 0 both for T_1 and T_2. On the contrary, when the NN model is trained with pipeline P2 the NN model shows an overall good agreement with the dictionary matching approach as highlighted by the high values of the Lin's coefficients, which are $\rho_c = 0.97$ for T_1 and $\rho_c = 0.98$ for T_2. The global error estimators show that the NN model produces T_1 values that are more in accordance with those obtained using the dictionary approach (RMSE = 104 ms, MPE = 0.1 % and MAPE = 2.7%) in respect to T_2 values, which show higher deviation from the GT values (RMSE = 21 ms, MPE = 5 % and MAPE = 7 %). An overall agreement between NN and dictionary approaches is also found when the NN is trained with pipeline P3, as proved by the Lin's coefficients, which are $\rho_c = 0.99$ for T_1 and $\rho_c = 0.95$ for T_2. As it is the case for pipeline P2, The global error estimators show that the NN model produces T_1 values more close to those obtained using the dictionary approach (RMSE = 44 ms, MPE = 0.2 % and MAPE = 1%) in respect to T_2 values, which show higher deviation from the GT values (RMSE = 37 ms, MPE = 10 % and MAPE = 14 %). Focusing on Fig. 3.5 (f), the pixel-wise APE allows one to identify the CSF region as the one presenting the highest deviations from the GT values for T_2.

3.3.2 Validation set: undersampled acquisition

The noise histogram of the temporal evolution of the signal within a background ROI is reported in Fig. 3.6 (a), along with its Q-Q plot (Fig. 3.6 (b)). The Q-Q plot suggests that the noise histogram does not well fit a Gaussian distribution, with tails of the histogram showing not negligible deviation from the normal distribution. The K-S test quantitatively confirms the graphical suggestion: the test rejects the null hypothesis (histogram coming from

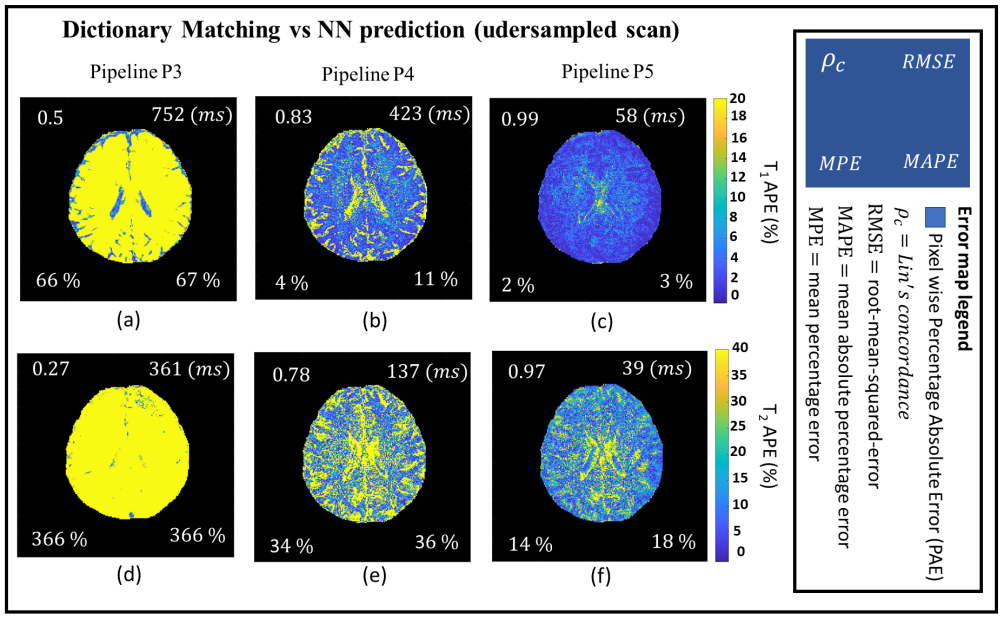


Figure 3.7: Performance estimation of the NN T_1 and T_2 predictions for the undersampled acquisition using different training pipelines, where T_1 and T_2 maps obtained with the dictionary approach are used as GT. The noise model used during NN trainig was complex Gaussian additive.

a Gaussian distribution) with 95% of confidence. The Fourier spectrum of the background temporal noise is reported in Fig. 3.6 (c), which clearly highlights specific frequencies mostly contributing to the noise spectrum.

In Fig. 3.7, T_1 and T_2 error maps estimated using NN models trained with training pipelines P3, P4 and P5, where complex white Gaussian additive noise has been used as noise model, are presented. The variances were generated by randomly sampling the SNR values. When the NN model is trained with pipeline P3, it shows overall poor performance in parameter estimation. Indeed, error maps (Fig. 3.7(a) and (d)) show severe deviations from the GT values, whit T_2 showing the highest error values (MAPE = 366% and RMSE = 361 ms) and the worst level of agreement with the dictionary estimation ($\rho_c = 0.27$). Using pipeline P4 (Fig. 3.7 (b) and (e)) there is an overall improvement of the NN performance, which is highlighted by the decreasing of MAPEs and RMSEs (T_1: MPE = 4 %, MAPE = 11 % and RMSE = 432; T_2: MPE = 34%, MAPE = 36% and RMSE = 137 ms) and the increasing of Lin's coefficients ($T_1\rho_c = 0.83$ and $T_2\rho_c = 0.78$). However, the NN model presents the best performance in parameter map reconstruction when trained with pipeline P5 (Fig. 3.7 (c) and (f)). In particular, T_1 shows a nearly perfect agreement with the GT map ($\rho_c = 0.99$, MPE = 2 %, MAPE = 3% and RMSE = 58 ms), whereas T_2 estimation shows higher errors (MPE = 14 %, MAPE = 18 % and RMSE = 39 ms) but high level of

agreement with the GT values as highlighted by the Lin's coefficient being equal to 0.97.

Figure 3.8 reports the analysis of the performance of the NN models in reconstructing T_1 and T_2 maps (from the undersampled acquisition) when they are trained with pipelines P5.1, P5.2 and P5.3. The upper box reports the error analysis in comparison with the GT maps reconstructed from the fully sampled acquisition using the dictionary approach. Hence, for the analysis reported in the upper box, the performance of the dictionary approach are also evaluated. Looking at the errors it is possible to observe that either the dictionary and the NN approaches produce T_1 and T_2 maps with considerable deviations from the GT values estimated from the fullysampled acquisition: MAPE is never below 11 % and 28 % for T_1 and T_2, respectively. Overall the deviations are higher for T_2 rather than T_1 estimations, with the CSF region reporting the highest deviations. However, among the NN models, the overall best agreement with GT values is reached when the NN is trained with pipeline P5.1, where only additive Gaussian noise is used (T_1: $\rho_c = 0.94$ and RMSE = 226 ms; T_2: $\rho_c = 0.79$ and RMSE = 88 ms). On the contrary, the worst performance are relatively to the NN trained with pipeline P5.2, where only Gaussian multiplicative noise is used (T_1: $\rho_c = 0.93$ and RMSE = 247 ms; T_2: $\rho_c = 0.69$ and RMSE = 127 ms). Pipeline P5.3, where the additive + multiplicative noise model is used, performs better than pipeline P5.2 and worst than pipeline P5.1 (T_1: $\rho_c = 0.94$ and RMSE = 235 ms; T_2: $\rho_c = 0.74$ and RMSE = 111 ms). Finally, the performance of the dictionary approach result nearly identical to those of the NN model trained with pipeline P5.1. This latter result is better highlighted when looking at the bottom box of Fig. 3.8, which reports the comparison between NN and dictionary T_1 and T_2 reconstructions performed on the undersampled acquisition. The agreement between dictionary and NN estimations is nearly perfect when the NN model is trained with pipeline P5.1. Lin's coefficient is 0.99 both for T_1 and T_2, with MAPE = 1.5 % and MAPE = 6% for T_1 and T_2, respectively. Moreover, the agreement is considerably higher, especially for T_2 , than the case in which the NN is trained with pipeline P5, where the additive Gaussian is the noise model used, but the variances are sampled by sampling the SNRs values from the uniform distribution in the range [2, 100] (See Fig. 3.7 (c) and (f)). In Fig. 3.9, T_1 and T_2 maps reconstructed from using the dictionary approach and the NN model trained with pipeline P5.1 is reported to visually appreciate the close agreement between the two methods.

3.3.3 Test set

In Figure 3.10, T_1 and T_2 errors maps between T_1 and T_2 maps reconstructed from the undersampled acquisition of volunteers numbers 2 and 3 using the NN model trained with pipeline P5.1 and those obtained using the dictionary approach are reported. Overall there is

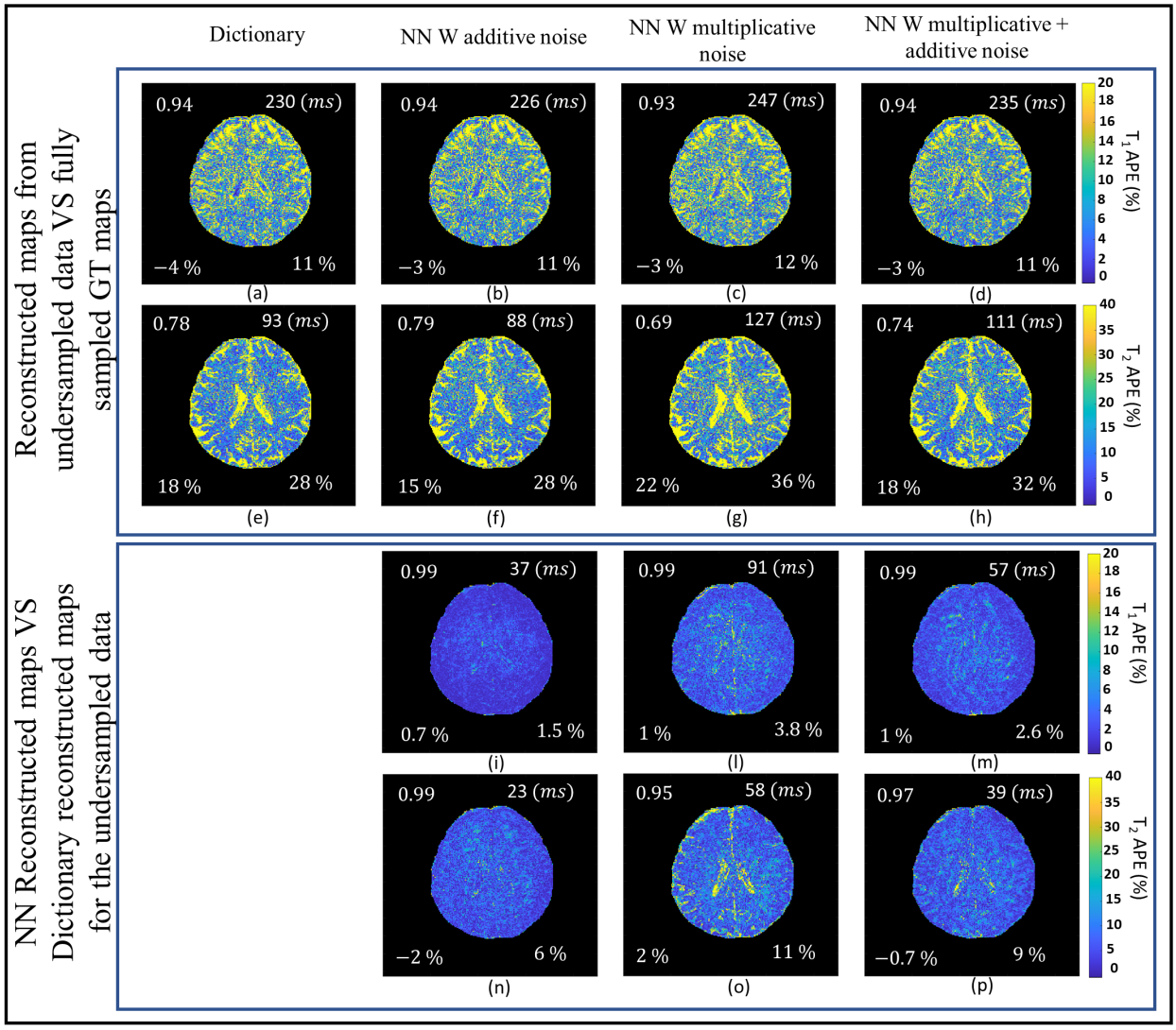


Figure 3.8: T_1 and T_2 parameter maps reconstructed from the undersampled acquisition of volunteer #1 compared to: (in the top box) the GT maps reconstruced from the fully sampled acquisition using the standard dictionary approach; and (in the bottom box) GT maps reconstructed from the undersampled acquisition using the standard dictionary approach. The NN reconstructions are performed with NN models trained with pipelines P5.1, P5.2 and P5.3 (i.e. three different noise models were used and the variances were directly sampled from a uniform distribution following the ranges reported in Tab. 3.2).

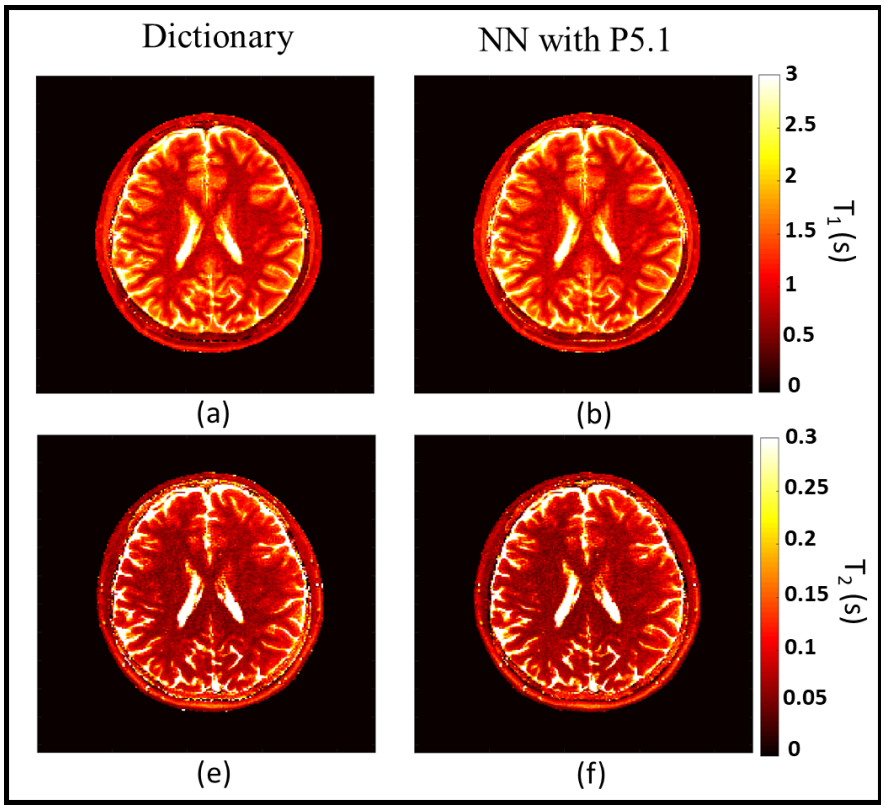


Figure 3.9: T_1 and T_2 maps reconstructed from the undersampled acquisition of volunteer number 1 using the dictionary approach (left column) and the NN model trained with pipeline P5.1 (right column).

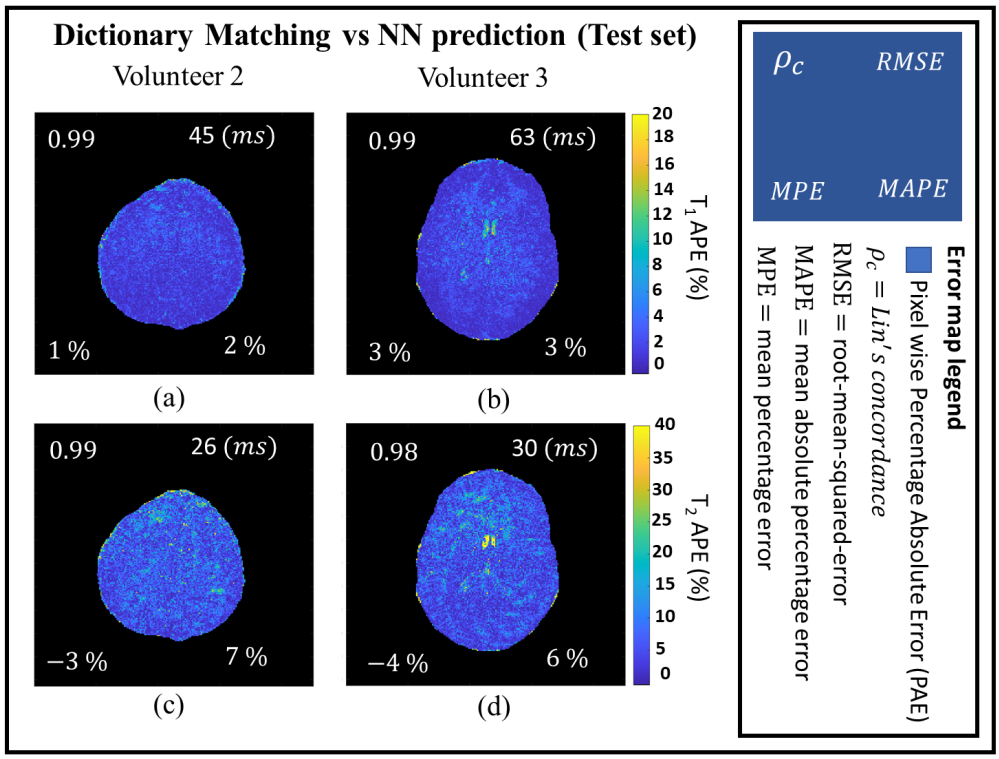


Figure 3.10: T_1 and T_2 errors maps between T_1 and T_2 maps reconstructed from the undersampled acquisition of volunteers number 2 and 3 using the NN model trained with pipeline P5.1 and those obtained using the dictionary approach.

a very high agreement between the NN and the dictionary reconstructions. Indeed, $\rho_c \geqslant 0.98$ for all the reconstructions, indicating very low bias between the two measurements. Moreover, MAPE is always less or equal than 3 % for T_1 and less or equal than 7 % for T_2, indicating relatively low dispersion.

3.4 Discussion

The results regarding the fullysampled acquisition show that when the complex MRF signal is used as input of the NN, a phase augmentation step during training is necessary to produce accurate results in the prediction task. With reference to Fig. 3.5, the NN model trained with pipeline P1, which does not include the phase augmentation step, shows very poor agreement with the GT dictionary matching approach (ρ_c = -0.06 and ρ_c = -0.006 for T_1 and T_2, respectively). On the contrary, with training pipeline P3 there is a strong agreement with the dictionary matching approach ($\rho_c = 0.99$ and $\rho_c = 0.95$ for T_1 and T_2, respectively). The best performance in terms of agreement and global errors against the GT maps is when the NN model is trained with pipeline P2, which uses the magnitude of the MRF signal and does not have to take into account the phase of the signal. Moreover, it is worth observing that even without using a de-noising filter as prepossessing step, the NN is able to implicitly build robustness to noise by adding Gaussian additive noise to clean simulated signals data during training. This result is somehow expected because, as shown in Fig. 3.4, white Gaussian noise can well represent the noise affecting real MRF signals when the k-space is fullysampled.

The results reported in Fig. 3.7 highlights that a de-noising filter is required when the NN is applied to undersampled data. As reported in Fig. 3.6, the noise affecting temporal signal evolution is not well described by white Gaussian noise, as expectable. Hence, NN model trained with pipeline P3 results in poor parameter estimation performance ($\rho_c = 0.5$ and 0.27 for T_1 and T_2, respectively). When SVD is added to the pipeline as prepossessing step, as it is the case for pipelines P4 and P5, the robustness to undersampling artifacts significantly improves. The best results are obtained when the NN model is trained with pipeline P5, where the complex signal is used as input instead of the magnitude, showing good level of agreement with the GT maps (T1: $\rho_c = 0.99$, MPE = 2 % and MAPE = 3 %; T2: ρ_c = 0.97, MPE = 14 % and MAPE = 18 %). When the magnitude is used, as in case of pipeline P4, the NN model presents higher errors. Although using the magnitude releases from the need to train the network with different phase offsets, the loss of information impacts the robustness to the noise introduced by the undersampling.

Although results so far highlighted that pipeline P5 is the best combination of phase augmentation and preprocessing steps, the errors in T_2 estimation were not negligible. The

attention then focused on how to best add the noise during training phase to increase robustness. Results regarding NN models trained with pipelines P5.1, P5.2 and P5.3, reported in Fig. 3.8, allow to highlight further considerations. First, noise added using a variable variance strategy where the probability of sampling high variances is higher than sampling low variances (i.e. higher probability to sample low SNRs instead of high SNRs) produces more robust NNs than sampling the SNRs from a uniform distribution. This is clearly evident when comparing Figures 3.8 (l) and (o) with Figures 3.7 (i) and (l). Both APE maps are computed comparing the NN and the dictionary reconstructions, where the NNs are trained with pipeline P5 and P5.1 for Fig. 3.7 and Fig. 3.8, respectively. Performance of pipeline P5.1 outperform those relative to pipeline P5. Since the only difference between those pipelines is the way in which the noise variances are sampled during training, this validates the above statement: sampling with higher probability low SNR values produce NN more robust to noisy data rather than sampling SNR values with a uniform distribution.

Second, results reported in 3.8 show that additive Gaussian noise is the best noise model to train the NN. Indeed, the NN trained with pipeline P5.1 outperforms the NN trained with pipeline P5.2 and P5.3. Moreover, the worst performance are registered for pipeline P5.2, that uses a Gaussian multiplicative noise during training. This result is surprising at first, because the noise affecting MRF signals in presence of undersampling in general present both an additive and a multiplicative components [62,97], with the multiplicative component being the one contributing the most. Hence, one would expect that augmenting the data during training with multiplicative noise, or a combination of additive and multiplicative noise, will produce the best generalization capabilities. The fact that this is not the case needs further considerations, since causes for such result can be dived in two main categories, which produce different interpretation of the result. One explanation has to do with the difference between the training and test probability distributions. Generally, fixed the NN model and the training algorithm, NN models trained with data that are drawn from the same distribution of the test data (the actual data that the NN will be working on) show better generalization performance than models trained using training data that follow a different distribution than the test data. In this case training data are simulated, and adding noise during the data augmentation step is aimed to make the training data distribution as similar as possible to the distribution of the real acquired data, acting as model regularizer during training. Following this reasoning, the result could highlight that using additive noise is a better regularizer than multiplicative noise, i.e. noise affecting real undersampled data is better modelled, for the aim of NN training, by additive noise rather than multiplicative noise. However, literature shows that multiplicative noise is a reasonable noise model for artifacts introduced in the temporal MRF signals from undersampling the k-space [62, 97].

However, the multiplicative model introduced in [62], which was used in this study, assumes an instantaneous correlation between the noise variance affecting the signal at time index TR_i, and the signal amplitude at that index. This assumption could be not correct, with the noise variance depending from states of the system at precedent TR, as it happens in financial time series, where AutoRegressive Conditional Heteroskedasticity (ARCH) models are used to model the change of volatility affecting the time series [38]. A different factor that could explain why additive noise resulted superior to multiplicative noise lies on the heteroscedasticity nature of multiplicative noise and the assumption of homoscedasticity that both SVD, used as preprocessing step, and MSE, used as loss function for model weights optimization, do. In the additive noise model, the variance is fixed and does not change in different time points of the MRF signal, which means the noise is homoscedastic, in contrast to multiplicative noise, which means the noise is heteroscedastic. Optimizing the weights of the NN model using the MSE as loss function is equivalent to maximizing the likelihood on a linear Gaussian model (i.e. noise is additive Gaussian) [47]. Hence, when the noise affecting the data on which the NN is trained is additive, the optimization algorithm performs the best when the MSE is used as loss function. When the noise introduced during training is multiplicative this equivalence is broken and the MSE is not guaranteed to perform as good as when the noise is additive, and can end up in worst local minima. Moreover, SVD is applied as pre-processing even during training which assumes homoscedasticity as well. In such a scenario it is possible that even if the real data are corrupted by a noise that is not additive, the model weights optimized using additive noise during training produce a NN model more able to generalize than using multiplicative noise during training.

Future analysis are needed to disambiguate what is causing the superiority in the prediction performance of the NN model trained using additive in respect to the NN model trained with multiplicative noise. Possible experiments is to study cross correlation between noise affecting MRF signals and the signal itself to assess if the assumption of instantaneous correlation is valid. At the same time, how heteroscedasticity affects the learning process has to be deeper studied.

However, the results regarding the test data set presented in Fig. 3.10, show that T_1 and T_2 maps predicted by the NN trained with pipeline P5.1 are in strong agreement with those obtained with the standard dictionary approach ($\rho_c \geqslant 0.98$ for all the reconstructions). Hence, the NN application here set up predicts MR parameteric maps that are consistent with those generated using the gold standard dictionary approach.

In summary, the best parameter reconstruction performance are obtained when the NN model takes the complex MRF signal as input. To make the NN able to correctly handle the phase of real MRF signals, clean simulated data have to be augmented by multiplying

them with different phase offsets. The undersampled k-space sampling scheme produces very noisy MRF signals. To promote robustness to these artifacts noise with different variances, witht Gaussian additive as noise model to be preferred over multiplicative noise, has to be added to the simulated signal. Moreover, a critical step is to add a de-noising prepossessing step, which in this case is k-SVD truncation. With this pipeline a NN to perform pixel-wise MR parameters estimation can be trained in such a way to produce parameter maps that are in strong agreement with those obtained with the conventional dictionary approach. At the same time the NN approach required only 1.5 s to compute the maps, whereas the dictionary approach required 115 s.

Because the way in which the NN handle the problem of phase offsets does not depend on a-priory knowledge of the expected phase offsets, which is the case for phase alignment solutions [46], the presented strategy is general enough to be applied also in MRF-bSSFP based pulse sequences, where the sensitivity to off-resonances does not allow phase alignment techniques without a previous estimation of the B0 map. Finally, it is worth remarking that no real data were necessary to reach such a good performance, meaning that the lack of training data will never be a problem for future applications of this approach to MRF problems characterized by a larger number of parameter to be retrieved. Moreover, since all the data augmentation steps are done online, one does not need to store augmented data either in the hard or flash memory.

3.5 Conclusions

In this chapter, a development of the deep learning application for MRF pixel-wise reconstruction presented in chapter 1 has been presented. The main goal was to investigate optimal training strategies to promote accuracy and robustness to undersampling artifacts during training using only simulated signals as training data. The presented experiments allowed to clearly identify the data augmentation and prepossessing steps that contribute the most in building accuracy and robustness in the NN model. The T_1 and T_2 maps predicted by the NN trained with the selected training pipeline are in a very strong agreement with those obtained with the standard dictionary approach for in-vivo test data.

In conclusion, we were able to train a NN for pixel-wise MRF reconstruction using a training pipeline that make use of only simulated data, being able to equal the standard dictionary reconstruction accuracy and robustness to artifacts produced by the undersampling. The presented results may help the developing of deep learning approaches for MRF, pushing fingerprinting pulse sequences design to encode and retrieve more meaningful MR parameters (sucha as T_2^*, B_1^+, transfer magnetization etc.) with no more limitations due to

the dictionary size.

Part II

Evaluation of the Trabecular Bone Volume Fraction through NMR relaxometry using Single-Sided Devices

Chapter 4

Introduction

4.1 General overview and outline of the problem

Diseases associated with an increase of bone fragility and susceptibility to fractures, like osteoporosis, are a major public health concern, with a high impact on the quality of human health [5, 49, 66]. They determine an impoverishment in the elderly people quality of life, and an increase in the costs of health care. Methods to improve the early detection of these diseases are objects of investigation for the national and international health organizations. Wide campaigns of screening of the population at risk are desirable to study diseases evolution and, ideally, to prevent the worsening of the bone conditions. This implies the development of techniques which can be applied easily, at low cost and based on portable devices.

Dual energy X-ray absorptiometry (DXA) is considered the gold standard for the diagnosis of osteoporosis. It employs ionizing radiations and provides a measure of the areal bone mineral density (BMD) but does not provide information about the micro-structure of the bone [14, 49] Nevertheless, the resilience of the trabecular bone depends not only on BMD but also on the architecture of the trabecular structure. Epidemiological studies have shown that BMD is a poor predictor of the fracture risk, being the low values of BMD able to explain about the 50 of the incident fracture cases [105].

Although micro computed tomography (micro-CT) is particularly suited to investigate the structure of calcified tissues at different scale lengths, it requires a higher X-ray dose than DXA and it is usually performed on small samples. Hence, micro-CT is not feasible in vivo.

NMR is well suited to study bone structural properties in vivo. It does not use ionizing radiations and laboratory studies have shown the ability of NMR to assess structural properties of bone with different techniques, ranging from relaxometry to diffusometry and Magic Angle Spinning [37,40,41,54,79,83,88,95,98,107]. Although clinical MRI allows one to assess bone structure in vivo [85,94,108], its cost limits the suitability of conventional MRI for rou-

tine bone screening. A way to perform more manageable MRI analyses has been proposed with the use of more compact and lower magnetic field scanners, for imaging trabecular bone structure on local body districts [42, 58]. Unfortunately, all these devices could hardly be employed for wide screening campaigns.

Portable NMR, based on low-field single-sided NMR scanners, is a promising and appealing approach to assess MR properties of biological tissues with the aim of medical applications. The main feature that differentiates a single-sided apparatus from others NMR scanners is the detection of the signal from a sensitive volume external to the magnet. Portable NMR has also the added advantages of low purchasing and running cost and low maintenance, largely due to the absence of superconducting magnets (they use permanent magnets). Portable NMR scanners, such as the NMR-MOUSE [15] (Magritek, New Zealand), are commercially available and have been used in a number of biomedical applications including testing of silicone breast implants [67] and studies of various biological tissues, including tendon [84], articular cartilage [91, 92], skin [12, 69], breast [2].

Recently we proposed a method to assess the bone-volume-to-total-volume ratio (BV/TV), also called bone volume fraction, an important parameter to evaluate the micro-structure of the trabecular bone, through single-sided NMR scanners [23]. The procedure is based on computing the ratio between the signals obtained from the trabecular bone and the signal obtained from a reference sample of bulk marrow. Further details of the procedure are given in the following sections. Although results reported in [23] show the feasibility of the technique, related limitations have to be addressed to push the technique toward in vivo applications.

One limitation relates to the presence of soft tissues other than trabecular bone that can be intercepted by the sensitive volume of the single-sided NMR scanner. In a real scenario, especially in appendicular skeleton sites (fingers, heel, distal radius), tissues other than the trabecular bone, such as muscle, may be present within the sensitive volume and affecting the estimation of the BV/TV parameter.

An important limitation is the need of a reference sample to compute the BV/TV ratio. If the volume of the reference sample differs from the volume of trabecular bone intercepted by the single-sided NMR scanner, the evaluation of BV/TV is prone to errors. Having a procedure able to obtain information about the micro-structure of the trabecular bone using single-sided NMR scanners without the need of a reference sample would improve the robustness and the in vivo feasibility of the methodology.

This second part of my Ph.D thesis aims to propose solutions for the two limitations stated above affecting the procedure presented in [23]. The part is divided in two chapters. Chapter 5 addresses the problem of detecting the signal from the trabecular bone marrow

and suppressing the signals from other tissues (muscle and cartilage), which may be present within the sensitive volume of a single-sided NMR scanner. The goal is achieved by a signal filter which exploits molecular diffusion. Part of the material of this chapter has been recently published [7]. Chapter 6 investigates the feasibility of a new methodology to assess the micro-structural properties of trabecular bone using single-sided NMR scanners without the need of a marrow reference sample.

In the following sections background is provided to allow the reader to familiarize with important concepts that have been just briefly introduced in this section.

4.2 Background

4.2.1 Trabecular bone and Osteoporosis

Bone is a mineralized tissue that is the building tissue of human skeleton [44]. It is a living composite bio-material composed of an organic substrate, consisting largely by type I collagen (around the 40% of volume) interspersed with mineral crystals composed of hydroxyapatite (around the 45% of the volume). Water occupies the remaining volume and it is either bound to collagen or it resides in the spaces of the lacuno-canalicular system. We can classify bone in cortical bone (CB) and trabecular bone (TB). Latter is predominant in the axial skeleton and near the joints of the long bones. It consists of a network of interconnected trabeculae, which are typically 100-150 μm thick, where the inter-trabecular spaces are filled with bone marrow. Thickness of CB varies between 1 and 5 mm. This combination confers to the bone its unique mechanical properties in terms of tense and compressive strength. Bone tissue experiences different loads in intensity and directions over time, which cause an adaptive response of the bone tissue, such as altering structural density and orientation (i.e. creating anisotropy) [89]. This adaptive response is possible because of bone remodeling, a process where bone tissue is removed by osteoclastic resorption and new bone is formed by osteoblasts. In the adult skeleton, the bone remodels itself in a process that implies a dynamic equilibrium between bone formation and bone resorption. With ageing and in the setting of osteoporosis, the balance of bone resorption and formation becomes negative. The bone loss in aged and osteoporotic bone is a consequence of imbalanced and excessive bone remodelling [93], which increases fracture risk.

Since bone remodelling occurs on bone surfaces, osteoporotic bone loss is a function of surface available for bone remodelling [111]. In individuals less than 65 years of age, the trabecular bone is the largest surface available for bone remodelling. In this population,trabecular bone provides only about 20% of the skeletal bone mass but it is responsible

for most of the turnover [89]. Hence, the bone loss in early osteoporosis is mainly due to trabecular bone. With increasing age, the cortical bone becomes more porous and, therefore, its endocortical surface increases. As a consequence, the largest loss of absolute bone mass due to osteoporosis occurs in cortical bone by intracortical rather than endocortical or trabecular remodelling. These considerations allow one to appreciate the importance of assessing bone structural properties, and the importance of having techniques able to conveniently gain those information to allow early detection of osteoporotic behaviour of the bone tissue. Some of the most common parameters for the characterization of the microarchitectural structure of trabecular bone are:

- BONE VOLUME / TOTAL VOLUME , BV/TV (%): % of bone volume in respect to the total volume of the analysed tissue;
- BONE SURFACE / TOTAL VOLUME, BS/TV (%): % of bone surface in respect to the total volume of the analyzed tissue;
- BONE SURFACE / BONE VOLUME, BS/BV (%): % of bone surface in respect to the bone volume of the analyzed tissue;
- TRABECULAR THICKNESS, Tb.Th (mm), is the mean trabecular thickness found in a sample;
- TRABECULAR SEPARATION, Tb.Sp (mm), is the mean space between trabeculae;
- TRABECULAR NUMBER, Tb.N (mm^{-1}), is the mean number of trabeculae per mm;
- TRABECULAR PATTERN FACTOR, Tb.Pf, evaluates the connectedness of bone structures, which can be described by the relation of convex to concave bone surfaces. A lot of concave surfaces represent a well connected spongy lattice, whereas a lot of convex surfaces indicate a badly connected trabecular lattice in two-dimensional sections [51];

4.3 Single-sided NMR scanners

Single-sided NMR devices detect the signal from a sensitive volume positioned above the surface of the magnet and the RF coil, which implies that there are not strict requirements about the size of the sample to be investigated. They use permanent magnets to generate the polarizing field B_0, which is normally in the order of hundreds of mT, with a significant small size compared to high-field NMR and whole body MRI scanners. All together, these

characteristics make single-sided NMR scanners portable, low-maintenance and suitable for *in-situ* measurements. In the works presented in this Ph.D thesis, two different single-sided NMR devices were used: the NMR-MOUSE PM10, produced by Magritek, and the NMR-MOLE, a prototype developed by Manz et al [80] and currently present only in the NMR group of the Victoria university of Wellington (New Zealand).
The main difference between the two scanners is the design of permanent magnet, which determines the polarizing magnetic field. The NMR-MOUSE PM10 has been designed in such a way that a strong constant field gradient (G) is present in the static magnetic field. This implies that the shape of the sensitive volume is in first approximation rectangular, with the thickness of the sensitive volume being straightforwardly related to the excitation bandwidth. Such sensitive volumes can be referred as "1D" sensitive volumes. The magnets array of the NMR-MOLE generates a static magnetic field with a sweet spot. In this case, the first-order components of the field (linear terms) are zero and strong quadratic and higher order terms dominate the spatial variations resulting in a complicated "3D" sensitive volume where a distribution of constant field gradient is present. The strength of this gradients is much smaller than the gradient of the NMR-MOUSE PM10, resulting in a bigger sensitive volume size. The characteristic of the two devices are discussed in details.

4.3.1 NMR-MOUSE PM10

The profile-NMR MOUSE (MObile Universal Surface Explorer) [15] is a single sided NMR scanner produced by Magritek.

Figure 4.1 presents a schematization of the NMR-MOUSE along with a picture of the real device. The scanner is composed by a U-shaped permanent magnet, which produces a polarizing magnetic field B_0 that in the sensitive volume (red box in the figure) is approximatrly 327 mT (Larmor frequency of 13.88 MHz). A constant gradient field (G) of 600 kHz/mm, orthogonal to the polarizing field force lines, is present inside the sensitive volume, which allows diffusivity measurements. A surface RF coil is used to both produce the pulsed magnetic field B_1, used to excite the spins of the protons present in the sample, and to detect the relaxation signal. Both B_0 and B_1 fields are highly inohomogeneous, a characteristic that has to be taken into account during designing experiments. For example, even for compounds with long T_2, such as bulk water, the Free Induction Decay (FID) decays too fastly to be acquired with the NMR-MOUSE, so that only echo-like sequences are used to acquire the signal, such as the CPMG sequence. The distance between the U-shaped permanent magnet and the sensitive volume is fixed, and it is 11 mm. The device has a lift that allows one to move down the permanent magnet, and so to move down the sensitive volume within the sample. This feature allows to perform measurements at different depths inside the sample obtaining

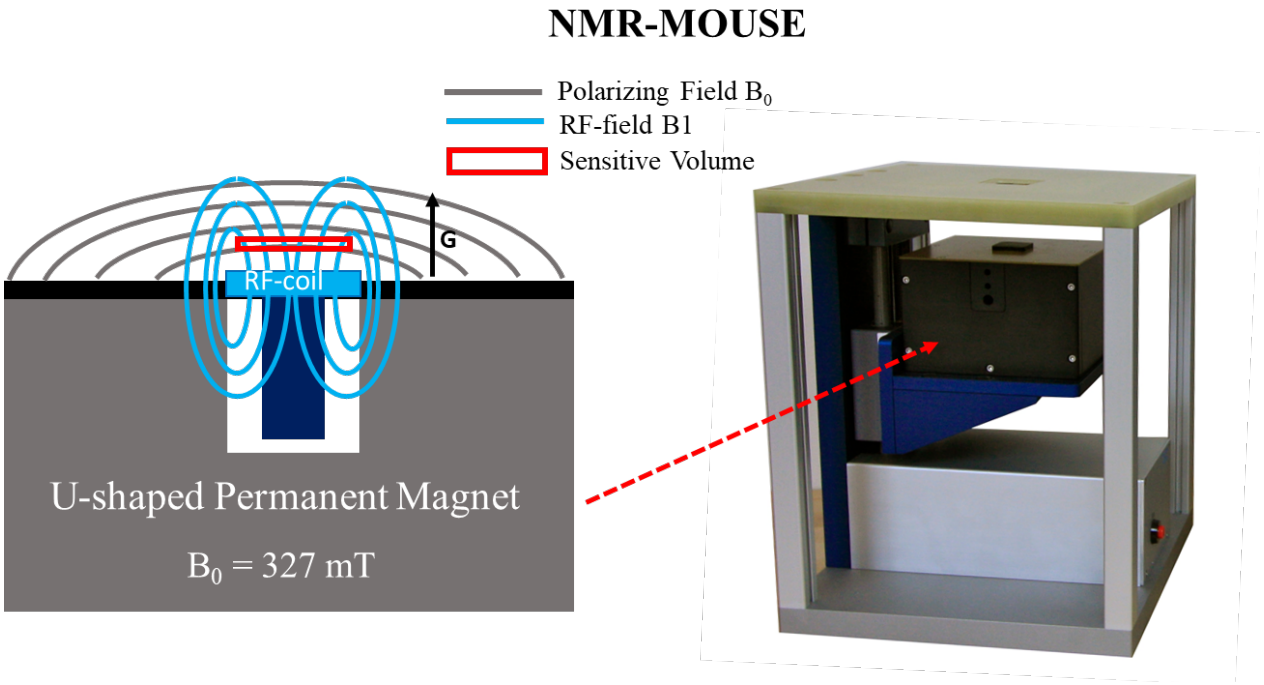


Figure 4.1: Schetck of the NMR-MOUSE design.

a profile of the analysed parameter (such as spin density or T_2 profiles). The MOUSE-PM10 comes with four external 2 mm thick spacers. In Fig. 4.2 the use of the spacers for the scanner is depicted. One can insert the spacers between the permanent magnet and RF coil to reduce the distance between the coil and the sensitive volume. Decreasing the distance between the sensitive volume and the RF coil increases SNR efficiency, but it decreases the maximum depth achievable inside the sample. The SNR efficiency is mainly determined by the thickness of the sensitive volume, which can be varied. Considering the size of the RF coil (surface of few cm^2) and the presence of a strong constant gradient along the direction perpendicular to B_0 force lines, the thickness of the sensitive volume mainly depends on the excitation bandwidth $\Delta\upsilon_{exc}$, indeed $\Delta z = \frac{\Delta\upsilon_{exc}}{\gamma G}$. $\Delta\upsilon_{exc}$ depends from the duration of the pulse, and since the efficiency of the B_1 excitation depends on the distance between the RF coil and the sensitive volume, the minimum pulse length to achieve a 90 pulse increases with the distance. In the 0 mm configuration the maximum sensitive volume thickness is around 100 μm, whereas it is around 300 μm in the 8 mm configuration.

4.3.2 NMR-MOLE

The NMR-MOLE (MObile Lateral Explorer) has been originally prototyped by Manz et al. [80]. For the experiments presented in this Ph.D thesis a slightly modified version of the prototype was used. Figure 4.3 presents a schematization of the NMR-MOLE along with

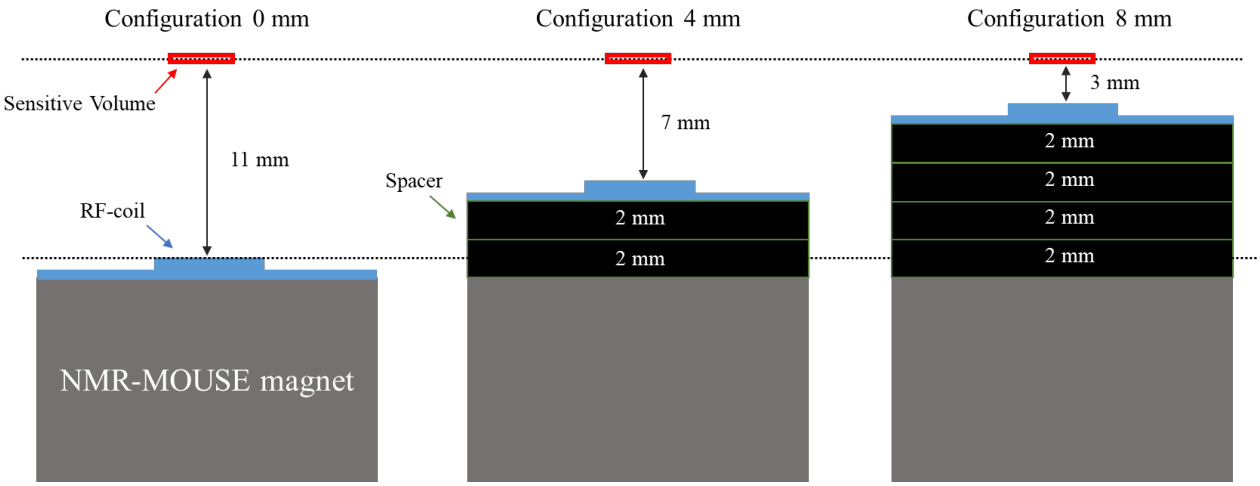


Figure 4.2: Example of spacer configurations for the NMR-MOUSE PM10

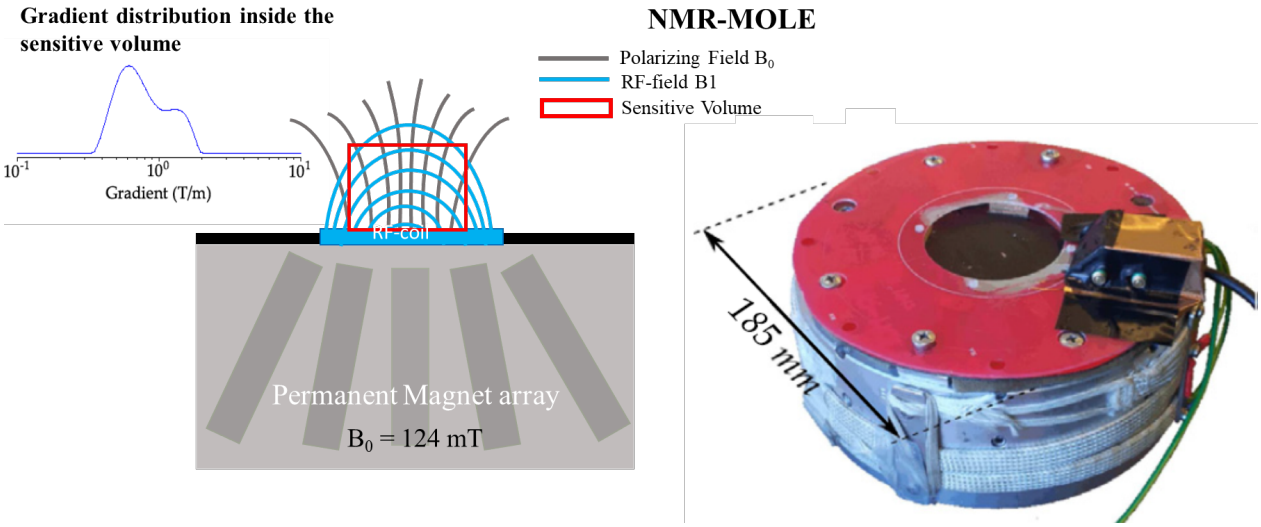
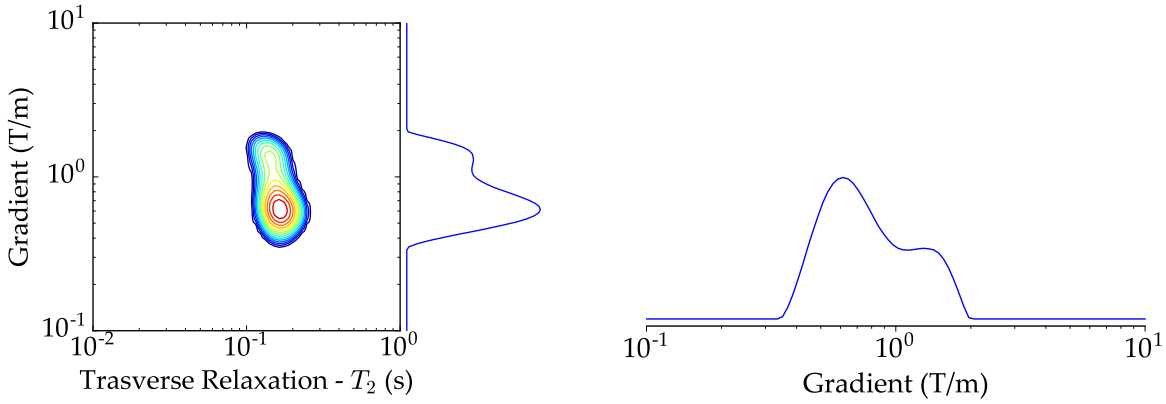


Figure 4.3: Schetck of the NMR-MOLE design

a picture of the real device. The NMR-MOLE presents a sweet spot in which the field is relatively homogeneous. Being a single-sided device, the magnetic field generated by the permanent magnets array remains highly inhomogeneous if compared with the non-single sided devices. The mean strength of the magnetic field is 124 mT at 32 C, corresponding to 5.26 MHz of Larmor's frequency of proton. The field gradient distribution was obtained using the method introduced by Nogueira et al [87].

The procedure relays on the assumption that the complicated field gradient shape, in first approximation, can be considered as a distribution of constant field gradients in one direction. A $D-T_2$ map of a reference sample, in this case bulk doped water, is then acquired as shown in Fig. 4.4. If just one constant gradient value is present, the resulting map would show a well defined single component (i.e. a little circle). When different constant gradient values



(a) $G - T_2$ after applying proper conversion from diffusion coefficient to gradient intensity of a $D - T_2$ map

(b) Gradient distribution

Figure 4.4: Determination of the gradient distribution of the NMR-MOLE via a Diffusion Editing CPMG pulse sequence.

are present, a broadening along the D axis will be observed. This can be considered as a signature of the device and using a customize kernel it is possible to compute the gradients distribution present in the volume occupied by the reference sample. From the distribution, the main field gradient, present in the sweet spot, is 0.6 T/m.

A PID system, which controls the current of a heat tape in contrast with the aluminium block, is used in order to keep the temperature of the NMR-MOLE stable.

4.3.3 Quantification of BV/TV ratio with single-sided NMR

In [23] we set up and validated in laboratory a procedure to assess the bone volume fraction of trabecular bone using single-sided NMR devices, NMR-MOUSE PM10 and NMR-MOLE in particular. In this section a brief review of the study is given, but reader is encouraged to read the publication for more details.

The key idea behind this technique is borrowed by a widely validated methodology in the field of NMR applied to the study of porous media. A way to estimate the porosity of a rock sample is to compute the ratio between two signal intensities. First, the rock sample under investigation is saturated with distilled water and a CPMG sequence is used to acquire the signal, which comes from the water inside the pores of the rock sample. Then, the same measurement is performed on a sample of bulk water characterized by the same volume of the rock sample. As long the volume of the samples are the same, the ratio of the two signals (where the signal is the total area under the T_2 distribution computed out of the CPMG

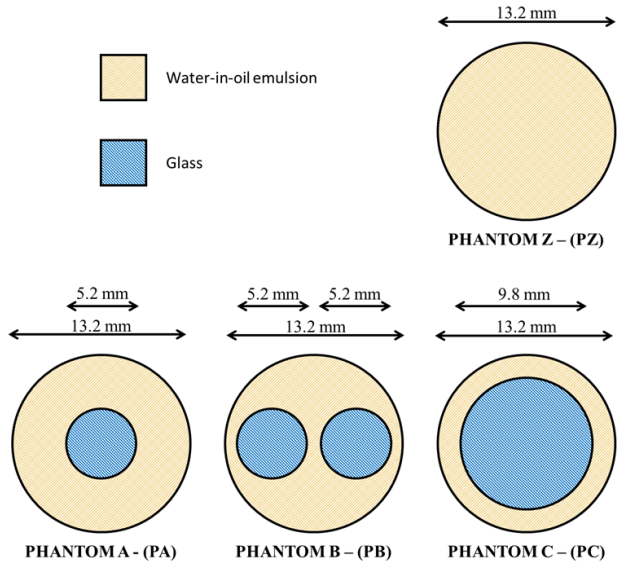


Figure 4.5: Sketch of the phantoms used for the calibration experiments in [23]

decay) is the percentage of water inside the rock sample, i.e the porosity of the sample. In trabecular bone the inter-trabecular spaces are filled with bone marrow so that the bone volume fraction can be evaluated according equation 4.1, where S_{bone} is the NMR signal from the marrow inside the bone sample and $S_{bulk\,marrow}$ is the NMR signal from a reference bulk marrow sample.

$$\frac{BV}{TV} = \left(1 - \frac{S_{bone}}{S_{bulk\,marrow}}\right) \times 100\,(\%) \tag{4.1}$$

In [23] the methodology was validated on a set of phantoms of which ground thruth BV/TV was easily computable by the geometry of the system, and on a set of six trabecular bone samples where the ground truth BV/TV was computed using micro-CT.

Four phantoms were created by filling a cylindrical glass container (13.20 $\pm$ 0.05 mm diameter) with a water-in-oil emulsion. Glass tubes (which are invisible to NMR) of different known diameters were embedded in three of the four phantoms, as depicted in Fig 4.5, to obtain samples with known fluid volume/total volume (complement to 100 of the simulated BV/TV) on a wide range of values. Six cylindrical samples of pig bone were used for validation, of approximate diameter 1 cm and height 1.5 cm, cored from different sites of the pig shoulder. Bulk marrow samples were extracted from the tibia of the pig. The pipeline of the experiment is summarized in Fig. 4.6. For each sample, the signal was acquired using a CPMG sequence and the signal intensity was evaluated as the total area under the T_2 distribution computed out of the echo train decay through an inverse tranform using the software UpenWin [21], which implements the inversion algorithm UPEN (Uniform PENalty) [18,19].

The results of the validation experiments presented in [23] are reported in Fig. 4.7.

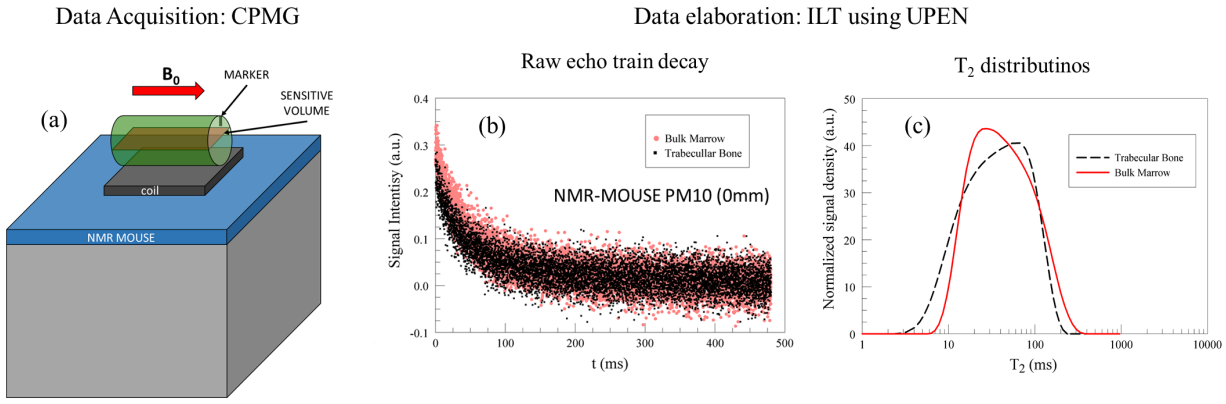


Figure 4.6: Sketch of the pipeline used to evaluate the BV/TV ratio of trabecular bone in [23]

showing high grade of correlation both for phantoms ($r^2 = 0.999$) and trabecular bone samples ($r^2 = 0.932$).

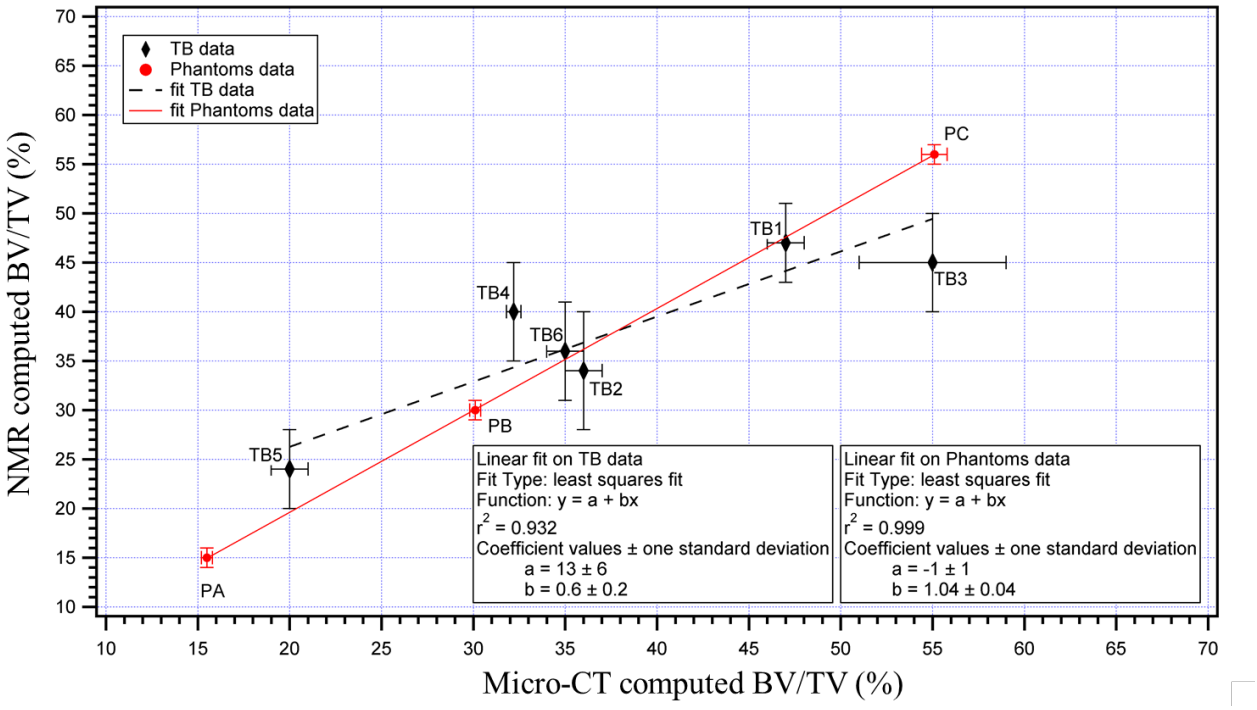


Figure 4.7: BV/TV values of TB and phantoms samples measured by the NMR-MOUSE PM10 vs the Nominal values. (∗) For the TB samples the nominal value are the BV/TV values determined by the micro-CT analysis, whereas for the phantoms they have been obtained by the geometry of the samples. Figure is taken from [23]

Chapter 5

Diffusion Weighted Pulse Sequences for the Estimation of Trabecular Bone Volume Fraction in the Presence of Muscle Tissue

5.1 Introduction

In [23] the bone volume fraction is estimated according to equation 4.1, here rewritten as equation 5.1:

$$\frac{BV}{TV} = \left(1 - \frac{S_{\text{Sample}}}{S_{\text{Reference}}}\right) \times 100\ (\%) \tag{5.1}$$

where S_{Sample} is the NMR signal from the material inside the sample (i.e. the marrow inside the inter-trabecular spaces of the trabecular bone) and $S_{\text{Reference}}$ is the NMR signal of a bulk marrow sample. Provided the intersections of these two volumes with the sensitive volume are the same, this procedure can be used in laboratory studies to measure the BV/TV of bone specimens in a non-invasive way. In *in-vivo* clinical applications, in the appendicular skeleton sites (fingers, heel, distal radius), tissues other than the trabecular bone, such as muscle, may be present within the sensitive volume, so that it is critical to isolate only the signal coming from the trabecular bone for a correct evaluation of the bone volume fraction. This chapter presents a modification of the procedure presented in [23] that allows detection of only the signal from the marrow in the TB suppressing signals from other tissues. The fundamental idea behind the proposed technique is to filter out the signal coming from the soft tissues during acquisition by exploiting the difference in their molecular diffusion coefficient.

NMR relaxometry and diffusometry measurements were first performed on biological tis-

sues samples to characterize their MR properties. Differences in their MR properties were then used to identify the most efficient way to isolate the signal contribution related with trabecular bone in a biological sample containing trabecular bone, articular cartilage and muscles. The new set up methodology was then validated with an *ad hoc* biological phantom, where the ground truth bone volume fraction was estimated using the technique presented in [23].

5.2 Background

5.2.1 2D correlation measurements

A powerful extension of the measurements of NMR relaxation and molecular self-diffusion in grossly inhomogeneous fields is the measurement of multi-dimensional distribution functions involving diffusion, transverse, and/or longitudinal relaxation [57]. The two-dimensional experiments allow one to extract inherently more information than the corresponding one-dimensional experiments, for example, in the case of multi-component systems, from the $T_1 - T_2$ map it is possible to extract the T_1/T_2 ratios of the single components, and from the $D - T_2$ maps it is possible distinguish components that otherwise would be undistinguishable by just performing the one-dimensional T_2 measurement, or one-dimensional D measurement. In this thesis two kinds of 2D correlation measurements have been used: the $T_1 - T_2$ correlation map and the $D - T_2$ correlation map. Because both the 2D experiments mathematically are based on the same concepts, i.e. the two-dimensional Inversion Laplace Transform (2D ILT), it is worthy to give a brief general introduction about the problem, and describe the pulse sequences used.

The measurement of 2D distribution functions of two quantities x_1 and x_2 (such as relaxation and diffusion) is based on sequences where two or more independent times are varied in such a way that the kernel of the integral reported below separates out. In that case, in fact, the measured magnetization $M(t_1, t_2)$ depends on the quantities of interest x_1 and x_2 through the equation 5.2 [57]

$$M(t_1, t_2) = \int \int f(x_1, x_2) K_1(x_1, t_1) K_2(x_2, t_2) dx_1 dx_2 \tag{5.2}$$

where $f(x_1, x_2)$ is the wanted distribution (also called *map*), and K_1 and K_2 are the kernels. Because for relaxation and diffusion NMR experiments the kernels are in the form of $\exp(-\alpha x_i t_i)$, the goal is to obtain the wanted function by a 2D ILT. The shape of the kernels depend on the specific pulse sequence, because, depending on the pulse sequence, different physical phenomena affect the spin dynamic.

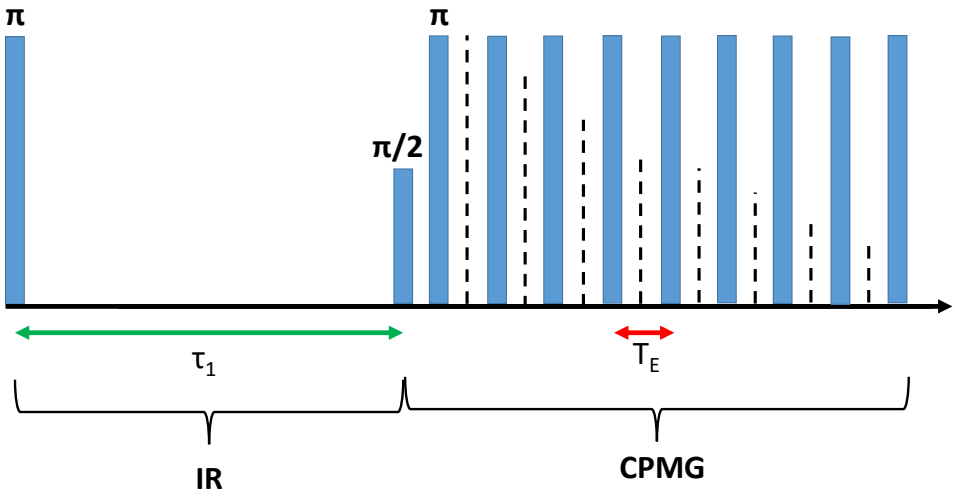


Figure 5.1: T_1 edited CPMG pulse sequence composed by an inversion recovery block followed by a CPMG signal detection block.

$T_1 - T_2$ correlation experiments

The $T_1 - T_2$ pulse sequence, or T_1 editing CPMG, can be though like a common CPMG in which the first 90 pulse is replaced by an editing sequence that allow one to make the system sensitive to the longitudinal relaxation. This can be done by stacking together an Inversion Recovery (IR) followed by a CPMG as sketched in 5.1, where the parameter τ_1 indicates the time left for T_1 recovery, i.e. the time between the 180 inversion pulse and the 90 pulse. The τ_1 is varied during the experiment. For each value of τ_1, a CPMG acquisition block is performed where T_E is kept constant among different acquisitions and as short as possible to avoid effect of diffusion during the CPMG decay. In case of $T_1 - T_2$ sequence, equation 5.2 becomes:

$$M(\tau_1, kt_E) = \int\int f(T_1, T_2) \left(1 - 2\exp\left(-\frac{\tau_1}{T_1}\right)\right) \exp\left(-\frac{kt_E}{T_2}\right) dT_1 dT_2 \tag{5.3}$$

and $f(T_1, T_2)$ is the 2D density distribution of the relaxation times, computable by the 2D ILT.

$D - T_2$ correlation experiments

The $D-T_2$, or diffusion edited CPMG, pulse sequence is represented in 5.2 and it is composed of a stimulated spin echo (SSE) followed by a CPMG detection block. The SSE is the editing sequence in which the effect of the diffusion is encoded. With the SSE block, the

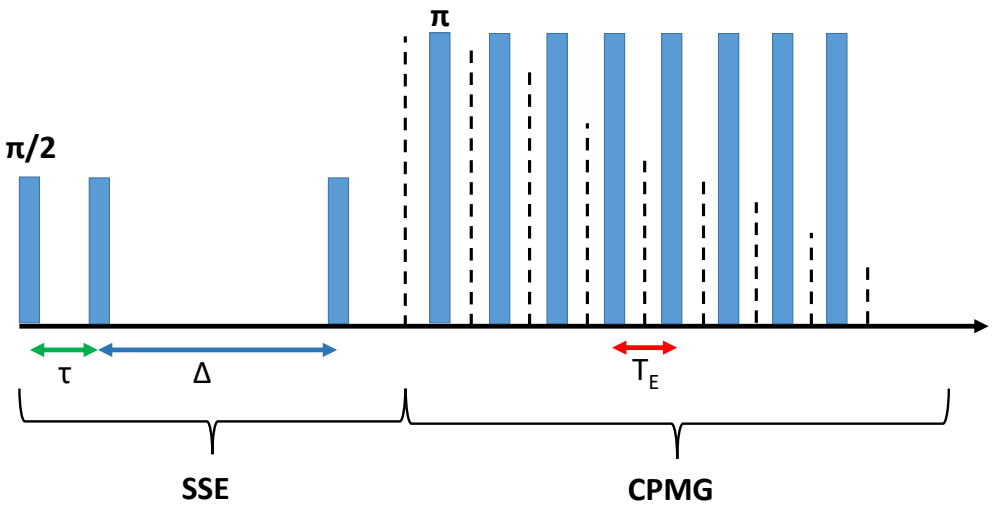


Figure 5.2: Diffusion Editing CPMG pulse sequence composed by a Stimulated Spin Echo block followed by a CPMG signal detection block.

magnetization is stored in z-direction during an evolution period Δ, where the magnetization decays due to T_1 and single spin components change position in the sensitive volume. Then, magnetization is reported in the transverse plane and the signal is detected using a CPMG sequence. In this case the equation 5.2 becomes:

$$M(\tau, kt_E) = \int\int f(D, T_2) \exp\left(-\gamma^2 G^2 D \tau^2 \left(\Delta - \frac{2}{3}\tau\right)\right) \exp\left(-\frac{kt_E}{T_2}\right) dD dT_2 \qquad (5.4)$$

and $f(D, T_2)$ is the 2D density distribution (or correlation function) of the diffusion coefficients and T_2 times, which can be computed by a 2D ILT.

Diffusion Weighted $T_1 - T_2$ correlation experiments

The Diffusion Weighted T_1 editing CPMG pulse sequence (DW $- T_1 - T_2$) is a here proposed modification of the classical $T_1 - T_2$ pulse sequence where the first two echoes of the CPMG are created using an echo time (named *First Echo Time*, T_{E1}) longer than the echo time (T_E) of the following 180 train. The pulse sequence time diagram is reported in Fig. 5.3. Provided that T_E is sufficiently short, T_{E1} is the time period the diffusion phenomenon mainly affects the signal decay. By setting different values for T_{E1}, the diffusion weights more or less the $T_1 - T_2$ correlation map. It is worth noticing that during T_{E1} transverse relaxation also contributes to the magnetization decay in addition to diffusion. Therefore, T_{E1} has to be short enough to not produce a significant decay of the transverse magnetization, otherwise, the echo attenuation would be due to the relaxation more than to the diffusion. To meet

this condition T_{E1} should be equal or less than the shortest T_2 of the compound under investigation.

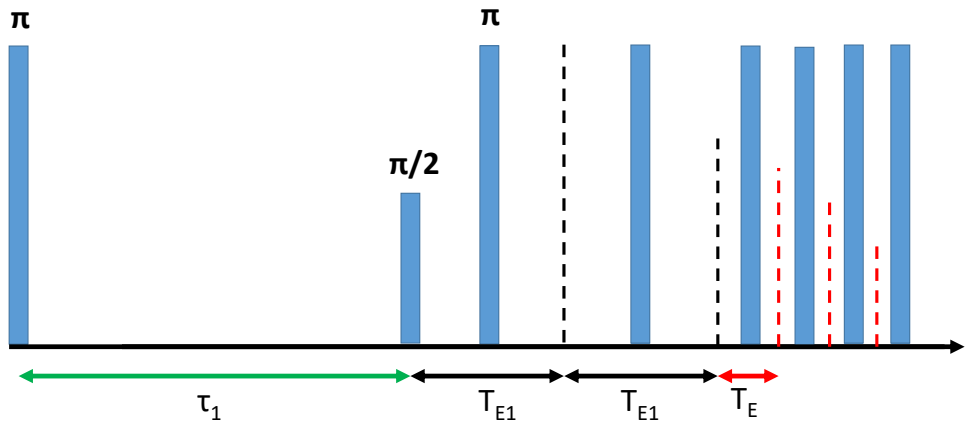


Figure 5.3: DW-T_1-T_2 pulse sequence. The sequence is an IR, where τ_1 is the recovery time, followed by a CPMG of echo time T_E. The first two π-pulses of the CPMG are separated by an adjustable first echo time (T_{E1}), that establishes the diffusion weighting. The dashed lines represent the echoes.

Performing DW $- T_1 - T_2$ with different T_{E1} can be helpful when analyzing sample composed by a variety of tissues. It gives a direct methodology to assess what spin components disappear when diffusion weights more and more the $T_1 - T_2$ map, which is related to the value of the coefficient of diffusion of the spin components.

5.3 Materials and methods

5.3.1 Samples

The samples and the corresponding experiments used in this chapter are summarized in Tab. 5.1. 2D NMR correlation experiments were performed on the set of samples S1, on sample BS1 and sample M1. The set of tissue samples, labeled S1, comprising three samples made of cartilage, muscle and marrow respectively, were used for $D - T_2$ experiments. The bone sample, labeled BS1, in which trabecular bone, cartilage and muscle were present together, and the bulk marrow sample, labeled M1, were used for performing Diffusion-Weighted $T_1 - T_2$ (DW $- T_1 - T_2$) experiments.

The trabecular bone sample, TB, the bulk marrow sample, M2, and the TB+Muscle

sample, composed of the trabecular bone sample TB with added muscle tissue were used to test the improved BV/TV evaluations procedure.

Label	Description	Experiment
S1	1 sample of Cartilage 1 sample of Muscle 1 sample of Marrow	$D - T_2$
BS1	Bone sample in which cartilage, muscle and marrow are present	DW-T_1-T_2
M1	Bulk marrow sample	
TB	Just trabecular bone	BV/TV
TB+Muscle	Trabecular bone with muscle added	
M2	Bulk marrow sample	

Table 5.1: Summary of the samples and their usage in the NMR experiments.

5.3.2 NMR Experiments

Single-sided devices

The experiments were performed with the NMR-MOUSE PM10 (Magritek, GmbH) and the NMR-MOLE. While the NMR-MOUSE PM10 is characterized by a constant magnetic field gradient (14 T/m), the NMR-MOLE presents a sweet spot in which the field is relatively homogeneous, with a distribution of field gradients (average value of 0.6 T/m). The sensitive volumes of the NMR-MOUSE PM10 and NMR-MOLE, with the pulse length used in these experiments, are $16 \times 16 \times 0.330\,\mathrm{mm}^3$ and $26 \times 26 \times 6\,\mathrm{mm}^3$ respectively. The characteristics of the two devices are described in detail in 4.3.

$T_1 - T_2$ correlation experiments

All the $T_1 - T_2$ correlation experiments were performed with the NMR-MOLE using a pulse length of 12 μs. τ_1 was varied 32 times in the range 1 ms to 1000 ms, with echo time (T_E) equals to 150 μs. The experiment was performed on S1 and BS1.

$D - T_2$ correlation experiments

All the $D - T_2$ experiments have been performed with the NMR-MOUSE PM10 using an rf-pulse length of 5 μs. Δ was set to 1 ms and τ was varied linearly over 32 values in the range from 0.02 ms to 2.2 ms or 0.02 ms to 0.15 ms, while the repetition time (TR) was set to 5 s. The first range was used for the measurement on the bulk marrow sample of the set S1, and the latter for the measurements on the cartilage and muscle samples of the set S1. The echo time T_E was set to 50 μs.

DW $- T_1 - T_2$ correlation experiments

All the DW$-T_1-T_2$ experiments were performed with the NMR-MOLE because of its higher SNR and larger sensitive volume compared to the NMR-MOUSE PM10. The sequence was performed on the sample BS1 with three different T_{E1}: 150 μs, 5000 μs and 10000 μs, and on M1 with $T_{E1} = 150$ μs. The inversion time τ_1 was varied over 32 values in the range from 1 ms to 1000 ms, with T_E set to 150 μs.

BV/TV evaluation experiments

The BV/TV evaluation experiments were performed on: TB, TB + Muscle and M2 samples. A sketch of the sample configurations used to perform the experiment is presented in Fig. 5.4.

Signals were acquired with the NMR-MOUSE PM10 by using the CPMG pulse sequence and the SSE+CPMG pulse sequence. The NMR-MOUSE PM10 was chosen because of its constant field gradient inside the sensitive volume. T_E was set to 50 μs for the CPMG and SSE+CPMG pulse sequence. In the SSE+CPMG sequence instead of varying τ to perfrom a $2D$ experiment, only one τ was used and it was set to 10 μs to perform a 1D experiment. For each sample, the pulse sequence was executed twice, one with $\Delta = 1$ ms and the other with $\Delta = 25$ ms. The BV/TV ratio was computed according to equation 5.1, where, in the first case S_{Sample} is the signal from the sample TB (Fig. 5.4 (a)), whereas in the second case S_{Sample} is the signal from the sample TB+Muscle (Fig. 5.4 (b)). For both cases, $S_{\text{Reference}}$ is the signal form the bulk marrow sample M2 (Fig. 5.4 (c)). The signals were evaluated by computing the total area under the T_2 distribution evaluated by the inversion of the CPMG data.

Data analysis

For the 1D experiments, all the data were inverted with the UPEN [18, 19] algorithm implemented in the UpenWin software [21]. A customized 2D Inverse Laplace Transform

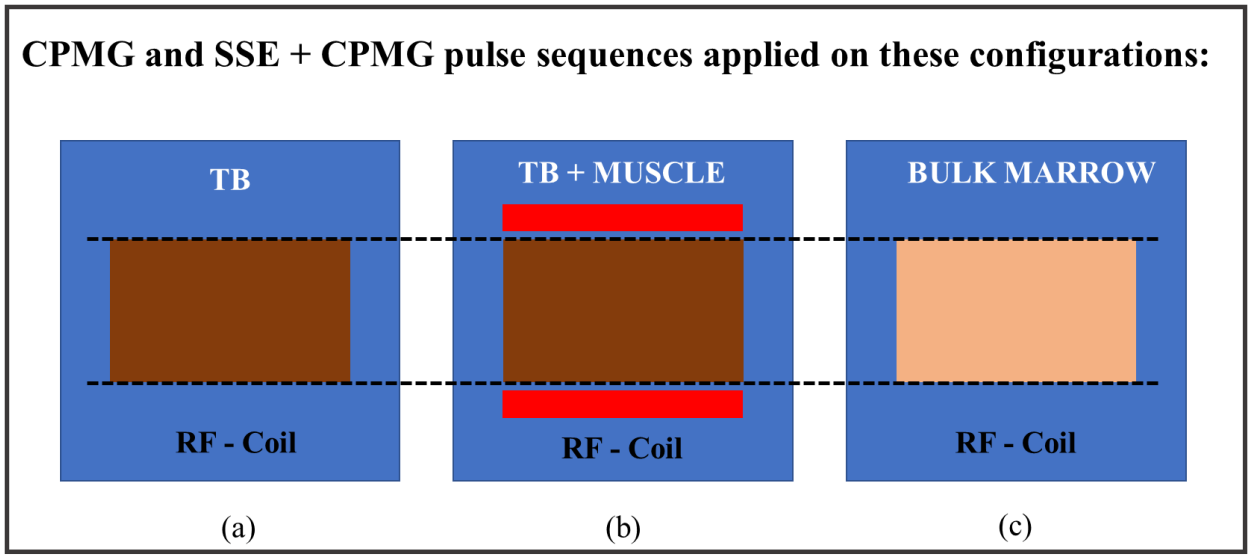


Figure 5.4: Sketch of the sample configurations used to perform the BV/TV evaluation experiment in presence of muscle tissue. (a) TB sample, (b) TB + muscle sample and (c) Bulk marrow sample (M2 sample). On all the samples a non-diffusion weighted CPMG is acquired to compute BV/TV according to the procedure described in [23] and a diffusion weighted CPMG pulse sequence (namely the SSE+CPMG) is performed to test feasibility of filtering out the muscle signal.

method [57] for 2D correlation experiments was implemented in Python [102] using the SciPy library [60].

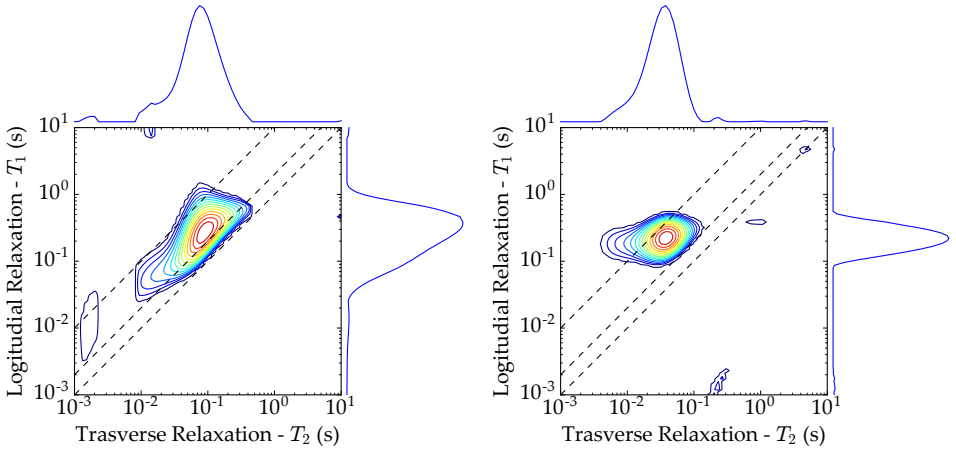
5.4 Results and Discussion

5.4.1 Tissues characterization

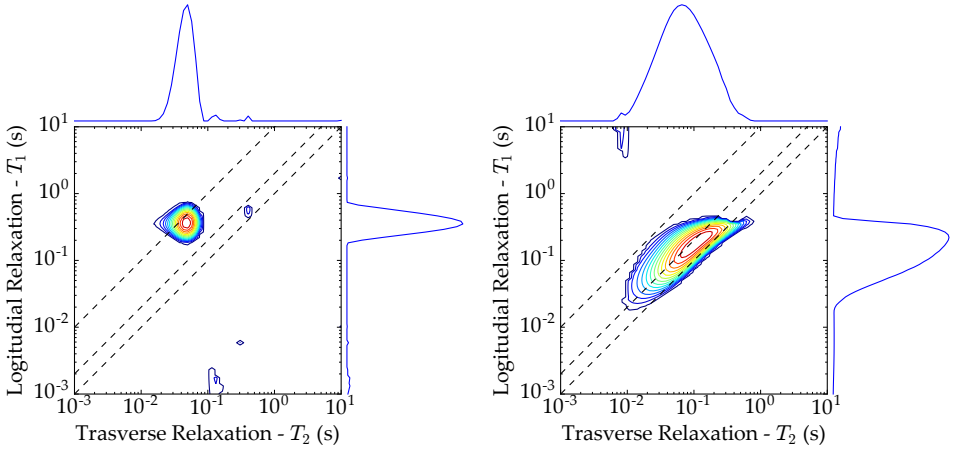
$T_1 - T_2$ correlation maps

First, the biological tissues were characterized through $T_1 - T_2$ correlation experiments.

The resulting maps are reported in Fig. 5.5. By looking at the maps of the tissues taken separately, it can be observed that the marrow (Fig. 5.5(d)) is characterized by a distribution in which T_1 is relatively close to T_2, the cartilage and the muscle (Fig. 5.5(b) and 5.5(c)), on the contrary, are characterized by distributions in which T_1 is much longer than T_2. Hence, although the map of the sample BS1 (Fig.5.5(a)), in which all the three tissues are present together, does not show separated classes, it is possible "to see" the presence of the cartilage and the muscle in addition to the marrow. This means, as expected, that one cannot neglect their contribution to the NMR signal when evaluating the BV/TV ratio in such a scenario, that is more like the *in-vivo* scenario. Moreover, by looking at the marginal distributions, it



(a) BS1 (b) Cartilage



(c) Muscle (d) Marrow

Figure 5.5: $T_1 - T_2$ correlation maps belonging to samples (a) BS1, (b)(c)(d) tissues of S1 set of samples. The three dashed lines represent, starting from the diagonal, $T_1 = T_2$, $T_1 = 2 \times T_2$ and $T_1 = 10 \times T_2$.

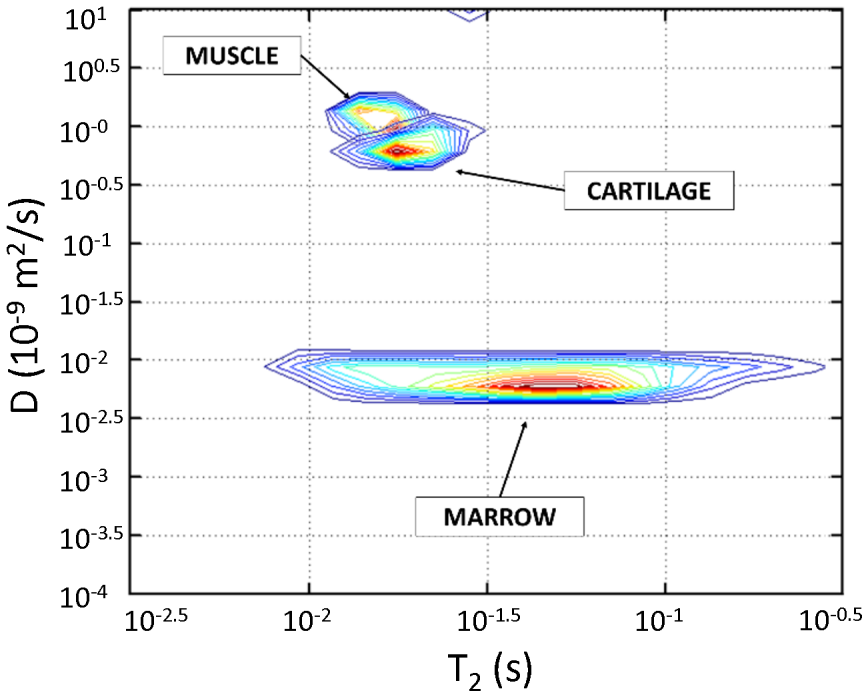


Figure 5.6: $D - T_2$ distributions of tissues belonging to the S1 set of samples (three separate muscle, cartilage and marrow samples). The map shows a superposition of the individually acquired $D - T_2$ maps.

is worthy to observe that this information is hidden if one performs the two 1D experiments, this because of the broad marrow's relaxation times distributions.

$D - T_2$ correlation maps of the tissues

The results of the $D - T_2$ correlation experiments are reported in Fig.5.6, which shows that fluids in the muscle and the cartilage have diffusion coefficient values about two orders of magnitude higher than fluid in the marrow, exhibiting a diffusion coefficient as low as $6 \times 10^{-3} \mu m^2/ms$.

This difference suggested developing a strategy to selectively suppress the signal coming from the cartilage and the muscle tissues, and to only detect the signal from the marrow. More specifically, by exploiting the intrinsic magnetic field gradient of single-sided NMR devices, and leaving sufficient time for molecular diffusion before signal acquisition, the signal from the more diffusive cartilage and muscle components can be reduced with respect to the less diffusive marrow component. The degree of signal suppression will depend on the time left for molecular diffusion.

Diffusion Weighted $T_1 - T_2$ experiments

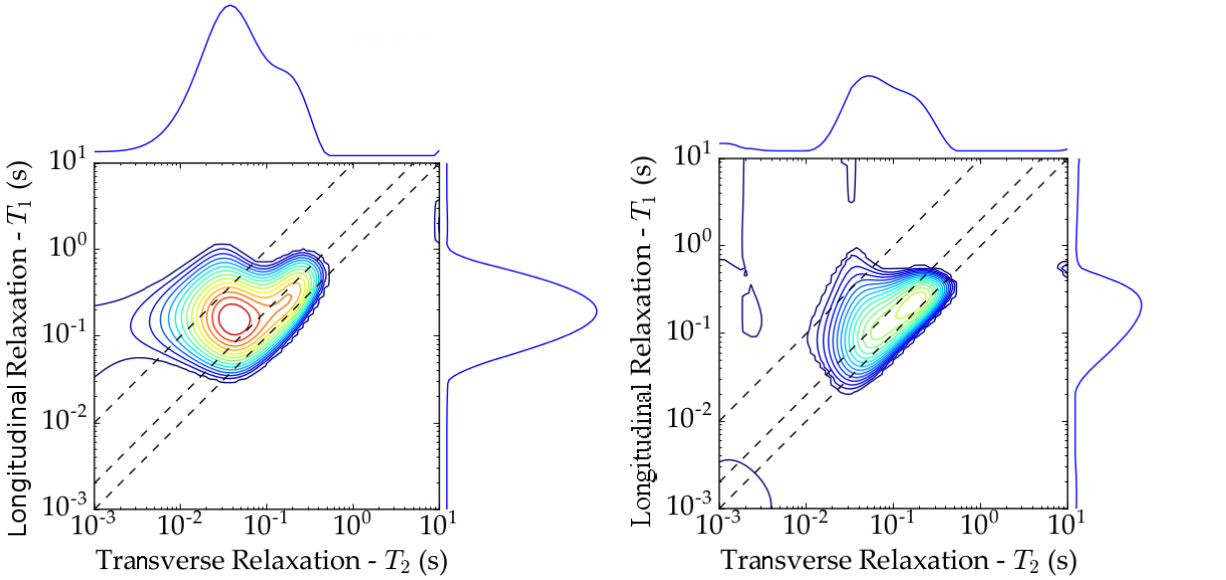
In order to verify the feasibility of the idea presented in the previous section, without looking for quantitative results, a Diffusion-Weighted $T_1 - T_2$ correlation experiment was implemented. To do this, the $T_1 - T_2$ pulse sequence was modified in order to produce a $\mathrm{DW} - T_1 - T_2$ pulse sequence, as described in section 5.3.2.

The results of the experiments performed on samples BS1 and M1, leaving different times for diffusion, are reported in Fig. 5.7. Figure 5.7(a) shows the $T_1 - T_2$ map of BS1 without diffusion weighting, in which case $T_{E1} = T_E$. All components in the sample were detected as expected. When T_{E1} is increased step by step (Fig. 5.7(b) and 5.7(c)), the most diffusive components (muscle and cartilage) disappear and only the marrow component remains, as is evident when the map 5.7(c) is compared with the pure bulk marrow map, Fig. 5.7(d). This experiment provides evidence that it is possible to selectively detect only the signal from the marrow tissue even when cartilage and muscle tissues are inside the sensitive volume by using the diffusion and intrinsic field gradient of single-sided NMR devices, in this case, the NMR-MOLE.

5.4.2 Trabecular bone BV/TB ratio evaluation in the presence of muscle tissue

The final step was to carry out a quantitative study in order to set-up and validate a procedure to correctly evaluate the BV/TV of trabecular bone in the presence of other tissues. Since the future goal of the research is to develop a clinical application, any procedure devised should be time saving. To this in mind, a procedure based on a 1D NMR pulse sequence was considered most appropriate. The selected pulse sequence was the SSE+CPMG where SEE establishes the diffusion filter, and the CPMG is for signal detection. This sequence required 8 minutes, while a full $D - T_2$ experiment required 270 minutes. It should be noted that a new method that vastly reduces the required data for successful 2D ILT has been recently introduced [11], paving the way for the acquisition of 2D correlation maps in clinically acceptable times.

The BV/TV was evaluated according to the procedure explained in section 5.3.2. The T_2 distributions of the samples TB and TB+Muscle are reported in Fig. 5.8. When no diffusion filter is used ($\Delta = 0$ ms) T_2 distributions present a short T_2 peak (300-500 μs) in addition to a peak at longer times (80 ms), as visible in Fig. 5.8 (a). The long T_2 peak is due to the marrow inside the inter-trabecular spaces in the TB sample, whereas in the TB+Muscle sample, muscle tissue also contributes in addition to the marrow. The short T_2 time peak is probably due to the intra-trabecular water, as reported in [40, 41]. As a result, the area



(a) BS1, $T_{E1} = 150$ μs

(b) BS1, $T_{E1} = 5000$ μs

(c) BS1, $T_{E1} = 10000$ μs

(d) Marrow, $T_{E1} = 150$ μs

Figure 5.7: (a),(b),(c) DW $- T_1 - T_2$ maps for different first echo times T_{E1} of the sample BS1, in which trabecular bone, muscle and cartilage tissues are present. (d) DW $- T_1 - T_2$ map of the bulk marrow sample M1 with no diffusion weighting.

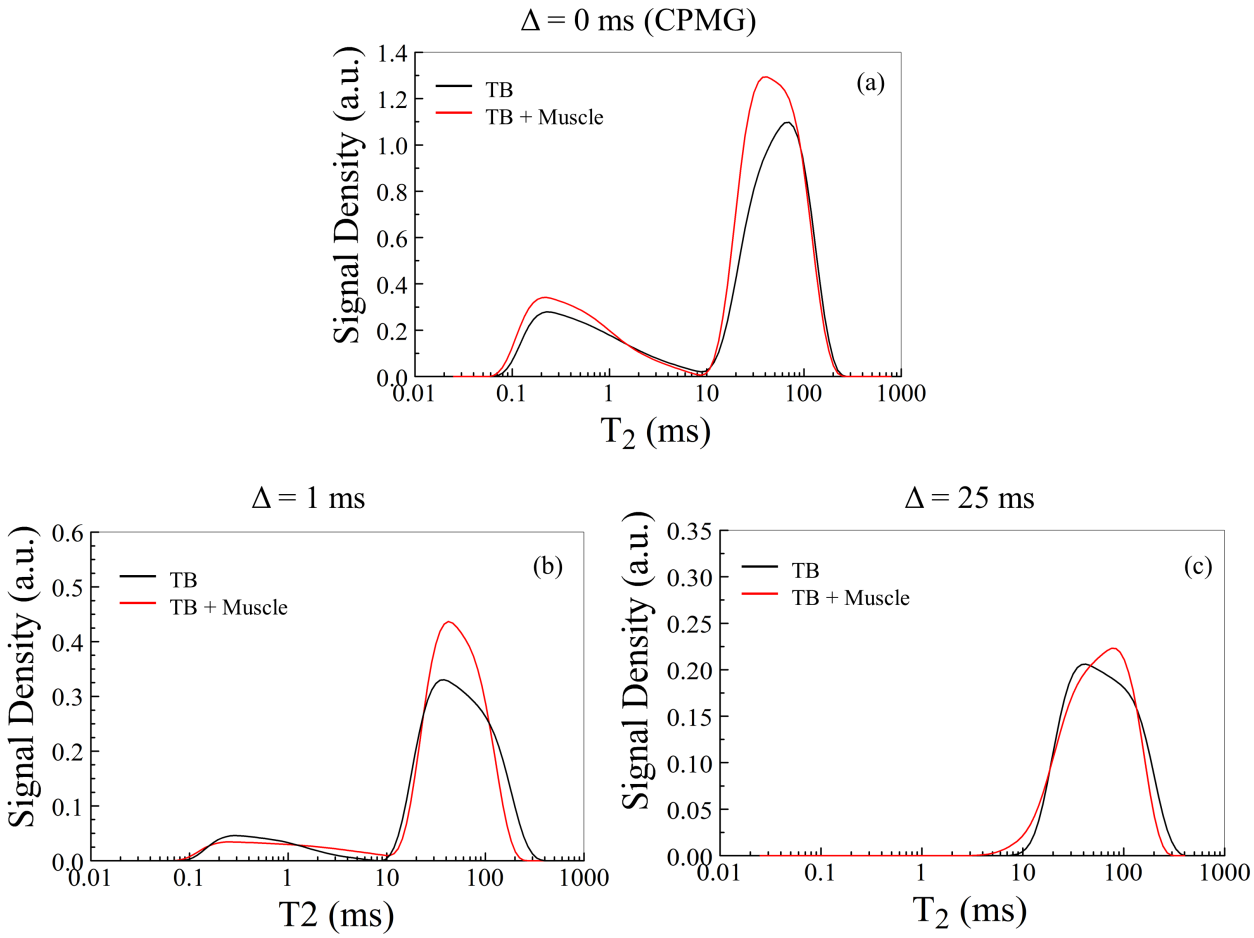


Figure 5.8: T_2 distributions of TB and TB + Muscle samples a a function of Δ time left for the diffusion in the SSE + CPMG pulse sequence. (a) $\Delta = 0$ ms (CPMG), (b) $\Delta = 1$ ms and (c) $\Delta = 25$ ms. As diffusion time increases the two T_2 distributions becomes more and more similar.

under this peak was not taken into account in computing the BV/TV ratios.

The BV/TV ratios for different diffusion weightings for the TB and TB+Muscle samples are reported in Tab. 5.2. The ground truth BV/TV value of the TB sample is the value associated with $\Delta = 0$ ms, which is equal to $(32 \pm 2)\%$. In fact, this is the value obtained with the procedure validated in [23], i.e. a simple CPMG without SSE filtering. It can be also observed that varying Δ, the BV/TV values of the TB sample are the same within experimental error. This means that the diffusion filter does not affect the accuracy of the proposed new procedure as compared to the procedure proposed in [23]. On the contrary, when the muscle is added to the TB sample (i.e. becomes a TB+Muscle sample) BV/TV values increase because the signal coming from the muscle, which is inside the sensitive volume, gradually decreases as a result of the increased diffusion time. By increasing the time left for diffusion, the BV/TV ratios of TB+Muslce sample approach those of the TB sample,

and, with $\Delta = 25$ ms, the BV/TV ratios of the two samples are the same. This provided the evidence that it is possible to detect the signal only from the marrow tissue present within the inter-trabecular spaces by suppressing the contribution of the muscle tissue. It should also be noted that if even in the presence of cartilage tissue, signal suppression would occur as well because cartilage has a self-diffusion coefficient close to that of muscle, as shown in section 5.4.1.

	$\Delta = 0$ ms	$\Delta = 1$ ms	$\Delta = 25$ ms
BV/TV of TB (%)	32±2	30±3	33±5
BV/TV of TB + Muscle (%)	16±1	26±2	33±5

Table 5.2: BV/TV values evaluated with different diffusion weighting for TB and TB+Muscle samples. $\Delta = 0$ ms values refer to the simple CPMG pulse sequence. The ground truth BV/TV is obtained when the signal from the muscle is completely suppressed.

5.5 Conclusions

The results obtained in [23] have demonstrated the feasibility of a NMR based technique to assess the BV/TV of trabecular bone using single-sided devices. Due to the low-cost and portability of these scanners, the method opens up new prospects for a cheap medical test to help the diagnosis of osteoporosis. These characteristics make the method eligible for the screening of the population at risk for osteoporosis. The reported results have demonstrated that it is possible to improve the technique presented in [23], suppressing the signal coming from muscle tissue, that may be within the sensitive volume of the single-sided NMR scanner in a clinical *scenario*.

The improvement was achieved by filtering the signal acquisition by molecular diffusion, as demonstrated by the results of the DW $- T_1 - T_2$ experiments. The pulse sequence was performed with the NMR-MOLE because of the larger sensitive volume and the higher Signal-to-Noise Ratio compared to the NMR-MOUSE PM10.

By using an SSE encoding sequence, with sufficient time for molecular diffusion to take place, followed by a CPMG as a signal reading sequence, a time-saving measurement was set up by using the NMR-MOUSE PM10. This procedure allowed correct BV/TV estimation of a trabecular bone sample even when muscle tissue was present within the sensitive volume. In order to make the procedure more general and robust for clinical application, a 1D diffusometry experiment could be applied prior to the relaxometry experiment. This would ensure that the correct diffusion-weighting is used in the SSE+CPMG pulse sequence.

These experiments exploiting diffusion-weighted pulse sequences may pave the way for the clinical application of a BV/TV evaluation procedure using single-sided scanners.

Chapter 6

Trabecular Bone Volume Fraction Estimation through Separation of Intra and Inter Trabecular Signals from T_2 Distribution Measured With a Single-Sided NMR Device

6.1 Introduction and theory

The goal of this chapter is to modify the methodology presented in [23] to be able to assess the BV/TV ratio of trabecular bones without the need of using a reference sample, while keeping all the key features of that technique, which are: to use a CPMG pulse sequence to acquire the data, and to use a single-sided NMR scanner.

6.1.1 Separating intra and inter trabecular signals with T_2 relaxation time distributions

The idea behind the new methodology is to take advantage of the fact that NMR signal coming from ^{1}H nuclei within the trabeculae are characterized by T_2 relaxation times that are shorter than the T_2 relaxation times characterizing the ^{1}H nuclei that are in the inter trabecular spaces. NMR relaxometry studies on cortical and trabecular bones have shown that T_2 relaxation times span a range that goes from hundreds of μs to hundreds of ms [24, 40, 41, 55, 94]. In Fig. 6.1 an example of a D-T_2 correlation map measured with the NMR-MOUSE PM10 of a cortical bone sample is reported along with a T_2 distribution

computed out from a CPMG decay acquired with a 20 MHz NMR spectrometer. The 2D map was computed using the software UPEN2D [20]. From the T_2 distribution it is possible

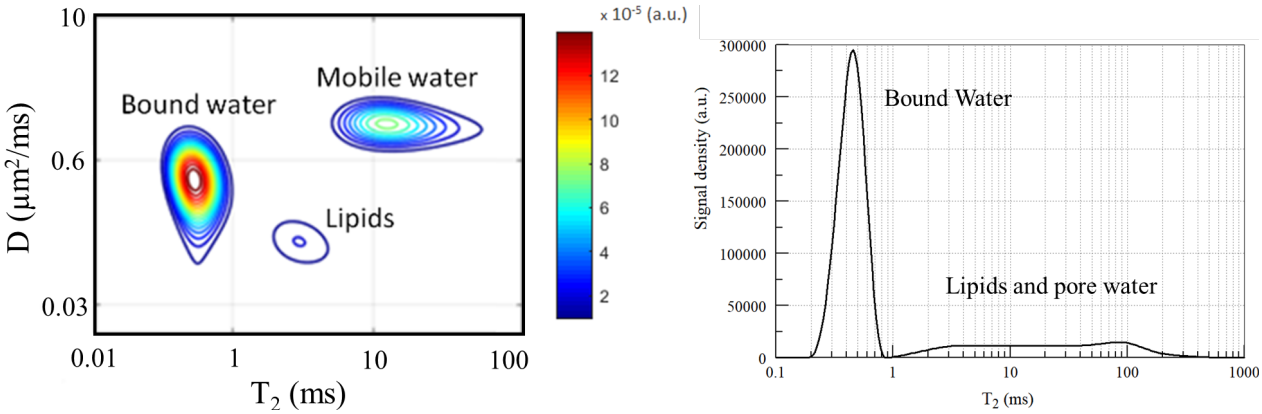


Figure 6.1: (a) D-T_2 correlation map of a cortical bone sample obtained using the NMR-MOUSE P10. (b) T_2 distribution of the same cortical bone sample obtained with a 20 MHz NMR spectrometer. Figures are taken from [24].

to observe that the ^{1}H nuclei of water bound to collagen proteins contributes the most to the total NMR signal of cortical bone, and their T_2 times range from 0.1 ms to 1 ms, with peak around 200-300 μs. The rest of the signal, a relative small part of the total signal, span T_2 times ranging from few ms to 200 ms. The D-T_2 correlation map allows one to identify lipids and mobile water inside Haversian canals and the lacuno-canicular system. Information regarding the NMR properties of cortical bone are helpful to investigate trabecular bone as well. Trabeculae have similarities with cortical bone tissue, with the exception that Haversian canals are not present. This means that in trabecular bones, bound water coming form inside the trabeculae is NMR detectable. Since the bone marrow is the main tissue that fills the inter trabecular spaces, and since it is characterized by T_2 times in the order of 80 - 100 ms, it is possible to separate intra and inter trabecular signal and to estimate the T_2 short intensity fraction. It is evaluated following equation 6.1, where S_{total} is the total area under the T_2 distribution (the total NMR signal) and $S_{\text{short } T_2}$ is the area under the peak representing the short T_2 component (the intra trabecular signal).

$$T_2 \text{ short intensity fraction} = \frac{S_{\text{short } T_2}}{S_{\text{total}}} \times 100(\%) \tag{6.1}$$

Since $S_{\text{short } T_2}$ is proportional to the volume of the bound water inside the trabeculae, the T_2 short intensity fraction is expected to be proportional to the BV/TV.

6.1.2 Estimating morphological parameters form the T_2 short intensity fraction

Estimation of BV/TV and BS/TV morphological parameters of trabecular bone from CPMG data is possible by evaluating the concentration of inter and intra trabecular signals using theoretical models.

BV/TV estimation

Considering the BV/TV ratio, the bone volume is equal to the volume occupied by the trabeculae, which can be divided in two main contribution: $V_{\text{trabeculae}} = \text{V}_{\text{bound water}} + \text{V}_{\text{mineralized tissue}}$. The total volume can be expressed as $V_{\text{marrow}} + V_{\text{trabeculae}}$. From a CPMG experimental perspective, the volume of the bound water, $\text{V}_{\text{bound water}}$, is proportional to $S_{\text{short } T_2}$, which represents the intra-trabecular signal. The volume occupied from the marrow, V_{marrow}, is proportional to $S_{\text{long } T_2}$, which represents the inter-trabecular signal. The volume of the mineralized trabecular tissue, $\text{V}_{\text{mineralized tissue}}$, is NMR insensitive and it cannot be estimated from the acquired CPMG data. However, Fantazzini et al [40] have studied the mean porosity of a trabecula by means of NMR relaxometry measurements performed on a set of bovine trabecular bones, Which was found to be $\phi_{\text{trabecula}} = (29 \pm 4)\%$. Estimating the intra trabecular porosity allows one to estimate a multiplying correction factor for the $S_{\text{short } T_2}$ to take into account for the mineralized tissue of the trabeculae. Using this correction factor is possible to estimate $\text{V}_{\text{trabeculae}}$ from $S_{\text{short } T_2}$. Indeed,

$$\phi_{\text{trabecula}} = \frac{V_{\text{bound water}}}{\text{V}_{\text{bound water}} + \text{V}_{\text{mineralized tissue}}} = \frac{V_{\text{bound water}}}{V_{\text{trabeculae}}} \tag{6.2}$$

which implies that the BV/TV ratio can be expressed, in terms of volumes, by equation 6.3,

$$\frac{BV}{TV} = \frac{\frac{V_{\text{bound water}}}{\phi_{\text{trabecula}}}}{V_{\text{marrow}} + \frac{V_{\text{bound water}}}{\phi_{\text{trabecula}}}} \cdot 100(\%) \tag{6.3}$$

which in terms of NMR signals can be rewritten in equation 6.4.

$$\frac{BV}{TV} = \frac{\frac{S_{\text{short } T_2}}{\phi_{\text{trabecula}}}}{S_{\text{long } T_2} + \frac{S_{\text{short } T_2}}{\phi_{\text{trabecula}}}} \cdot 100(\%) \tag{6.4}$$

BS/TV estimation

The BS/TV ratio can be estimated using equation 6.5 when trabeculae are modelled as cylinders.

$$\frac{BS}{TV} = \frac{2 \cdot \frac{V_{\text{bound water}}}{\phi_{\text{trabecula}}}}{r} \cdot \frac{1}{V_{\text{marrow}} + \frac{V_{\text{bound water}}}{\phi_{\text{trabecula}}}} \cdot 100(\%) \tag{6.5}$$

The first multiplicative term on the right side of the equation is the expression of the bone surface, which comes from expressing $V_{\text{trabecula}} = \pi r^2 h$, where r and h are the radius and the height of the cylinder, respectively. The height can be written in terms of $V_{\text{trabecula}}$, so that the surface of the cylinder takes the form reported in equation 6.5. Since the signal intensities of the long and short T_2 components are proportional to the volumes of inter and inter trabecular compounds, the BS/TV ratio can be written in terms of NMR signals, as expressed in equation 6.6, where r was set equals to the average trabecular thickness of the set of samples under investigation, which was estimated from the micro-CT measurements.

$$\frac{BS}{TV} = \frac{2 \cdot \frac{S_{\text{short } T_2}}{\phi_{\text{trabecula}}}}{r} \cdot \frac{1}{S_{\text{long } T_2} + \frac{S_{\text{short } T_2}}{\phi_{\text{trabecula}}}} \cdot 100(\%) \tag{6.6}$$

6.2 Materials and Methods

6.2.1 Trabecular bone samples

Eleven cylindrical trabecular bone samples, of approximate diameter 1 cm and height 1.5 cm, where cored from different sites of the pig shoulder, placed inside test tubes and immediately deep frozen. NMR measurements were performed before micro-CT analysis since the NMR protocol acquires the signal from the soft tissues inside the trabecular bone, which are more prone to degradation than calcified tissues. During preliminary measurements on a test sample a loss of NMR signal was observed after micro-CT measurements due to the evaporation of the liquid component.

6.2.2 NMR measurements

The NMR measurements were performed using the NMR-MOUSE PM10 scanner (Magritek, New Zealand). The CPMG pulse sequence was used to acquire the NMR signal. Four spacers were used to optimize SNR efficiency (8 mm spacer configuration), implying that the sensitive volume was at 3 mm from the surface coil. Echo time (T_E) was set equal to 50 μs, using a pulse length of 4.75 μs, and 2000 echoes were used to acquire the CPMG decay. The sensitive volume is around 300 μm thick. To increase SNR, acquisition was repeated 128

times and averaged using a repetition time of 6 s to allow complete longitudinal magnetization recovery between consecutive scans. Per each sample, the total acquisition time was around 13 minutes. Figure 6.2 sketches the NMR-MOUSE PM10 configuration used for acquiring the NMR signal.

Since the height of trabecular bone samples is considerably higher than the maximum achievable depth from the NMR-MOUSE PM10 using the 8 mm configuration, a marker was sketched on one of the circular bases of each cylindrical sample. The NMR signal was acquired both having the circular base containing the marker facing the surface coil (i.e. marker configuration) and both having the circular base not containing the marker facing the surface coil (i.e. no-marker configuration). Hence, the total number of NMR acquisitions is 22. Trabecular bone tissue can be highly heterogeneous at different depths, and having measurements from different location can increase the range of variation of the trabecular micro structure, which is helpful to test the proposed procedure. For these reasons the 22 NMR measurements were considered as independent.

The measurement procedure for each sample was composed of the following steps. (1) the sample is left defrosting for 45 minutes inside the test tube. Then (2) a CPMG measurement was preformed in one of the two configuration (i.e. marker or no-marker configuration), and (3) replaced in the refrigerator before conducting the CPMG acquisition on the remaining configuration. The procedure is repeated for other samples. By using this protocol CPMG measurements are randomized, with this meaning that the marker and no-marker parts of a sample are not scanned consecutively. The rationale for such a procedure is to avoid possible systematic bias that can possibly arise from degradation of the sample while it undergoes the NMR acquisition.

6.2.3 Micro-CT and morphological parameters evaluation

Micro-CT investigations were performed after NMR analysis on a Skyscan micro-CT mod. 1072 (Bruker MicroCT, Konthich, Belgium) at the Laboratory of Medical Technology of the Istituto Ortopedico Rizzoli of Bologna. The X-ray source was set at 50 kV and 200 μA with a resolution (pixel size) of 20 μm. Each scan is performed with a sample rotation of 180 degree on its axis, with a rotation step of 0.9 degree producing 206 projections. Cross-sectional images are obtained from the projections by use of NRecon reconstruction software ver. 1.6.8.0 (Bruker MicroCT), with the application of a beam hardening correction of 24% to reduce artifacts. The reconstruction process produced a stack of images in 8-bit bmp format. A global thresholding is applied to binarize the images, using the Otsu's method to find the threshold, and the software CtAnalyser ver. 1.14.4.1 (Bruker MicroCT) is used to calculate morphological parameters of trabecular bone samples. The morphological

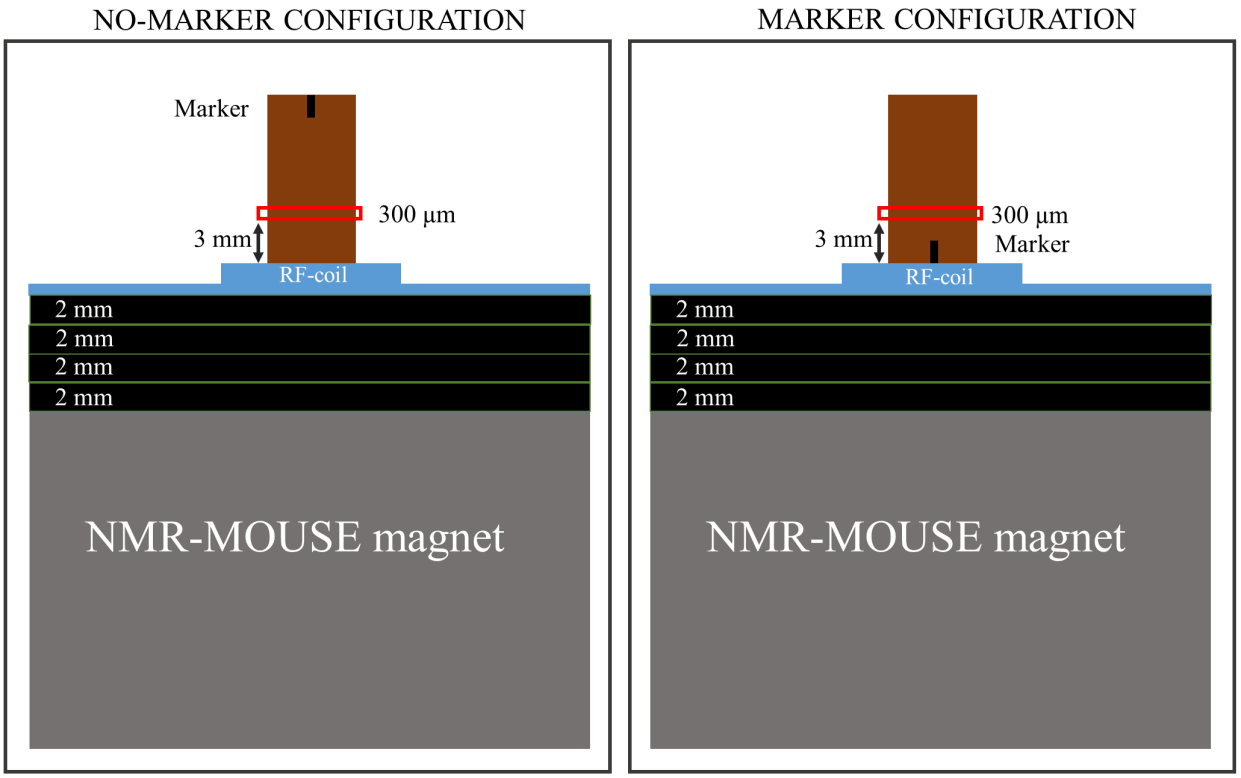


Figure 6.2: Sketch of the signal acquisition procedure using the NMR-MOUSE PM10. Four spacers (2mm thick each) were used to optimize SNR efficiency, meaning that the sensitive volume was at 3 mm from the surface coil having a thickness of 300 μm. Considering the height of the trabecular samples NMR measurements were carried on two parts of the sample: one measurement whit the face containing the marker being scanned (marker configuration) and one with the face not containing the marker being scanned (no-marker configuration).

parameters that were computed are: the bone volume fraction (BV/TV (%)), the bone surface density (BS/TV (mm^{-1})), the trabecular number (Tb.Nb (mm^{-1})), the trabecular thickness (Tb.Th (mm)), the bone surface to bone volume ratio (BS/BV (mm^{-1})) and the trabecular pattern factor (Tb.Pf (mm^{-1})).

With the aim of making a correct comparison between NMR and micro-CT measurements, the morphological parameters were evaluated in two regions of interest (ROI) of the sample, which were selected to match as good as possible the regions intercepted by the sensitive volume of the NMR-MOUSE PM10 during NMR measurements. The two 3D ROIs were found using the reference sign marked on each sample, and knowing that NMR signal was obtained from a 300 μm thick slice placed at 3 mm depth inside the sample. Figure 6.3 reports a coronal view of a micro-CT scan of a sample along with a schematization of the analysed ROIs and axial images representing a slice from each ROI. Although ROIs selection was performed with care, it cannot be excluded that the positions of the slices could have been slightly different (within the resolution of the micro-CT) from the ones intercepted by the sensitive volume of the NMR-MOUSE PM10.

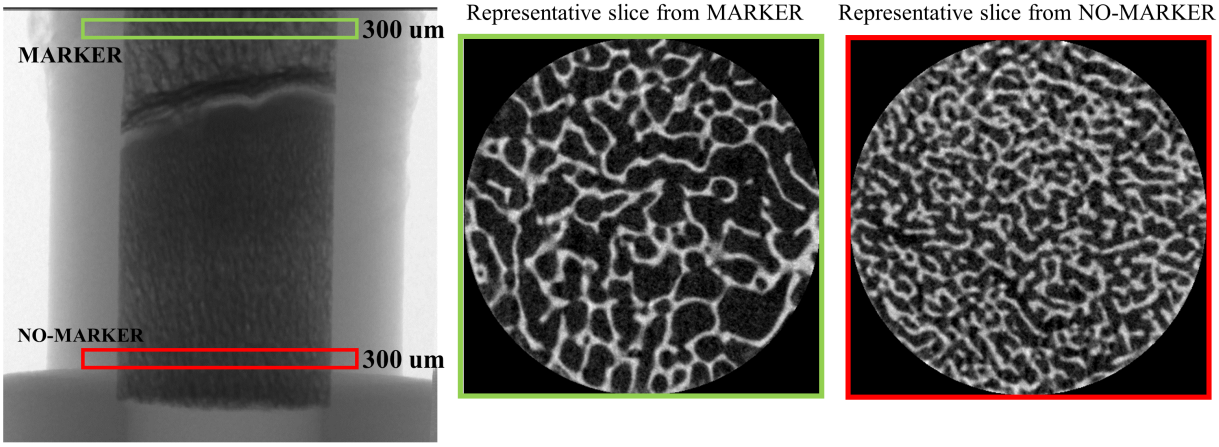


Figure 6.3: micro-CT example.

Error associated with the micro-CT measurements were evaluated considering that the main source of uncertainty is the selection of the threshold value for bone segmentation. Hence, as done in [23], morphological parameters were evaluated using different threshold values in the interval = "optimal threshold value" $\pm$ 5%.

6.2.4 NMR data analysis

CPMG transverse magnetization decay curves were analyzed by the UpenWin software [21] that implements the inverse Laplace transfor algorithm UPEN [18, 19] to obtain quasi-continuous distribution of transverse relaxation times. Figure 6.4(a) reports the pipeline

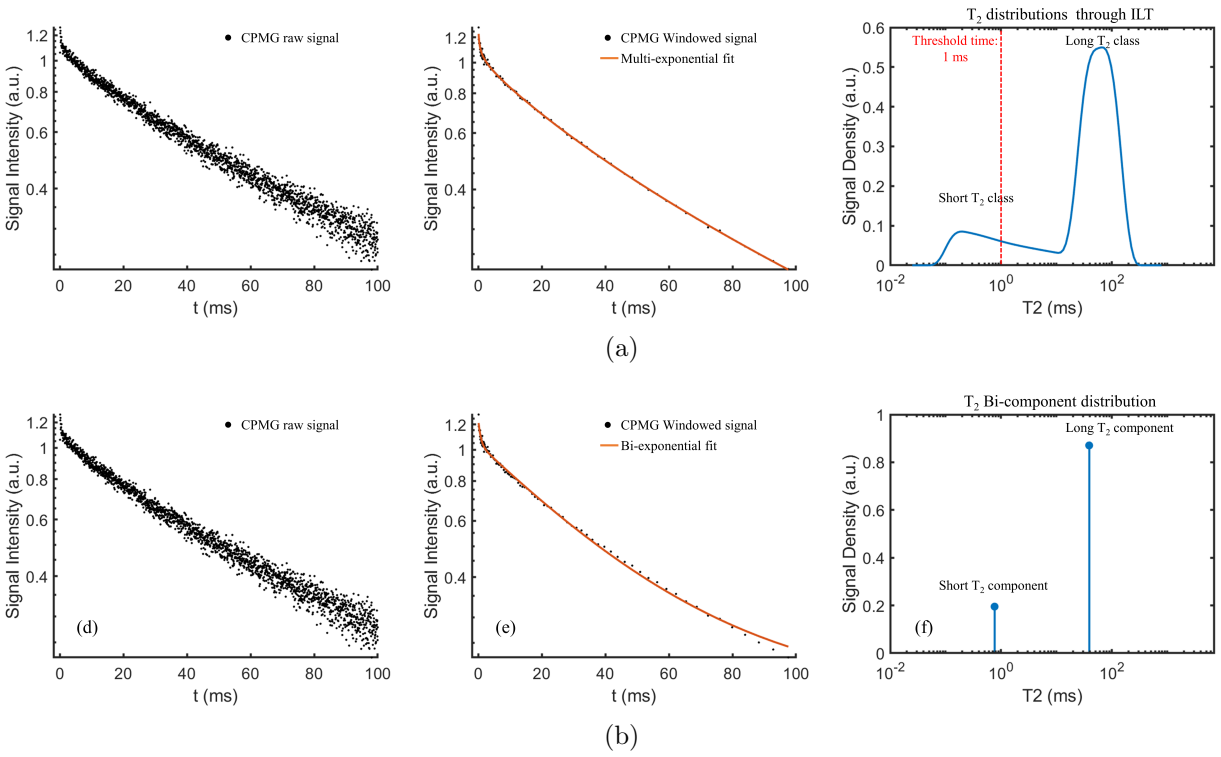


Figure 6.4: Pipeline for analysisg CPMG decay curves using UpenWin software along with (a) UPEN algorithm and (b) bi-exponential fit for T_2 times estimation. Noise reduction through windowing of raw CPMG data is performed both for UPEN and bi-exponential approaches.

used for obtaining the T_2 distribution out of the CPMG raw data. Raw data (Fig. 6.4(a) (a)) are first de-noised using a sliding variable window average filter, i.e. raw data within a certain interval of times are averaged together to mitigate noise. For the first few time points of the decay no average is performed to avoid suppression of fast decay processes, i.e. the window has length one. The window is then translated through the decay curve and the length of the window is progressively increased until it reaches a fixed maximum length. After the windowing procedure, the UPEN algorithm perform the ILT (Fig. 6.4(a) (b)) and the T_2 distribution is obtained (Fig. 6.4(a) (c)).

The T_2 short fraction is evaluated following equation 6.1, where S_{total} is the total area under the T_2 distribution and $S_{\text{short } T_2}$ is the area under the peak representing the short T_2 component. For quasi-continuous distributions, a threshold time is used to separate long and short T_2 classes. In the context of separating intra trabecular signal, the threshold time can be set on the basis of reported literature results regarding T_2 measurements on cortical and trabecular bones performed in homogeneous fields [41, 55, 94]. Inside the trabeculae, the bound water is characterized by one distinct peak around 200 - 300 μs with a cut-off at 1 ms, with lipids and pore water belonging to bone tissue having T_2 of a few ms. Bone marrow has a T_2 peak around 80 ms, but it is characterized by a tail to shorter T_2s as little as a few ms. Hence, in the T_2 region going from 1 ms to few unit of ms the signal comes both from bone and marrow tissues, and the marrow is expected to contribute for most of the signal. This reasoning brought to set the threshold time for separating intra and inter signal contributions to T_2 equals to 1 ms. A standard bi-exponential fit of windowed CPMG raw data was also used on data to evaluate the quantification of $S_{\text{short } T_2}$ and $S_{\text{long } T_2}$ as shown in Fig 6.4(b). The T_2 short fraction of trabecular bone samples under investigation was compared with the morphological parameters obtained from the micro-CT and the Pearson's coefficient was used to assess correlation between parameters measured with the NMR-MOUSE PM10 and the micro-CT.

The BV/TV ratio is estimated following eq. 6.4. To perform such estimation a value for the $\phi_{\text{trabecula}}$ has to be used. Following the methodology reported in [40], the mean intra trabecular porosity of a trabecular bone sample can be estimated from its T_2 distribution and knowing its total porosity ϕ using the following equation:

$$\phi_{\text{trabecula}} = \phi S_{\text{short } T_2}/(1 - \phi + S_{\text{short } T_2}\phi) \tag{6.7}$$

Following a similar criterion introduced in [23], for each sample an estimation of the total porosity is found by dividing the total signal coming from the bone sample and the total signal coming from a reference bone marrow sample having the same volume of the trabecular

bone sample. The intra trabecular porosity for each sample is then computed. The values of $\phi_{\text{trabecula}}$ from this study are then compared with the values found from Fantazzini et al [40] to assess agreement between the two findings. For the purpose of BV/TV estimation, computed using equation 6.4, the mean $\phi_{\text{trabecula}}$ of the set of samples is used for all the samples instead of using the sample specific $\phi_{\text{trabecula}}$. This is because the aim of the proposed methodology is to overcome the need of a reference sample to perform BV/TV estimation using single-sided NMR. However, for the purpose of comparison, the BV/TV was also evaluated using the estimated $\phi_{\text{trabecula}}$ for each sample. This latter method is equivalent to use the method introduced in [23].

The BS/TV ratio is estimated using equation 6.6, where r was set equals to the half mean trabecular thickness of the set of samples, which was estimated from the micro-CT measurements.

The BV/TV and BS/TV ratios of bone samples obtained using equations 6.4 and 6.6, respectively, were compared with the ratios evaluated from the micro-CT, which are here taken as the ground-truth values. The Pearson's and Lin's correlation coefficients were used to assess correlation and agreement between parameters measured with the NMR-MOUSE PM10 and the micro-CT.

Errors associated with BV/TV and BS/TV parameters were estimated with the following reasoning. For the analysis with UPEN, the main source of uncertainty is the estimation of the total area under the T_2 distribution that UPEN computes, whereas we consider negligible the error that the algorithm does in assigning intensities of each T_2 component in the quasi-continuous distribution. Because performing repeated measurements on trabecular bone samples is not ideal since the samples can undergo degradation during measurements, an inorganic sample made of rubber was considered and tailored in such a way to obtain the same SNR level as measurements on bones. Thirty repeated measurements were performed, and mean and standard deviation were used to evaluate the relative error to be associate to S_{total}. Relative errors for BV/TV and BS/TV were then computed using error propagation.

6.3 Results

6.3.1 Micro-CT analysis

The morphological parameters evaluated using the micro-CT are reported in Tab. 6.1. The data show good amount of variability in bone microstructure among samples, as highlighted by the BV/TV that ranges from 24% to 53%. Variability in bone microstructure is not only found among samples cored in different regions of the pig shoulder, but also within samples

cored from the same sites but analyzed at different depths. For example, the BV/TV of the trabecular cylinder M1-1 is 38% at the bottom (M1-1 No Marker) and 28% at the top (M1-1 Marker). Rapresentative slices of cored trabecular cylinder M1-1 are reported in Fig. 6.3.

Sample	Orientation	BV/TV (%)	BS/TV (%)	Tb.Nb (mm^{-1})	BS/BV (mm^{-1})	Tb.Th (mm)	Tb.Pf (mm^{-1})
M1-1	No Marker	38	9.1	2.26	23.7	0.171	5.3
M1-1	Marker	28	6.4	1.65	23.0	0.170	6.0
M1-2	No Marker	41	9.2	2.31	22.4	0.178	4.2
M1-2	Marker	36	7.7	1.99	21.6	0.180	3.1
M1-3	No Marker	45	9.9	2.54	22.0	0.177	2.4
M1-3	Marker	31	6.2	1.59	20.6	0.189	5.6
M1-4	No Marker	44	10.7	2.70	23.9	0.166	-4.5
M1-4	Marker	32	7.5	1.92	23.1	0.170	4.3
M1-5	No Marker	35	8.1	2.03	23.7	0.168	-1.4
M1-5	Marker	26	6.1	1.61	23.6	0.160	5.5
M2-1	No Marker	34	9.0	2.23	26.5	0.152	4.7
M2-1	Marker	29	7.6	1.94	25.5	0.154	0.1
M2-2	No Marker	51	13.7	3.62	26.3	0.144	-27.8
M2-2	Marker	33	7.4	1.83	22.9	0.177	6.5
M2-3	No Marker	24	6.7	1.61	27.9	0.148	9.9
M2-3	Marker	38	7.5	1.95	20.0	0.193	2.0
M2-4	No Marker	28	7.0	1.70	25.0	0.164	7.9
M2-4	Marker	53	12.8	3.60	24.9	0.143	-18.8
M3-1	No Marker	44	9.0	2.22	20.3	0.201	1.5
M3-1	Marker	41	8.5	2.16	20.4	0.192	1.5
M3-2	No Marker	45	11.1	2.77	24.7	0.162	0.1
M3-2	Marker	34	8.1	2.16	23.6	0.160	4.3

Table 6.1: Morphological parameters of trabecular bone samples measured through micro-CT.

Visually analysing the acquired images, 6 samples characterized by tissue heterogeneity were found. Representative slices of these samples are reported in Fig 6.5, where presence of tissues other than trabecular bone is visible. These tissues are growth cartilage and bone tissue characterized by lower mineralization than surrounding trabecular bone.

Principal component analysis (PCA) was applied on bone samples using the micro-CT computed morphological parameters as feature vectors. Figure 6.6 reports the projection of the bone feature vectors on the 2D space spanned by the first two principal components, where the red marked dots refers to the 6 samples in which tissue heterogeneity was found. Even without performing a numerical cluster analysis, two clusters are clearly separable. One cluster contain 20 out of 22 samples, whereas one cluster is composed of samples "M2-2 No Marker" and "M2-4 Marker", which are the samples characterized by the most high tissue heterogeneity (See Fig. 6.5 (d) and (f)).

Figure 6.7 reports image analysis for three bone samples, that are: M1-1-No-Marker, M1-4-No-Marker and M2-2-No-Marker. For each sample, a micro-CT image of an acquired slice is reported along with its histogram and binary image obtained after global thresholding

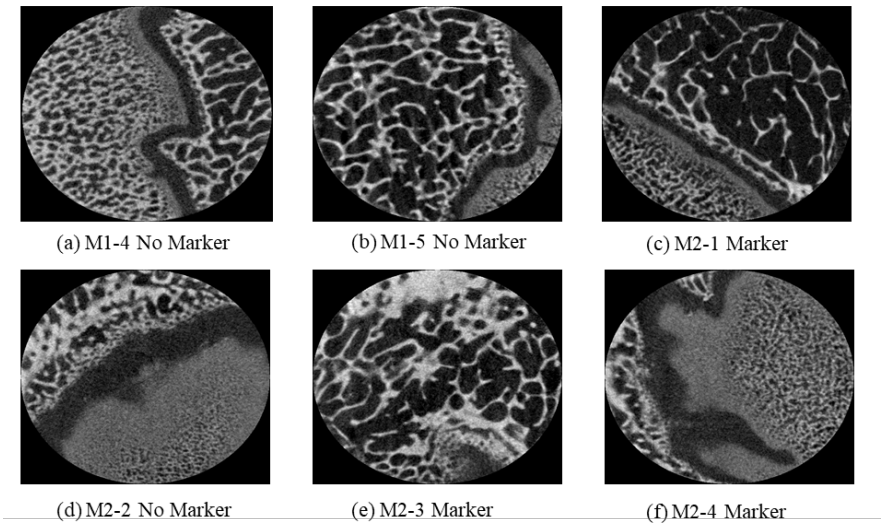


Figure 6.5: Example slices from micro-CT images of the samples in which tissues other than bone marrow and bone is found. These other tissues are: growth cartilage, which has the same gray level of the bone marrow, and less mineralized bone tissue, appearing less intense than bone in the images.

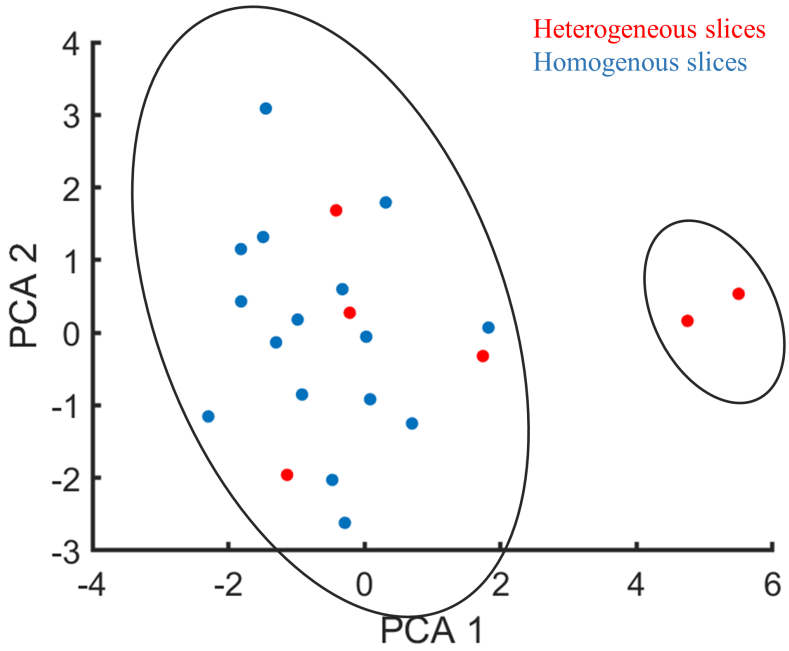


Figure 6.6: PCA analysis of the trabecular bone samples where each sample is characterized by a feature vector composed of morphological parameters evaluated from the micro-CT images. Before PCA, the data set has been normalized. Red dots mark the samples in which growth cartilage and less mineralized bone tissue are present.

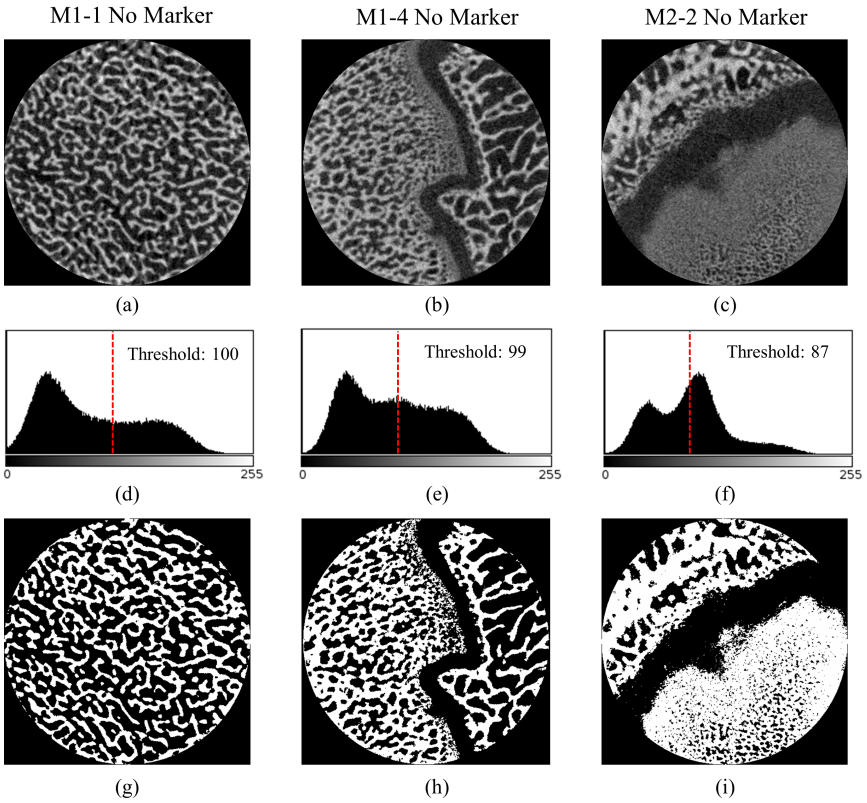


Figure 6.7: Examples of micro-CT images (top row), image histograms (middle raw) and thresholded images (bottom row) for three slices that represent three different conditions found in our set of bone samples. (a), (d) and (g) refers to a sample in which only bone and marrow tissues are present. (b), (e) and (h) refers to a sample in which growth cartilage and less mineralized bone tissues are present in addition to marrow and bone tissues. (c), (f) and (i) refers to a sample in which growth cartilage and less mineralized bone tissues are present in addition to marrow and bone tissues, and the less mineralized bone dominates the image.

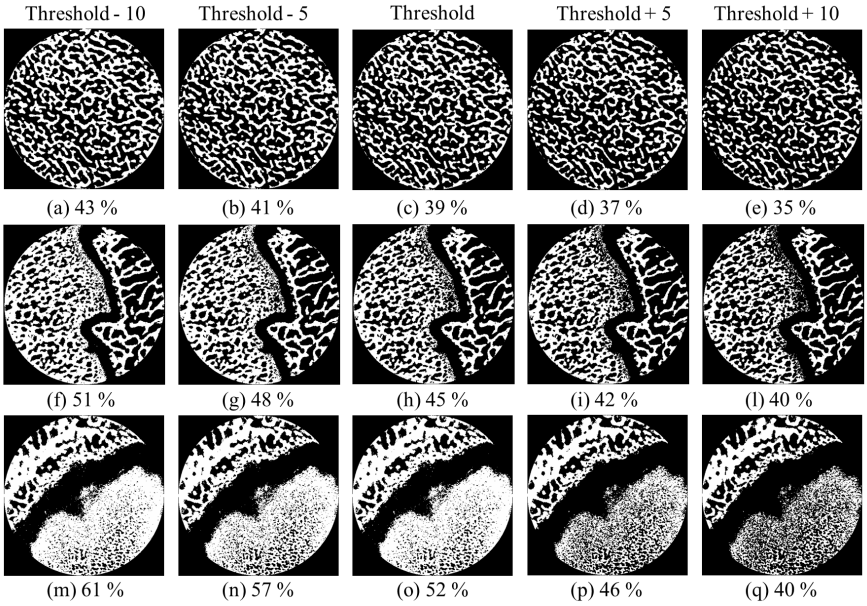


Figure 6.8: Examples showing the effect on BV/TV estimation of choosing threshold values that differ few gray levels from the threshold found by the Otsu's method.

with the threshold value automatically estimated using the Otsu's method. The micro-CT images (Fig. 6.7 (a), (b) and (c)) allow to visually observe that sample M1-4-No-Marker and M2-2-No-Marker have other tissues other than trabecular bone and marrow, whereas sample M1-1-No-Marker only presents these latter tissues. The image histogram of sample M1-1-No-Marker shows one peak at low intensity value, which represents the bone marrow, a shoulder at high intensity values, which represents the bone tissue, and a flat transition in between. For sample M2-2-No-Marker, the low intensity value peak embeds both bone marrow and growth cartilage (appearing dark in the micro-CT image), and instead of a flat transition between low intensity peak and high intensity shoulder, a relatively high populated third peak is present. This latter peak belongs to bone characterized by less mineralization compared to fully calcified bone, as highlighted from the lower gray level values they have. In sample M1-4-No-Marker, where this latter tissue is less present, the height of this peak make it barley visible between low-intensity values peak and high-intensity values shoulder, and the histogram still appears quite flat in-between. Bone volume fraction was evaluated in these three samples as a function of the global threshold used for segment the bone tissue. Global threshold values that were 10 gray levels less and greater than the Otsu's threshold value were used. The bone volume fraction was then evaluated for the segmented images. In Fig. 6.8 the segmented images are reported for different threshold values along with the computed BV/TV ratio. For samples M1-1-No-Marker and M1-4-No-Marker there is a linear relationship between variation in threshold value and the corresponding BV/TV value. On

the contrary, for sample M2-2-No-Marker, where less mineralized bone is vastly present, the variation is not linear because using a threshold value that is the 10 gray levels lower than the Otsu's value it increases the BV/TV of about 12%, whereas using a threshold value that is 10 gray levels greater than the Otsu's value it decreases BV/TV of about 9%. This latter trend was found also for sample M2-4-Marker, even if not shown, which is the other sample that presents a relative high amount of less mineralized bone tissue.

For the purpose of comparing NMR and micro-CT results, the samples M2-2-No-Marker and M2-4-Marker were excluded for two main reasons. First, the PCA analysis clearly differentiated those samples from the others. Second, because of the relative high presence of less mineralized bone tissue for which the NMR properties have not been studied. However, the other four samples that presented growth cartilage and a relative low amount of less mineralized bone tissue were not excluded in the comparison because, on the basis of PCA analysis, they were considered similar to the other samples.

6.3.2 T_2 distributions from NMR measurements

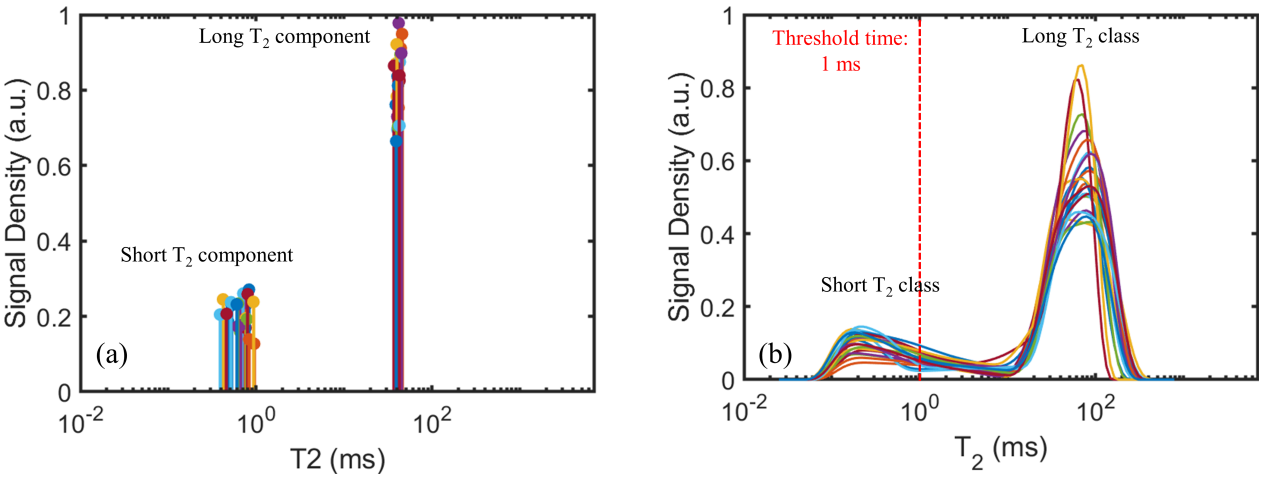


Figure 6.9: T_2 distributions of trabecular bone samples. The bi-component distributions evaluated with the bi-exponential fit are reported on the left, while on the right the quasi-continuous T_2 distributions computed with UPEN are reported. In this latter case the threshold time to distinguish the intra and inter trabecular signals, which are the short and long T_2 classes, respectively, is set to 1 ms.

In Fig. 6.9 the quasi-continuous T_2 distributions computed with UPEN are reported for the bone samples along with the bi-component distributions evaluated with the bi-exponential fit. For almost all the samples, the T_2 distributions show two peaks: one peak has T_2 value near 80 ms, the other has T_2 peak value around 200 -300 μs. A wide transition region between long and short T_2 peaks is present between 1 ms and 10 ms. The bi-exponential fit (Fig.

6.9 (b)) allows clear separation between short and long T_2 components with the drawback of loosing information about the full spectrum of transverse relaxation times. While for the long T_2 component, the estimated T_2 time is around 40 - 50 ms for all the bone samples, the estimated T_2 time of the short T_2 component span a wider range of values among samples, with values ranging from 0.4 to 1 ms. Residuals analysis performed on the fits performed by UPEN and the bi-exponential fit shows that the fit errors are comparable between the two algorithms. Fit mean percentage error (MPE) is (-0.04 ± 0.03) percentage units and (-0.06 ± 0.02) percentage units for UPEN and bi-exponential fits, respectively. Fit root mean squared errors (RMSE) is 0.014 ± 0.001 and 0.016 ± 0.001 for UPEN and bi-exponential fits, respectively.

6.3.3 Data analysis using quasi-continuous T_2 distributions computed with UPEN

T_2 short intensity fraction: preliminary assessing correlation with morphological parameters

The NMR computed T_2 short intensity fraction as a function of micro-CT computed morphological parameters is reported in Fig. 6.10 for the trabecular bone samples, using the quasi-continuous T_2 distributions for estimating the signal intensity of the spin class characterized by the short T_2, i.e. $S_{\text{short } T_2}$. Samples M2-2-No-Marker and M2-4-Marker, which belong to a separate cluster in respect to the other bone samples in the PCA analysis, were excluded from the analysis. The four samples that, on the basis of micro-CT image visualization, presented growth cartilage but did not appeared to be significantly different from other samples on the basis of the PCA analysis were included in the comparison, and they are marked with red dots in the plots. For each plot, a linear fit was evaluated and the Pearson's coefficient is used to measure the degree of the correlation. T_2 short intensity fraction is strongly positively correlated with BV/TV, BS/TV and Tb.Nb, being the correlation coefficients equals to 0.94, 0.81, and 0.81, respectively. A moderate positive correlation is found with Tb.Th, being r equals to 0.51, and moderate negative correlation is found with the Tb.Pf parameter, r equals to -0.52. Moderate negative correlation is also found between T_2 short intensity fraction and BS/BV, with the correlation coefficient equals to -0.45.

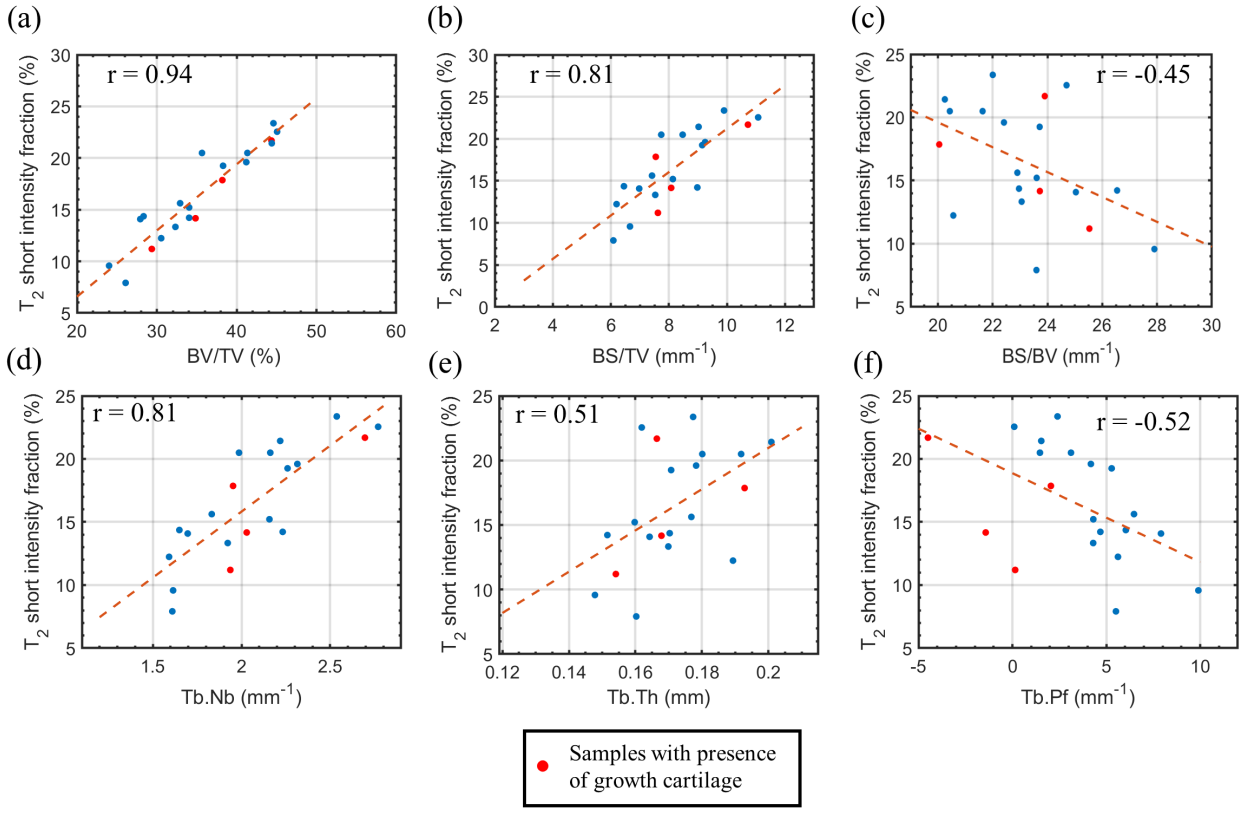


Figure 6.10: Correlation between T_2 short intensity fraction, evaluated using the quasi-continuous T_2 distributions, and the morphological parameters evaluated using the micro-CT. (a) BV/TV, (b) BS/TV, (c) BS/BV, (d) Tb.Nb, (e) Tb.Th and (f) Tb.Pf

Intra trabecular porosity

The intra trabecular porosity of bone samples estimated using equation 6.7 are reported in Tab. 6.2 along with the intra trabecular porosity reported in the study of Fantazzini et al [40]. Each measurements is affected by an error in the order of few percentage units on the reported values. The mean intra trabecular porosity of the samples under investigation was found to be (33 ± 5) percentage units, which is comparable to the mean value reported in [40], which is (29 ± 5) percentage units.

This study			Fantazzini et al. [40]	
Sample label	Orientation	$\phi_{trabecula}$ (%)	Sample label	$\phi_{trabecula}$ (%)
M1-1	Marker	31	1	31
M1-1	No marker	34	2	30
M1-2	Marker	40	3	29
M1-2	No marker	32	4	31
M1-3	Marker	25	5	30
M1-3	No marker	38	6	34
M1-4	Marker	27	7	26
M1-4	No marker	41	8	25
M1-5	Marker	20	9	26
M1-5	No marker	34	10	26
M2-1	Marker	37	12	25
M2-1	No marker	32	13	36
M2-2	Marker	39	14	22
M2-3	Marker	37	15	38
M2-3	No marker	27		
M2-4	No marker	37		
M3-1	Marker	34		
M3-1	No marker	34		
M3-2	Marker	33		
M3-2	No marker	32		
	Mean ± Std			Mean ± Std
	33 ± 5			29 ± 5

Table 6.2: on the left, intra trabecular porosity of the trabecular bone samples considered in this study evaluated using a signal coming from a reference sample having the same volume of the samples but composed of only bone marrow. Signal was analysed using UPEN. On the right: reported intra trabecular porosity found by Fantazzini et al in [40]. Each reported intra porosity value is affected by an error that is in the order of few percentage units.

The ANOVA test was applied between the group of measurements to assess if the difference between the mean values found in the two groups were significantly different. It is worth mentioning that all the samples in which growth cartilage and less mineralized bone tissue were present other than bone and marrow tissues were excluded from the test, which are 4 out of 20 samples. This is because these tissues were not present in any of the samples

used in [40]. The test is graphically reported in Fig. 6.11. No significance difference was found between the two groups of measurements (p =0.12).

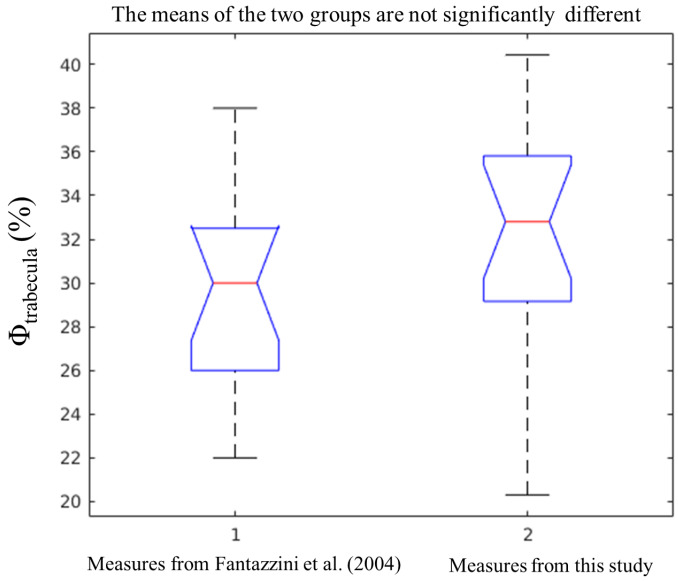
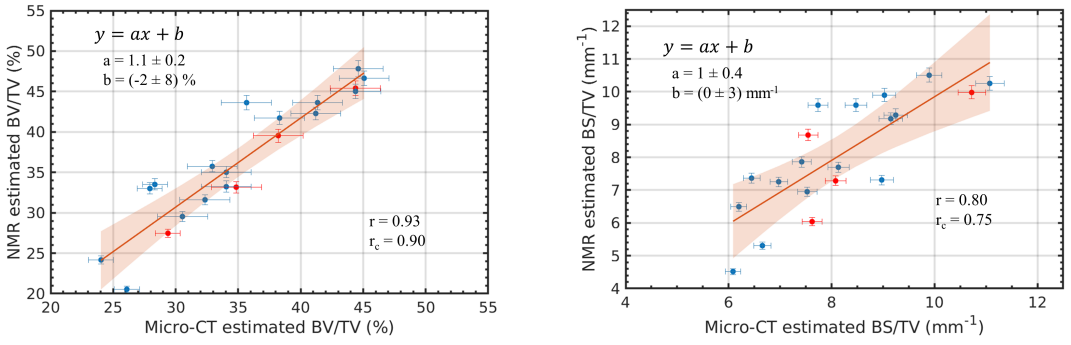


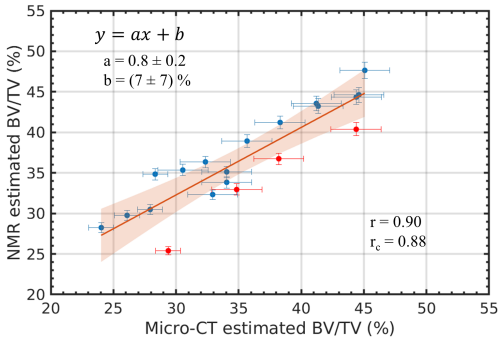
Figure 6.11: Box plots of the intra-trabecular porosity values reported b et al [40] and those found in this study using the same procedure described in ref. [40]. The means of the two groups of measurements are consistent on the basis of a t-test (p =0.12).

NMR computed BV/TV and BS/TV ratios: comparison with micro-CT

Figure 6.12 reports the comparison between the BV/TV and BS/TV ratios estimated from NMR measurements, using eq. 6.4 and 6.6, and the corresponding ratios evaluated with the micro-CT analysis. Figure 6.12(a) reports the NMR estimation of BV/TV where for all the sample the intra trabecular porosity group mean was used as value of $\phi_{\text{trabecula}}$ in equation 6.4. A strong correlation and agreement between NMR and micro-CT computed BV/TV ratios is reported, as testified by the Pearson's coefficient being equals to 0.93 and the Lin's coefficient, which is equals to 0.90. The linear fit reports a slope of 1.1 $\pm$ 0.2, and a bias term of (-2 $\pm$ 8)%. In Fig. 6.12(b), the NMR computed BS/TV ratios were evaluated using the group means of intra trabecular porosity and trabecular thickness as values of $\phi_{\text{trabecula}}$ and r in eq. 6.6. A moderate strong correlation is found between NMR and micro-CT estimated values, with the Pearson's coefficient being equal to 0.80, and moderate agreement is found between the two methods. Indeed, the Lin's coefficients is 0.75. The linear fit reports a slope of 0.6 $\pm$ 0.2, and a bias term of (0 $\pm$ 3) mm^{-1}. Finally, Fig. 6.12(c) reports the comparison between NMR and micro-CT computed BV/TV where the sample specific $\phi_{\text{trabecula}}$ (See Tab. 6.2) was used instead of using the group mean value for all the samples in eq. 6.4.



(a) (b)



(c) NMR values estimated using the methodology introduced in [23].

Figure 6.12: Comparison between NMR and micro-CT computed parameters. In (a) and (b) for each sample, the sample group means of intra trabecular porosity and half trabecular thickness were used as values of $\phi_{\text{trabecula}}$ and r in eq. 6.4 and 6.6 to compute BV/TV and BS/TV ratios, respectively. In (c) the NMR computed BV/TV values were estimated using eq. 6.4 using the sample specific $\phi_{\text{trabecula}}$, as reported in Tab. 6.2, which is equivalent to use the method presented in [23], and affected by the necessity of using the reference signal.

It is worth reminding that this is equivalent to use the methodology introduced in [23]. A strong correlation and agreement between NMR and micro-CT BV/TV values is reported, as demonstrated by the Pearson's and Lin's coefficients, which are equal to 0.90 and 0.88, respectively. The linear fit reports a slope of 0.8 ± 0.2, and a bias term of (7 ± 7) percentage units, values that are consistent to what found in [23] (See Fig. 4.7).

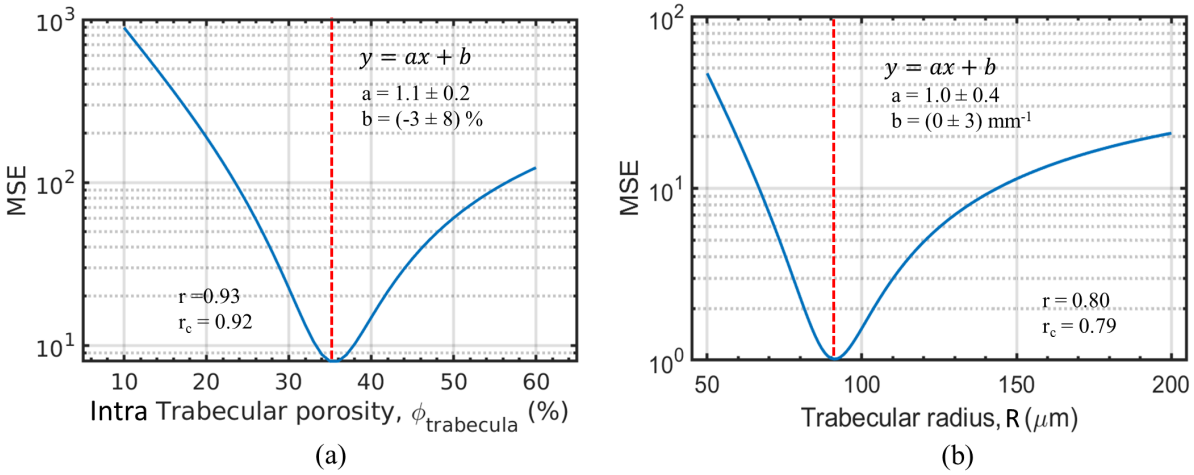


Figure 6.13: On the left, Mean Squared Error (MSE) between NMR and micro-CT computed BV/TV as a function of the $\phi_{\text{trabecula}}$ parameter put in eq. 6.4. On the right, MSE between NMR and micro-CT computed BS/TV as a function of the trabecular radius, r, put in eq. 6.6.

To retrospectively find the $\phi_{\text{trabecula}}$ and trabecular radius that have to be used in eq. 6.4 and 6.6, respectively, to minimize the mean squared error (MSE) between NMR and micro-CT estimated values of BV/TV and BS/TV ratios, different values of $\phi_{\text{trabecula}}$ and trabecular radii were tried and MSE computed for each trial. The MSEs between NMR and micro-CT parameters for BV/TV and BS/TV parameters as a function of $\phi_{\text{trabecula}}$ and trabecular radius, respectively, are are reported in Fig. 6.13. The best $\phi_{\text{trabecula}}$ is found to be 35%, which is consistent with the measured mean intra trabecular porosity of the set of samples. Linear model coefficients do not significantly differ from those reported in Fig. 6.12(a). The best trabecular radius is found to be around 90 μm, which is less than all the trabecular thickness measured from the micro-CT, but it is still a realistic value for trabecular thickness. With this value, the Lin's coefficient goes up to 0.90, meaning that setting this value brings to a strong agreement between NMR and micro-CT BS/TV values. Linear model coefficients significantly differ from those reported in Fig. 6.12(b): the slope is equal to 1.0 ± 0.4 and the bias term is equal to (0 ± 3) percentage units.

6.3.4 Data analysis using bi-component T_2 distributions computed with a bi-exponential fit

Intra trabecular porosity

The comparison between the group of intra trabecular porosity measured for the samples under investigation, and the porosity reported in [40] for trabecular bones coming from pig and cow femurs is reported in Fig. 6.14.

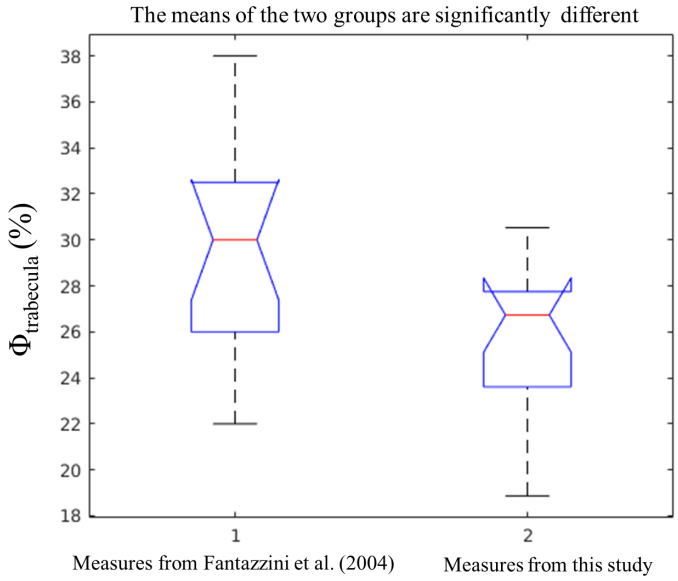


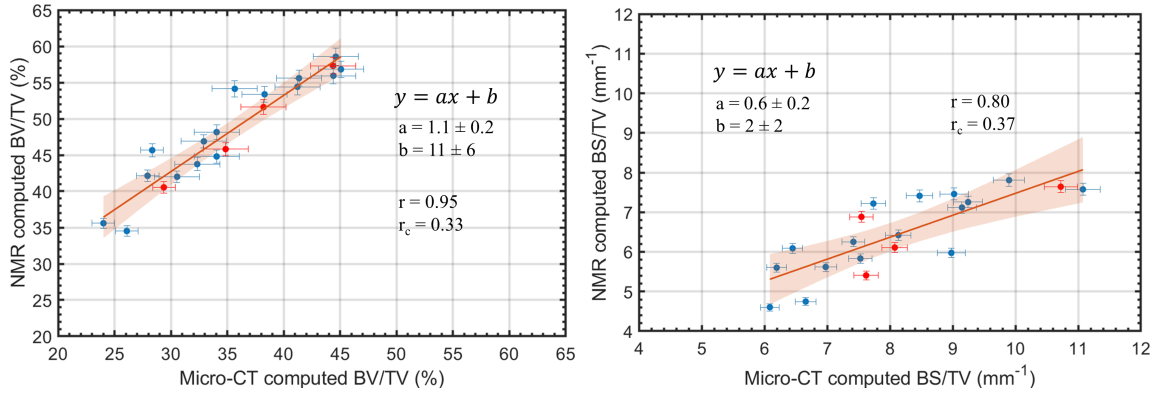
Figure 6.14: Box plots of the intra-trabecular porosity values reported b et al [40] and those found in this study using the bi-exponential fit with the same procedure described in ref. [40]. The means of the two groups of measurements are not consistent on the basis of the t-test (p =0.01).

The mean intra trabecular porosity in the group of measurements from this study was found to be (27 ± 3)%, which is not consistent with the mean value found in [40], which is (29 ± 5) percentage units, as verified on the basis of a t-test (p =0.01).

NMR computed BV/TV and BS/TV ratios: comparison with micro-CT

Fig. 6.15 reports the comparison between NMR and micro-CT computed parameters using the bi-exponential fit to estimate the signal intensity of the spin class characterized by the short T_2 in the CPMG decay. Although strong correlation is found for both BV/TV and BS/TV with ground-truth values, the agreement with micro-CT values is only weak to moderate. For BV/TV r is equals to 0.95, but Lin's coefficient is 0.33, which is considerably less than the Lin's coefficient found when the quasi-continuous T_2 distribution is used to separate

intra and inter trabecular signals. The low moderate agreement is also highlighted by the bias term of the linear fit, which is (11 ± 6) percentage units, whereas the slope is 1.1 ± 0.2. For the BS/TV ratio, the Pearson's coefficient is 0.90 and the Lin's coefficient is 0.37. The slope of the linear fit is equal to 0.6 ± 0.2, whereas the bias term is equal to (2 ± 2) percentage units.



(a) (b)

Figure 6.15: Comparison between NMR and micro-CT computed parameters. Comparison between NMR and micro-CT computed parameters. For each sample, the sample group means of intra trabecular porosity and half trabecular thickness were used as values of $\phi_{\text{trabecula}}$ and r in eq. 6.4 and 6.6 to compute BV/TV and BS/TV ratios, respectively.

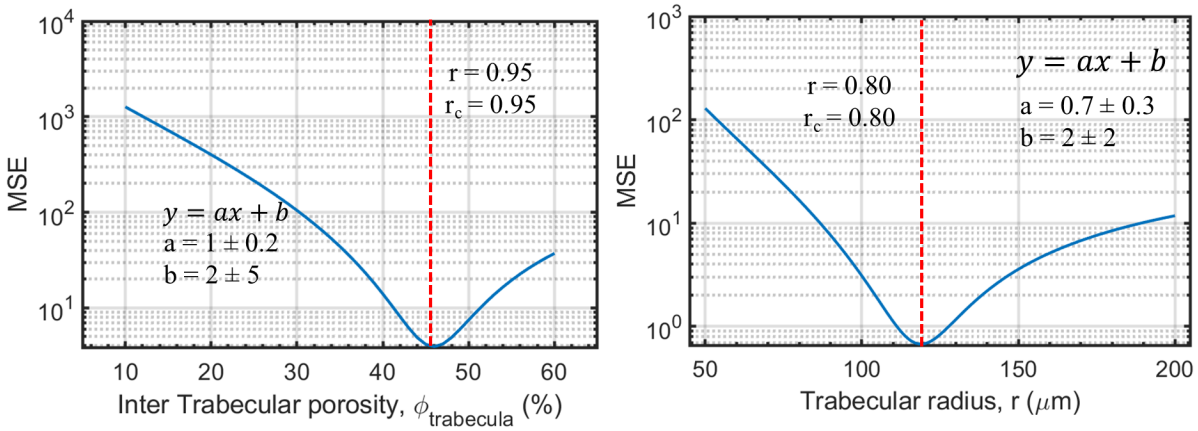


Figure 6.16: On the left, Mean Squared Error (MSE) between NMR and micro-CT computed BV/TV as a function of the $\phi_{\text{trabecula}}$ parameter put in eq. 6.4. On the right, MSE between NMR and micro-CT computed BS/TV as a function of the trabecular radius, r, put in eq. 6.6.

The MSEs between NMR and micro-CT parameters for BV/TV and BS/TV parameters as a function of the $\phi_{\text{trabecula}}$ and the trabecular radius, respectively, are reported in Fig.

6.16. The best $\phi_{\text{trabecula}}$ is found to be 46%, which is considerably higher than the measured mean intra trabecular porosity, and it is also considerably higher than the value reported in [40]. With this value, the BV/TV Lin's coefficient passes from 0.33 to 0.95. The best trabecular radius is found to be around 130 μm, which is in the order of the trabecular thickness measured from the micro-CT. With this value, the Lin's coefficient is 0.80, meaning a moderate strong agreement between NMR and micro-CT estimated BS/TV ratios.

6.4 Discussion

The findings here presented show that the quantification of BV/TV and BS/TV ratios of trabecular bone samples using a single-sided NMR scanner, without the need of a reference sample, strongly correlate and agree with the parameter values obtained using the gold standard micro-CT analysis (See Fig. 6.12). Moreover, the study also evaluated the correlation between the T_2 short intensity fraction and 6 morphological parameters obtained from micro-CT images. Strong correlations was found with BV/TV, moderate strong correlation was found with BS/TV and Tb.Nb, and moderate correlation was found with BS/BV, Tb.Th and Tb.Pf.

The assessment of trabecular bone morphological parameters using NMR relies on the capability of distinguishing the signal coming from inside the trabeculae and the signal coming from the inter trabecular spaces. Although the capability of NMR relaxometry to provide assessment of bone micro-structure have been already demonstrated in the literature using NMR and MRI [37, 40, 41, 54, 78, 79, 83, 88, 95, 98, 107], most of the studies have focused on cortical bones. Studies regarding the trabecular bone have said little to nothing about the relationship between the T_2 short intensity fraction and the morphological bone parameters. This is mainly because studies carried out using MRI aimed to directly assess the morphological parameters by taking advantage of the spacial resolution of the NMR signal [65, 78, 95]. Studies regarding the application of NMR relaxometry to trabecular bone focused on the evaluation of trabecular bone porosity using a reference sample [23, 40, 41]. The demonstration of the possibility to assess trabecular bone micro-structure using a single-sided NMR scanner with the ability to quantify the absolute value of BV/TV and BS/TV ratios without the need of using a reference sample, is, to our best knowledge, an original contribution of this work.

Laplace inversion of CPMG decays allowed one to estimate quasi continuous T_2 distributions and to quantify the intensity fraction of the signal coming from the intra trabecular bound water. The high correlation between the T_2 short intensity fraction and the micro-CT computed BV/TV ratio confirms the goodness of attributing that signal component to intra

trabecular bound water, as one could have expected from the studies conduced on cortical bones [55]. With the use of theoretical models, we were able to perform the estimation of the absolute values of BV/TV and BS/TV ratios, as discussed in section 6.1.2, without the need of using a reference sample. Although these models are a simplification of the reality, the strong level of correlation and the moderate strong agreement with the micro-CT estimations suggest that these models are reasonable representation for the purpose of parameters estimation.

Regarding BV/TV quantification, the correct estimation of the fraction of signal within the trabeculae is a critical point of the whole quantification procedure. A threshold time of 1 ms in the quasi-continuous T_2 distribution computed by UPEN was demonstrated to be very reasonable. This value was in origin selected from reasoning based on the available knowledge found in the literature about the T_2 relaxation times of the water compartments in trabecular bones. The finding that the intra trabecular porosity $\phi_{\text{trabecula}}$ measured for our set of bone samples were found to be not significantly different from the values obtained in [40] is an empirical evidence of the goodness of this choice. The results obtained using the bi-exponential fit are a further confirmation that 1 ms is a reasonable threshold for separatating intra and inter trabecular signals. When bi-exponential fit is used to analyze the CPMG decays and distinguish intra and inter trabecular signals results are much less in agreement with literature values, for the intra trabecular porosity $\phi_{\text{trabecula}}$, and with micro-CT analysis for BV/TV quantification. As highlighted in Fig. 6.9, the bi-compartment model is an oversimplification of the reality, and indeed the ILT performed by UPEN finds quasi-continuous T_2 times between 1 ms and 10 ms. With the bi-exponential fit some of these components are assigned to the short T_2 component, with the effect of assigning signals that comes from inter trabecular spaces to intra trabecular compartments. This leads to an overestimation of the bone volume, which explains why NMR BV/TV values overestimate the micro-CT values when the the bi-exponential fit is used (See Fig. 6.15(a)). It is also interesting to notice that the method does not fail even when the trabecular sample has cartilage inside. This is because T_2 relaxation times of the cartilage are in the order of tens of ms, which is the same order of magnitude of the T_2 relaxation times of the marrow. Hence, the cartilage is considered as inter trabecular signal, since is part of the long T_2 component, so that does not contribute to the bone volume estimation. In micro-CT images, growth cartilage has gray level intensities similar to bone marrow, so that also in micro-CT analysis growth cartilage is not considered to estimate the bone volume.

Regarding the BS/TV ratio, although the model (See eq. 6.6) makes the simplified assumption that all the trabeculae are cylindrical, a good agreement is found between the NMR and micro-CT estimated BS/TV ratios. Moreover, the trabecular radius found using

the optimization method, around 90 μm, is very consistent to what measured by the micro-CT.

Regarding the correlation between T_2 short intensity fraction and morphological parameters, these kind of correlation have been investigated in literature, but mainly in cortical bone [94]. This bio-marker has been found promising to assess bone quality. In MRI studies on cortical bones, a well know shortcoming of in vivo evaluation of T_2 short intensity fraction is that it requires a CPMG measurement, which is not clinically feasible due to SAR limits and the necessity to use very short echo times and 180 pulse lengths [94]. Using the T_2^* instead of T_2 allows in vivo quantification, but the strength of the correlation with morphological parameters rapidly decreases with the field strength due to the T_2^* shortening of the long T_2 component because of the induced high internal field gradients, which reduces separation between bound and pore water T_2 relaxation times (i.e. short and long T_2 component). The reported results allow to think about the possibility to transfer this methodology, here validated for trabecular bone, to cortical bone. Moreover, the ability of distinguishing bound and pore water in cortical bone though single-sided NMR has been recently demonstrated [24]. In combination with the characteristics of single-side NMR scanners the methodology seems promising to envision in vivo assessment of bound water fraction in cortical bone.

Finally, intra trabecular porosity found in this work are significantly consistent with values found by Fantazzini et al. [40], on a basis of an ANOVA test. Therefore, we are confident in using an intra trabecular porosity of around 30% in eq. 6.4 for BV/TV estimation even for new trabecular samples. However, this confidence holds for healthy trabecular bones. Literature have not clearly established at what extent diseases affecting the bone tissue, such as osteoporosis, modifies the structure at the level of the single trabecula, so that parameters like the porosity of the trabecula could vary. More specifically, it is not clear if the imbalanced and excessive bone remodelling that occurs in osteoporosis only affects the trabecular network, by removing some trabeculae, or it also affects the porosity of the trabeculae. Having established a method to investigate trabecular bone micro structure with single-sided NMR, even at the trabecular level, future studies can focus on assessing if bone diseases affect the intra trabecular porosity.

6.5 Conclusion

In this chapter, a modification of the methodology introduced in [23] for assessing the BV/TV ratio of trabecular bone using using single-sided NMR scanners without the need of using a marrow reference sample is proposed. The key idea behind the technique is to evaluate the BV/TV ratio by separating the CPMG signal contributions coming from the intra and inter

trabecular compartments.

A set of trabecular bone samples, cored from different regions of pig's shoulders, was first measured using a CPMG sequence, and successively micro-CT images were acquired. To separate the intra and inter trabecular signals, the CPMG signal was inverted by computing the associated T_2 distribution using both UPEN, an algorithm for ILT that retrieves quasi-continuous distributions, and a bi-exponential fit. For each sample, the T_2 short intensity fraction was directly evaluated, whereas BV/TV and BS/TV ratios were estimated using theoretical models, which require an esteem of the intra trabecular porosity $\phi_{\text{trabecula}}$ and of the mean trabecular radius (trabeculae are modelled as cylinders). Using NMR the $\phi_{\text{trabecula}}$ was evaluated for each sample after evaluating its total porosity using a reference sample, while micro-CT allowed to estimate the mean trabecular thickness for each sample. The sample group means of $\phi_{\text{trabecula}}$ and trabecular thickness were used in the theoretical models. Comparison of $\phi_{\text{trabecula}}$ values with literature reported values was also performed. Correlation between T_2 short intensity fraction and morphological parameters estimated using micro-CT was assessed, and agreement between NMR and micro-CT computed BV/TV and BS/TV ratios was investigated.

Results have shown that the mean intra trabecular porosity measured in this study, which is (33 ± 5) percentage units, does not significantly differ from the value reported in [40], where the same type of trabecular bones were analysed, on the basis of an ANOVA test. The T_2 short intensity fraction resulted to be strongly correlated with BV/TV and from moderate to moderate strongly correlated with the other morphological parameters. NMR computed BV/TV and BS/TV ratios strongly correlate and agree with the values evaluated using micro-CT.

Overall, the findings here presented showed that is possible to successfully quantify BV/TV and BS/TV ratios of trabecular bone using a single-sided NMR scanner without the need of a reference sample. The proposed technique is a step forward the application of single-sided NMR scanners for in vivo assessing of bone micro-structure, which could be used for wide spread screening campaigns of the population at risk of bone related diseases, such as osteoporosis.

Summary

This Ph.D thesis presented scientific researches that aimed to improve the usefulness of quantitative NMR methodologies for medical applications. The manuscript focused on two different topics, which are expressions of two directions that can be followed to improve clinical applicability of quantitative NMR and therefore to enhance its impact for medical purposes. The work presented in the previous chapters is here summarized.

Part I: Circumventing the Curse of Dimensionality in Magnetic Resonance Fingerprinting through a Deep Learning Approach

The first part of the thesis focused on the set-up of a deep learning approach based on deep Fully Connected Networks for pixel-wise MR parameter prediction task in MRF as solution to overcome the curse of dimensionality affecting the gold standard dictionary approach. Another key feature of this deep learning application was the ability to effectively train the neural network with only simulated data to be able to generalize well on real acquired MRF signals, which are corrupted by noise caused by the k-space undersampling.

Chapter 1

The fundamental concepts of the MRF methodology were introduced and a brief literature review was provided to highlight the main areas that the research community have been investigating to improve applicability of MRF in a clinical setting. The review showed that important efforts have been focusing on the problem of handling the size of the dictionary, which scales exponentially with the number of MR parameters one includes in the Bloch simulations when computing the dictionary. This generates efficiency problems both for the storage of the dictionary and for the computational time inefficiency during template matching. Although several solutions have been reviewed, the curse of dimensionality in MRF

was highlighted as an intrinsic limitation of dictionary based approaches. Deep NNs were then spotlighted as a possible solution for conveniently handle MRF applications disregarding the problem of the size of the dictionary, reviewing the current state of the art of deep learning applied to MRF. The advantage and disadvantages of FCN for parameter map reconstruction in MRF were also discussed. From now on FCN will be referred just as NN, since the deep learning application here presented focused on this architecture.

Chapter 2

The NN model was first set up and by means of simulations we investigated how the NN approach performance scales with the number of parameters to be retrieved in comparison with the standard dictionary approach. An optimal training procedure was also investigated by including different strategies of noise addition and parameter space sampling to achieve better accuracy and robustness to additive Gaussian noise. Four MRF sequences were considered: IR-FISP, IR-bSSFP, IR-FISP-B_1 and IR-bSSFP-B_1, the latter two designed to be more specific for B_1^+ parameter encoding. Preliminary robustness to undersampling artifacts was also assessed. Results demonstrated that training with random sampling and different levels of noise variance yielded the best performance. NN performance was at least as good as the dictionary-based approach in reconstructing MR parameter maps using Gaussian noise as source of artifacts: the difference in performance increased with the number of estimated parameters, because the dictionary method suffers from the coarse resolution of the MR parameter space sampling. The investigation of robustness to undersampling artifacts allowed us to characterize the limits whithin which the current form of the NN approach is less robust to undersampling artifacts than the dictionary approach.

Chapter 3

The major goal of the chapter was to modify the training pipeline set-up in chapter 2 to correctly train the NN, using simulated data, so that it is accurate and robust to artifacts affecting real MRF signals, such as those arising from undersampling. Using an ablation approach, we trained the NN using simulated MRF data and different training pipelines and tested their performance on both fullysampled and undersampled in-vivo data acquired with the IR-FISP MRF sequence. Each pipeline used a different combination of data augmentation steps (that are online noise adding and phase offset multiplication), preprocessing steps (that are k-SVD de-noising filter and normalization) and input type selection (i.e. either the magnitude of the signal or the complex signal). Moreover, three models of noise were tested: additive Gaussian noise, multiplicative Gaussian noise and a combination of additive and

multiplicative Gaussian noises. The performance of the NN was assessed by comparison with the T_1 and T_2 maps estimated using the standard dictionary approach. Results allowed to heuristically determine the optimal training strategy to train a NN able to predict T_1 and T_2 maps that are in strong agreement with those obtained with the dictionary approach. To promote accuracy and robustness to undersampling artifacts, the complex signal should be used as input of the network, and during training online noise adding and phase offsets multiplication have to be applied as data augmentation steps. Moreover, k-SVD de-noising is necessary as preprocessing step to handle undersampled data. Moreover, although additive Gaussian noise does not constitute a realistic model for noise arising from undersampling, it has been found to be the best noise model to train the NN. This last finding needs further investigations, since it could be related to the heteroskedasticity of the multiplicative noise, which could determine lower performance in k-SVD and/or NN model weights optimization.

In conclusion, a NN approach, trained with only simulated data, able to predict MR parametric maps that are very consistent with those generated using the gold standard dictionary approach was effectively set-up and validated.

Part II: Evaluating the Trabecular Bone Volume Fraction through NMR relaxometry using Single-Sided Devices

The second part of this thesis aimed to propose a new quantitative NMR methodology to estimate the BV/TV ratio, also called bone volume fraction, using portable single-sided devices.

Chapter 4

First, background and motivation for the work were provided. The inability of DEXA of providing sufficient insights about fracture risk in the population at risk of Osteoporosis is presented as the main motivation for the need of clinically feasible methodologies for assessing bone micro-structure parameters. However, DEXA remains the gold standard diagnostic tool for Osteoporosis because alternative methodologies are either not feasible in vivo, such as micro-CT, or too costly to allow wide screening campaigns, such as high-resolution MRI. Portable NMR, based on low-field single-sided NMR scanners, is then proposed as a promising and appealing approach to assess bone morphological parameters combining the advantage of NMR methods and the cost related advantages coming with portable NMR devices.

The methodology recently proposed in [23] to assess the BV/TV, an important parameter to evaluate the micro-structure of the trabecular bone, through single-sided NMR scanners was then presented. The procedure is based on computing the ratio between the CPMG signals obtained from the trabecular bone and the CPMG signal obtained from a reference sample of bulk marrow. Further details of the procedure were given and its limitations were discussed. One limitation relates to the presence of soft tissues other than trabecular bone that can be intercepted by the sensitive volume of the single-sided NMR scanner in a real scenario, which would affect the estimation of the BV/TV parameter. The need of using a reference sample to compute the BV/TV ratio is also spotlighted as the main limitation of the current procedure, lowering its versatility and in vivo feasibility.

The following chapters aimed to overcome the two limitations here highlighted. Finally, the characteristic of the single-sided NMR devices used in the experiments were presented, which are the NMR-MOUSE PM10 and the NMR-MOLE.

Chapter 5

Based on the procedure presented in [23], a modification is proposed to allow selective detection of the signal from the marrow in the trabecular bone suppressing signals from muscle and cartilage tissue present within the sensitive volume of the single-sided scanner. The improvement was achieved weighting the signal acquisition by molecular diffusion. Experiments based on an *ad-hoc* designed Diffusion Weighted $T_1 - T_2$ pulse sequence, demonstrated that muscle and cartilage signals can be suppressed by using diffusion weighting, because marrow has a much lower diffusion coefficient than the diffusion coefficient of cartilage and muscle. On the basis of those experiments, a 1D measurement was established using a diffusion prepared CPMG acquisition that allowed correct estimation of trabecular bone volume fraction in a sample where also muscle tissue was present within the sensitive volume of the single-sided scanner.

Chapter 6

The main goal of the study presented in this chapter was to effectively evaluate the BV/TV ratio of trabecular bone without the need of using a reference sample, while keeping all the key features of the technique presented in [23], which are: to use a CPMG pulse sequence to acquire the data, and to use a single-sided NMR scanner. The key idea to reach the goal was to exploit the information about the transverse relaxation times T_2 of the trabecular bone sample coming from the CPMG measurements to separate the intra trabecular signal (mainly due to water bound to collagen) and the signal coming from the inter trabecular

spaces (mainly due to bone marrow). The intra and inter trabecular signals are proportional to the volumes occupied by the ^{1}H nuclei inside the trabeculae and the inter trabecular spaces, respectively. A set of trabecular bone samples, cored from pig shoulder, were analysed with NMR and with the gold standard micro-CT to assess ground truth parameters. The T_2 short intensity fraction was directly evaluated, whereas BV/TV and BS/TV ratios were estimated using theoretical models, which require an esteem of the intra trabecular porosity $\phi_{\text{trabecula}}$ and of the mean trabecular radius (trabeculae are modelled as cylinders). Comparison between NMR and micro-CT showed high level of agreement for BV/TV and high correlation with moderate agreement with BS/TV. The intra trabecular porosity measured in the work presented in this chapter was significantly consistent with values found by Fantazzini et al. [40], and this makes us confident in using an intra trabecular porosity of around 30% as esteem for BV/TV estimation even for new trabecular samples. However, this confidence holds for healthy trabecular bones. Literature has not clearly established at what extent diseases affecting the bone tissue, such as osteoporosis, modifies the structure at the level of the single trabecula, so that parameters like the porosity of the trabecula could vary. Future studies can focus on assessing if bone diseases affect the intra trabecular porosity.

Conclusions

In conclusion, although the works reported in this Ph.D thesis requires further steps to be used in clinical applications, they may be considered as deliverables in the direction forward the advance of the role of quantitative NMR in medicine.

Bibliography

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Tonima S. Ali, Monique C. Tourell, Honor J. Hugo, Chris Pyke, Samuel Yang, Thomas Lloyd, Erik W. Thompson, and Konstantin I. Momot. Transverse relaxation-based assessment of mammographic density and breast tissue composition by single-sided portable NMR. *Magn. Reson. Med.*, 82(3):1199–1213, 2019.

[3] John Ashburner and Karl J. Friston. Voxel-based morphometry–the methods. *NeuroImage*, 11(6):805 – 821, 2000.

[4] Jakob Assländer, Martijn A. Cloos, Florian Knoll, Daniel K. Sodickson, Jürgen Hennig, and Riccardo Lattanzi. Low rank alternating direction method of multipliers reconstruction for mr fingerprinting. *Magnetic Resonance in Medicine*, 79(1):83–96, 2018.

[5] Louis V Avioli. Significance of osteoporosis: A growing international health care problem. *Calcif. Tissue Int.*, 49(1):S5–S7, jan 1991.

[6] F. Balsiger, Konar A. Shridhar, S. Chikop, V. Chandran, O. Scheidegger, S. Geethanath, and M. Reyes. Magnetic Resonance Fingerprinting Reconstruction via Spatiotemporal Convolutional Neural Networks. *ArXiv e-prints*, July 2018.

[7] M. Barbieri, L. Brizi, V. Bortolotti, P. Fantazzini, M. Nogueira d'Eurydice, S. Obruchkov, H. Liu, and P. Galvosas. Single-sided NMR for the diagnosis of os-

teoporosis: Diffusion weighted pulse sequences for the estimation of trabecular bone volume fraction in the presence of muscle tissue. *Microporous Mesoporous Mater.*, 269:166–170, oct 2018.

[8] Marco Barbieri, Leonardo Brizi, Enrico Giampieri, Francesco Solera, Gastone Castellani, Claudia Testa, and Daniel Remondini. Circumventing the curse of dimensionality in magnetic resonance fingerprinting through a deep learning approach. *ArXiv e-prints*, abs/1811.11477v1, 2018.

[9] Marco Barbieri, Philip K. Lee, Leonardo Brizi, Enrico Giampieri, Alexander R. Toews, Gastone Castellani, Daniel Remondini, Brian A. Hargreaves, and Claudia Testa. *"Deep Learning for Magnetic Resonance Fingerprinting: Data Augmentation with Phase Encoding and SVD Preprocessing for Accurate Parameter Reconstruction of FISP Data."*, 2019. Proc. Intl. Soc. Mag. Reson. Med. 27 (2019).

[10] R. Bellman. *Adaptive control processes: a guided tour*. Princeton University Press, 1961.

[11] Dan Benjamini and Peter J. Basser. Use of marginal distributions constrained optimization (madco) for accelerated 2d mri relaxometry and diffusometry. *J. Magn. Reson.*, 271:40 – 45, 2016.

[12] Elad Bergman, Yifat Sarda, Noa Ritz, Edmond Sabo, Gil Navon, Reuven Bergman, and Uri Nevo. In vivo assessment of aged human skin with a unilateral NMR scanner. *NMR Biomed.*, 28(6):656–666, 2015.

[13] Bhairav Bipin Mehta, Simone Coppo, Debra Frances McGivney, Jesse Ian Hamilton, Yong Chen, Yun Jiang, Dan Ma, Nicole Seiberlich, Vikas Gulani, and Mark Alan Griswold. Magnetic resonance fingerprinting: a technical review. *Magnetic Resonance in Medicine*, 81(1):25–46, 2019.

[14] Glen M Blake and Ignac Fogelman. The role of dxa bone density scans in the diagnosis and treatment of osteoporosis. *Postgrad. Med. J.*, 83(982):509–517, 2007.

[15] B Blümich, P Blümler, G Eidmann, A Guthausen, R Haken, U Schmitz, K Saito, and G Zimmer. The NMR-mouse: construction, excitation, and applications. *Magn. Reson. Imaging*, 16(5-6):479–484, jun 1998.

[16] Zhao Bo, Setsompop Kawin, Adalsteinsson Elfar, Gagoski Borjan, Ye Huihui, Ma Dan, Y. Jiang, E. P. Grant, M. A. Griswold, and W. L. Lawrence. Improved magnetic

resonance fingerprinting reconstruction with low-rank and subspace modeling. *Magn. Reson. Med.*, 79(2):933–942, April 2017.

[17] G. C. Borgia, R. J. S. Brown, and P. Fantazzini. Uniform-penalty inversion of multiexponential decay data. *J Magn Reson*, (132):65– 77, 1998.

[18] G.C. Borgia, R.J.S. Brown, and P. Fantazzini. Uniform-penalty inversion of multiexponential decay data. *Journal of Magnetic Resonance*, 132(1):65 – 77, 1998.

[19] G.C Borgia, R.J.S Brown, and P Fantazzini. Uniform-penalty inversion of multiexponential decay data: Ii data spacing, t2 data, systematic data errors, and diagnostics. *Journal of Magnetic Resonance*, 147(2):273 – 285, 2000.

[20] V. Bortolotti, L. Brizi, P. Fantazzini, G. Landi, and F. Zama. Filtering techniques for efficient inversion of two-dimensional nuclear magnetic resonance data. *Journal of Physics: Conference Series*, 904(1):012005, 2017.

[21] V. Bortolotti, R.J.S. Brown, and P. Fantazzini. *"A software for inversion of multiexponential decay data for Windows system"*, 2012. Available at `http://software.dicam.unibo.it/upenwin`.

[22] M. Boudreau, C. L. Tardif, N. Stikov, J. G. Sled, W. Lee, and G. B. Pike. B1 mapping for bias-correction in quantitative t1 imaging of the brain at 3t using standard pulse sequences. *J. Magn. Reson. Imaging*, 46(6):1673–1682, 2017.

[23] Leonardo Brizi, Marco Barbieri, Fabio Baruffaldi, Villiam Bortolotti, Chiara Fersini, Huabing Liu, Marcel Nogueira d'Eurydice, Sergei Obruchkov, Fangrong Zong, Petrik Galvosas, and Paola Fantazzini. Bone volume–to–total volume ratio measured in trabecular bone by single-sided NMR devices. *Magn. Reson. Med.*, 79(1):501–510, 2018.

[24] Leonardo Brizi, Marco Barbieri, Claudia Testa, and Paola Fantazzini. *1H nuclei compartmentalization, exchange and self-diffusion in cortical bone by one- and two-dimensions NMR in homogeneous and inhomogeneous fields*, 2019. Proc. Intl. Soc. Mag. Reson. Med. 27.

[25] Guido Buonincontri and Stephen J Sawiak. Mr fingerprinting with simultaneous b1 estimation. *Magn. Reson. Med.*, 76(4):1127–1135, September 2015.

[26] Jean-François Cabana, Ye Gu, Mathieu Boudreau, Ives R Levesque, Yaaseen Atchia, John G. Sled, Sridar Narayanan, Douglas L. Arnold, G. Bruce Pike, Julien Cohen-Adad, Tanguy Duval, Manh-Tung Vuong, and Nikola Stikov. Quantitative magnetization

transfer imaging made easy with qmtlab: Software for data simulation, analysis, and visualization. *Concepts in Magnetic Resonance Part A*, 44A(5):263–277, 2015.

[27] Xiaozhi Cao, Congyu Liao, Zhixing Wang, Ying Chen, Huihui Ye, Hongjian He, and Jianhui Zhong. Robust sliding-window reconstruction for accelerating the acquisition of mr fingerprinting. *Magnetic Resonance in Medicine*, 78(4):1579–1588, 2017.

[28] Stephen F. Cauley, Kawin Setsompop, Dan Ma, Yun Jiang, Huihui Ye, Elfar Adalsteinsson, Mark A. Griswold, and Lawrence L. Wald. Fast group matching for mr fingerprinting reconstruction. *Magnetic Resonance in Medicine*, 74(2):523–528, 2015.

[29] Y. Chen, Y. Jiang, S. Pahwa, D. Ma, L. Lu, M. D. Twieg, K. L. Wright, N. Seiberlich, M. A. Griswold, and V. Gulani. Mr fingerprinting for rapid quantitative abdominal imaging. *Radiology*, 279(1):278–286, 2016. PMID: 26794935.

[30] Yong Chen, Ananya Panda, Shivani Pahwa, Jesse I. Hamilton, Sara Dastmalchian, Debra F. McGivney, Dan Ma, Joshua Batesole, Nicole Seiberlich, Mark A. Griswold, Donna Plecha, and Vikas Gulani. Three-dimensional mr fingerprinting for quantitative breast imaging. *Radiology*, 2018. PMID: 30375925.

[31] Christopher C. Cline, Xiao Chen, Boris Mailhe, Qiu Wang, Josef Pfeuffer, Mathias Nittka, Mark A. Griswold, Peter Speier, and Mariappan S. Nadar. Air-mrf: Accelerated iterative reconstruction for magnetic resonance fingerprinting. *Magn. Reson. Imaging*, 41:29–40, 2017.

[32] Martijn A. Cloos, Florian Knoll, Tiejun0 Zhao, Kai T. Block, Mary Bruno, Graham C. Wiggins, and Daniel K. Sodickson. Multiparametric imaging with heterogeneous radiofrequency fields. *Nature Communications*, 7:12445, Aug 2016. Article.

[33] O. Cohen, B. Zhu, and M. S. Rosen. Mr fingerprinting deep reconstruction network (drone). *Magn. Reson. Med*, 80(3):885–894, April 2018.

[34] Ouri Cohen and Matthew S. Rosen. Algorithm comparison for schedule optimization in mr fingerprinting. *Magnetic Resonance Imaging*, 41:15 – 21, 2017.

[35] Raymond Damadian. Tumor detection by nuclear magnetic resonance. *Science*, 171(3976):1151–1153, 1971.

[36] Sean C.L. Deoni, Brian K. Rutt, and Terry M. Peters. Rapid combined t1 and t2 mapping using gradient recalled acquisition in the steady state. *Magnetic Resonance in Medicine*, 49(3):515–526, 2003.

[37] Jiang Du and Graeme M. Bydder. Qualitative and quantitative ultrashort-TE MRI of cortical bone. *NMR Biomed.*, 26(5):489–506, may 2013.

[38] Robert Engle. Garch 101: The use of arch/garch models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168, December 2001.

[39] Z. Fang, Y. Chen, M. Liu, L. Xiang, Q. Zhang, Q. Wang, W. Lin, and D. Shen. Deep learning for fast and spatially-constrained tissue quantification from highly-accelerated data in magnetic resonance fingerprinting. *IEEE Transactions on Medical Imaging*, pages 1–1, 2019.

[40] Paola Fantazzini, Villiam Bortolotti, Robert J.S. Brown, Mara Camaiti, Carla Garavaglia, Rossella Viola, and Gianluca Giavaresi. Two 1 H-nuclear magnetic resonance methods to measure internal porosity of bone trabeculae: By solid-liquid signal separation and by longitudinal relaxation. *J. Appl. Phys.*, 95(1):339–343, jan 2004.

[41] Paola Fantazzini, Robert James Sidford Brown, and Giulio Cesare Borgia. Bone tissue and porous media: Common features and differences studied by NMR relaxation. *Magn. Reson. Imaging*, 21(3-4):227–234, 2003.

[42] M A Fernández-Seara, H K Song, and F W Wehrli. Trabecular Bone Volume Fraction Mapping by Low-Resolution MRI. *Magn. Reson. Med.*, 46(1):103–111, 2001.

[43] J. A. Fessler and B. P. Sutton. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, Feb 2003.

[44] I. Fogelman and Gnanasegran G. *Radionuclide and Hybrid Bone Imaging*, chapter 2, pages 29–57. Springer, 2012.

[45] J. H. Friedman. On bias, variance, 0/1–loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, 1(1):55–77, 1997.

[46] Mohammad Golbabaee, Carolin M. Pirkl, Marion I. Menzel, Guido Buonincontri, and Pedro A. Gomez. *"Deep MR Fingerprinting with total-variation and low-rank subspace priors"*, 2019. Proc. Intl. Soc. Mag. Reson. Med. 27 (2019).

[47] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning, Chapter 5.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[48] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning, Chapter 7.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[49] Mathilde Granke, Alexander J Makowski, Sasidhar Uppuganti, Mark D Does, and Jeffry S Nyman. Identifying Novel Clinical Surrogates to Assess Human Bone Fracture Toughness. *J. Bone Miner. Res.*, 30(7):1290–1300, 2015.

[50] H. Gudbjartsson and S. Patz. The rician distribution of noisy mri data. *Magn. Reson. Med.*, 34(6):910–914, 1995.

[51] M. Hahn, M. Vogel, M. Pompesius-Kempa, and G. Delling. Trabecular bone pattern factor—a new parameter for simple quantification of bone microarchitecture. *Bone*, 13(4):327–330, jul 1992.

[52] Jesse I. Hamilton, Yun Jiang, Yong Chen, Dan Ma, Wei-Ching Lo, Mark Griswold, and Nicole Seiberlich. Mr fingerprinting for rapid quantification of myocardial t1, t2, and proton spin density. *Magnetic Resonance in Medicine*, 77(4):1446–1458, 2017.

[53] Elisabeth Hoppe, Gregor Körzdörfer, Tobias Würfl, Jens Wetzl, Felix Lugauer, Josef Pfeuffer, and Andreas K. Maier. Deep learning for magnetic resonance fingerprinting: A new approach for predicting quantitative parameter values from time series. *Studies in health technology and informatics*, 243:202–206, 2017.

[54] R. Adam Horch, Daniel F. Gochberg, Jeffry S. Nyman, and Mark D. Does. Clinically compatible MRI strategies for discriminating bound and pore water in cortical bone. *Magn. Reson. Med.*, 68(6):1774–1784, dec 2012.

[55] R. Adam Horch, Jeffry S. Nyman, Daniel F. Gochberg, Richard D. Dortch, and Mark D. Does. Characterization of1H NMR signal in human cortical bone for magnetic resonance imaging. *Magn. Reson. Med.*, 64(3):680–687, sep 2010.

[56] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257, 1991.

[57] M.D Hürlimann and L Venkataramanan. Quantitative measurement of two-dimensional distribution functions of diffusion and relaxation in grossly inhomogeneous fields. *Journal of Magnetic Resonance*, 157(1):31 – 42, 2002.

[58] Nachiko Iita, Shinya Handa, Sadanori Tomiha, and Katsumi Kose. Development of a compact MRI system for measuring the trabecular bone microstructure of the finger. *Magn. Reson. Med.*, 57(2):272–277, feb 2007.

[59] Y. Jiang, D. Ma, N. Seiberlich, V. Gulani, and M. A. Griswold. Mr fingerprinting using fast imaging with steady state precession (fisp) with spiral readout. *Magn. Reson. Med.*, 74(6):1621–1631, December 2014.

[60] E. Jones, T. Oliphant, P. Peterson, and et al. 2001. `http://www.scipy.org/`.

[61] H. Jürgen. Echoes—how to generate, recognize, use or avoid them in mr-imaging sequences. part i: Fundamental and not so fundamental properties of spin echoes. *Concepts in Magnetic Resonance*, 3(3):125–143, 1991.

[62] Danielle Kara, Mingdong Fan, Jesse Hamilton, Mark Griswold, Nicole Seiberlich, and Robert Brown. Parameter map error due to normal noise and aliasing artifacts in mr fingerprinting. *Magnetic Resonance in Medicine*, 81(5):3108–3123, 2019.

[63] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, December 2014.

[64] Feliks Kogan, Evan Levine, Akshay S. Chaudhari, Uchechukwuka D. Monu, Kevin Epperson, Edwin H.G. Oei, Garry E. Gold, and Brian A. Hargreaves. Simultaneous bilateral-knee mr imaging. *Magnetic Resonance in Medicine*, 80(2):529–537, 2018.

[65] Katsumi Kose, Yoshimasa Matsuda, Takeaki Kurimoto, Seitarou Hashimoto, Yukako Yamazaki, Tomoyuki Haishi, Shin Utsuzawa, Hiroshi Yoshioka, Shigemasu Okada, Masaaki Aoki, and Tsuyoshi Tsuzaki. Development of a compact MRI system for trabecular bone volume fraction measurements. *Magn. Reson. Med.*, 52(2):440–444, aug 2004.

[66] Roland Krug, Andrew J. Burghardt, Sharmila Majumdar, and Thomas M. Link. High-Resolution Imaging Techniques for the Assessment of Osteoporosis. *Radiol. Clin. North Am.*, 48(3):601–621, may 2010.

[67] Mirko Krüger, Annett Schwarz, and Bernhard Blümich. Investigations of silicone breast implants with the NMR-MOUSE. *Magn. Reson. Imaging*, 25(2):215–218, feb 2007.

[68] Alexey Kulpanovich and Assaf Tal. The application of magnetic resonance fingerprinting to single voxel proton spectroscopy. *NMR in Biomedicine*, 31(11):e4001, 2018. e4001 NBM-18-0030.R2.

[69] Maxime Van Landeghem, Ernesto Danieli, Juan Perlo, Bernhard Blümich, and Federico Casanova. Low-gradient single-sided NMR sensor for one-shot profiling of human skin. *J. Magn. Reson.*, 215:74–84, feb 2012.

[70] Bennett A. Landman, Alan J. Huang, Aliya Gifford, Deepti S. Vikram, Issel Anne L. Lim, Jonathan A.D. Farrell, John A. Bogovic, Jun Hua, Min Chen, Samson Jarso, Seth A. Smith, Suresh Joel, Susumu Mori, James J. Pekar, Peter B. Barker, Jerry L.

Prince, and Peter C.M. van Zijl. Multi-parametric neuroimaging reproducibility: A 3-t resource study. *NeuroImage*, 54(4):2854 – 2866, 2011.

[71] P. C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190 – 191, 1973.

[72] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, May 2015.

[73] B. Lemasson, N. Pannetier, N. Coquery, L. S. B. Boisserand, N. Collomb, N. Schuff, M. Moseley, G. Zaharchuk, E. L. Barbier, and E. L. Christen. Mr vascular fingerprinting in stroke and brain tumors models. *Sci. Rep.*, 6:37071, November 2016.

[74] Lawrence I-Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–268, 1989.

[75] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold. Magnetic resonance fingerprinting. *Nature*, 495:187–192, March 2013.

[76] Dan Ma, Simone Coppo, Yong Chen, Debra F. McGivney, Yun Jiang, Shivani Pahwa, Vikas Gulani, and Mark A. Griswold. Slice profile and b1 corrections in 2d magnetic resonance fingerprinting. *Magnetic Resonance in Medicine*, 78(5):1781–1789, 2017.

[77] Alex MacKay, Cornelia Laule, Irene Vavasour, Thorarin Bjarnason, Shannon Kolind, and Burkhard Madler. Insights into brain microstructure from the t2 distribution. *Magnetic Resonance Imaging*, 24(4):515 – 525, 2006.

[78] S Majumdar, H K Genant, S Grampp, D C Newitt, V.-H Truong, J C Lin, and A Mathur. Correlation of Trabecular Bone Structure with Age, Bone Mineral Density, and Osteoporotic Status: In Vivo Studies in the Distal Radius Using High Resolution Magnetic Resonance Imaging. *J. Bone Miner. Res.*, 12(1):111–118, 1997.

[79] Mary Kate Manhard, R. Adam Horch, Kevin D. Harkins, Daniel F. Gochberg, Jeffry S. Nyman, and Mark D. Does. Validation of quantitative bound- and pore-water imaging in cortical bone. *Magn. Reson. Med.*, 71(6):2166–2171, 2014.

[80] B. Manz, A. Coy, R. Dykstra, C.D. Eccles, M.W. Hunter, B.J. Parkinson, and P.T. Callaghan. A mobile one-sided nmr sensor with a homogeneous magnetic field: The nmr-mole. *Journal of Magnetic Resonance*, 183(1):25 – 31, 2006.

[81] G. Mazor, L. Weizman, A. Tal, and Y. C. Eldar. Low rank magnetic resonance fingerprinting. *ArXiv e-prints*, January 2017.

[82] D. F. McGivney, E. Pierre, D. Ma, Y. Jiang, H. Saybasili, V. Gulani, and M. A. Griswold. Svd compression for magnetic resonance fingerprinting in the time domain. *IEEE Trans. Med. Imaging*, 33(12):2311–2322, Dec. 2014.

[83] Kamal H. Mroue, Yusuke Nishiyama, Manoj Kumar Pandey, Bo Gong, Erin McNerny, David H. Kohn, Michael D. Morris, and Ayyalusamy Ramamoorthy. Proton-detected solid-state NMR spectroscopy of bone with ultrafast magic angle spinning. *Sci. Rep.*, 5, jul 2015.

[84] G. Navon, U. Eliav, D. E. Demco, and B. Blümich. Study of order and dynamic processes in tendon by NMR and MRI. *J. Magn. Reson. Imaging*, 25(2):362–380, 2007.

[85] Qingwen Ni, J. Derwin King, and Xiaodu Wang. The characterization of human compact bone structure changes by low-field nuclear magnetic resonance. *Meas. Sci. Technol.*, 15(1):58–66, 2004.

[86] M.J. Nissi, J. Rieppo, J. Toyras, M.S. Laasanen, I. Kiviranta, J.S. Jurvelin, and M.T. Nieminen. T2 relaxation time mapping reveals age- and species-related diversity of collagen network architecture in articular cartilage. *Osteoarthritis and Cartilage*, 14(12):1265 – 1271, 2006.

[87] Marcel Nogueira d'Eurydice and Petrik Galvosas. Measuring diffusion–relaxation correlation maps using non-uniform field gradients of single-sided NMR devices. *J. Magn. Reson.*, 248:137–145, 2014.

[88] Henry H Ong, Alexander C Wright, and Felix W Wehrli. Deuterium nuclear magnetic resonance unambiguously quantifies pore and collagen-bound water in cortical bone. *J. Bone Miner. Res.*, 27(12):2573–2581, 2012.

[89] Georg Osterhoff, Elise F. Morgan, Sandra J. Shefelbine, Lamya Karim, and Laoise M. McNamara. Bone mechanical properties and changes with osteoporosis. *Injury*, 47:S11–S20, jun 2016.

[90] Diego G. Loyola R, Mattia Pedergnana, and Sebastian Gimeno Garcia. Smart sampling and incremental function learning for very large high dimensional data. *Neural Networks*, 78:75 – 87, 2016. Special Issue on Neural Network Learning in Big Data.

[91] Erik Rössler, Carlos Mattea, Simo Saarakkala, Petri Lehenkari, Mikko Finnilä, Lassi Rieppo, Sakari Karhula, Miika T. Nieminen, and Siegfried Stapf. Correlations of low-field NMR and variable-field NMR parameters with osteoarthritis in human articular cartilage under load. *NMR Biomed.*, 30(8):1–14, 2017.

[92] Erik Rössler, Carlos Mattea, and Siegfried Stapf. Feasibility of high-resolution one-dimensional relaxation imaging at low magnetic field using a single-sided NMR scanner applied to articular cartilage. *J. Magn. Reson.*, 251:43–51, feb 2015.

[93] Ego Seeman. Bone quality: The material and structural basis of bone strength. *J. Bone Miner. Metab.*, 26(1):1–8, 2008.

[94] Alan C. Seifert, Suzanne L. Wehrli, and Felix W. Wehrli. Bi-component T<inf>2</inf>* analysis of bound and pore bone water fractions fails at high field strengths. *NMR Biomed.*, 28(7):861–872, 2015.

[95] Eric E. Sigmund, H. Cho, and Y. Q. Song. High-resolution MRI of internal field diffusion-weighting in trabecular bone. *NMR Biomed.*, 22(4):436–448, 2009.

[96] J. R. Singer. Blood flow rates by nuclear magnetic resonance measurements. *Science*, 130(3389):1652–1653, 1959.

[97] K. Sommer, T. Amthor, M. Doneva, P. Koken, J. Meineke, and P. BÃ¶rnert. Towards predicting the encoding capability of mr fingerprinting sequences. *Magnetic Resonance Imaging*, 41:7 – 14, 2017.

[98] Sara M. Sprinkhuizen, Jerome L. Ackerman, and Yi Qiao Song. Influence of bone marrow composition on measurements of trabecular microstructure using decay due to diffusion in the internal field MRI: Simulations and clinical studies. *Magn. Reson. Med.*, 72(6):1499–1508, dec 2014.

[99] Christiaan C. Stolk and Alessandro Sbrizzi. *"Understanding the combined effect of k-space undersampling and transient states excitation in MR Fingerprinting reconstructions: the role of incoherence"*, 2019. Proc. Intl. Soc. Mag. Reson. Med. 27 (2019).

[100] Fei Tong and Xila Liu. Samples selection for artificial neural network training in preliminary structural design. *Tsinghua Science & Technology*, 10(2):233 – 239, 2005.

[101] Md. Nasir Uddin, Teresa D. Figley, Kevin G. Solar, Anwar S. Shatil, and Chase R. Figley. Comparisons between multi-component myelin water fraction, t1w/t2w ratio, and diffusion tensor imaging measures in healthy human brain structures. *Scientific Reports*, 9(1), 2019.

[102] G. van Rossum. 1995. `http://www.python.org/`.

[103] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA, 2008. ACM.

[104] P. Virtue, S. X. Yu, and M. Lustig. Better than real: Complex-valued neural nets for mri fingerprinting. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3953–3957, Sep. 2017.

[105] Stacey A Wainwright, Lynn M Marshall, Kristine E Ensrud, Jane A Cauley, Dennis M Black, Teresa A Hillier, Marc C Hochberg, Molly T Vogt, and Eric S Orwoll. Hip Fracture in Women without Osteoporosis. *J. Clin. Endocrinol. Metab.*, 90(5):2787–2793, oct 2005.

[106] J. Wang, W. Mao, M. Qiu, M. B. Smith, and R. T. Constable. Factors influencing flip angle mapping in mri: Rf pulse shape, slice-select gradients, off-resonance excitation, and b0 inhomogeneities. *Magn. Reson. Med.*, 56(2):463–468, 2006.

[107] Felix W. Wehrli. Magnetic resonance of calcified tissues. *J. Magn. Reson.*, 229:35–48, 2013.

[108] Felix W. Wehrli, Hee Kwon Song, Punam K. Saha, and Alexander C. Wright. Quantitative MRI for the assessment of bone structure and function. *NMR Biomed.*, 19(7):731–764, nov 2006.

[109] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel. An optimal radial profile order based on the golden ratio for time-resolved mri. *IEEE Transactions on Medical Imaging*, 26(1):68–76, Jan 2007.

[110] M. Yang, D. Ma, Y. Jiang, J. Hamilton, N. Seiberlich, M. A.Griswold, and D. McGivney. Low rank approximation methods for mr fingerprinting with large scale dictionaries. *Magn. Reson. Med*, 79(4):2392–2400, August 2017.

[111] Roger MD Zebaze, Ali Ghasem-Zadeh, Ann Bohte, Sandra Iuliano-Burns, Michiko Mirams, Roger Ian Price, Eleanor J Mackie, and Ego Seeman. Intracortical remodelling and porosity in the distal radius and post-mortem femurs of women: a cross-sectional study. *Lancet*, 375(9727):1729–1736, may 2010.

[112] B. Zhao, K. Setsompop, H. Ye, S. F. Cauley, and L. L. Wald. Maximum likelihood reconstruction for magnetic resonance fingerprinting. *IEEE Trans. Med. Imaging*, 35(8):1812–1823, Aug. 2016.