Multi-component MRI transverse-relaxation parameter estimation to detect and monitor neuromuscular disease

Nikolaos Zafeiropoulos

UCL, Queen Square Institute of Neurology

Thesis submitted for the degree of Doctor of Philosophy

Declaration

I, Nikolaos Zafeiropoulos confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

We aimed to optimise the estimation of skeletal muscle-water spin-spin relaxation time (T_{2m}), and fat fraction estimated from multi-echo MRI, as potential biomarkers, by accounting for instrumental factors such as B_1 errors, non-Gaussian noise and non-ideal echo train evolution.

A multi-component slice-profile-compensated extended phase graph (sEPG) model for multi-echo Carr-Purcell-Meiboom-Gill (CPMG) spin-echo sequence signals was implemented, modelling the fat signal as two empirically calibrated sEPG components with fixed parameters, and the remaining unknown parameters (B_1 field factor, T_{2m}, fat fraction (ff_a), global amplitude and Rician noise SD) determined by maximum likelihood estimation. After validation using a calibrated test object the algorithm was used to analyse clinical muscle study data from patient groups with amyotrophic lateral sclerosis (ALS), Kennedy's disease (KD) and Duchenne muscular dystrophy (DMD) and matched healthy controls. Parameter maps were generated using quality control steps to reject pixels failing fit quality or physical meaningfulness criteria. Muscle fat-fraction was also determined independently by 3-point Dixon MRI (ff_d).

In ALS and KD median T_{2m} were significantly elevated compared with healthy controls in varied patterns and time courses, whereas it was decreased in DMD; other T_{2m} distribution histogram metrics such as the skewness and full width at quarter maximum also differed significantly between patients and healthy volunteers. Quantitative comparison of ff_a and ff_d in the same muscles revealed a monotonic

relationship deviating from linearity due to differing deviations from the assumed ideal signal behaviour in each method. Finally, the effects upon estimation accuracy and precision of practically realisable pulse sequence parameter choices were explored in simulations and with real data. Recommendations are presented for optimal choices.

Clinically practical conventional CPMG sequences, combined with an appropriate signal model and parameter estimation method can provide robust T_{2m} and ff_a measures which change in disease and may sensitively reflect different aspects of neuromuscular pathology.

Impact Statement

The overall aim of this thesis was to optimise sensitive and objective biomarkers to detect and monitor the changes occurring on skeletal muscle due to disease. This facilitates drug development and can lead to availability of new treatments for previously untreatable conditions. The proposed method is already in use in three international natural history studies, validating MRI clinical outcome measures.

With regards to clinical trials, the potential application to individual examinations may also improve diagnosis and treatment monitoring using clinical MRI. Due to the wide use of T_2 relaxometry in hospital healthcare (e.g. NHS) these advances can also benefit other workers and researchers, as they can potentially be implemented at any centre with a standard modern scanner.

There is also direct impact to the academic community by jointly creating and exchanging expertise in the study of neuromuscular diseases – the main scope of this work – and further applications to liver and brain imaging and potentially also other fields.

The economic impact due to the associated financial implications can also be substantial. First, by faster and more efficient drug development. Second, by reducing scan duration and cohort sizes. Third by reducing the need for using additional or alternative – MRI or other – diagnostic methods and modalities. An additional benefit is to avoid lengthy, invasive (e.g. biopsies) or unreliable (e.g. functional tests) methods and the associated patient discomfort or difficulty to comply.

Such benefits can have immediate effect in the translational research or individual examinations taking place in our department and collaborative institutions as university hospitals etc. This can be extended to a national or international level in accordance with the legislation in place. Any acquired expertise can be also brought about through disseminating outputs (e.g. scholarly journals), conferences and academic or entrepreneurial associations.

Acknowledgements

This work was funded by an EPSRC CASE studentship award with industrial partner GSK (GlaxoSmithKline, London, United Kingdom). The analysis was conducted at the image analysis centre of the Neuroradiological Academic Unit (NAU), UCL Queen Square Institute of Neurology, University College London. The patient studies described were funded separately by the UCLH NIHR BRC (Biomedical Research Centre) (ALS and KD) and GSK (DMD).

Affiliations:

MRC Centre for Neuromuscular Diseases, London, United Kingdom

Dubowitz Neuromuscular Centre, UCL Institute of Child Health, London, United Kingdom

In the generation of this work I owe sincere gratitude to many people for their substantial and decisive contributions:

Primary supervisor: Dr John S. Thornton

Subsidiary supervisors: Dr Christopher D. J. Sinclair, Dr Robert L. Janiczek (GSK), Dr Enrico De Vita

Professor Tarek A. Yousry and Professor Mary M. Reilly. All colleagues from the Clinical MRI Physics group, Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery. Everyone else at the 3rd floor (image analysis centre) at 8 – 11 Queen Square.

Table of contents

List of Figures

List of Tables

Presentations

Posters

"Determination of tissue-water T_2 of fat infiltrated upper limb skeletal muscle with MRI in Duchenne muscular dystrophy", Poster, Ninth UK Neuromuscular Translational Research Conference, Oxford, 22-23 March 2016.

"A simplified method to determine tissue-water T_2 from CPMG image data in fat infiltrated skeletal muscle: application in the forearm in Duchenne muscular dystrophy", ISMRM annual meeting, Singapore, 7-13 May 2016.

"Slice profile-corrected maximum likelihood estimation of muscle water T_2", Eleventh UK Neuromuscular Translational Research Conference, Cambridge, 19-20 April 2018.

"Improved muscle T_2 estimation by maximum-likelihood parameter estimation using an extended-phase-graph signal model with locally estimated Rician noise levels", ISMRM annual meeting, Paris, 16-21 June 2018.

"Determination of tissue-water T_2 of fat infiltrated upper and lower limb skeletal muscle with MRI in amyotrophic lateral sclerosis, Kennedy's disease and Duchenne muscular dystrophy", Twelfth UK Neuromuscular Translational Research Conference, Newcastle, 4-5 April 2019.

"Extended phase graph model based tissue-water T_2 estimation from CPMG image data in fat-infiltrated skeletal muscle: application in amyotrophic lateral sclerosis and Kennedy's disease", ISMRM annual meeting, Montreal, 11-16 May 2019.

"Addressing sources of error in estimating CPMG parameters in muscle MRI", British Chapter of the International Society for Magnetic Resonance in Medicine, Sheffield, 16-18 September 2019.

"Tissue-water CPMG T_2 and fat fraction mapping of upper and lower limb skeletal muscle in amyotrophic lateral sclerosis, Kennedy's disease and Duchenne muscular dystrophy", Imaging in Neuromuscular Disease Conference, Berlin, 17-19 November 2019.

"Stirring STIR contrast", ISMRM annual meeting, Virtual Conference, 8-14 August 2020.

"Reducing parameter estimate bias due to B_1 field error ambiguity in CPMG MRI T_2 relaxometry", ISMRM annual meeting, Virtual Conference, 8-14 August 2020.

Pitch presentations

"Skeletal muscle T_2 relaxometry in the forearm at 3T: feasibility and normative data in healthy teenage boys", Poster, British Chapter of the International Society for Magnetic Resonance in Medicine (ISMRM) annual meeting, London, 10-11 September 2015.

"Two-component muscle T_2 quantification by maximum-likelihood estimation using an extended-phase-graph signal model with locally evaluated Rician noise", British Chapter of the International Society for Magnetic Resonance in Medicine, Oxford, September 2018.

Oral presentations

"Improved multi-component muscle T_2 mapping using maximum likelihood estimation", Imaging in Neuromuscular Disease Conference, Berlin, 19-21 November 2017.

Chapter 1. Introduction

Many medical imaging methods such as computerised tomography (CT) and ultrasound (US) imaging rely conceptually upon a single type of image contrast. For CT this is the degree of absorption of x-rays as they interact with the electron cloud surrounding atoms at the molecular level, and for US the degree of reflection of sound waves at structural interfaces within the human body (Zaidman and Van Alfen, 2016). Magnetic resonance imaging (MRI) on the other hand, takes advantage of multiple aspects of the magnetization properties of atomic nuclei in tissue, offering multi-modal contrasts probing different physical properties at a deep structural level (McRobbie et al., 2006). In conventional MRI the properties proton density, T_1 relaxation time and T_2 relaxation time provide the main MRI contrasts. These offer versatile ways to distinguish different tissue types, and are sensitive to pathology (Wattjes et al., 2010).

Inherited neuromuscular disorders represent a wide spectrum of diseases, including genetically distinct forms of myopathies, neuropathies, and mitochondrial disorders that have been increasingly identified in recent decades. Their prevalence in the UK is approximately 9.2 cases per million children under 18 years of age. Cases in the general population are thought to be between 1 in 3500 and 1 in 3000 (Emery, 1991). These diseases all share the clinical feature of fatigable weakness, however age at onset, concomitant symptoms, weakness patterns, long-term progression and treatment responses differ depending on the molecular mechanism resulting from the genetic defect (Cruz et al., 2014). In studies of neuromuscular disease there are needs to better characterise subtle or subclinical changes (e.g. relating to fat

infiltration or oedema which may precede symptomatic change), achieve early detection, and sensitively monitor their progression, including response to therapies (Lovitt et al., 2006); this is the overall translational purpose of the work described in this thesis. The need for sensitive and responsive imaging biomarkers for neuromuscular disorders is motivated by factors including their rarity, complex and varying time courses, various underlying processes affecting imaging observations, and current lack of effective treatments (Klickovic et al., 2019). Specific diseases investigated in this thesis are amyotrophic lateral sclerosis and Kennedy's disease (both neuropathies with consequent muscle involvement), and Duchenne muscular dystrophy (a congenital muscular dystrophy). There is growing evidence that MRI may provide both quantitative disease biomarkers for trials, and that in clinical practice muscle MRI may help in selecting appropriate genetic and biochemical diagnostic investigations, as well as suggesting muscle biopsy targets for pathological studies (Mercuri et al., 2007).

Based on the main MRI contrast mechanisms, MRI parameter maps can be constructed. For example, as will be described in detail later, by recording the T_2-weighted image intensity with increasing echo time at each pixel and fitting a suitable signal model, T_2 maps can be generated and summary histogram metrics then extracted. It has been suggested that T_2 mapping may be used as a non-invasive and sensitive biomarker for the quantification of early and subtle muscle changes caused by disease, or changes due to therapeutic interventions in neuromuscular diseases (Arpan et al., 2013). Such quantitative MRI methods offer the ability to assess large and complete anatomical structures – for instance entire limb cross-sections and

their individual muscles – in contrast to established methods like biopsy which is invasive and restricted to small tissue sample volumes. Additionally, as will be explored in this work, T_2-based mapping can potentially also provide fat quantification indices as well as information regarding instrumental imperfections (Azzabou et al., 2015). More generally, as a potential source of biomarkers, MRI does not involve ionizing radiation in contrast with CT, is not complicated by participant or observer performance variation, as is the case with clinical functional assessment, and data are stored and analysed retrospectively. The latter allows for greater flexibility in the analysis as well as the option of updating the analysis according to advancing method developments.

The above observations support the use of MRI, which in the case of parameter mapping is commonly referred to as quantitative MRI, to provide outcome measures for neuromuscular disease treatment trials (Huang et al., 1994). However reliability must be established in terms of the precision, accuracy and reproducibility of measurement results (Kirshner and Guyatt, 1985). Precision refers to the variability observed in an outcome measure after replicate measurements from the same (repeatability precision) or similar (reproducibility precision) specified experimental conditions (Sullivan et al., 2015). If this variability is larger than the true disease-resultant differences between subjects or groups of subjects, the measurement method cannot be considered reliable. Typically lack of precision impacts the power of comparisons using statistical methods such as analyses of variance (ANOVA) or student t-tests etc. Accuracy refers to the ability of a method to produce measured values comparable to those of a ‘gold standard’, i.e. our best estimate of the

theoretical ground truth. Put differently, accuracy assesses the extent to which the method in question is biased, i.e. introduces a difference between the measured and reference – i.e. true, or best available approximation to true – values. In MRI this is typically investigated using test objects with known values (i.e. phantoms) and by comparing *in vivo* parameter estimates with those obtained with alternative methods such as biopsy or MR spectroscopy.

The aims of this thesis are threefold. First, to optimise methods for transverse relaxometry parameter estimation based on multi-echo MRI in order to extract potential biomarkers from clinically realisable examinations. Second, to test the performance of the proposed methods both in terms of test object data and data from *in vivo* patient cohort studies, comparing results with those obtained with alternative established methods. Third, by empirical investigation of acquisition protocol designs constrained to pragmatically realisable conventional scanner parameter settings, to develop useful recommendations and guidelines for future acquisition optimisation.

The first task involved adopting an improved physics model describing the signal behaviour in multi echo MRI acquisitions, and an underlying statistical approach to parameter estimation appropriate to the experimental conditions encountered in real-world acquisitions. In Chapter 2 the basic theoretical background pertaining to both is given and in Chapter 3 the experimental work that led to optimisation of the method for practical application is described. Also in Chapter 3, validation using a calibrated test object (phantom), part of the second aim, is presented. Chapter 4 continues discussion of the second aim where an application of the proposed method

to analysis of muscle MRI data from a motor neuron disease cohort study is presented, including comparisons with results from alternative MRI methods. Chapter 5 constitutes the final part of the second aim where an application in a more challenging dataset from a natural history study of Duchene muscular dystrophy patients is presented. Again validation is explored via comparison with existing methods. Finally the third aim is developed in Chapter 6 where common multi-echo sequence acquisition design and tailoring strategies are presented and discussed. The thesis ends in Chapter 7 with an overall discussion and conclusions bringing together results from all the previous chapters.

Chapter 2. Theory

2.1 MRI Background

2.1.1 Brief general description of MRI theory

Magnetic resonance relies on the effect that a magnetic field has on the spin of specific nuclei. Spin is a quantum mechanical property of a nucleus which takes discrete values dependent upon the number of protons and neutrons of which it is composed. The hydrogen nucleus spin can take either a positive or an equal negative value and can therefore be considered as a microscopic magnet with a spin vector aligned along or against the direction of an external magnetic field. While there are other nuclei with non-zero spin, the hydrogen nucleus is the most important in medical magnetic resonance as hydrogen is a component of water and lipids which are highly abundant in the human body. In equilibrium a small surplus of spins are aligned with, as opposed to against, the applied magnetic field B_0 which produces an equilibrium total magnetization M_0 parallel to the field (Brown and Semelka, 2005).

By convention the direction of M_0 (and B_0) defines the z-axis of the coordinate system used to describe this phenomenon. If a second rotating magnetic field B_1 perpendicular to z is applied, it will force M_0 to depart from the z direction towards the x-y plane. The frequency at which B_1 rotates is equal or close to the same frequency by which individual spins are caused to rotate by B_0, called the resonant or Larmor frequency for historical reasons, and is different for different nuclear species and magnetic field strengths. The resonance frequency ω_0 for a specific nucleus (e.g. a proton in this case) in a magnetic field (B_0) is given by the simple relationship

$$\omega_0 = \gamma B_0 \qquad \text{Equation 1}$$

where γ is the gyromagnetic ratio that characterises the nucleus in question. Its value for protons is $2.67 \cdot 10^8$ radians$\cdot s^{-1} \cdot T^{-1}$, or equivalently 42.56 MHz$\cdot T^{-1}$. If the duration for which of B_1 is applied is chosen carefully, the net magnetization M_0 can be made to rotate entirely into the *x*-*y* plane. B_1 is produced by transmitting modulated radio frequency (RF) energy to the system, and thus the application of B_1 in this manner is referred to as a 90° RF pulse (Figure 2.1). In quantum mechanical terms the resonance phenomenon couples the RF energy at this specific frequency to the energy difference between the field-aligned and the field-opposed spins, tending to equalize the two populations, while also bringing the spins into phase. Following the 90° RF pulse, the net magnetization of the system now entirely in the x-y plane can be referred to as M_{xy}, and because it is rotating about B_0 it can induce a voltage signal in a suitably orientated RF receiver coil. The detection and quantification of this signal induced by the resonant magnetization is the principle behind MRI (Brown and Semelka, 2005).

The above process of generating transverse magnetization is referred to as excitation. It is followed by relaxation, because spins will gradually depart from the coherent state imposed on them by the 90° RF pulse, starting immediately after it is switched off. This process has two aspects; firstly spins exchange energy with their environment (the so called “lattice”) so that the initial equilibrium of more spins aligned with B_0 is gradually restored, with M_z re-approaching M_0. Secondly spins also exchange energy amongst themselves so that their phase coherence in the *x*-*y* plane also reduces with time. The first mechanism affects the return of the net

magnetization to the z-axis (M_0), while both mechanisms affect the gradual reduction of M_{xy}, which therefore decays faster than M_0 recovers. These are exponential processes described quantitatively by the time constants T_1 and T_2: M_0 recovery is governed by an exponential regrowth with time constant T_1 while M_{xy} decreases by exponential decay with time constant T_2 (Figure 2.2) (McRobbie et al., 2006).

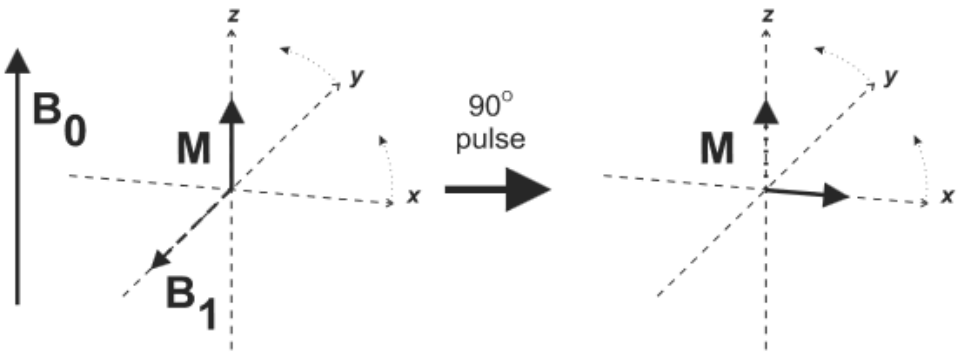


Figure 2.1: An RF pulse centred at the resonant frequency can be seen in the rotating reference frame as an additional magnetic field B_1 oriented perpendicular to B_0. When energy is applied at the appropriate frequency, the protons absorb it and M rotates into the transverse plane. M rotates into a direction perpendicular to both B_0 and B_1 (Brown and Semelka, 2005).

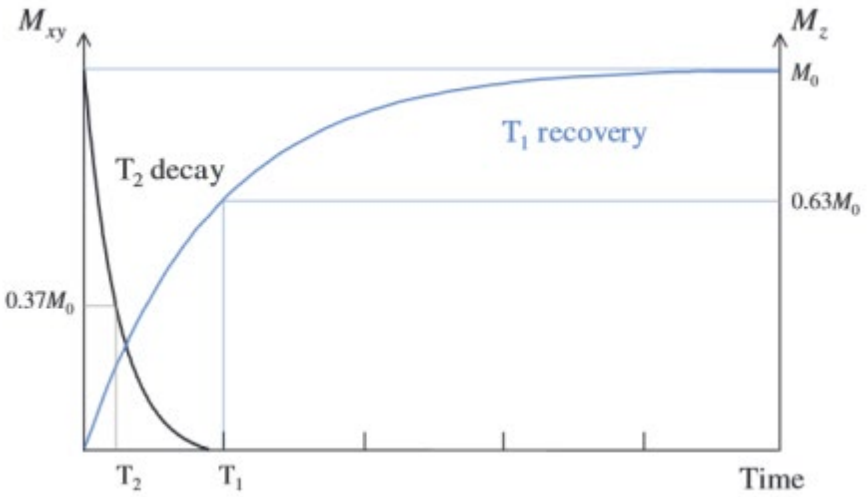


Figure 2.2: T_1 and T_2 relaxation processes take place in parallel, but T_2 is more rapid (McRobbie et al., 2006)

Under practical measurement conditions additional factors affect the decay of M_{xy}, predominantly the inhomogeneity of the main magnetic field B_0. In reality this field will vary slightly at different spatial locations within the magnet, altering the precise frequency at which protons rotate at each location and thus causing them to become out of phase with each other more quickly. This causes the net transverse magnetization to decrease faster than it otherwise would, with its decay now described by a new time constant T_2^* which is shorter than T_2. It is possible to reverse this additional decay with the use of a second B_1 RF pulse with a duration such as to rotate any magnetization about its axis in the rotating frame through 180°. This effectively causes spins which have lost phase coherence from each other to regain it (Figure 2.3). At a time twice as long as the time that the 180° was applied after the initial 90° pulse, the signal returns to a maximum when the transverse net magnetization rephases, and the peak signal at this time will have decayed only by the T_2 time constant. The signal at that point is known as a spin echo (Hahn, 1950) and the time at which it occurs as the echo time (*TE*).

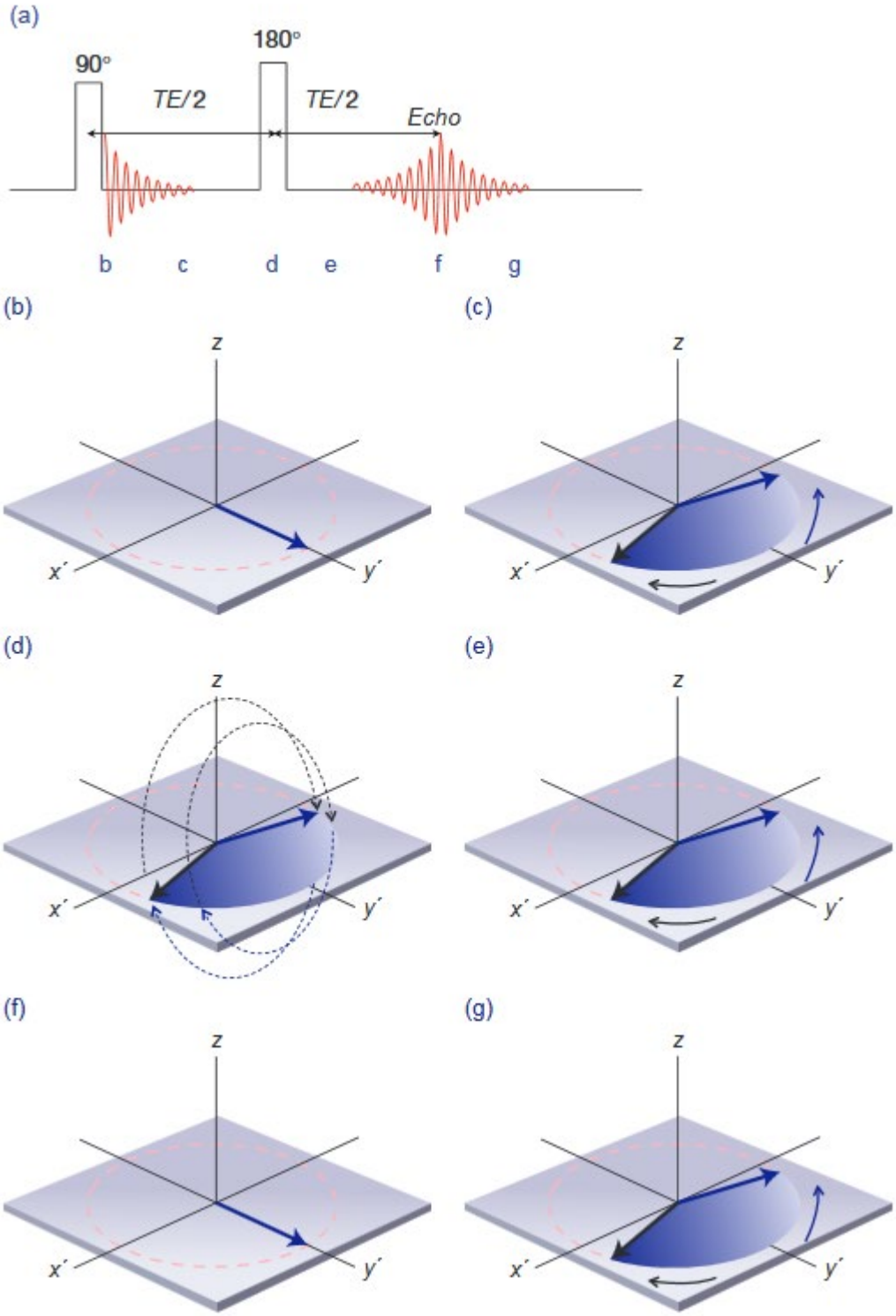


Figure 2.3: (a) Spin-echo pulse sequence. Spins initially in phase (b) dephase naturally (c) until the 180˚ pulse is applied. (d) Immediately after the pulse their phases are reversed, but they continue to dephase in the same direction (e) forming an echo (f) and then dephasing again (g) (McRobbie et al., 2006)

In addition to the externally applied B_0, the local magnetic field experienced by protons is influenced also by the magnetic effect of the electron cloud of their molecular environment. The slightly different magnetic fields experienced by water and lipid protons is an example; molecular water has two hydrogen atoms bonded to one oxygen atom, whereas lipid has many hydrogen atoms bonded to a long-chain carbon framework. As a consequence the electron cloud densities proximal to water and lipid protons differ, meaning that they experience small corresponding differences in their local magnetic fields. This local field difference is known as chemical shielding and is proportional to the main magnetic field strength.

$$B_i = B_0(1 - \sigma_i) \qquad \text{Equation 2}$$

where σ_i is the shielding term for the specific proton. Chemical shielding causes different resonant frequencies for fat and water protons in the same externally applied main magnetic field. Because these frequency differences are very small, and linearly dependent on B_0, they are typically expressed as parts in a million (ppm), denoting the resonant frequency of the proton of interest relative to a reference frequency (Figure 2.4).

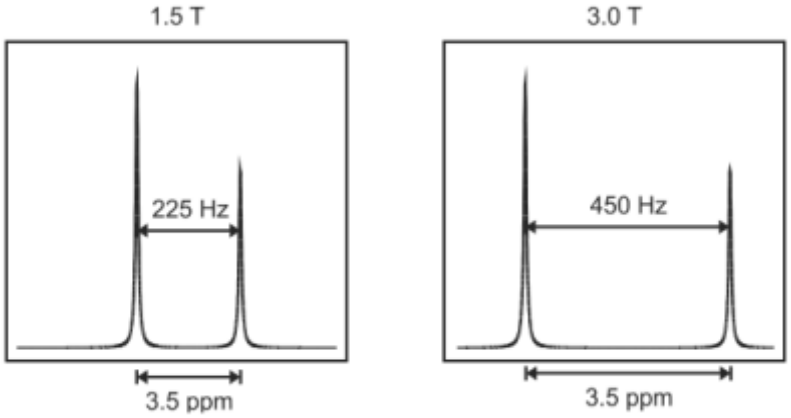


Figure 2.4 Nuclear magnetic resonance spectrum of water and fat at 1.5 T (a) and 3.0 T (b). The resonant frequencies for water and fat are separated by approximately 3.5 ppm, which corresponds to an absolute frequency difference of 220 Hz for a 1.5 T magnetic field (63MHz) or 450 Hz at a magnetic field of 3.0 T (126 MHz) (Brown and Semelka, 2005)

Frequency differences expressed in this form are known as chemical shifts (Brown and Semelka, 2005).

2.1.2 Dixon fat water imaging

The chemical shift between water and lipid resonant frequencies can be used to determine their respective distributions quantitatively. This can be achieved by encoding this frequency difference into gradient echo images with different *TE* shifts. In a time Δt water and fat proton spins in the *x-y* plane will dephase by an angle $\alpha=\gamma \cdot B_0 \cdot \sigma \cdot \Delta t$, where σ is here the fat-water chemical shift difference, so by adjusting Δt by correctly specifying *TE* during the frequency encoding period (section 2.1.5), for a given B_0, we can acquire images with the fat and water protons precessing in phase (0°) or with opposite phases (180°). The image intensity will be based on the individual water and fat signal intensities in phase $S_0=W+F$, and in phase opposition $S_1=W-F$, where *W*, *F* are the complex signals emanating from water and fat protons

respectively. The water and fat signal levels can then be calculated as $W=0.5\cdot(S_0+S_1)$ and $F=0.5\cdot(S_0-S_1)$.

Such methods are called “Dixon” techniques (Dixon, 1984). The above described version is commonly referred to as the “two point Dixon technique”, referencing the two “points” in time (*TE*) at which the signal was sampled. However this is only the simplest implementation illustrating the basic principle. A drawback to this simple implementation is the effect of field inhomogeneity ΔB_0 which may result in phase-wrapping and water–fat “swapping” in the image. Furthermore, fat is known to have a multi-peak spectrum with only the main peak frequency (methylene group) 3.5 ppm away from water.

Various methods have been proposed to correct for the phase wrapping predominantly due to field inhomogeneity (ΔB_0), to achieve ‘phase unwrapping’. This is inherent to all Dixon water–fat separation methods and necessitates complex correction algorithms to avoid the concomitant fat-water swapping. The three point Dixon technique achieves this by encoding the phase differences due to inhomogeneity and other imperfections, in addition to the required chemical shift difference, by recording two gradient-echo images in which the fat signal is in opposite phase to the water – instead of only one that would be sufficient in the absence of ΔB_0. Then with the use of computational phase unwrapping techniques the error phase differences can be determined (Glover and Schneider, 1991) leading to a more accurate estimation of the fat and water signal and their ratio.

Most unwrapping algorithms exploit the fact that B_0 field inhomogeneities vary smoothly across the image and are continuous. In the Dixon method used in the work described here (to generate Dixon fat fraction maps in the results sections) the "Prelude" phase unwrapping tool from FSL software was used (FSL, 2015). Fat fraction (*ff*) is calculated as $ff = F/(W+F)\cdot 100\%$, where W and F are the water and fat signal levels respectively, as previously described, and fat fraction maps are thus constructed. Subsequently regions of interest (ROIs) are commonly manually defined (to avoid operator bias often with the use of an "auxiliary" T_1 weighted, or raw Dixon acquisition image) to extract mean *ff* values for analysis.

This type of quantitative MRI (qMRI) offers an observer independent measurement which is highly repeatable and can be performed at most clinical centres (Burakiewicz et al., 2017). Additional correction for the multiple peaks of the lipid spectrum may assist in obtaining more accurate fat fractions (Wokke et al., 2013).

2.1.3 Factors affecting T_1 and T_2 relaxation times

The process of relaxation relies on the energy transfer from an excited proton (spin) to either another spin (T_2 effects) or to its surrounding environment (T_1 and T_2 effects) (section 2.1.1). This is mediated by fluctuations of the local magnetic field that spins experience, acting as an additional magnetic field B_R effectively perturbing spins from their in-phase (*x*-*y* plane coherence) and higher energy (aligned to the opposite direction of the main B_0 field) state. These fluctuations are in general caused by movement of neighbouring magnetic moment-possessing particles, such as protons in the same molecule or orbital electrons in neighbouring molecules. Since water is

the most common MRI detected molecule, the movement of the dipole consisting of its two protons is one of the main sources of relaxation. Magnetic field fluctuations thus depend on temperature, and on the microscopic environment that affects the motion of the particles in question, and as a result comprise of a wide range of frequency components. For example unconstrained ('free') water molecules will generate high frequency local field fluctuations while strongly bound protons in large molecules will cause low frequency fluctuations (Foster and Hutchison, 1987).

The effects of the aforementioned local B_R field fluctuations can be better understood by decomposing them into transverse and longitudinal components. The transverse B_R field components have two effects. First, in a way similar to the action of B_1 they affect the energy state of spins causing transitions (flipping) from the antiparallel to the parallel state (towards M_0), leading to T_1 relaxation. Second, this alteration in the lifetime of the spin states causes ambiguity in their energy difference ΔE, which according to quantum mechanical physics (Heisenberg uncertainty principle, $\hbar \approx \Delta E \cdot \Delta t$) introduces also uncertainty Δt in the duration that spins remain at each state; this relates to the width of the spectral distribution (Lorentzian) expressing the transverse magnetization decay in the frequency domain and thus T_2 relaxation (Figure 2.5).

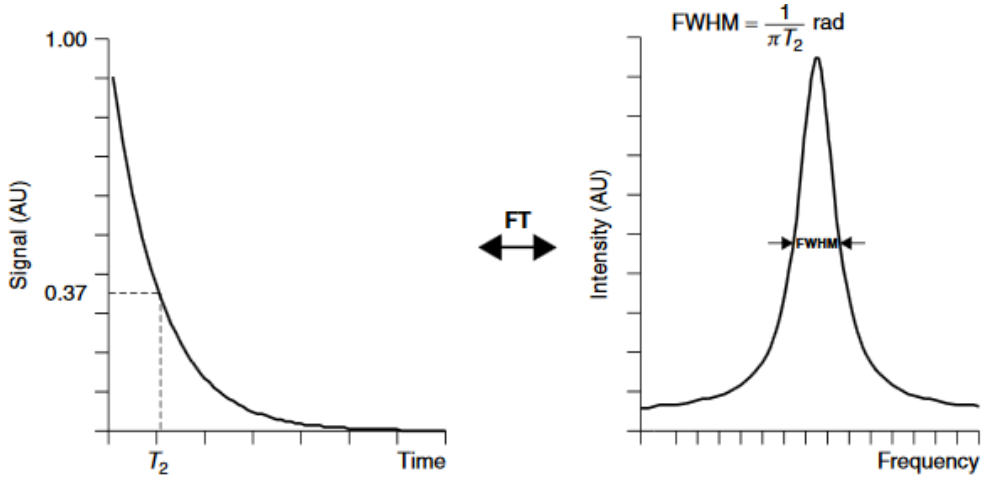


Figure 2.5: The exponential T_2 decay corresponds to a Lorentzian distribution in the frequency domain (Fourier transform), with the width of the distribution inversely proportional to the T_2 decay constant of the time domain (Tofts, 2004)

The longitudinal B_R field components introduce alterations to the main B_0 field by adding or subtracting to it, causing temporally random departures from the nominal (Larmor) frequency locally, which gradually brings spins out of phase coherence thus contributing to T_2 relaxation (Tofts, 2004). Since longitudinal B_R field components are related to the (longitudinally) static B_0 field they appear (average) as fluctuations close to zero frequencies (ω=0), while transverse B_R field components are added to and therefore centred around the Larmor (B_0-dependent) frequency ω_0. The first play a key role in T_2 relaxation, which due to its influence by both types of B_R field components is also affected by the latter, while T_1 relaxation depends only on the latter (local magnetic field fluctuations around the ω_0 frequency) (Foster and Hutchison, 1987).

To describe the above processes in the context of the biophysical molecular environments present in typical MR experiments, the concepts of spectral density function $J(\omega)$ and correlation time τ_c have been widely used (McRobbie et al., 2006).

The first describes the distribution of the frequencies influencing relaxation of a particular collection of molecules. The second is a measure of the time it takes for interactions to take place within such a collection and therefore describes the mobility and motion of its contents. In that light T_1 and T_2 relaxation behaviour can be discussed in terms of representative biophysical environment scenarios. A common approach considers water molecules – as the main source of MRI signal – to exist in a range of 'binding' conditions, as presented in Figure 2.6 (Foster and Hutchison, 1987).

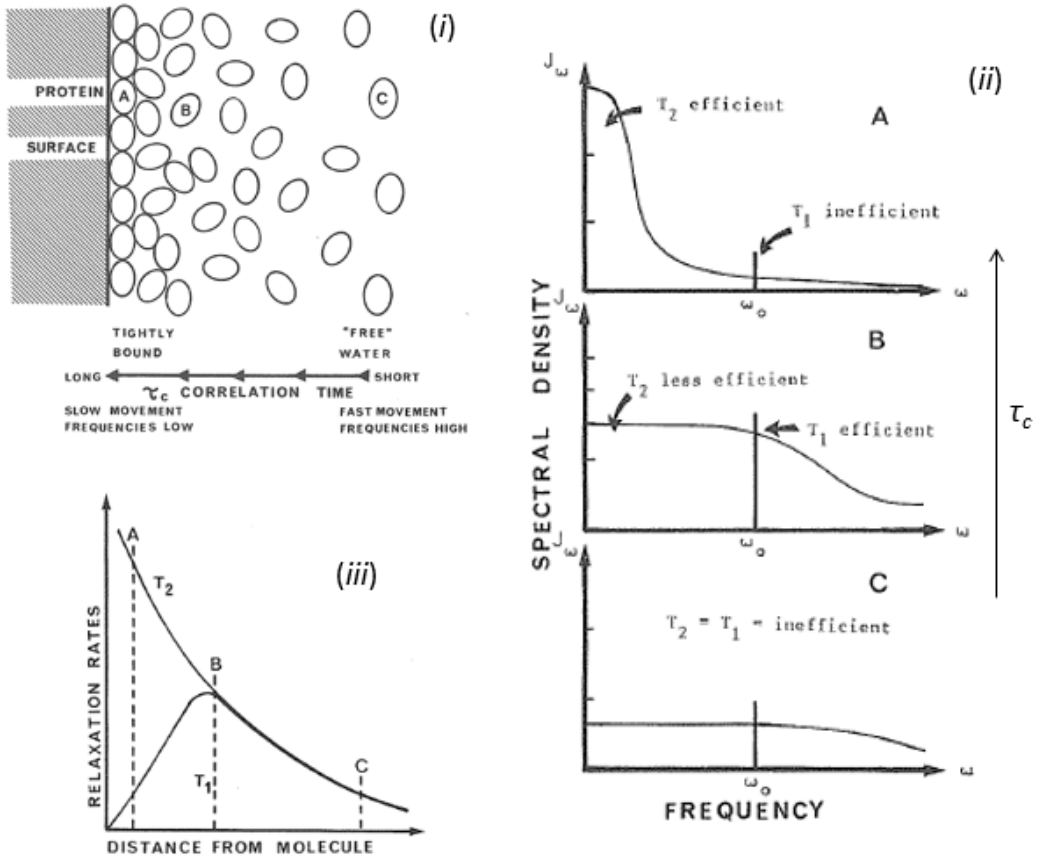


Figure 2.6: (*i*) Water molecules can be tightly bound to a large molecule (molecule A, as part of an hydration layer), in an intermediate distance (molecule B) or at a larger distance (molecule C, as free water). (*ii*) The corresponding expected spectral density distributions and associated relaxation behaviour for the previous types of water molecule attachment. (*iii*) The corresponding relaxation rates $1/T_1$ and $1/T_2$ as a function of the distance (degree of attachment) of the molecules (Foster and Hutchison, 1987)

Tightly bound water molecules are considered to reside within the hydration layers at the surface of large molecules such as polysaccharides and proteins, as molecule A in this example. Thereafter water molecules are continuously distributed with distance, from intermediately bound as molecule B to completely free as molecule C [panel (*i*)]. Tightly bound molecules like A do not readily interact with other molecules hence their correlation time τ_c is long. The corresponding spectral density function distribution is plot A [panel (*ii*)] showing high intensities at low frequency field fluctuations for ω close to zero and markedly decreasing thereafter before ω_0. This translates to efficient T_2 but inefficient T_1 relaxation processes. For less tightly bound molecules like B – at larger distances from the membrane – correlation time τ_c is relatively shorter with spectral density as shown in plot B, where T_2 processes become less dominant and T_1 processes more efficient. For free water molecules like C correlation time τ_c is much shorter with spectral density relatively uniform along the frequency ω spectrum, and both relaxation processes become inefficient. The relationship between T_2 and T_1 relaxation and distance from the membrane as a measure of binding efficacy are shown in panel (*iii*). The strong dependence of T_2 relaxation upon low local field frequency components is the reason that T_1 is in general longer than T_2. It also follows that T_1 relaxation is more frequency-dependent (and hence B_0 dependent) than T_2 relaxation.

A foundational assumption in the above discussion is that only field fluctuations due to particle rotation and translation ('tumbling') are considered. However there are also strong exchange mechanisms due to redistributions taking place in real biophysical situations. A water proton may over time change environment between

tightly bound, intermediate binding and free water conditions. The proportion of time at each state is considered to be in average equal to the proportional occupancy of each state, and the observed relaxation rates to be a weighted mean of the rates for the participating water fractions:

$$\frac{1}{T_{1,2}^{observed}} = \frac{f^{bound}}{T_{1,2}^{bound}} + \frac{f^{interm.}}{T_{1,2}^{interm.}} + \frac{f^{free}}{T_{1,2}^{free}} \qquad \text{Equation 3}$$

where $f^{bound} + f^{interm.} + f^{free} = 1$ (McRobbie et al., 2006). Another simplifying assumption is that only large hydrophilic molecules such as proteins attract water molecules, whereas there is in reality much greater variety in the number and types of binding sites in particular tissues. For instance carboxylic and other ionic groups within proteins (or polysaccharides, nucleic acids etc.) attract water while methyl groups are hydrophobic. Such differences are a potential reason for the differences in the relaxation times observed between different tissues, for instance brain, muscle, fat or liver tissue. Finally the assumption so far of considering mainly water protons is not always a good approximation due to the potential presence of mobile lipid molecules depending on tissue type and disease status. These molecules are in the form of triglycerides, consisting of relatively small glycerol molecules combined with usually unbranched fatty acids. The central part of such acids is an alkyl chain – with high hydrogen concentration – which retains certain freedom of motion within the triglyceride molecule. In adipose or fat infiltrated tissue triglyceride molecules tend to be closely packed and due to these mobility and structural properties show high proton density in MR images. In such cases protons are much more constrained in general than in water and hence exhibit shorter T_1 and significantly longer T_2. Due to their dense packing triglycerides can form a significant component of fatty cells,

and therefore of pixel volume, while due to their hydrophobic nature they can be considered to be compartmentalised from water content (Foster and Hutchison, 1987).

2.1.4 Neuromuscular pathology and multi-component T_2 relaxation

Since the principal source of biological MRI signal is from fat and water protons, abnormalities causing a change in fat or water content result in the modification of proton density and T_1 and T_2 relaxation times. Changes in properties such as free vs. bound water fractions and intracellular vs. extracellular water distribution also influence the observed relaxation behaviour. In addition to compartmentalisation between fat and water components, tissue water itself appears also to often be distributed between non-exchanging compartments. The latter has been demonstrated by the multi-exponential nature of T_2 decay, or equivalently the multi-peak T_2 distribution spectrum seen in skeletal muscle data if the appropriate imaging technique is used, in contrast to the single relatively short T_2 (typically 30-40ms) yielded with more conventional methods (Saab et al., 1999, Sharafi et al., 2017). There appears to be at most only rather slow exchange between intra- and extracellular water compartments. The latter is a relatively small water fraction of about 10 to 20% and has a relatively long T_2 above 100ms. Most of the water is thought to be in the intracellular compartments – around 70% or more – and exhibits various decay rates. These are thought to be related to the restricted water molecular motion caused by the relatively large surfaces by which the cytoplasm is permeated; additional decay processes may be due to water contained in membrane-bound structures and differences in the muscle fibre intracellular space due to twitching.

Other metabolites might convolute these findings (Krssák et al., 2004). Additional short (below 5ms) or long (above 200ms) components have also been reported, in relatively small proportions, and have been related to the hydration shell of macromolecules and small amounts of relatively free water in tissue or vascular blood, respectively (Sharafi et al., 2017, Saab et al., 1999).

Nevertheless, on the conventional T_2 weighted images widely used radiologically healthy muscle appears to have a single short T_2 relaxation time, while both oedema-related water and fat have longer T_2 relaxation times. Therefore T_2 weighted images are sensitive to both fat and water changes within muscle. T_2 weighted images may show abnormalities early in the disease course of many myopathic conditions, likely reflecting increased water content due to inflammation or abnormal blood flow. These contrast properties may identify regions with a higher probability of yielding in a useful pathologic diagnosis when biopsied. The underlying cause of T_2-weighted hyper-intensities in muscle disease is still not entirely understood. Reversal of these findings in inflammatory myopathy is consistent with both improvement of strength and Creatine Kinase level. These findings have been characterised as relating to 'inflammation' or 'muscle oedema', by different authors. However, it has also been reported that inflammatory myopathy often does not result in detectable oedema on muscle biopsy. In addition, denervation causes muscle T_2 weighted hyper-intensity, despite the lack of oedema and inflammation in most cases of denervation (Lovitt et al., 2006).

When there are two separate tissue components in the same volume which are not in exchange, then two corresponding separate T_2 decays are observed. This is the

case in fat infiltrated muscle which contains separate water and lipid components. Various strategies can be used to separate or independently measure the T_2 of these 2 components.

Fat suppression

In short tau inversion recovery (STIR) imaging the signal originating from fat is expected to be nulled, allowing unequivocal visualization of oedema and water. This method takes advantage of the difference in longitudinal magnetization recovery rates between water (muscle) and fat (Figure 2.7). Fat nulling is achieved with the introduction of a 180^o inversion pulse at the beginning of the imaging pulse sequence, and with a subsequent delay before application of a 90^o pulse to generate the image signal at a time when the longitudinal magnetization due to fat protons is zero. STIR images are however only qualitative so cannot be used on their own to generate the quantitative T_2 values required for an MRI outcome measure. Other limitations are the inability to acquire multi-slice images data across large volumes in a single scan due to the constraint of the repetition times (*TR*), the sometimes complex contrast dependence on both T_1 and T_2 (section 4.4), and a difficulty in discriminating tissues with similar T_1 relaxation times, for instance when using contrast agents (Brown and Semelka, 2005).

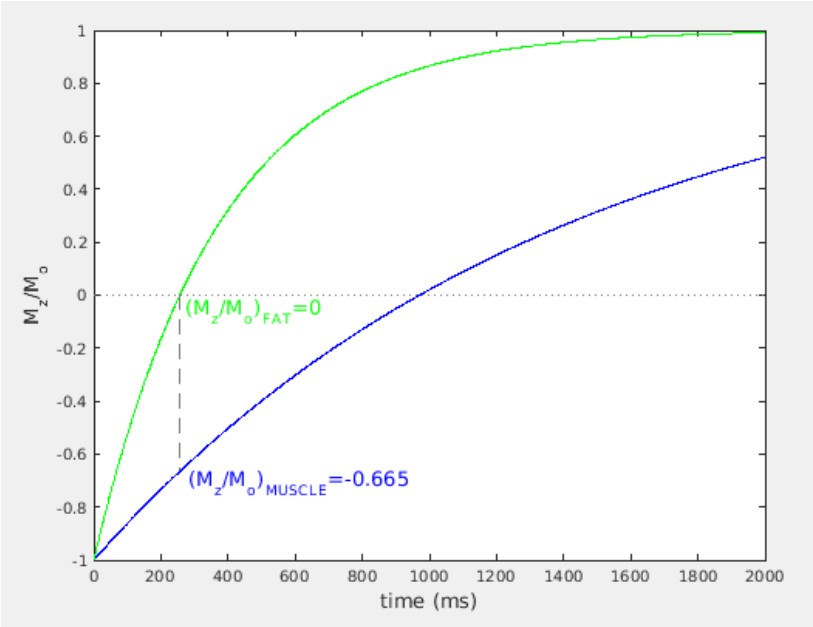


Figure 2.7: T_1 recovery curves for inversion recovery sequences. The 180° inversion pulse inverts the net magnetization for both fat and water at time zero. If the 90° pulse is applied when the fat longitudinal magnetization crosses zero (dashed line), echoes will be formed from signal emanating only from muscle.

An alternative to the STIR approach is frequency selective fat saturation, a technique based on the chemical shift difference between fat and water. In this method again a preparation RF pulse is used, in this case having a narrow bandwidth with centre frequency chosen in order to excite only proton spins in the fat resonance (fat peak) so that they are tipped into the transverse plane and then dephased - before the rest of the pulse sequence, so that no remaining image signal originates from the fat. However under certain conditions there may be difficulty in successfully saturating the fat resonances, e.g. at low B_0 fields because they overlap with water proton frequencies and at higher fields due to the multiple peaks apparent in the fat spectrum. In addition inhomogeneity in either the main magnetic field B_0 or the RF pulse amplitude can cause incomplete fat suppression (Brown and Semelka, 2005, McRobbie et al., 2006).

Post hoc multicomponent modelling

In a different approach, explored in this thesis, methods have been developed based on quantifying multiple components in the decay of T_2-weighted image signal intensity with *TE* to separate the water and fat signal contributions (Kan et al., 2009). In this approach fat and water protons co-existent in a particular imaging volume, are assumed to be physically separate (i.e. not in exchange) and therefore their composite signal decay can be decomposed into their respective individual contributions illustrated in the theoretical curves presented in Figure 2.8.

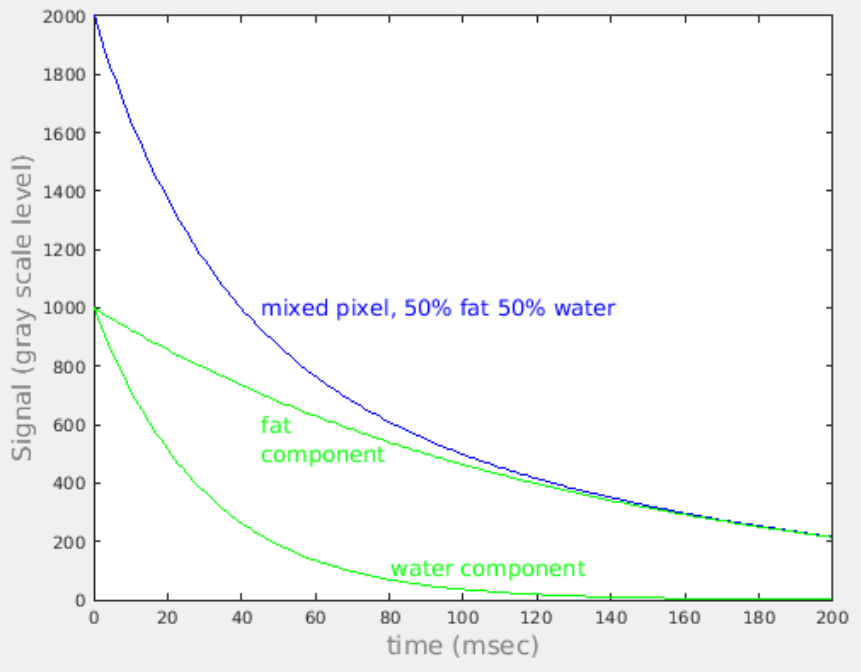


Figure 2.8: Simulation of a decay of T_2 weighted signal comprising of mono-exponential fat and water components, assuming equal proton densities.

After signal decomposition, post-processing can include correction factors for component-specific T_1 weighting and proton density bias (Yao and Gai, 2012), or allowance for transmit field inhomogeneities (Azzabou et al., 2015). The latter publication suggested that such quantitative MR protocols can offer an easily

implemented, straight forward alternative to existing Dixon techniques. The significance of this is that by separating the fat from the muscle signal, the latter will show T_2 value alterations corresponding to oedema and related changes, independent of fat signal intensity changes, which are potentially of great importance since they are potentially early indicators of reversible myopathologies.

Techniques which integrate multi-component and Dixon-type measurements have also been proposed, such as IDEAL – CPMG combined signal decomposition (Janiczek et al., 2011), or fat correction of the conventional T_2 values with use of the Dixon acquired muscle – fat fraction (Yao and Gai, 2012).

From the above discussion it emerges that a useful distinction can be drawn between the so called global T_2, referring to a T_2 estimated assuming a mono-exponential signal decay even where this may not actually be so, and multicomponent T_2 values estimated with a model accounting for different compartments existing in *in vivo* skeletal muscle (Carlier, 2014).

2.1.5 Spatial localisation, slice selection and sources of artefacts

The signal induced by the resonant magnetization (section 2.1.1) must be detected and analysed in a way that leads to image formation, i.e. to provide spatial localisation. While in the Larmor equation (Equation 1, section 2.1.1) the value of γ for a given nucleus is fixed, the value of the magnetic field strength can be varied linearly along any spatial direction by applying a magnetic field gradient. If such a gradient is applied along the *x* direction this equation becomes

$$\omega_0(x) = \gamma(B_0 + xG_x)$$ Equation 4

where G_x is the gradient along that direction, the magnitude of magnetic field strength added or subtracted from B_0 per unit distance from the origin. The main field B_0 direction is by convention along the *z*-axis (along the magnet bore of the MRI scanner) therefore gradients are linearly varying added or subtracted *z*-components of magnetic field. This means that the total field in the *z* direction that nuclei experience depends on their position, in this case along the *x*-axis, and for that reason their resonance frequency is also spatially dependent. Conversely determination of the frequency of a proton's MR signal determines its position in space. This principle is illustrated for a single dimensional sample consisting of different numbers of protons at each position in Figure 2.9 below.

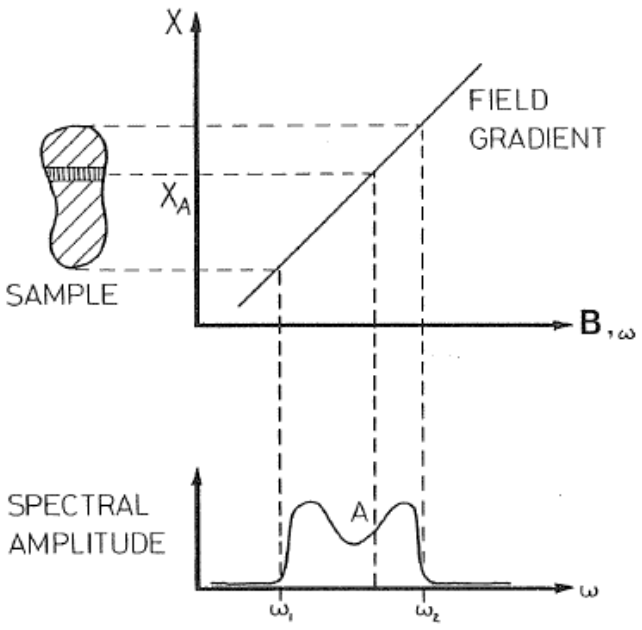


Figure 2.9: Illustration of the use of a magnetic field gradient to produce a spectrum of frequencies corresponding to a density of spins in a given sample, along a single direction (Foster and Hutchison, 1987)

By sequentially applying three gradients G_x, G_y and G_z, in each orthogonal direction, 3-dimensional spatial localisation may be achieved. The simplest 2D version of MRI is tomographic, in which the first gradient applied in any acquisition is G_z, because it has to be combined with the signal excitation RF pulse in order to image the desired slice. With the use of a relatively strong G_z gradient the resonant band of frequencies (or bandwidth) defined by the selective RF pulse can be centred closely around the image isocentre as shown in Figure 2.10 so that signal originates only from a relatively narrow slice.

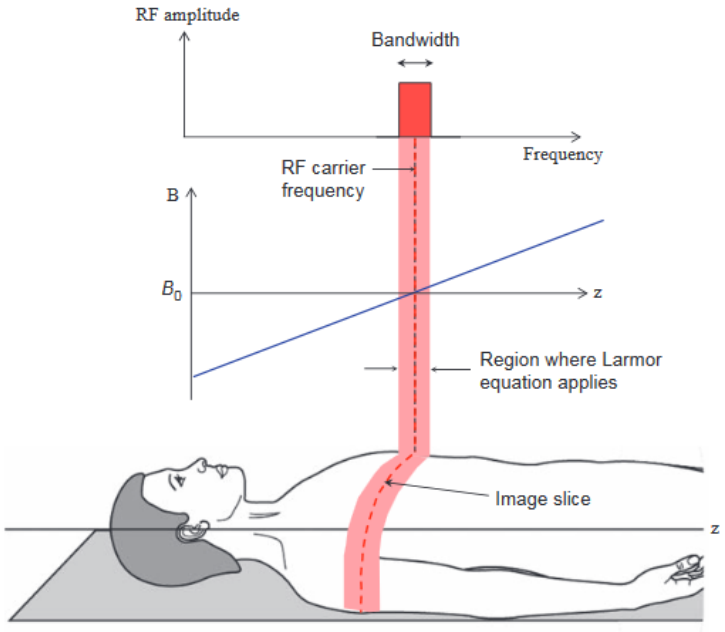


Figure 2.10: Slice selection by combining the bandwidth-limited excitation RF pulse with a G_z gradient (McRobbie et al., 2006).

A representation of the RF pulse and the G_z gradient (or G_{ss}, slice selective gradient) in the time domain is shown in Figure 2.11 below.

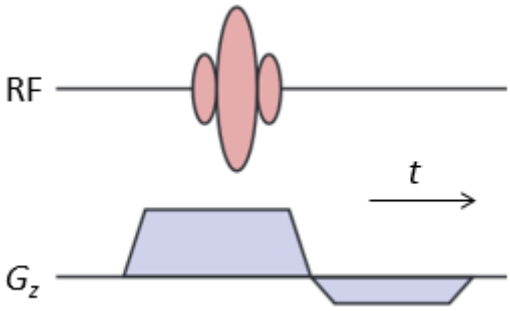


Figure 2.11: Representation of the generic shape of the excitation RF pulse and the slice selective gradient with time (McRobbie et al., 2006)

The RF pulse amplitude as a function of time (the "pulse-shape") usually approximates a sinc function as, assuming linear time-frequency behaviour, this corresponds to a constant (square) excitation frequency spectrum producing a rectangular slice profile as suggested in Figure 2.10. Mathematically this is because the Fourier transform of a sinc function (time domain) is a rectangle or 'top hat' function in the frequency domain; this ensures that the necessary spatial frequency components (which translate to image contrast) are adequately represented. Figure 2.11 also shows that the applied gradient has a secondary negative lobe. Its role is to bring back into the same phase all the excited protons across the selected slice to correct for the dispersion in resonant frequency that the applied G_z gradient caused such that the final phase of the protons is close to zero after the application of both G_z lobes.

At the same time that the refocusing part of the slice selective gradient is applied the in-plane frequency encoding commences with the application of another gradient in either the *x* or the *y* direction. Its role is to assign a range of frequencies along the length of the respective dimension. The first, negative frequency encoding gradient

results in dephasing of the signals along that dimension, with the phase at each position corresponding to its specific Larmor frequency. By reversing the applied gradient's polarity, signals start to rephase until they all coincide back at the same initial phase, as shown in Figure 2.12.

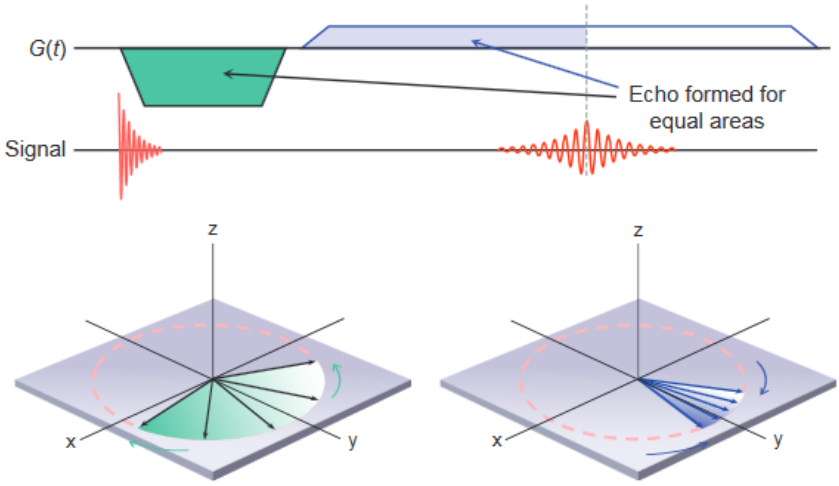


Figure 2.12: Dephasing of signal due to the range of frequencies imposed by the gradient along the y direction, shown as different phase angles in green (bottom left); reversing the polarity of the gradient eventually brings all spins along y at the same phase (bottom right) in the form of a gradient echo (McRobbie et al., 2006)

This coherence recovery is called a gradient echo and this echo is the signal finally detected. It is a function of the resonant frequency range used and the gradient strength (Equation 4). For that reason it can be written as a weighted sum of spatial (localisation dependent) frequencies that are linear functions of the gradient strength. This expression consists mathematically of an inverse Fourier transform of these weights with respect to the above spatial frequencies as parameters. These weights describe in fact the spin density along that direction because the application of the gradient has captured the spatial distribution of the excited spins. Thus by

performing a (forward) Fourier transform of the acquired signal we can recover the spins spatial distribution as a single dimensional image.

Localisation in the second in-plane dimension is typically achieved with use of the phase encoding gradient G_{PE} oriented with the remaining (third) *y* or *x* dimension. This is done by repeating the previously described frequency encoding process, starting for each repetition from a different initial phase introduced by the phase encoding gradient, as illustrated in the example in Figure 2.13.

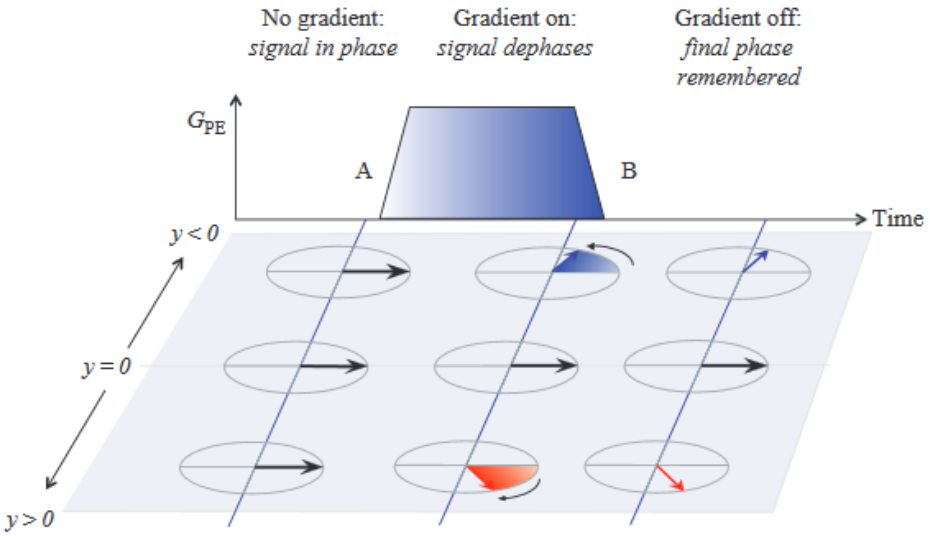


Figure 2.13: Phase encoding returns the signal to the Larmor frequency but with phase changes dependent on the position along the phase encoding axis (McRobbie et al., 2006)

This phase will then be preserved in the frequency encoded gradient echo coherence (Figure 2.12) and will be unique for each encoding step and (via the G_{PE} gradient) for each position along the phase encoding direction. To achieve complete localisation the phase encoding gradient is applied in steps of increasing amplitude for each sequence repetition. Due to Equation 4 this means that for each phase encoding gradient increment the phase differences between spins in different locations are

maintained between each phase encoding step. Therefore the spatial distribution of spin densities will be captured as a function of the phase steps, or equivalently the G_{PE} gradient. The acquired signal in terms of both phase and frequency encoding evolves in time as shown in Figure 2.14.

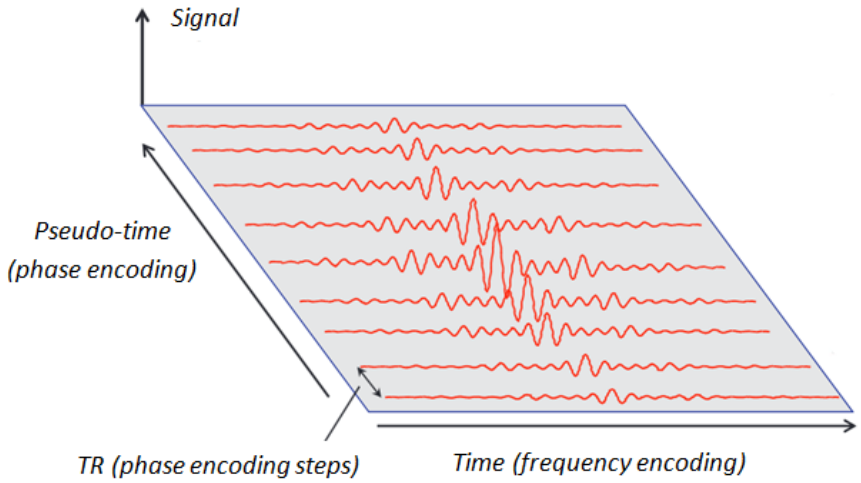


Figure 2.14: Representation of the signal acquisition with time. Frequency encoding takes place in real time while phase encoding is implemented in discrete time intervals or pseudo-time (McRobbie et al., 2006)

As can be seen there is equivalence between the two axes with increasing signal towards the centre which corresponds to spins exactly on resonance (in the frequency axis) with zero phase (in the phase axis). In extension to the previous explanation for the one dimensional image recovered from the frequency encoded signal, a 2-dimensional Fourier transform of the type of signal presented in Figure 2.14 will yield the spin density distributions in space, i.e. the desired 2D image of the sample. The entire acquisition scheme described above is shown graphically in Figure 2.15 below as a 'pulse sequence'.

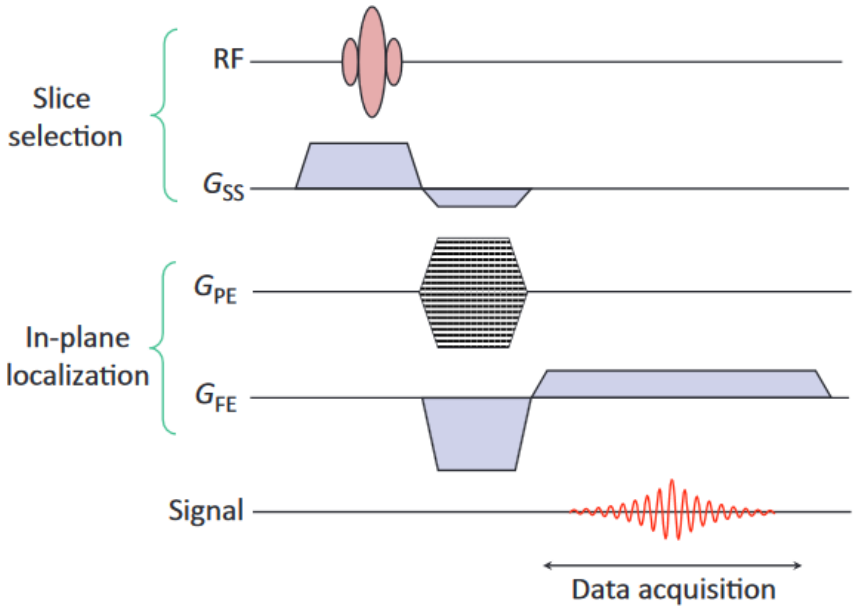


Figure 2.15: A generic gradient echo sequence. Time is shown horizontally and amplitudes vertically; G_{SS} is the slice selective gradient, G_{FE} the frequency encoding gradient and G_{PE} the phase encoding gradient (McRobbie et al., 2006)

The spin echo sequence described in section 2.1.1 (Figure 2.3), relies on the use of an additional RF pulse to refocus and acquire the signal as a spin echo which is an alternative echo-formation method to the gradient reversal used in gradient echo sequences. Spin echo sequences are more pertinent to the scope of this work and are further explored in later sections, although the Dixon technique (section 2.1.2) is usually based on differently timed gradient echoes.

The necessity in conventional imaging for relatively long time intervals ('pseudo-time') between phase encoding steps, i.e. the frequency encoding repetition time *TR*, renders it susceptible to motion artefacts, one of the most common sources of MR imaging imperfections. A special case of motion is blood flow in vessels. Both types of motion occur relatively slowly in comparison to the frequency encoding timescale but much faster than that of phase encoding, thus between the phase encoding steps

either anatomy or blood movement can cause erroneous signal registration (image 'ghosts') along that direction (Figure 2.16).

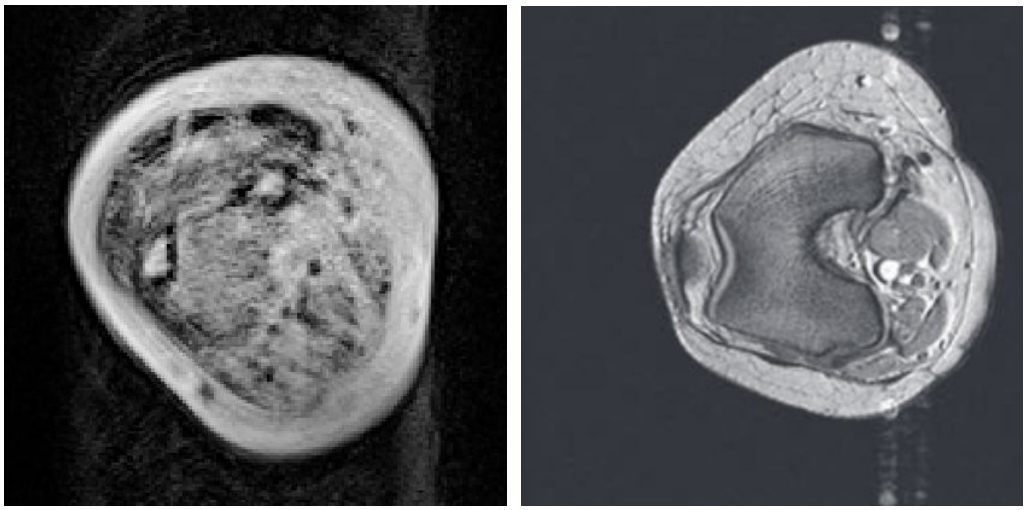


Figure 2.16: Left: example of bulk motion artefact in a forearm image (from the datasets analysed in chapter 5). Right: artefact from flowing blood in a knee image (McRobbie et al., 2006)

The chemical shift phenomenon (section 2.1.1) can have an effect in the frequency encoding process when the imaged anatomy contains a mix of fat and water protons, dependent upon the receive bandwidth per pixel (the range of frequencies corresponding to the pixel size). If the latter in a given acquisition is comparable to the absolute difference in the resonant frequencies of fat and water, chemical shift artefacts might appear along the frequency encoding direction, typically in the form of dark bands (void signal) or bright bands (superimposed signal) at the interface between different tissue types (i.e. muscle and fat) or the background (Figure 2.17).

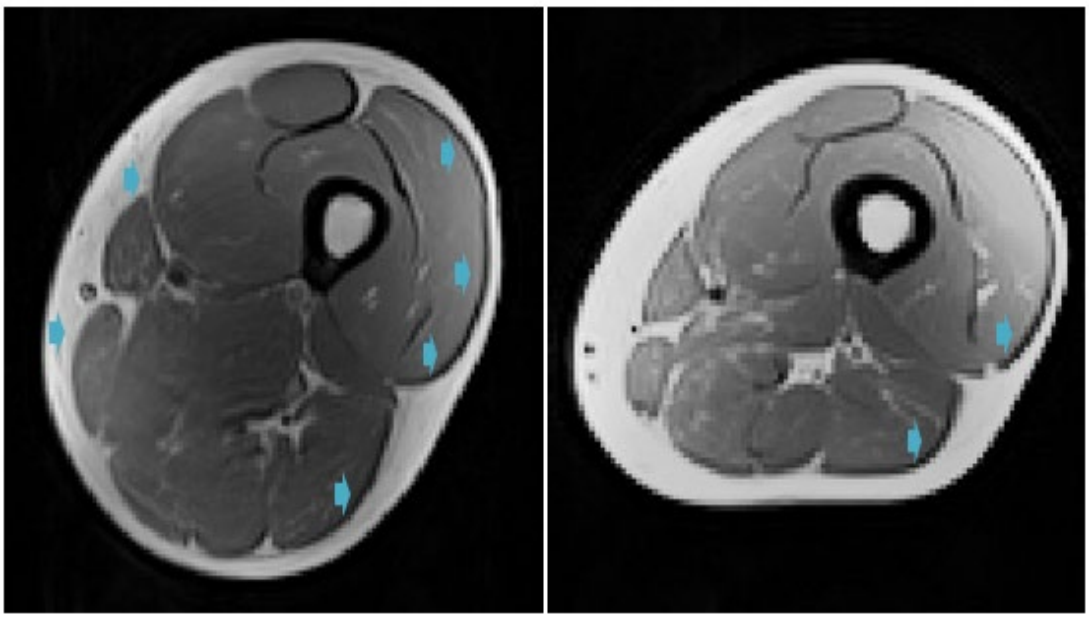


Figure 2.17: Examples of chemical shift artefacts (frequency encoding direction) in thigh images (from the datasets analysed in chapter 4)

2.1.6 Signal behaviour in spin echo sequences; the limitations of exponential signal models

By applying successive additional 180˚ RF refocusing pulses following excitation, a series of spin-echoes may be generated in what is known as a multi-echo spin echo (MESE) sequence (Carr and Purcell, 1954), the generic design of which in an imaging context is shown in Figure 2.18.

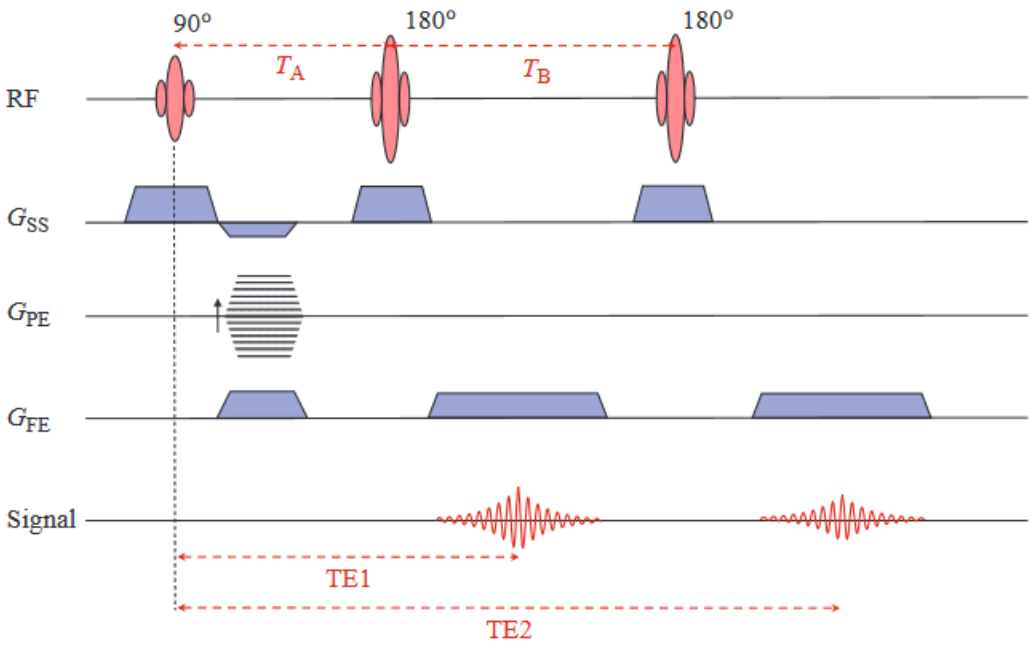


Figure 2.18: A schematic representation of the generic multi-echo spin echo sequence design (here shown for 2 echoes). The exact shape or slope, duration and timing of the RF pulses and the gradients is responsible for the variations in the spin echo modalities used in practical applications (McRobbie et al., 2006)

A modification commonly used in practice is the Carr-Purcell-Meiboom-Gill (CPMG) sequence (Meiboom and Gill, 1958) which is a special case of the MESE method. The main advantage of the CPMG sequence lies in its ability to partially compensate for imperfect 180˚ refocusing pulses. Such (imperfect) pulses leave the spin populations unbalanced, with lost phase coherence at the echo time so that their vector sum refocuses slightly above or below the transverse plane (Figure 2.19); this leaves a longitudinal magnetization component which can accumulate over the echo train. The improvement implemented in the CPMG sequence is to apply the 180˚ pulses with a 90˚ phase shift relative to the initial 90˚ B_1 RF pulse so that any residual longitudinal magnetization will be cancelled at every other echo as shown in Figure 2.19.

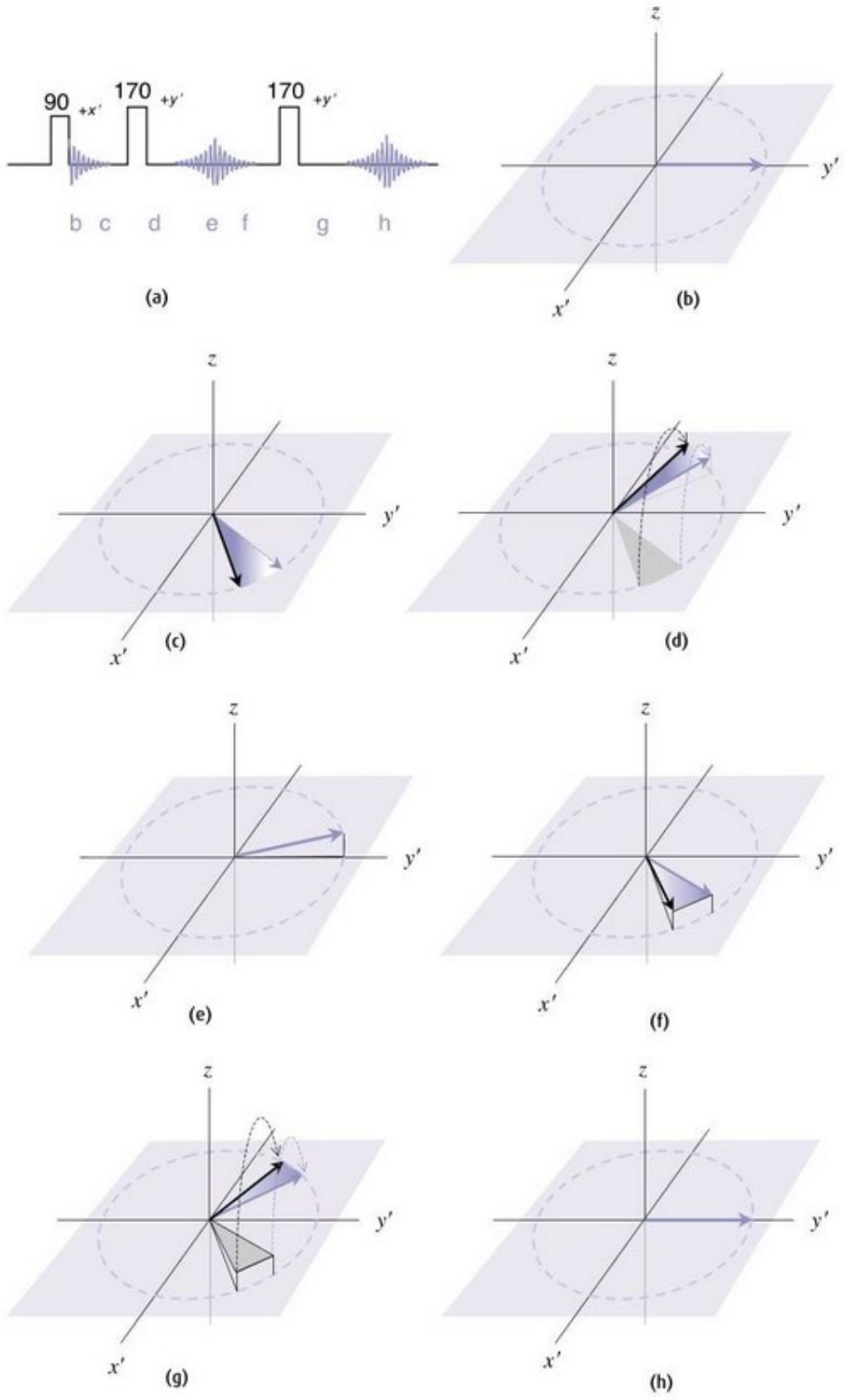


Figure 2.19: (a) In the CPMG sequence, the 180° pulses are applied along the y' axis instead of the x' axis. If the 180° pulse is imperfect, e.g. only 170° (c), the first spin echo will rephase above the transverse plane (e), giving a slightly reduced echo height. However, the second 170° brings the spins exactly back to the transverse plane (g) so that the even echoes are always correct (h) (McRobbie et al., 2006)

The decay in successive spin-echo signal magnitudes in an MESE sequence is heavily dependent upon T_2 as well as other factors; thus if an appropriate model for the signal decay with parameters including T_2 is available, then fitting this model to the data in some optimal way should yield accurate estimates for those parameters, including T_2. If applied to imaging data sets this will yield maps indicating the tissue T_2 values at each pixel location.

Historically T_2 mapping was attempted assuming the signal decay in CPMG imaging experiments follows a simple exponential decay; in practice it was found that such imaging data is better fitted by a model comprising a mono-exponential decay plus a constant (Milford et al., 2015). The reason for including a constant has been stated as the existence of a rectified background noise floor or a build-up of a persistent background signal due to imperfect refocusing. Both of these factors will be addressed in full detail later in this thesis, but in brief, this background signal typically occurs in MESE sequences when spins partially refocus at the sampling (echo) times, or when echoes originating from previous pulses overlap with subsequent ones and accumulate with time. The former are known as Hahn echoes and the latter as stimulated echoes (Hahn, 1950). B_1 field inhomogeneity resulting in flip-angle errors is a major contributor to this background signal (Milford et al., 2015).

In this case the equation modelling the signal evolution in such a sequence is

$$s_n = \alpha e^{-\frac{nTE}{T_2}} + c \qquad \text{Equation 5}$$

where s_n is the signal acquired at echo number n, α is the signal amplitude extrapolated to time zero, TE the echo time, T_2 the decay time constant and c the constant accounting for noise and any background signal. In the conventional fitting process the signal acquired is compared to the values predicted by the model at each TE, iteratively varying the model parameters in a systematic manner to obtain the "best fit" by means of a least squares minimization, which aims to minimize the sum of the squared residuals (differences) between the acquired values (signal) and the predicted ones.

However, inspection of real-world MRI CPMG signal trajectories suggests that the signal behaviour is far from mono-exponential, with the first echo amplitude commonly less than the second, an alternating amplitude modulation between even and odd echoes, as well as the build-up of an apparent baseline signal contribution. Figure 2.20 exemplifies the stimulated-echo coherence and alternate echoes effects in signals acquired with CPMG sequences, clearly demonstrating divergence from simple exponential decay behaviour.

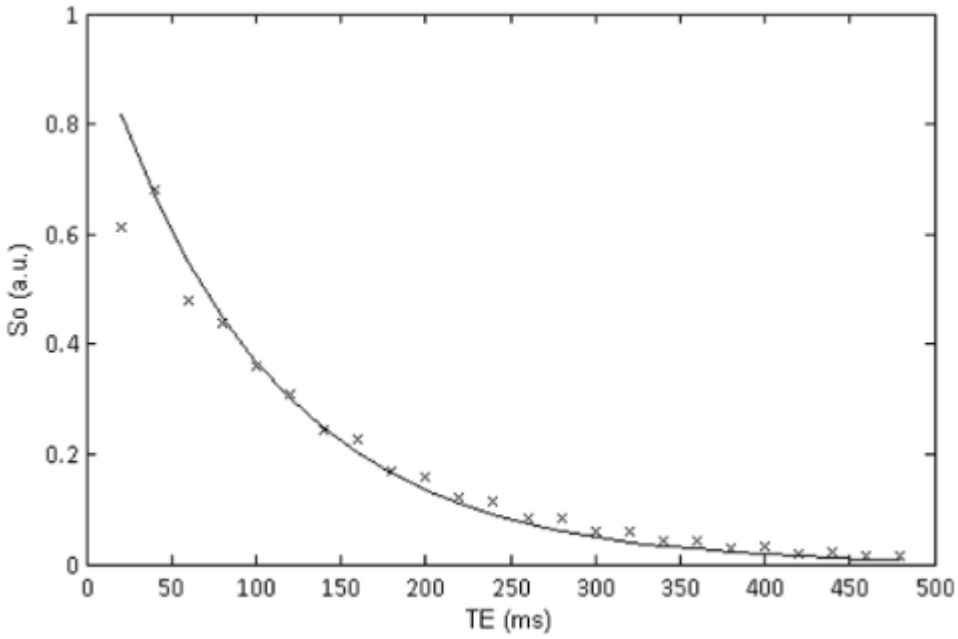


Figure 2.20: Simulation of the acquired signal using a MESE sequence in the case of imperfect RF pulses. The signal here is fitted to an exponential model excluding the first echo due to the stimulated echo effect, however the alternate echo uneven amplitude effect is clearly visible and not accounted for by this signal model (Milford et al., 2015).

In order to achieve reliable parameter estimation using exponential models, additional tactics have typically been adopted, for instance inclusion of a constant and omission of the uneven echoes – or at least the first echo (Milford et al., 2015, McRobbie et al., 2006).

2.1.7 EPG formalism

Assuming that the tissue of interest is properly characterised by a single T_2 value, these apparent deviations from simple exponential T_2 decay arise largely due to deviations in the refocusing RF pulse flip angles from their ideal values (the effects of noise on the parameter estimation process will be addressed separately below). The CPMG sequence introduced in section 2.1.1 attempts to compensate errors in MESE acquisitions where while the initial excitation angle is 90˚, the effects of imperfect

refocusing pulses are partially corrected at even echo times. However under practical conditions MR imaging experiments involve various causes of non-ideal flip angles including potentially a combination of globally mis-calibrated, or locally deviating B_1 fields, a deliberate choice to reduce flip angles e.g. to manage the SAR – the specific absorption rate (Guérin et al., 2014), and more fundamentally the interaction between frequency-selective excitation and refocusing RF pulses and the field gradients used for slice selection. A characteristic of the latter is that the theoretically required infinite sinc function-shaped RF pulse – the Fourier transform of a perfectly square ('top hat') slice profile – cannot be achieved in practice. Additional effects commonly causing alterations in the slice profile shape and amplitude for the excited and refocused magnetization are magnetic field inhomogeneities, gradient non-linearity as well as tissue-field interactions. The extended phase graph (EPG) formalism (Hennig, 1991) can be used to more accurately predict the magnetization behaviour in real world scanners where such imperfections are typically present, accurately quantifying the expected signal at successive echo times. The two fundamental EPG concepts are its representation of dephasing and its description of the RF pulse action. Its implementation is then based on the combination of these two concepts.

The dephasing caused by the action of gradients (Figure 2.21) can be quantified using a wavenumber, essentially quantifying the number of cycles of dephasing that has occurred while a particular gradient was switched on. The motion of all protons of magnetization *M* at position *r* from the isocenter will be described as $M_x(r) = \cos(kr)$ and $M_y(r) = \sin(kr)$, where the wavenumber *k* describes the

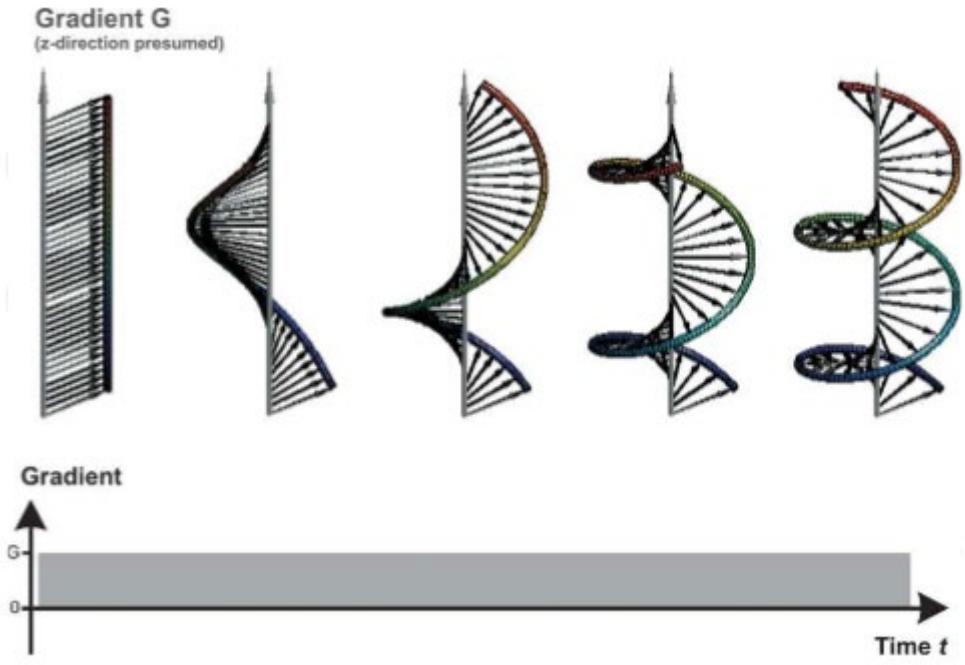


Figure 2.21: The dephasing effect due to an assumed z-direction gradient exhibited by initially in-phase transverse magnetization, evolving with time (left to right) and with distance from isocenter (bottom to top), denoting increasing frequency (coloured external helix) to both directions (Weigel, 2015).

action of the gradient $k(t) = \gamma \int_0^t G(t')dt'$ after time t and x, y are the transverse plane axes. At this stage a change of coordinates transforms the previous equations to more convenient expressions:

$M_+(r) = M_x(r) + iM_y(r) = Me^{ikr} = (M_-)^*$ and

$M_-(r) = M_x(r) - iM_y(r) = Me^{-ikr} = (M_+)^*$

where i is the imaginary unit and the asterisk denotes the complex conjugate operation. The latter is equivalent to a 180˚ rotation of M_+ to M_- around the x axis and vice versa, therefore it mimics the action of a 180˚ RF pulse by means of a simple algebraic operation. This is a core feature of the EPG formalism, essentially linking the spin dephasing aspect with RF pulse rotation effects. Depending on the action of gradients and RF pulses, M_+ will be the magnetization component left to be dephased (moving away from the x axis) while M_- will be the magnetization component to be

refocused (moving towards the *x* axis). A second important characteristic of the complex notation is that it can be generalised to include all participating signals – with arbitrary frequencies or positions – the entire available transverse magnetization is represented as the integral $\int M e^{ikr} d^3r$. This integral may be evaluated in the case of M_+ over a volume *V* according to

$$\tilde{F}_+(k) = \int_V \left(M_x(r) + iM_y(r)\right) e^{-ikr}\, d^3r = \int_V M_+(r) e^{-ikr} d^3r$$

$$\overset{FT}{\Leftrightarrow} \quad M_x(r) + iM_y(r) = M_+(r) = \int_V \tilde{F}_+(k) e^{ikr} d^3k$$ which is the Fourier transform ($\tilde{F}$) of each *k*-dependent dephasing package (state) of dephased magnetization ('isochromat'). Thus $\tilde{F}_+$ will describe the distribution (scale factor in terms of Fourier decomposition) of magnetization among different dephasing states *k* and $M_+$the magnetization dependence on distance *r* (position space) for each *k*, within the specific volume *V*. This basic relationship in frequency and space is demonstrated in Figure 2.22 (Weigel, 2015).

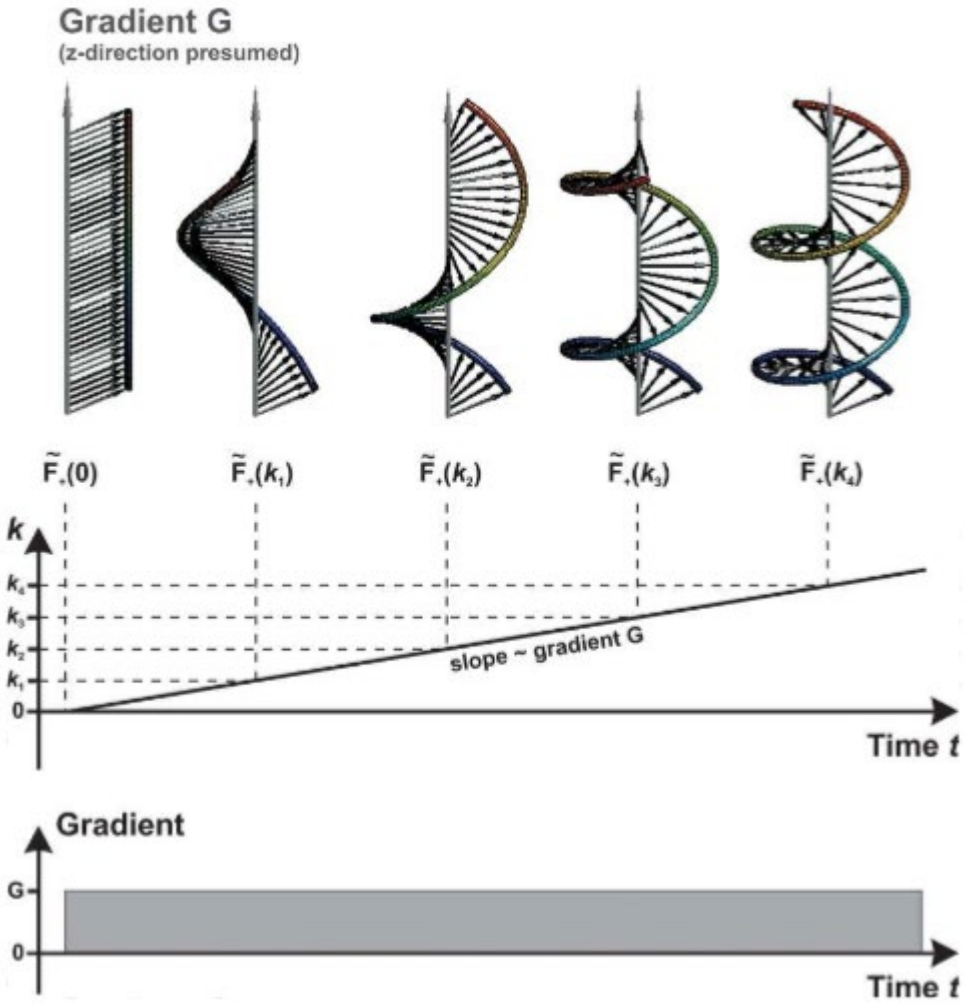


Figure 2.22: The z-direction gradient dephasing effects of Figure 2.21 linked to the dephasing stages corresponding to *k* evolution (Weigel, 2015).

In a similar fashion $\tilde{F}_-$ will describe the Fourier decomposition of rephasing magnetization as a function of *k*. While dephasing occurs due to the action of gradients, rephasing occurs after the application of refocusing RF pulses – which will be mathematically expressed by the conjugate operation. However RF pulses in the general case may cause rotation through any angle – as for example in excitation, inversion and any intermediate/imperfect refocusing pulse – and this will cause the existence and modulation also of longitudinal z-component of the magnetization. For that reason it is necessary to introduce also the Fourier term $\tilde{Z}(k)$ denoting the distribution of longitudinal magnetization with *k*.

The above principle of the *k*-dependent $\tilde{F}_+$, $\tilde{F}_-$ and $\tilde{Z}$ configuration states is also the framework used to describe the action of the RF pulses. The action of RF and gradient pulses can be mathematically expressed by the action of the rotation matrices

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \quad \text{and} \quad R_z(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{on a}$$

magnetization vector of the form $\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix}$. R_x expresses in general a rotation of angle α around the *x* axis, e.g. the effect of an RF pulse applied along the *x* axis on a magnetization vector. R_z expresses rotation of angle φ around the *z* axis, e.g. the dephasing caused by the gradients. Therefore if an RF pulse is applied with a (transverse plane) phase φ with respect to the *x* axis, its action on the above magnetization vector will be given by the rotation matrix $R_\varphi(\alpha) = R_z(\varphi)R_x(\alpha)R_z(-\varphi)$ (Hahn, 1950, Carr and Purcell, 1954). In order to express this action in terms of the dephasing and rephasing *k* states prerequisite in the EPG approach, we need to use the same transformation to complex notation coordinates as previously. The magnetization vector then becomes $\begin{bmatrix} \tilde{F}_+(k) \\ \tilde{F}_-(k) \\ \tilde{Z}(k) \end{bmatrix}$, expressing the three different magnetization dephasing and rephrasing components per *k* state, and the R_x and R_z matrices according to linear algebra become the rotation matrices

$$T_x(\alpha) = \begin{bmatrix} \cos^2\frac{\alpha}{2} & \sin^2\frac{\alpha}{2} & -i\sin\alpha \\ \sin^2\frac{\alpha}{2} & \cos^2\frac{\alpha}{2} & +i\sin\alpha \\ -\frac{i}{2}\sin\alpha & +\frac{i}{2}\sin\alpha & \cos\alpha \end{bmatrix} \text{ and } T_z(\varphi) = \begin{bmatrix} e^{+i\varphi} & 0 & 0 \\ 0 & e^{-i\varphi} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ around}$$

the x and z axis respectively. Therefore in this transformed coordinate system the

effect of an RF pulse with a phase difference φ from the *x* axis causing rotation of the magnetization by an angle α will be given by the matrix $T_\varphi(\alpha) = T_z(\varphi)T_x(\alpha)T_z(-\varphi)$. Taking together the above, the expression describing the action of a generic RF pulse (inserted at any angle in the transverse plane) on a complex *k*-state dependent magnetization vector, will be:

$$\begin{bmatrix} \tilde{F}_+(k) \\ \tilde{F}_-(k) \\ \tilde{Z}(k) \end{bmatrix}^+ = \begin{bmatrix} \cos^2\frac{\alpha}{2} & e^{2i\varphi}\sin^2\frac{\alpha}{2} & -ie^{i\varphi}\sin\alpha \\ e^{-2i\varphi}\sin^2\alpha & \cos^2\alpha & ie^{-i\varphi}\sin\alpha \\ -\frac{i}{2}e^{-i\varphi}\sin\alpha & \frac{i}{2}e^{i\varphi}\sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} \tilde{F}_+(k) \\ \tilde{F}_-(k) \\ \tilde{Z}(k) \end{bmatrix}^- \quad \text{Equation 6}$$

where the symbols "-" and "+" denote respectively the magnetization before and after application of the RF pulse. The above equation describes fully the effects of the application of gradients and RF pulses on the magnetization's constituent parts as these are modulated in both phase and amplitude. For that reason it has also been termed a partition state effect or method (Kaiser et al., 1974), and because it is applied to all Fourier components for each state and partition of the magnetization, it provides the basis of the EPG model. For clarity, the term state is commonly used for each of the three configurations $\tilde{F}_+$, $\tilde{F}_-$ and $\tilde{Z}$ and for each dephasing order *k* – thus for $\tilde{F}_+(k)$, $\tilde{F}_-(k)$ and $\tilde{Z}(k)$. It is useful to emphasize that any magnetization in the $\tilde{Z}(k)$ state does not undergo phase modulation from the RF pulses' action (its *k* remains constant) since this occurs only on the transverse plane, however it continues to carry the position encoding phase information via the Fourier transform and decomposition.

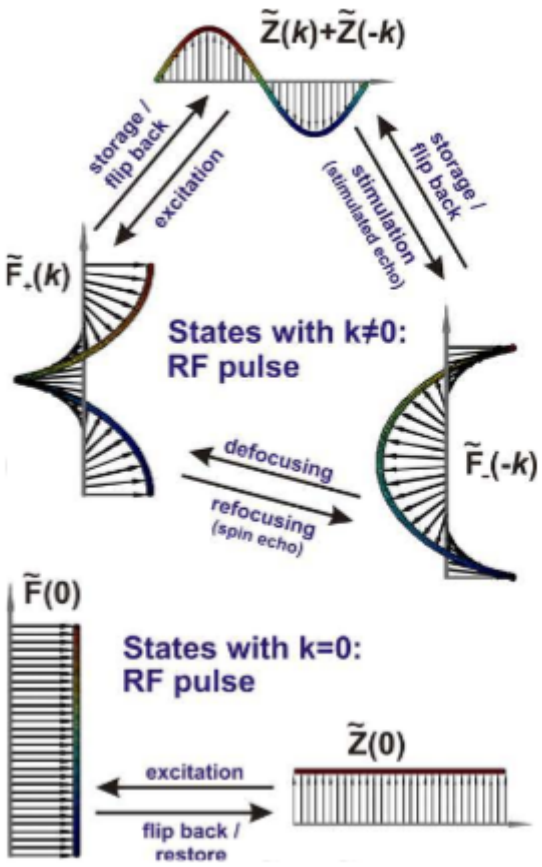


Figure 2.23: A schematic representation of the mechanisms of magnetization exchange described by EPGs – taking place for equal dephasing orders *k* only. Notably for exchanges when k=0 $\tilde{F}_+(0)$ and $\tilde{F}_-(0)$ express the same physical quantity ($\tilde{F}(0)$), and exchange between $\tilde{Z}(0)$ and these denotes signal induction. For k≠0 conversions occur between $\tilde{F}_+$, $\tilde{F}_-$ and $\tilde{Z}$ and of interest is any $\tilde{F}_-$ magnetization as this can lead to signal generation via either a spin or a stimulated echo (Weigel, 2015).

For that reason $\tilde{Z}$ states are also commonly referred to as stored magnetization. An important special case is for *k*=0, the initial time point of the described phase history, where $\tilde{Z}(0)$ is converted to $\tilde{F}_+(0)$ which is the same as $\tilde{F}_-(0)$ because they are at the same zero initial phase, or state $\tilde{F}(0)$. This is simply the excitation stage (section 2.1.1) as notated within the EPG formalism. All consequent $\tilde{F}(0)$ (zero phase) states are the result of refocusing components and will lead to signal generation either in the form of a spin echo or a stimulated echo, while all other (*k*≠0) $\tilde{F}_+$ and $\tilde{F}_-$ states are considered to be of completely dephased magnetization. The latter is a key

assumption of practical importance in applications such as simulating the MESE and CPMG types of sequences examined in this work. For these applications *k* is as an integer, a discrete number denoting this full dephasing between echoes in units of 2π. In that light *k* represents the number of helical turns or wave cycles in the graphs in Figure 2.21 through Figure 2.23 (Weigel, 2015).

The above description of dephasing between RF pulses, described as *k*-states succession has a direct beneficial impact in the practical implementation of the EPG concept in spin echo applications such as the CPMG sequence. It means that dephasing can be captured quantitatively as a function of (discretized) *k* in between *TE*s, and at each *TE* it is only affected by the action of the RF pulse. While this is not exactly what is happening in the real MR system (dephasing is a continuous and not quantized process), it is nevertheless a convenient and numerically correct way to track refocusing (coherent) magnetization and thus the accumulated signal in realistic scenarios. It is also the basis on which actual EPGs are plotted as demonstrated in Figure 2.24, and as commonly found in the related literature (Scheffler, 1999).

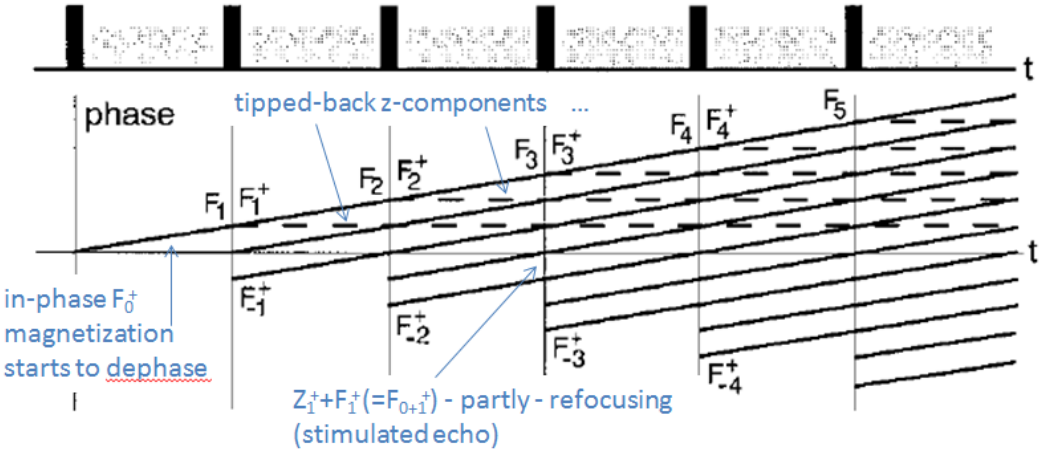


Figure 2.24: Extended phase graph (EPG) for a generic spin echo sequence with constant gradient and equal RF pulses and echo spacing. For simplicity the more compact F^{+}_{-k}-style in notation is used for the refocused magnetization states after each RF pulse. Magnetization components are classified (binned) in *k*-dependent envelopes (states) and at each RF application are redistributed depending on the flip angle (not shown) (Scheffler, 1999).

It should be noted that the term 'extended' underlines the fact that these phase graphs refer to the entire spatial distribution of spins that through the Fourier transform are converted to *k*-state components and redistributed accordingly. However these graphs are only a schematic depiction and in this form do not provide a quantitative analysis of the signal evolution. A common way to achieve this is with the use of algebraic matrices and operations, as an extension of the concept of rotation matrices presented already with respect to RF pulse action. The history of phase evolution according to the above *k*-state classification or binning can be tracked as summarised in the P matrix manipulations shown in Figure 2.25 below, which in fact is the information of a 'classic' EPG (as in Figure 2.24) converted into a matrix operations description (Hargreaves, 2012).

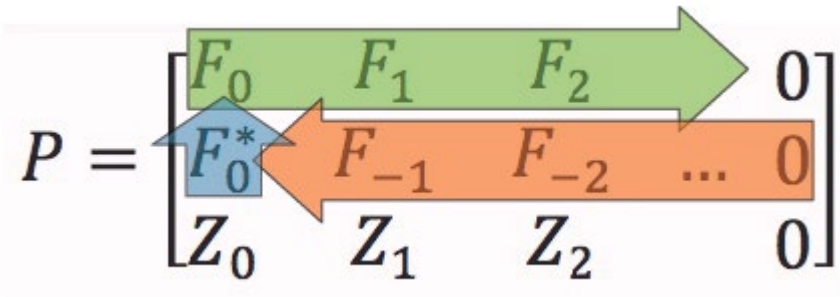


Figure 2.25: Schematic representation of a generic EPG as a phase history matrix P. Here all *k*-states are symbolized in a compact manner as F_k for dephasing components (green arrow envelope), F_{-k} for rephrasing components (red arrow envelope) and Z_k for longitudinal 'stored' magnetization components. The transition between the two first is achieved at *k*=0 (zero phase state) with use of the complex conjugate operation (blue arrow envelope) (Hargreaves, 2012).

It can be seen that the phase history matrix P consists of the top row dephasing components moving to higher *k*-states (to the right), the middle row rephrasing components moving to lower *k*-states – since they have been reversed by the RF pulse action – and the stable *k*-states of longitudinal magnetization in the bottom row. Importantly the complex conjugate operation mentioned previously completes the zero phase crossing where coherent magnetization can be quantified and rephasing (middle row) becomes dephasing (top row).

Apart from gradient and RF pulse effects the signal evolution in real acquisitions is affected also by the tissue-specific T_2 and T_1 decay rates (section 2.1.1). These effects are dependent upon a geometric distinction between transverse and longitudinal magnetization which is maintained within the EPG model through the $\tilde{F}$ and $\tilde{Z}$ states, and they affect equally all *k* states with the exception of the $\tilde{Z}(0)$ state (*k*=0, coherent $\tilde{Z}$ magnetization) where T_1 recovery will also be taking place. In terms of matrix operations decay effects can therefore be expressed as

$$\begin{bmatrix} F_+ \\ F_- \\ Z \end{bmatrix}^+ = \begin{bmatrix} e^{-\frac{t}{T_2}} & 0 & 0 \\ 0 & e^{-\frac{t}{T_2}} & 0 \\ 0 & 0 & e^{-\frac{t}{T_1}} \end{bmatrix} \cdot \begin{bmatrix} F_+ \\ F_- \\ Z \end{bmatrix}^- + \begin{bmatrix} 0 \\ 0 \\ 1 - e^{-\frac{t}{T_1}} \end{bmatrix}$$, where t is the elapsed time interval. For spin echo sequences this is typically split into two halves of the echo spacing time to accommodate for the RF pulse action as explained further below. Signal quantification can finally be achieved by combining the operation of the rotation and decay matrices introduced so far with the phase history matrix. This takes the form of successive operator matrix applications on the *k*-dependant states of magnetization, with the generic expression

$[F_+, F_-, Z(k+1)] = E \cdot S \cdot T \cdot [F_+, F_-, Z(k)] \cdot E \cdot S$, where E and T are the decay and RF rotation matrices respectively, and S the phase history matrix operation. This is an iterative process as lower *k*-order states participate in – or are modulated to – higher *k*-order states, and is demonstrated in Figure 2.26 linked to the phase evolution framework representation of Figure 2.25.

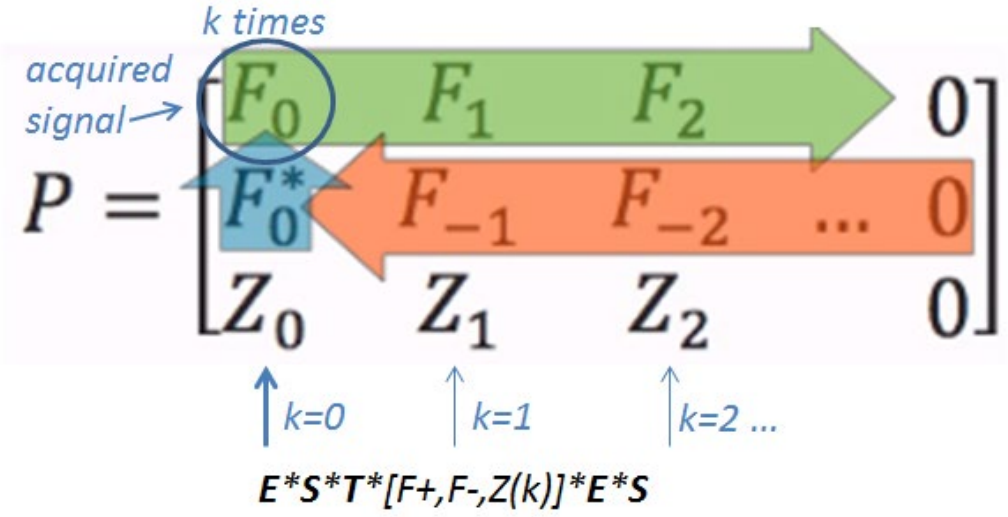


Figure 2.26: Schematic representation of the quantitative version of the phase evolution matrix of Figure 2.25. After inclusion of the decay E, rotation T and state transition (or phase evolution) S matrices action on the magnetization states, F_0 will now quantify the acquired signal at each echo time, and the entire matrix can be renamed to the (states) population P matrix.

Due to the existence of the RF pulse the relaxation decay effects are split in two parts for $t = TE/2$ each before and after its application, and also the *k*-dependent state transition has to occur twice in order to correctly distribute populations. This is demonstrated practically in the numerical example shown in Figure 2.27, where echo formation with use of the EPG formalism is calculated for a hypothesized CPMG sequence using intermediate (≠90˚ or 180˚) flip angles.

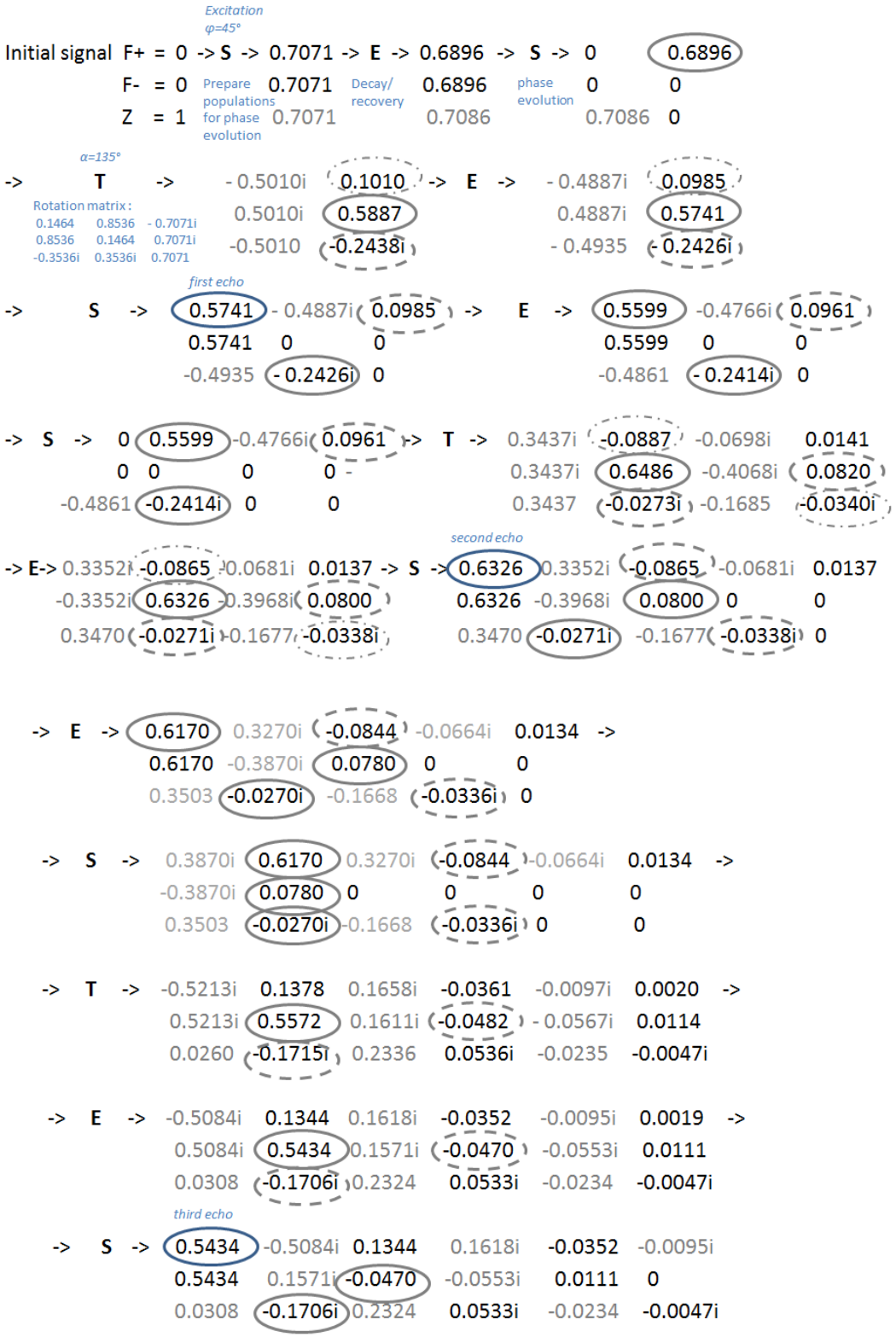


Figure 2.27: A numerical example demonstrating the action of the E, S and T matrix operators participating in the magnetization states populations matrix P, according to the quantitative matrix-based EPG representation described in the previous text and figures, for excitation angle 45˚, refocusing angle 135˚, T_2=200ms and T_1=1000ms, for the 3 first echoes (assuming unit initial magnetization vector).

Calculations were performed as a succession of simple default Matlab functions and commands executing the respective mathematical operations (e.g. matrix rotations etc.), and form in essence the basic computational implementation of the EPG formalism (Figure 2.28), in this case written in Matlab 2015a (MathWorks, Massachusetts, United States).

function se = EPG (p, echoes)

parameters p: excitation angle φ, refocusing angle α, echo spacing TE,
T_2, T_1, RF phase ϑ, amplitude

Initiate se — *initiate signal vector, will assume equal echoes in length*

Initiate P — *initiate populations matrix P; will be a 3·(2·length of echoes) in size because*
E and s operators act from both sides of P (signal calculated at uneven states)

$$E = \begin{bmatrix} e^{-\frac{t}{T_2}} & 0 & 0 \\ 0 & e^{-\frac{t}{T_2}} & 0 \\ 0 & 0 & e^{-\frac{t}{T_1}} \end{bmatrix}$$ — *define decay matrix (T_1 recovery applied separately)*

$$T = \begin{bmatrix} \cos^2\frac{\alpha}{2} & e^{2i\vartheta}\sin^2\frac{\alpha}{2} & -ie^{i\vartheta}\sin\alpha \\ e^{-2i\vartheta}\sin^2\alpha & \cos^2\alpha & ie^{-i\vartheta}\sin\alpha \\ -\frac{i}{2}e^{-i\vartheta}\sin\alpha & \frac{i}{2}e^{i\vartheta}\sin\alpha & \cos\alpha \end{bmatrix}$$ — *define RF rotation matrix T*

P (1st column, 1st entry) = P (1st column, 2nd entry) = sin(φ); — *excited signal*

P (1st column, 3rd entry) = cos(φ);

*Perform E*S*T*E*S multiplication:*

For k = 1:length(echoes) — *start P states evolution for-loop*

P = E * P; — *E operator: decay/recovery of magnetization, including*

P (1st column, 3rd entry) + $\left(1 - e^{-\frac{t}{T_1}}\right)$ — *T_1 recovery contribution through state zero*

$$S = \begin{Bmatrix} move\ P\ 1^{st}\ row\ to\ the\ right\ (1^{st}\ entry\ zero) \\ move\ P\ 2^{nd}\ row\ to\ the\ left\ (last\ entry\ zero) \\ P\ 1^{st}\ row\ 1^{st}\ entry = P\ 2^{nd}\ row\ 1^{st}\ entry\ conjugate \end{Bmatrix}$$ — *S operator phase evolution*

P = T * P; — *T operator: rotation, mixing states of equal dephasing order k*

P = E * P; P (1st column, 3rd entry) + $\left(1 - e^{-\frac{t}{T_1}}\right)$ — *E operator action for the 2nd half of TE*

P = S * P — *S operator action for the 2nd half of TE*

se (echoes) = P (1st column, 1st entry); — *signal at echo time*

end

Figure 2.28: The generic structure ('pseudo-code') of the EPG function as a computer program

To summarise, EPGs rely on:

- Capturing the magnetization phase history as packages of complete dephasing between events (RF action, decay etc.), denoted by a corresponding integer *k*-number. For instance, if no other actions take place, signals after excitation dephase forming a disc on the transverse plane with zero vector sum. This phase evolution is encapsulated in the action of the *S*-operator.
- Action of the RF pulse, described by matrix operations – the *T*-operator – which provide a closed form solution for quantification of any refocused, defocused and longitudinal (z-axis) stored remnant magnetization. A typical example is Hahn's experiment: the action of a second 90° pulse on the disc of dephased spins mentioned above, transforming it subsequently to the 'eight-ball' shape with vector sum half the initially available magnetization.
- Decay effects – T_1 and T_2 relaxation and T_1 recovery – described by action of the *E*-operator along the phase history, readjusting the magnetization sum vector accordingly.
- For CPMG sequences in particular, decay and dephasing continues after RF pulse action and this has to be added to the phase history description, forming, after the initial excitation, a *E*-*S*-*T*-*E*-*S* pattern of operators fully describing the effects of the applied sequence.

Finally, EPGs also vary along the slice selection direction due to the slice-selection gradient, causing location-dependent resonance offsets, which has to be accounted for, as will now be discussed.

2.1.8 Slice profile effects

Real MR experimental conditions involve in general imperfect slice profiles: a basic reason is that the theoretically required infinite-duration sinc function-shaped RF pulse – the Fourier transform of a perfectly square ('top hat') slice profile – cannot be implemented in practice. Also the selective excitation process in MR is non-linear except for a close approximation at low flip angles. In real sequences the excited and subsequently refocused magnetization will therefore in general show variation in both magnitude and phase across the selected slice. This is due to inevitable design compromises in the selective RF pulse shapes, and their associated field gradients, complicated by possible global and local B_1 calibration errors. Hence the EPG formalism presented above must be extended to account for these variations across the slice profile. Under this approach the signal is considered as the aggregate decay curve representing EPG-generated echo train amplitudes at successive locations integrated over the slice profile, where the effective excitation and refocusing angle for each location across the slice can be calculated from the applied RF pulse and gradient characteristics, assuming these are known *a priori*. This is described by the action of the B_1 field factor B_1f, which is a percentage scale factor accounting for B_1 transmit inhomogeneity, having a value of 100% for a slice profile with 180° refocusing angle at its centre. B_1f determines the precise excitation and refocusing flip angles.

The program design ('pseudocode') corresponding to the above description of the slice profile corrected EPG function (sEPG) is shown in Figure 2.29. In the first step the magnetization immediately following the excitation pulse is calculated for each

position across the slice profile, and this is defined as the initial magnetization at each position, in terms of its *x*-*y* and *z* projections and angle with respect to the *x'* axis. This is based on the Bloch simulation of the rotations caused by the gradient and excitation RF pulse action (including its phase) per time step and slice position (Hargreaves, 2016). In this example all time steps are equal, the RF pulse duration vector *t* is the same for both excitation and refocusing, and the refocusing RF pulse is of double amplitude than the excitation pulse, however in the general case this might not necessarily be so. Similarly, for the subsequent refocusing pulses the effective refocusing angle is calculated from the *z*-component (arc of cosine) after action of the refocusing RF pulse and gradient on an initial unitary z-axis-only magnetization. This is because the EPG function will fully describe the remaining effects of the above refocusing rotation on the excited magnetization at each slice position – since these are integrated in its complex notation, through zero phase *x*-*y* plane crossings and *z*-axis residuals – and also account for T_1 and T_2 decay effects.

function [sEPGsignal, sEPGtrains] = sEPG (t, G, fRF, B1f, x, df, p, echoes)

t: RF pulse duration vector (time domain) *G: gradient strength in Gauss*

fRF: RF function type (e.g. 'sync') *B1f: B1 field strength*

x: slice width (frequency domain) *df: phase of the incoming RF pulse*

p: EPG model parameters: echo spacing TE, T_2, T_1, RF phase ϑ, amplitude

echoes: echo train length

grad = G (size t vector) *initiate gradient vector (length of t vector)*

rf = B1f*fRF*t + zeros (size t vector) *initiate RF vector according to fRF function (e.g. sinc) scaled by B1f plus 2nd part of zeros when refocusing gradient acts*

rfrefoc = 2*B1f*fRF*t *RF refocusing pulse, usually double than excitation without rephasing*

refocratio = -.52; gradref = grad + refocratio*grad; tr = t + t; *define refocusing gradient vector and respective time vector*

sEPGtrains = zeros (size x vector * size echoes vector); *initiate sEPG echo trains across slice profile matrix*

[M_{xy}exc, Mexc] = sliceprofile(rf, gradref, tr, Inf, Inf, x, df); *slice profile of the excited signal*

T_1 and T_2 decay will be accounted for by the EPG function (below)

Sliceprofile module

M = [0;0;1]; Δt = t step, e.g. t(2)-t(1) *define initial magnetization and time step*

M_{xy}exc = zeros size x, Mexc zeros 3-size x *initiate M_{xy}exc & Mexc (Mexc the 3D magnetization populations vs x)*

for length x *start for-loop for all x positions*

for length rf *start for-loop for all RF pulse time steps*

rot$_z$ M for df *where* $rot_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$

rot$_z$ M for gradref (Δt/2) *where gradref = 2π*γ*x*gradref*Δt/2*

rotate M for rf *this will be a rotation around x' and z' based on rotation matrices as above (e.g. rot$_z$ & rot$_x$) and rf time steps*

rot$_z$ M for df & gradref (Δt/2) *second part of df and gradient action*

end

[Mrefoc] = sliceprofile(rfrefoc, grad, t, Inf, Inf, x, df); *for unit-M_z flipping only: M_{xy} refocusing is accounted for by the EPG function (decompose the RF action to flipping + refocusing)*

refoc_angle=acos(Mrefoc 3rd entry)*180/π; *calculate the effective refocusing angle, will be arc of cosine of the Mrefoc z' projection*

for length x *start for-loop for all x positions*

sEPGtrains = EPG ([refoc_angle, TE, T_2, T_1, M_{xy}exc, Mexc(z), Mexc(angle), amplitude], echoes);

end *EPG echo trains vs x based on the above excited and refocused signal slice profiles (from 'sliceprofile' function)*

sEPGsignal = mean(sEPGtrains); *final 'projection' of the slice profile signal*

end

Figure 2.29: The generic structure ('pseudo-code') of the sEPG function as a computer program.

In Figure 2.30 an example of simulated sEPG echo trains according to through-slice position are shown, for the same parameters that were used in the example of Figure 2.27: the three first echo heights of the solid blue middle slice echo train are the EPG values calculated in that example.

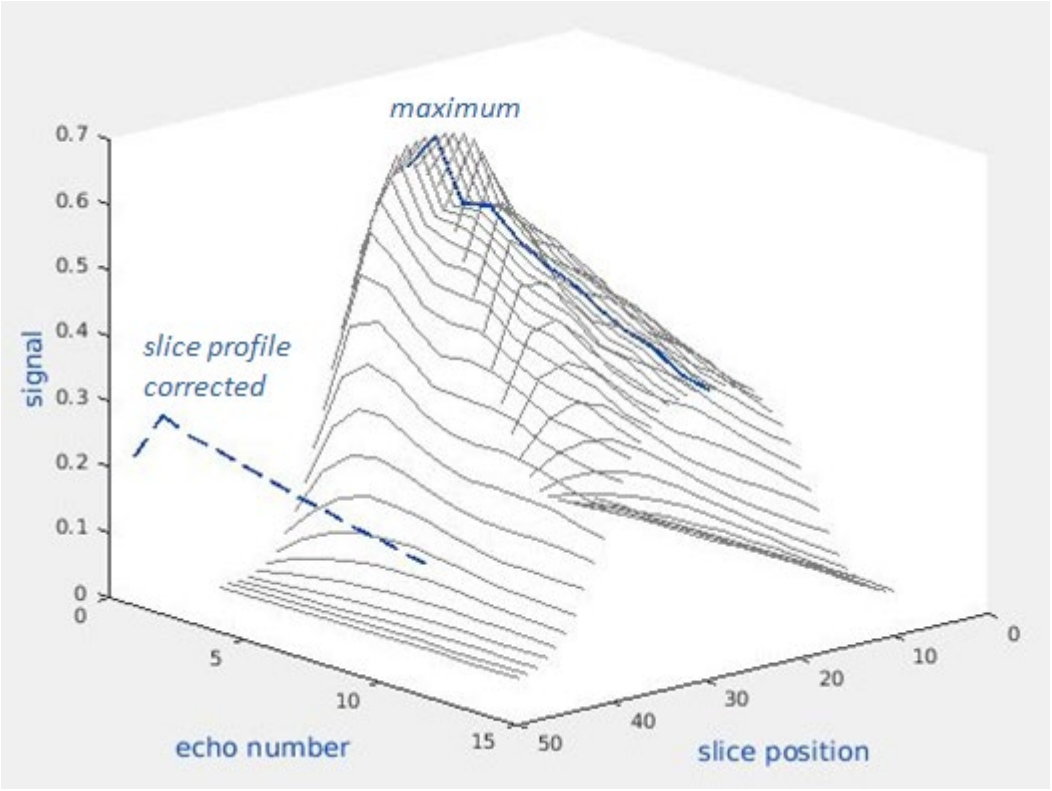


Figure 2.30: Simulation of the signal behaviour according to the EPG formalism at different positions across the slice profile for successive CPMG echoes. The mean of these signals represents the CPMG signal observed in real acquisitions (labelled as “slice profile corrected”) whereas the signal at the centre of the slice (labelled as “maximum” in this case) is typically higher.

Both the excitation and refocusing pulses were selective – typical of those implemented on our Siemens 3T scanners used in this work – with different pulse shapes and appropriate slice gradient amplitudes. The dashed blue line echo train projected on the signal-to-echo number plane is the mean of the individual sEPG contributions, representing (with a normalised amplitude) the signal that would be

observed in a practical imaging experiment; as it can be seen slice profile effects cause both signal reduction (away from the centre of the slice) and echo train modulation. This suggests that it is essential to take the slice profile into consideration when estimating transverse relaxation parameters from conventional CPMG data.

2.1.9 Noise in MR images

The levels of (thermal or electronic) noise in MR images depend upon the receiver bandwidth, the pixel size, the field strength, the number of averages taken and the number of coil elements. Most of these factors are operator controlled and therefore may be selected to optimise image quality in terms of SNR, notwithstanding their influence upon resolution and scan time. A generic expression describing their relationship is $SNR \propto \Delta x \cdot \Delta y \cdot \Delta z \cdot F_{seq}\sqrt{\frac{NSA \cdot N_{PE} \cdot N_{FE}}{BW}}$, where *NSA* is the number of signal averages, N_{PE} and N_{FE} the phase encode and frequency encode matrix size respectively, *BW* the receive bandwidth, $\Delta x \cdot \Delta y \cdot \Delta z$ the pixel dimensions and F_{seq} is a sequence dependent relaxation factor; for spin echo (SE) sequences for example F_{seq} takes the form $F_{SE} \propto \left(1 - e^{-\frac{TR}{T_1}}\right) \cdot e^{-\frac{TE}{T_2}}$ in the simplest version, where *TR* is the sequence repetition time, *TE* the inter-echo spacing, and T_1, T_2 the longitudinal and transverse magnetization decay rates respectively. The factors in the square root are related to the noise while the rest determine the available signal (McRobbie, 1996).

NMR signals are detected using quadrature detection providing real and imaginary components; following Fourier transformation the image data are magnitude

reconstructed, meaning that any phase information is disregarded. The noise in each quadrature channel can be assumed to be normally distributed – ignoring at this stage multi-coil combination effects, and assuming noise to be spatially stationary (Hardy and Andersen, 2009). However calculating the magnitude from the complex source data is a non-linear operation and the noise in the resulting magnitude signal therefore not itself normally distributed. If *A* is the true image pixel intensity in the absence of noise and *M* the measured signal, it can be shown that the probability distribution p_M for *M* in the presence of noise is given by

$$p_M(M) = \frac{M}{\sigma^2} e^{-\frac{M^2+A^2}{2\sigma^2}} I_0\left(\frac{A \cdot M}{\sigma^2}\right) \qquad \text{Equation 7}$$

where I_0 is the modified Bessel function of 0th order and σ the standard deviation of the assumed equal Gaussian noise in both the real and imaginary channels (Gudbjartsson and Patz, 1995). This expression is known as the Rician distribution following its original introduction for electrical engineering applications (Rice, 1944). The dependence of p_M on SNR (A/σ) is shown in Figure 2.31.

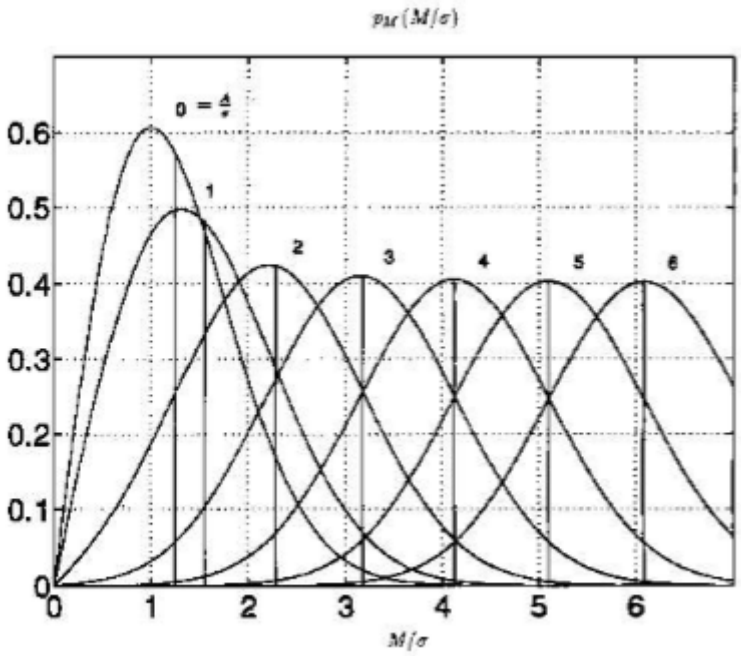


Figure 2.31: The Rician distribution of the measured signal *M* and the resulting mean, with increasing signal to noise ratio *A/σ* (Gudbjartsson and Patz, 1995)

It can be seen that for lower SNRs the probability distribution for the observed signal (*M*) and therefore the corresponding mean of *M* deviate increasingly from a Gaussian distribution and the expected signal (*A*) respectively. This bias in the mean value of *M* relative to *A* is the result of rectification of negative noise contributions, contributing especially at low signal levels, through taking the magnitude of the real and the imaginary parts of the acquired signal. In image regions where only noise exists, *A* is zero, and Equation 7 reduces to

$$p_M(M) = \frac{M}{\sigma^2} e^{-\frac{M^2}{2\sigma^2}} \qquad \text{Equation 8}$$

known as the Rayleigh distribution (Papoulis, 2002). In that case it can be shown that the mean signal observed and its standard deviation are given by the analytical expressions

$\bar{M} = \sigma\sqrt{\pi/2}$ and $\sigma_M^2 = (2 - \pi/2)\sigma^2$ Equations 9

which can be used to estimate the true noise σ from the background signal in magnitude reconstructed images (section 3.1) (Gudbjartsson and Patz, 1995, Papoulis, 2002).

2.2 Statistical modelling – Maximum likelihood estimation

2.2.1 Introduction

Given a parametrised signal model, such as an exponential function, or sEPG formulation, and experimental observations, such as the signal decay in a MSME acquisition for a particular pixel, the task is to adjust the model parameters until the signal predicted by the model is in some sense most compatible with the observed signal; the parameter estimates will in that case be taken as our best estimates of the (unknown) ground truth values.

The process of fitting a data sample of measurements M(**x**) containing N entries then involves iterative comparison of data and a theoretical model S(**x**,**p**) assumed to describe those data. The **x** are discriminating variables to be used in the fit, taking the value x_i for the i^{th} measurement point, and **p** is the vector of parameters participating in the theoretical model S. The parameters are allowed to vary in the fit to data, i.e. each iteration of the optimisation process uses a different value for **p**, in order to improve the compatibility of data and model. This process requires the definition of a test statistic that is used to quantify how well the model and data agree and to then vary **p** to try and obtain an improved model description of the data. So in order to

perform a fit to data we perform an optimisation process in the parameter space which involves minimising the sum ST of the chosen test statistic T, dependent on the sample of observations $M(x_i)$ and the model $S(\mathbf{x},\mathbf{p})$, over the data. Therefore for each iteration of the optimisation process **p** are varied in order to improve compatibility between $M(x_i)$ and $S(\mathbf{x},\mathbf{p})$, which is quantified by T.

In order to converge on a solution, one starts with an initial estimate of the parameter set which is used to evaluate ST. Having done this a new estimate of **p** is then calculated following a pre-defined rule, and at this and each subsequent step each corresponding ST is evaluated. From an initial set of estimates of the parameters **p**, it is normally possible for the algorithm to determine in which direction in the parameter space a more optimal set lies. Having done this, the algorithm will perform another search starting from a point **p'** yielding an ST hopefully closer to the assumed minimum than the previous one. This process is repeated until the optimisation algorithm yields a sufficiently small step size $\boldsymbol{\delta} = |\mathbf{p} - \mathbf{p}'|$. When **δ** is smaller than some pre-defined minimum distance in the parameter space, the optimisation is said to have converged on a minimum value $\mathbf{p}_{min}$. In practice a set of convergence criteria is defined, examples of which may be thresholds with regards to the change in the values of **p**, T or the number of iterations. Other criteria may be analogous to the condition that the gradient of T must be zero at a minimum. After finding the value $\mathbf{p}_{min}$ corresponding to a minimum the final step usually is for the algorithm to determine the corresponding uncertainty $\delta\mathbf{p}_{min}$, which depends on the test statistic that is being minimised.

Two different approaches to parameter estimation, each using a different form of test statistic, are considered in this thesis: least squares fitting (LSQ), and the method of maximum likelihood estimation (MLE).

2.2.2 Least squares fitting

Least squares fitting aims to minimise the sum of the squared residuals (differences) between the acquired signal values and those predicted by the model. Various computational approaches have been proposed to achieve least squares minimization (or nonlinear regression) for nonlinear models such as Equation 5; the Levenberg-Marquardt method (Bates, 1988) is commonly used being with some variations and improvements implemented within numerous image processing software and statistical packages. It achieves a fast and efficient least squares minimization by a smooth transition from using the steepest descent direction of the gradient of the residuals function of the parameters, i.e. the 'chi-squared', far away from the minimum in the error space, to the second order expansion terms (derivatives) path near the minimum (where the first order terms will be zero). The basis for this is the ability to approximate any function (here of the residuals) at a given point by a function of its derivatives at that point (i.e. a Taylor series expansion). Usually fitting proceeds iteratively, by updating for each step the provisional solution for the parameters, so that eventually the point corresponding to the minimum of the sum of the squared residuals is reached. Thus a global convergence is hopefully achieved rapidly with the use of the steepest descent for the gradient, and then fast local convergence with the use of the second order terms' steps.

2.2.3 Maximum likelihood parameter estimation

The most obvious performance measure to consider in a parameter estimation optimisation experiment is bias: the difference between the average, or expectation value of a parameter estimate (estimator) and its true value (Morris et al., 2017). Since in investigational (e.g. patient) study datasets the latter is unknown, bias is itself subject of estimation. From statistical theory it is known that if for a particular experiment there exists an unbiased estimator for which the variance can attain the lowest possible value, it is obtained by the maximum likelihood (ML) method; for that reason ML estimators are said to be asymptotically more precise than conventional estimators (Sijbers et al., 1998a). For a random sample of N observations **x** from a distribution that depends on one or more unknown parameters **p** with probability density (or mass) function $l(\mathbf{x}, \mathbf{p})$, the joint probability density (or mass) function of **x** with regards to **p**, $L(\mathbf{p}) = \prod_{i=1}^{N} l(\mathbf{x}, \mathbf{p})$, is called the likelihood function; the set of parameters that maximizes L is the maximum likelihood estimator of **p** (PennState, 2018). The main test statistic used in this work is therefore *likelihood* based; the likelihood L of an event to occur is proportional to the probability density function (PDF) related to the probability of the event occurring at a given point in the parameter space (Bevan, 2013). For that reason it is not necessarily normalised but can be useful in investigating the relative expectations of an event or experiment, whereas probabilities and PDFs typically refer to the absolute expectation of an outcome. As a means of systematic evaluation a method based on such a statistic will be compared here with least squares minimization results (considered a χ^2 statistic,

i.e. for least squares based methods the test statistic is the residuals function of the parameters).

This process in the case of least-squares minimisation was described in general terms in section 2.2.2. In the case of maximum likelihood estimation, L represents the likelihood of the parameters **p** given the observed data **x**, and therefore is a function of **p**. After data collection and the likelihood function of a model given the data has been calculated – based on the conditions and nature of the particular experiment – statistical inferences about the data sample can be made. The aim is to find the parameters and therefore the specific PDF, that make the observed data most likely. It is computationally more convenient to, rather than maximise L, minimise the negative logarithm of L, $-\log L(x|p)$ (i.e. minus the logarithm of "the likelihood of the parameters **p** given the data **x**"), justified since L and $-\log L(x|p)$ are monotonically related. If $-\log L(x|p)$ is differentiable and $\mathbf{p}_{min}$ exists, the following partial differential equation

$$\frac{\partial[-\log L(x|p)]}{\partial p} = 0 \qquad \text{Equation 10}$$

must hold for each parameter in **p**. This corresponds to the fact that the first derivatives of a continuous and differentiable function vanish at the locations of its minimum or maximum. Equation 10 is known as the likelihood equation (Myung, 2003). An additional condition that must also be satisfied is that the second derivative of $-\log L(x|p)$ is negative:

$$\frac{\partial^2[-\log L(x|p)]}{\partial^2 p} < 0 \qquad \text{Equation 11}$$

This ensures that the neighbourhood of $\mathbf{p}_{min}$ in the parameter space is concave, corresponding to a trough of the test statistic function $-\log L$, and not convex which would correspond to a peak. The fact that the (negative) logarithm of L behaves smoothly and continuously in such regions facilitates the search for $\mathbf{p}_{min}$, and is a basic reason for choosing it as the quantity to be minimised. An additional reason is related to the joint PDF of the parameters (given the data) which assuming statistically independent observations (Bevan, 2013), is given by the product of the PDFs of the individual observations, as in the expression

$$L(x|p) = L_1(x_1|p) \cdot L_2(x_2|p) \cdot \ldots \cdot L_N(x_N|p) \qquad \text{Equation 12}$$

for *N* observations (data points). By taking the logarithm of L this product becomes a sum which is also easier to handle computationally.

Numerical example

Practical application of the above can be demonstrated by the following example, taken from a review (Myung, 2003) in the field of psychology. The aim was to compare two different models, an exponential and a power model, tested in an experiment of 'forgetting data' where the proportion of correct answers recalled as a function of time is recorded, for a cohort of subjects. The two models are defined as

power model: $S(p,t) = p_1 \cdot t^{-p_2}\ (p_1, p_2 > 0)$ Equation 13a

exponential model: $S(p,t) = p_1 \cdot e^{-p_2 \cdot t}\ (p_1, p_2 > 0)$ Equation 13b

where *S* are the data consisting of the observed proportion of correctly recalled answers at time *t*, p_1 and p_2 the parameter vector **p** in each model, and the data will be in the form $\mathbf{x}=x_1, x_2, \ldots x_N$ for *N* recordings. Application of MLE requires specification of the PDF governing the data for each model. Each proportion of correctly recalled answers x_i recorded is the ratio between the number of correct responses y_i and the total number of trials *n*, $x_i = y_i/n$ $(0 \leq x_i \leq 1)$, and follows a binomial distribution $P(y|n,x) = \frac{n!}{y!(n-y)!}x^y(1-x)^{n-y}$ where $0 \leq x \leq 1$ and $y=0, 1, \ldots, n$. Therefore the PDF corresponding to each model will be

power model: $P(y_i|n,p) = \frac{n!}{y!(n-y)!}(p_1 \cdot t^{-p_2})^{y_i}(1-p_1 \cdot t^{-p_2})^{n-y_i}$ Equation 14a

exp. model: $P(y_i|n,p) = \frac{n!}{y!(n-y)!}(p_1 \cdot e^{-p_2 \cdot t})^{y_i}(1-p_1 \cdot e^{-p_2 \cdot t})^{n-y_i}$ Equation 14b

according to how each observation x_i is modelled via Equations 13, where $y_i=0, 1, \ldots, n$ and $i=0, 1, \ldots, N$. Since x_i is related to y_i by a scaling constant $1/n$ any statistical inference regarding y_i applies also to x_i, taking into consideration this scale transformation, by substituting y_i in $P(y_i|n,p)$ with nx_i. By combining Equation 12 with Equations 14 the desired log-likelihood expression can be produced. For the case of the power model it will be

$$-\log L(y_i|n,p) = -\log[P(y_1|n,p) \cdot P(y_2|n,p) \cdot \ldots \cdot P(y_N|n,p)]$$

$$= -\sum_{i=1}^{N} \log P(y_i|n,p) = -\sum_{i=1}^{N}[y_i \log(p_1 \cdot t^{-p_2}) + (n-y_i)\log(1-p_1 \cdot t^{-p_2}) + \log n! - \log(n-y_i)! - \log y_i!].$$

Minimising this quantity with respect to the parameters p_1 and p_2 in this case is not affected by the last three terms as they are not parameter-dependent and therefore can be omitted. The expressions to be minimised corresponding to the two models therefore are

$$-\log L(y_i|n,p) = -\sum_{i=1}^{N}[y_i \log(p_1 \cdot t^{-p_2}) + (n - y_i)\log(1 - p_1 \cdot t^{-p_2})] \text{ and}$$

$$-\log L(y_i|n,p) = -\sum_{i=1}^{N}[y_i \log(p_1 \cdot e^{-p_2 \cdot t}) + (n - y_i)\log(1 - p_1 \cdot e^{-p_2 \cdot t})].$$

Minimising such functions is a common requirement in many fields of research and tools in various programming languages exist, including Matlab as in this paradigm – and also as in the work presented throughout this thesis. A constrained minimisation was performed with inputs each of the above expressions as the objective function, random start points and bounds according to the above constraints (e.g. $0 \leq x_i \leq 1$), and default optimisation options. In the particular experiment correct recalls were recorded for six time intervals t_i = 1, 3, 6, 9, 12 and 18 s (N=6) and were found to be y_i = 94, 77, 40, 26, 24, 16 out of n=100 repetitions. The yielded $-\log L$ was 313.37 and 305.31 for the power and the exponential model respectively, indicating that the latter fits better the data (Figure 2.32).

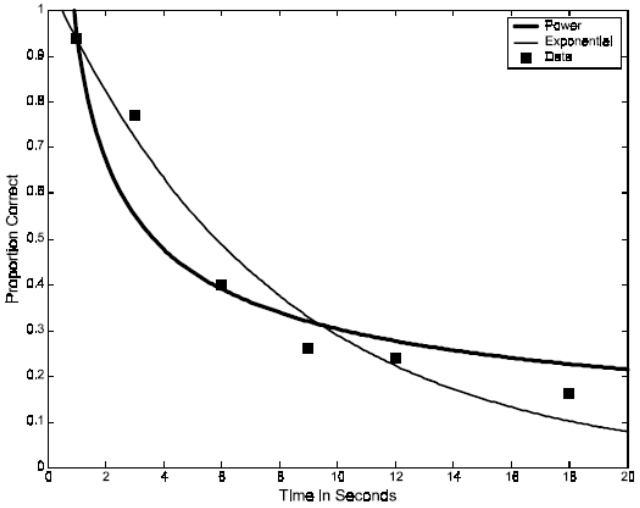


Figure 2.32: The fitting of the proportion correct observations with the use of a power (thick curve) and an exponential (thin curve) model by the MLE parameters yielded for binomially distributed data (Myung, 2003).

Application to T_2-decay models

The above MLE based fitting is directly applicable to signals describing transverse magnetization decay as is the scope of this work. Objective functions under MLE in general incorporate a model and a PDF as shown in Figure 2.33.

```
function loglik = mle ( p, x, t)
                                p: parameter vector, x: data vector, t: independent variable
  M = f( p, t);                 define the objective function, the model M applying to the given experiment (e.g. exponential)
  loglikvector = -log [ P( x| M, t) ];   define the PDF P applying to the problem and take its –log; will be a vector equal in size to the
                                data vector (comparing the modelled data M to the observations x)
  loglik = sum( loglikvector);  estimated log-likelihood (test statistic) as a scalar
end
```

Figure 2.33: The generic structure (‘pseudo-code’) of the computer function (code) computing the log-likelihood, to be minimised in the MLE process. The parameters (‘p’) and the model describing the given experiment (‘M’) are passed as arguments to this function via the PDF (‘P’) applying to the particular problem.

To adapt the MLE method to the estimation of transverse magnetisation decay model parameters, as in the present work, candidate physical models could include the exponential, EPG or sEPG functions introduced in previous sections, and candidate PDFs the Rician or Gaussian distribution. For example if fitting CPMG data with an sEPG model where the underlying noise is assumed to be Rician distributed, the quantity to be minimised under MLE would take the form

$$-\log L =$$

$$\sum_{echoes}\left\{-\log\left[\frac{M(echoes)}{\sigma^2}\,e^{-\frac{M(echoes)^2+sEPG(p,echoes)^2}{2\sigma^2}}I_0\left(\frac{sEPG(p,echoes)\cdot M(echoes)}{\sigma^2}\right)\right]\right\},$$

where *M(echoes)* is the signal measured, *sEPG(p,echoes)* the signal modelled by the sEPG model, *p* the parameters to be estimated, *'echoes'* represents echoes as a vector of integers (1, 2, …, echo-train length; echo spacing is one of the fixed model parameters) and *σ* the underlying noise SD in each acquisition channel. The above expression is an application of the general expression of the Rician distribution (Equation 7) to the particular model; this will thus be the objective function taken as an argument by the minimisation function (i.e. Matlab 'fmincon') to iteratively determine the MLE estimated parameters *p*. In that case a more specific version of the pseudocode corresponding to the pseudocode to that of Figure 2.33 would be:

```
function logliksepg = sepg_mle (tex, tref, rfex, rfref, p, s, echoes)

        s0 = sepg (tex, tref, rfex, rfref, p(1:end-1), echoes);

        logliksepgvec = - log(s) + 2·log(p(end)) + (s0.^2+s.^2)./(2·p(end)^2)

                        - abs(real(s0.·s./p(end)^2)) - log(besseli(0,s0.·s./p(end)^2,1));

        logliksepg = sum(logliksepgvec);

end;
```

where logliksepg is the resulting $-\log L$ (a scalar), tex and rfex the time and amplitude vector of the excitation RF pulse, tref and rfref the time and corresponding amplitude vector of the refocusing RF pulse, s the measured signal along the echo train (echoes), s0 the signal modelled by the sEPG model and logliksepgvec the vector of partial log-likelihoods per echo. The parameters (*p*) are the B_1 field factor ($B_1 f$), the echo spacing (*TE*), the T_2 and T_1 decay constants, signal amplitude α (intercept) and σ. A period (dot) after a symbol (e.g. 's0.') denotes vector operations. *TE* is usually known from the acquisition (sequence) settings and T_1 is commonly fixed (Chapter 3, section 3.2.4) therefore the parameters to be estimated in *p* are $B_1 f$, T_2, α and σ.

Implementation considerations

There are a few additional points to be made regarding this particular implementation of the minimisation algorithm. First, it is necessary in practice to distinguish the time and amplitude vectors for each RF pulse whereas previously

(sections 2.1.7 and 2.1.8) the refocusing pulse was assumed to have exactly double amplitude and half the duration (does not include rephasing) of the excitation pulse (Figure 2.27). This change reflects the way the CPMG MESE sequence is practically implemented on the system used (Siemens MAGNETOM Prisma) where the RF excitation and refocusing pulse, with their associated gradient pulses are individually programmed in terms of shape and amplitude. This proprietary pulse shape information may be provided by the MRI system vendor in the context of a collaboration agreement with a given institution. Second, σ is under this approach one of the parameters to be estimated, rather than assumed or measured *a priori* (Chapter 3, section 3.2.5); σ appears as the last end element of p 'p(end)' in the pseudo code above, and is treated separately from the remaining estimated parameters as it appears only in the Rician PDF expression and not in the sEPG model. Its role is to account for the rectified background noise floor (sections 2.1.5 and 2.1.8), while the sEPG model accounts for the magnetization behaviour including the slice profile-corrected stimulated and alternate echoes coherences (sections 2.1.6 and 2.1.7). Third, after taking logarithms for the Rician PDF there appears to be an extra term before the logarithm of the Bessel function (the last but one term). The reason for this is that 'besseli', the default Matlab function for I_0 (the modified Bessel function of the 0^{th} order) tends to return infinite results for large values of its argument, namely the product of the measured and modelled signal over the noise SD, due to numerical overload. This can be avoided by calling a normalised version of the function which is also available (MathWorks, Massachusetts, United States), necessitating use of the exponential of the negative argument within the function, reversed by taking the exponential of the positive argument outside (before the

function is executed), which leads to the additional term in the log-likelihood expression. The above issue is a consequence of the algebraic construction of the Rician PDF, which is a combination of an increasing with signal Bessel-function term and an exponentially decreasing term (Rice, 1944).

Common generic issues with iterative minimisation methods are that they do not always converge to a minimum and that a global minimum is not always guaranteed. Approaches to address these problems include for the former to empirically experiment with the start points and step sizes, and for the latter to perform a parameter scan by repeating the process for a grid of parameter start points (Bevan, 2013). Both these techniques were used in the experimental work presented here as will be seen in particular later sections.

2.3 Review of literature on quantitative T_2 MRI

Table 2.1 below is a comprehensive list of published work related to quantitative T_2 mapping for the most part pertaining to musculoskeletal disease and imaging, although directly relevant papers from other related fields which are referenced later in this thesis are included. The literature reflects the history that the simplest and historically commonly used T_2 estimation method with a mono-exponential model and least-squares fitting has over time been challenged and replaced by improved alternatives. These have included multi-component analyses, physical models more advanced than a simple exponential decay, and more accurate statistical descriptions.

Out of 53 papers included here, 22 pertain to total or global T_2 mapping (section 2.1.4), i.e. estimating T_2 with a mono-exponential decay model (Adams et al., 1992, Poon et al., 1992, Jehenson et al., 1993, Huang et al., 1994, Phoenix et al., 1996, Hatakenaka et al., 2001, Livingston et al., 2001, Phillips et al., 2001, Dardzinski et al., 2002, Maillard et al., 2004, Nygren, 2006, Sesto et al., 2008, Shafer-Crane et al., 2005, Kim et al., 2010a, Gloor et al., 2011, Friedman et al., 2012, Arpan et al., 2013, Saleh Elessawy et al., 2013, Arpan et al., 2014, Kim et al., 2010b, Forbes et al., 2014, Bryan et al., 1998). One paper was based on T_2 weighted signal intensity comparisons (Jenkins et al., 2018).

Thirteen of the papers refer to multi-component models of T_2 relaxation behaviour which take into account multi-compartment tissue structure at the pixel level (water, fat, oedema etc.) (Kamman et al., 1987, Cole et al., 1993, Saab et al., 1999, Gambarota et al., 2001, Kan et al., 2009, Prasloski et al., 2012, Yao and Gai, 2012, Araujo et al., 2014, Azzabou et al., 2015, Wary et al., 2015, Marty et al., 2016, Mankodi et al., 2017, Keene et al., 2020). In most of these reports multi-component exponential models were used, except one case where a 2-component EPG model was used (Marty et al., 2016). One of the papers was particularly focused on the effect of monitoring treatment by T_2 obtained quantitatively in diseased tissue (Arpan et al., 2014). However the implications for treatment monitoring were also part of the discussion in other papers (Wary et al., 2015). Two papers involved a comparison between T_2 relaxometry mapping and MR spectroscopy (Kim et al., 2014, Kim et al., 2015). Finally there were seventeen papers of generic content, addressing common shortcomings of the method and practical issues (Majumdar et al., 1986, Gold et al., 2004, Jordan

et al., 2013) and novel analysis methods (Sijbers et al., 1998b, Hardy and Andersen, 2009, Walker-Samuel et al., 2009, Lebel and Wilman, 2010, Prasloski et al., 2012, Neumann et al., 2014, Ben-Eliezer et al., 2015, Petrovic et al., 2015, Milford et al., 2015, McPhee and Wilman, 2017, Milford et al., 2018, Lankford and Does, 2018, Basiri et al., 2019, Poot et al., 2013). These papers will be referenced at various points in the work to follow.

Table 2.1: Quantitative T_2 literature overview

Guide (themes):	*Mono-exponential T_2 behaviour with disease*	*Multi-exponential T_2 decay analysis*	*Quantitative T_2 mapping vs MR spectroscopy*	*(Mono-exponential) T_2 behaviour with treatment*	*Generic / method & modality dependent / artefacts*
Abbreviations: *CPMG: Carr-Purcell-Meiboom-Gill sequence* *LSQ: Least squares* *NNLS: non-negative least squares* *EPG: extended phase graph (model)* *FF: fat fraction* *SLR: Shinar-Le Roux method* *SE: spin echo* *MLE: Maximum likelihood estimation* *IR: inversion recovery* *sEPG: slice profile corrected EPG* *Bloch: Bloch equations* *FT: Fourier transform*					
Title	**Author / Reference / Year**	**Body part**	**Acquisition method / sequence**	**Analysis method**	**Summary / Results**
Errors in the Measurements of T_2 Using Multiple-Echo MRI Techniques I. Effects of Radiofrequency Pulse Imperfections	(Majumdar et al., 1986)	general / phantom	Quantitative T_2	Simulation and experiment based	Signal imperfections due to RF pulsing
Multi-exponential relaxation analysis with MR imaging and NMR spectroscopy using fat-water systems	(Kamman et al., 1987)	general / phantom	CPMG – fat/water phantom	Multi-exponential modelling	Multi-compartment exponential behaviour of tissue in T_2 estimation
Magnetic resonance imaging and electromyography as indexes of muscle function	(Adams et al., 1992)	forearm	Multi echo SE	Signa Software	Increased T_2 with exercise
Quantitative Magnetic-Resonance-Imaging Parameters and Their Relationship to Mammographic Pattern	(Poon et al., 1992)	breast	Multi echo SE	Three-parameter Levenberg-Marquardt, Sun workstation	T_2 histograms more informative than nominal T_2
The origin of biexponential T_2 relaxation in muscle water	(Cole et al., 1993)	animal study	Variable τ single spin echo	Nonlinear least squares / Macintosh	Bi-exponential T_2 decay of muscle represents anatomical compartmentation of tissue water
MR imaging as a potential diagnostic test for metabolic myopathies: importance of variations in the T_2 of muscle with exercise	(Jehenson et al., 1993)	forearm	Double echo SE	Intensity ratio calculations	Correlation of T_2 and end-exercise pH in McArdle's disease

Quantitative MR relaxometry study of muscle composition and function in Duchenne muscular dystrophy	(Huang et al., 1994)	calf	Double echo SE	Intensity ratio logarithms, SUN Sparc workstation	Increased T_2 / non uniform distribution histogram
Objective quantification of muscle and fat in human dystrophic muscle by magnetic resonance image analysis	(Phoenix et al., 1996)	calf	Multi echo SE	log-linear least squares, SUN Sparc workstation	Prolonged T_2, wider and inhomogeneous distribution
Magnetic resonance imaging of muscle in amyotrophic lateral sclerosis	(Bryan et al., 1998)	calf	Multi echo SE	Mono-exponential fitting	Strong negative correlation between muscle T_2 and physiological tests
Optimal estimation of T_2 maps from magnitude MR images	(Sijbers et al., 1998b)	simulations/ brain	6 echo SE	MLE (vs LSQ) / mono-exp.	Superior accuracy with MLE compared to LSQ
Multicomponent T_2 relaxation of in vivo skeletal muscle	(Saab et al., 1999)	forearm	Volume localized CPMG	Non-negative least squares	Multi component decay of the cellular water organization
Osmotic effects on the T_2 relaxation decay of in vivo muscle	(Gambarota et al., 2001)	animal study	CPMG	NNLS, Matlab / ProFit	Intra/extra-cellular water quantification by bi-exponential T_2 fitting
Effects of Aging on Muscle T_2 Relaxation Time Difference Between Fast- and Slow-Twitch Muscles	(Hatakenaka et al., 2001)	calf	Multi echo SE / IR fat suppression	Mono-exponential least squares fitting	T_2 increases in fast twitch muscles with age
Functional activation of the extensor carpi radialis muscles in humans	(Livingston et al., 2001)	forearm	Multi echo SE	Mono-exponential fitting / Scion Image software	T_2 monitors differential forearm muscle activation
Patterns of muscle involvement in inclusion body myositis: clinical and magnetic resonance imaging study	(Phillips et al., 2001)	thigh / calf / forearm	Multi echo SE	T_2 relaxometry	T_2 findings of preferential muscle involvement
Mapping T_2 relaxation time in the pediatric knee: feasibility with a clinical 1.5-T MR imaging system	(Dardzinski et al., 2002)	knee	Multi echo SE	Mono-exponential linear least squares / IDL	Spatial variation and increase in T_2
Musculoskeletal MRI at 3.0T: Relaxation times and image contrast	(Gold et al., 2004)	knee	Multi echo SE	Mono-exponential T_2 relaxometry	Decreased T_2 from 1.5T to 3T
Quantitative assessment of MRI T_2 relaxation time of thigh muscles in juvenile dermatomyositis	(Maillard et al., 2004)	thigh	CPMG	T_2 relaxometry	T_2 measures inflammation and correlates with functional tests
Shortened T_2 after exercise in ischemic skeletal muscle	(Nygren, 2006)	calf	Multi echo SE	Mono-exponential least squares fitting	Ischemia detection
Mechanical and magnetic resonance imaging changes following eccentric or concentric exertions	(Sesto et al., 2008)	forearm	3 echoes SE	T_2 relaxometry / MR Vision	Increase in T_2 of forearm muscles with exercise
Effect of occupational typing on MRI of forearm muscles in subjects with and without symptoms of carpal tunnel syndrome	(Shafer-Crane et al., 2005)	forearm	Double echo SE	Intensity ratio logarithms	Increase of T_2 in forearm flexors with typing

Calculating T_2 in images from a phased array receiver	(Hardy and Andersen, 2009)	simulations/ phantom	Multi-echo SE	log-linear, nonlinear LSQ, MLE fitting	MLE yields the most accurate T_2 in the presence of phased array coil bias
Quantitative MR imaging of individual muscle involvement in facioscapulohumeral muscular dystrophy	(Kan et al., 2009)	calf	Multi-echo SE	Bi-exponential T_2 mapping / IDL	Quantitative assessment of fatty infiltration
Robust estimation of the apparent diffusion coefficient (ADC) in heterogeneous solid tumours	(Walker-Samuel et al., 2009)	simulations / mice prostate	Di*ff*usion sequence & simulation	LSQ / MLE fitting	MLE recommended over LSQ for robust ADC estimation
T_2 mapping in Duchenne muscular dystrophy: distribution of disease activity and correlation with clinical assessments	(Kim et al., 2010b)	pelvis and thighs	Multi-echo SE	Linear least squares curve fitting	Elevated T_2 with disease
T_2 relaxation time changes in distal femoral articular cartilage in children with juvenile idiopathic arthritis: a 3-year longitudinal study	(Kim et al., 2010a)	knee	Multi-echo SE	AUC of the T_2 relaxation time profile	Increase of T_2 with distance
Transverse relaxometry with stimulated echo compensation	(Lebel and Wilman, 2010)	simulations/ phantom/ brain	Multi-echo SE	FT/sEPG nonlinear least-squares fitting	accurate transverse relaxometry in heterogeneous transmit fields
Quantification of fat infiltration in oculopharyngeal muscular dystrophy: comparison of three MR imaging methods	(Gloor et al., 2011)	thigh and calf	Multi-echo TSE	nonlinear least-squares fitting	Elevated T_2 with disease
Musculoskeletal MRI at 3.0 T and 7.0 T: A comparison of relaxation times and image contrast	(Jordan et al., 2013)	knee	Multi-echo SE	nonlinear least squares regression / Matlab	Decreased T_2 from 3T to 7T
The magnetic resonance imaging spectrum of facioscapulohumeral muscular dystrophy	(Friedman et al., 2012)	knee	32 echo SE	NNLS / Matlab	T_2 increase due to oedema
Applications of Stimulated Echo Correction to Multicomponent T_2 Analysis	(Prasloski et al., 2012)	simulations / brain	3D multi-echo SE	EPG simulated database / NNLS	Stimulated Echo Correction improves Multicomponent T_2 Analysis
Fat-Corrected T_2 Measurement as a Marker of Active Muscle Disease in Inflammatory Myopathy	(Yao and Gai, 2012)	thigh	Multi-echo SE	Bi-exponential nonlinear least-squares fitting	Fat-corrected T_2 values lower and more widely distributed
T_2 mapping provides multiple approaches for the characterization of muscle involvement in neuromuscular diseases: a cross-sectional study of lower leg muscles in 5-15-year-old boys with Duchenne muscular dystrophy	(Arpan et al., 2013)	calf	Multi-echo SE	Mono-exponential fitting / IDL	Increased heterogeneity in T_2 histograms for DMD patients
Bias correction of maximum likelihood estimation in quantitative MRI	(Poot et al., 2013)	simulations	Exponential Multi-echo plus σ (noise)	bias-corr. MLE, MLE, LSQ, nonlinear LSQ	MLE bias can be corrected

The role of MRI in the evaluation of muscle diseases	(Saleh Elessawy et al., 2013)	pelvis, thigh and calf	Multi-echo SE	T_2 relaxometry	Increased T_2 with fat infiltration
New Insights on Human Skeletal Muscle Tissue Compartments Revealed by In Vivo T_2 NMR Relaxometry	(Araujo et al., 2014)	calf (soleus muscle)	Volume localized (ISIS) CPMG	NNLS / Matlab	biexponential T_2 relaxation models reflecting anatomical compartmentation of myowater
Examination of effects of corticosteroids on skeletal muscles of boys with DMD using MRI and MRS	(Arpan et al., 2014)	thigh and calf	Multi-echo SE / MRS	T_2 relaxometry / spectroscopy	corticosteroids reduce inflammation in DMD
Validation of a generic approach to muscle water T_2 determination at 3T in fat-infiltrated skeletal muscle	(Azzabou et al., 2015)	thigh	17 echo SE	Tri-exponential model, nonlinear least squares	Quantification of water T_2 changes independently of fat infiltration
Magnetic Resonance Imaging and Spectroscopy Assessment of Lower Extremity Skeletal Muscles in Boys with Duchenne Muscular Dystrophy: A Multicentre Cross-Sectional Study	(Forbes et al., 2014)	thigh and calf	Multi-echo SE	Mono-exponential fitting / IDL	T_2 increases with DMD disease and with age
Objective measurement of minimal fat in normal skeletal muscles of healthy children using T_2 relaxation time mapping (T_2 maps) and MR spectroscopy	(Kim et al., 2014)	pelvis	Multi-echo SE / MR spectroscopy	Mono-exponential linear least squares	Elevated T_2 due to fat infiltration correlates with spectroscopy acquired fat fraction
Simple recipe for accurate T_2 quantification with multi spin-echo acquisitions	(Neumann et al., 2014)	simulations / phantom / brain	Multi-echo SE	Mono-exponential fitting / EPG	Stimulated echoes correction in mono-exponential fitting
Longitudinal measurements of MRI-T_2 in boys with Duchenne muscular dystrophy: Effects of age and disease progression	(Arpan et al., 2014)	calf	Multi-echo SE	Mono-exponential fitting / IDL	T_2 and its histogram width increases with time in DMD patients
Rapid and Accurate T_2 Mapping from Multi–Spin-Echo Data Using Bloch-Simulation-Based Reconstruction	(Ben-Eliezer et al., 2015)	simulations / brain / prostate	Multi-echo SE, single SE	Simulated sEPG signals database comparison (L2 norm)	Bloch simulated sEPG model improves T_2 mapping accuracy
Quantitative Skeletal Muscle MRI: Part 2, MR Spectroscopy and T_2 Relaxation Time Mapping-Comparison Between Boys with Duchenne Muscular Dystrophy and Healthy Boys	(Kim et al., 2015)	pelvis and thigh	Multi-echo SE / MR spectroscopy	Mono-exponential nonlinear least squares fitting / ISS	T_2 mapping more accurate in separating patients and normals based on fat infiltration
Closed-Form Solution for T_2 Mapping with Nonideal Refocusing of Slice Selective CPMG Sequences	(Petrovic et al., 2015)	simulations / phantom / brain	Multi-echo SE	Generating functions (GF) / nonlinear least squares	A solution for stimulated echoes correction in order to to compute comparable T_2 values
Quantitative NMRI and NMRS identify augmented disease progression after loss of ambulation in forearms of boys with Duchenne muscular dystrophy	(Wary et al., 2015)	forearm	Multi-echo SE	Tri-exponential model, nonlinear least squares	inflammation/oedema and fat infiltration progress in opposite directions with disease evolution

Mono-Exponential Fitting in T_2-Relaxometry: Relevance of Offset and First Echo	(Milford et al., 2015)	phantom	Simulations / Multi-echo SE	Mono-exponential fitting / EPG	Effect of the refocusing pulse, echo spacing, echo train length and offset on T_2 error
Simultaneous muscle water T_2 and fat fraction mapping using transverse relaxometry with stimulated echo compensation	(Marty et al., 2016)	thigh	Multi-echo SE / Dixon *FF*	NNLS / dictionary based sEPG	Sensitive and fast simultaneous muscle water T_2 and fat fraction mapping
Skeletal muscle water T_2 as a biomarker of disease status and exercise effects in patients with Duchenne muscular dystrophy	(Mankodi et al., 2017)	lower leg	Multi-echo SE	Tri-exponential model, nonlinear least squares	Tri-exponential method muscle water T_2 sensitive to the underlying disease status in DMD
Transverse Relaxation and Flip Angle Mapping: Evaluation of Simultaneous and Independent Methods Using Multiple Spin Echoes	(McPhee and Wilman, 2017)	simulations / phantom / brain	Multi-echo SE	Bloch / SLR / FT -sEPG and nonlinear least-squares	Bloch-sEPG fitting improves T_2 estimation but is limited by flip angle redundancy
Imaging muscle as a potential biomarker of denervation in motor neuron disease	(Jenkins et al., 2018)	Whole body	T_2 weighted fast spin echo	Signal intensity comparisons	Higher relative T_2 signal associated with physiological findings
A novel method for T_2 quantification in presence of B_1 inhomogeneities	(Milford et al., 2018)	simulations / phantom	Multi-echo SE	sEPG – custom minimisation technique	Improved T_2, proton density and flip angle estimation
Propagation of error from parameter constraints in quantitative MRI: Example application of multiple spin echo T_2 mapping	(Lankford and Does, 2018)	simulations	Multi-echo SE	EPG algorithm	Flip-angle measurement and constraint could be beneficial to T_2 accuracy and precision in specific T_2 mapping applications
Transverse relaxometry with transmit field-constrained stimulated echo compensation	(Basiri et al., 2019)	simulations / *in vivo* (brain)	Multi-echo SE	sEPG - nonlinear least-squares fitting	More reliable and reproducible quantitative T_2 maps
T_2 relaxation-time mapping in healthy and diseased skeletal muscle using extended phase graph algorithms	(Keene et al., 2020)	simulations / *in vivo*	Multi-echo SE	NNLS / dictionary based sEPG	Recommended sEPG slice flip-angle profile correction for through-plane chemical-shift displacements
Guide (themes):	*Mono-exponential T_2 behaviour with disease*	*Multi-exponential T_2 decay analysis*	*Quantitative T_2 mapping vs MR spectroscopy*	*(Mono-exponential) T_2 behaviour with treatment*	*Generic / method & modality dependent / artefacts*

Abbreviations:

CPMG: Carr-Purcell-Meiboom-Gill sequence
LSQ: Least squares
NNLS: non-negative least squares
EPG: extended phase graph (model)
FF: fat fraction
SLR: Shinar-Le Roux method
SE: spin echo
MLE: Maximum likelihood estimation
IR: inversion recovery
sEPG: slice profile corrected EPG
Bloch: Bloch equations
FT: Fourier transform

Chapter 3. Practical implementation, optimisation and validation of T_2 estimation algorithms

The work presented in this chapter aimed to develop, implement and characterise robust methods to estimate muscle-water T_2 values from multi-echo imaging data, using computer simulated data, with validation and practical performance assessment in exemplar test object and *in vivo* data. In this work it was necessary to address the challenge of developing tools which could reliably estimate T_2-related parameters from real-world MRI data which in general may have not been acquired under ideal experimental conditions. In subsequent chapters the methods' performance is further tested by application in two different neuromuscular disease natural history studies.

3.1 Noise and SNR measurements

In realistic experimental conditions the presence of noise combined with low signal intensities, i.e. low signal-to-noise ratios (SNRs) may compromise the effectiveness of our analysis methods. Before proceeding it is therefore instructive to obtain a quantitative indication of the SNR available in typical MESE datasets.

There are a number of alternative methods for quantifying the noise levels in MRI images: one of the established techniques is the signal-background method, where we delineate regions of interest (ROIs) avoiding motion and filter roll-off artefacts, usually square ROIs near the image edges as shown in Figure 3.1 and then calculate the mean signal intensity of the included pixels (McRobbie et al., 2006).

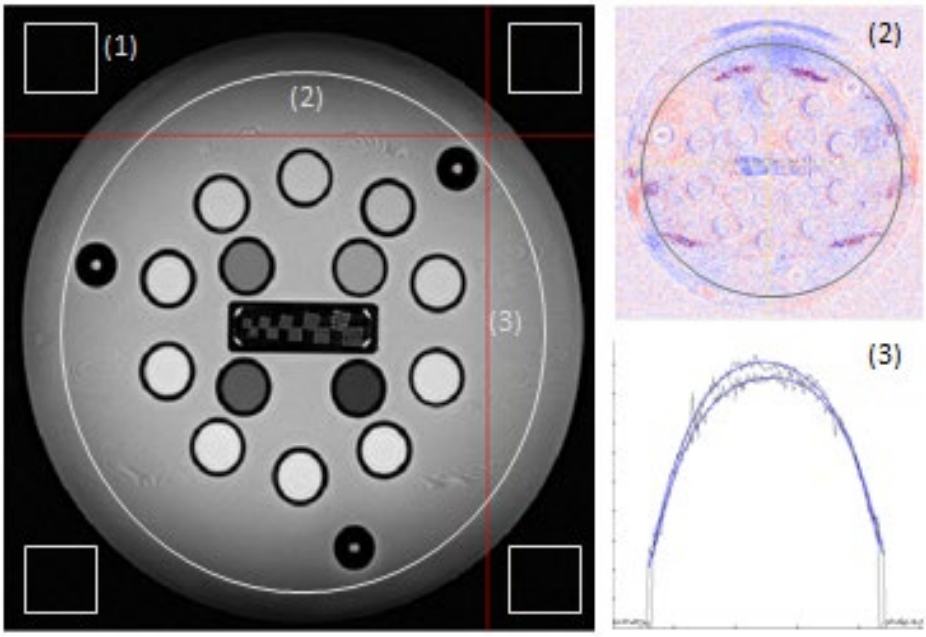


Figure 3.1: Measuring noise or SNR from phantom image data. With the signal-background method the noise will be the corrected mean (see text) of the signal in uniform square background areas (1). With the NEMA method the SNR will be the corrected ratio of the signal in a large ROI of the phantom (area within the white circle) to the noise SD from the subtraction of two identical images from that ROI (2). A third method is to fit a curve on the signal from smoothly varying image profiles, so that the standard deviation of the residual following subtraction of this curve from the image line profile will give an estimate for the image noise (3). The ground truth T_2 values (per manufacturer) of the regions of interest (spheres) used during the relaxometry experiments presented in later sections were between 32 and 133ms (details in section 3.2.10), to be comparable with *in vivo* datasets.

The underlying σ_N (the Gaussian generating-distribution SD in the Rician noise PDF) is recovered by dividing the mean magnitude image background signal intensity by a factor of 1.25 which comes from the analytical expression for the mean of a Rayleigh distribution (Equations 9, section 2.1.9), which is the special case of the Rician distribution in areas with no signal (Gudbjartsson and Patz, 1995).

As an example, CPMG imaging data were obtained from an ISMRM/NIST System Phantom (High Precision Devices, Inc. 1668 Valtec Lane Suite C, Boulder, CO, 80301) examined at 3T (Siemens Magnetom Prisma) using a transmit/receive ('TxRx') head coil with a multi-echo spin-echo sequence (TR= 3000ms, 17 TEs from 9.9-168.3ms

with 9.9ms interval, a single 2 mm slice at the T_2 spheres level, matrix 210x210, in-plane resolution 0.8x0.8 mm). The mean value yielded from several background ROIs (squares in Figure 3.1) in a test-image was to 2 significant figures 36 greyscale units translating to an image noise σ_N of 29 units. Mean signal intensities for ROIs within the test object body, e.g. within the T_2 spheres, were typically above 1800 units for the highest intensity echo image (the 2nd in this case), yielding an SNR exceeding 50 with this noise estimation approach. Commonly – and also throughout this work – SNR for multi-echo signals is defined as the highest magnitude echo intensity to noise σ_N ratio, except where otherwise stated.

In a similar method proposed by the National Electrical Manufacturers Association (NEMA) the mean signal from a large uniform ROI from within the phantom is divided by the SD from the same ROI in a subtraction of two identically acquired images as the noise estimate, multiplied by $\sqrt{2}$ to correct for the noise propagation in the subtraction of two images, to provide an SNR estimate (area within the circle marked '(2)' in the left main panel of Figure 3.1 and the same area in the top right panel for the subtraction image) (NEMA, 2015). These calculations produced an SNR of 40 for this example, with a noise σ_N estimate of 36 greyscale units in absolute value, a factor of 1.25 higher than the previous signal-background method estimate. Plausible reasons for this difference in the noise value might include scanner instabilities, vibration, temperature variations or fluid flow within the test object.

A third practical method to estimate the noise σ_N is by fitting a smoothly varying image profile (line) from relatively uniform areas with a suitable function (typically a polynomial) and find the SD of the subtraction of the true signal from the signal

modelled by the function (Figure 3.1, bottom right). The mean estimated noise σ_N from such calculations was 40 greyscale units, translating to a similar SNR estimate with the previous two methods.

The last two methods could not be applied to *in vivo* images from the patient-study datasets available - the second due to the lack of repeat-acquisition images and the third due to difficulty in modelling the underlying true signal variation due to the dominance of anatomical structures with rapidly varying spatial features (e.g. nerves, fascia etc.). However application of the first method to representative images from the dataset analysed in chapter 4, showed noise levels of 3 to 5 units for the background noise σ_N, where the mean signal intensity for most anatomical ROIs was roughly 250 to 300 units (SNR range from 35 to 100). Correspondingly, images from the dataset analysed in chapter 5 produced an estimate for the noise σ_N between 20 and 30 units while intensities in ROIs within the imaged anatomy were seen to range between 1000 and 2000 units (SNR range from 50 to 100) (Figure 3.2).

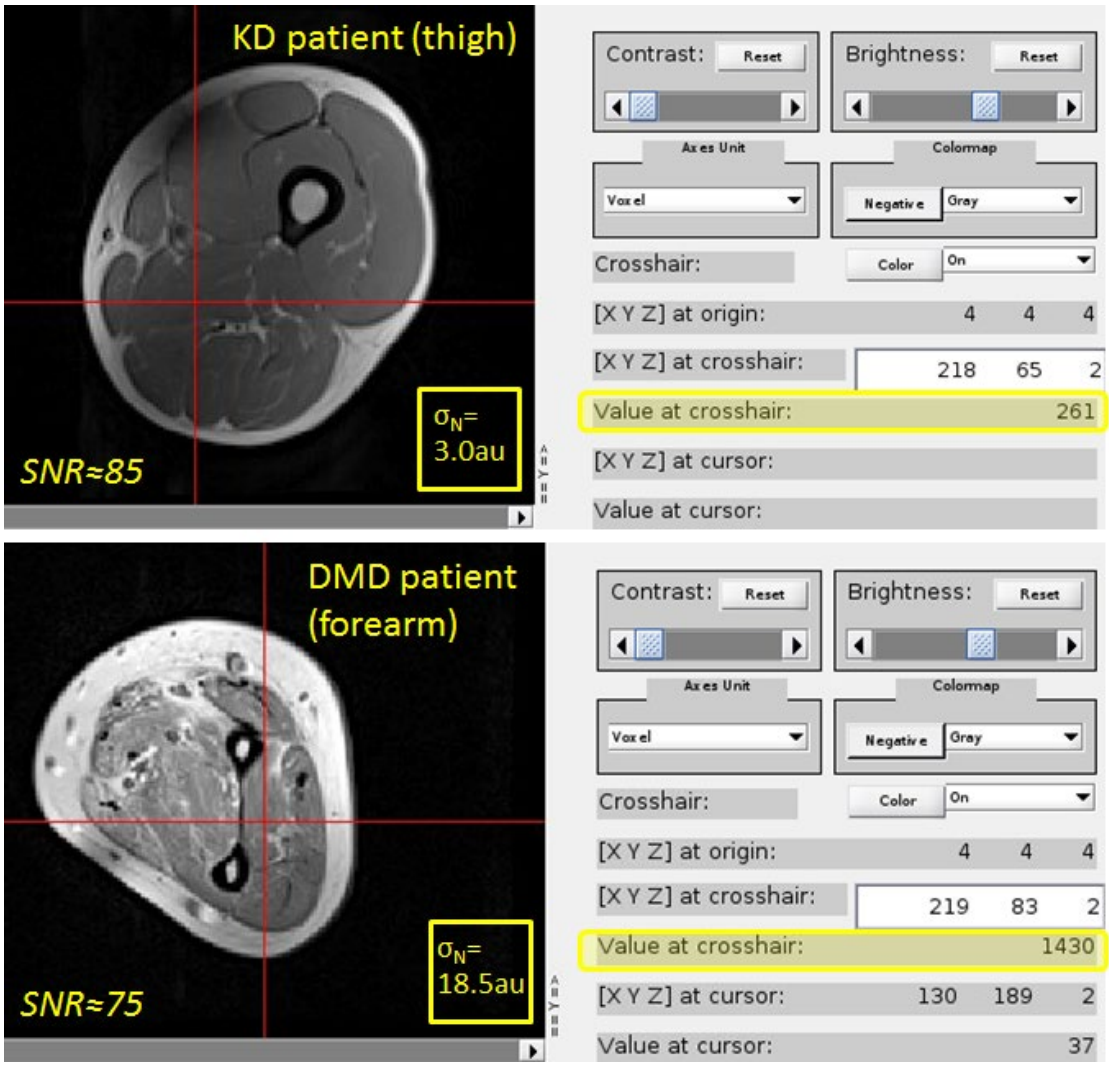


Figure 3.2: Applying the signal-background method (McRobbie et al., 2006) to estimate the SNR in representative images from patient datasets (acquisition characteristics described in sections 4.2 and 5.2).

Therefore SNR in *in vivo* images analysed here lay generally between 35 and 100, broadly consistent with values from the phantom images.

3.2 Numerical Simulations

In order to test the effectiveness of T_2 and other transverse magnetization relaxation parameters estimation across different signal models and fitting methods, we adopted the approach of numerical simulations where the ground truth parameter

values are defined *a priori*. Monte Carlo simulations is a technique used widely to characterise methods for parameter estimation in quantitative MRI (Hardy and Andersen, 2009, Walker-Samuel et al., 2009, Lebel and Wilman, 2010, McPhee and Wilman, 2017). A large number of replicates of the modelled signal are generated, each with independently generated random noise added conforming to the known underlying statistical distribution. The model in question is then fitted to each replicate, the combined results from these allowing a statistical assessment of the precision and accuracy (bias) of the parameter estimates. In the context of this work the modelled signal reflects the spin behaviour in multi-pulse MRI sequences and the aim is to investigate the accuracy of estimated T_2 and other parameters, their associated uncertainties, and the impact of image noise levels on their estimation.

3.2.1 Number of repetitions

In order to obtain reliable simulation experiments it is important to determine the number of replicates required to adequately characterise the fitting process in each case. Typically powers of 10 are used (e.g. 100 or 1000 repetitions) and commonly their number is justified in terms of the statistical properties of the simulated signal replicates, which are considered to be a sample from an infinitely large population of events or trials. The mean and standard deviation of the parameter estimates across this sample can be used to estimate the number of repetitions required to obtain a specified uncertainty in the parameter estimates. These sample-obtained quantities are considered to represent the parent population: the mean of the means of a large number of such samples is expected to converge to the parent population's mean. The uncertainty of this estimate with respect to the true population mean is

estimated via the sample σ_N: for instance we can be 95% confident that a sample's mean will be within ± 1.96 times the σ_N of the population mean, which in turn is estimated by the sample σ_N divided by the square root of the sample size (in this case number of replicates). The proportion of the estimated population σ_N corresponding to a specified confidence level is called the confidence interval and in the general case is given as $z_c \frac{\sigma_N}{\sqrt{N}}$, where z_c (the confidence level) is the interval width in multiples of the standard deviation, σ_N the sample's standard deviation and N its size. In order to determine the required number of repetitions, a maximum acceptable percentage error for the mean can be specified via the above expression. By considering the (two-sided) confidence interval to represent this maximum error, this will be

$error_{max} = z_c \frac{\sigma_N}{\sqrt{N}}$, or written as the percentage error of the mean ε, it becomes

$\varepsilon = \frac{100\ z_c\ \sigma_N}{\text{mean}\ \sqrt{N}}$ (Naval Postgraduate School, 2004). Solving for *N* this gives

$$N = \left(\frac{100\ z_c\ \sigma_N}{\varepsilon\ mean}\right)^2$$

As will be shown later, typical values for muscle-water T_2 – the most important of the MRI parameters investigated in this work – lie in the range 30 to 40ms with a typically-obtained standard deviation of 1 to 2ms, both in real data and simulations. If an error of 0.29%, which would translate into a 0.1ms error in estimating a true T_2 of 35ms was considered acceptable, for a 1.96 confidence level factor and σ_N of 1ms the estimated number of repetitions required would be 384. This theoretical estimate can also be justified empirically via simulations with varying numbers of replicates. An example using the physics and statistical model adopted in the final stages of this

work presented in section 3.2.5 (single component case) is given here. In this test noise was included in the simulations at a level such that SNR (highest echo amplitude / noise standard deviation) was 35, close to that observed in datasets from real acquisitions (section 3.1). The experiment was repeated with an increased number of repetitions each time in order to reach an asymptotically optimised estimation (Sijbers et al., 1998b), producing a T_2 estimate of 34.3ms.

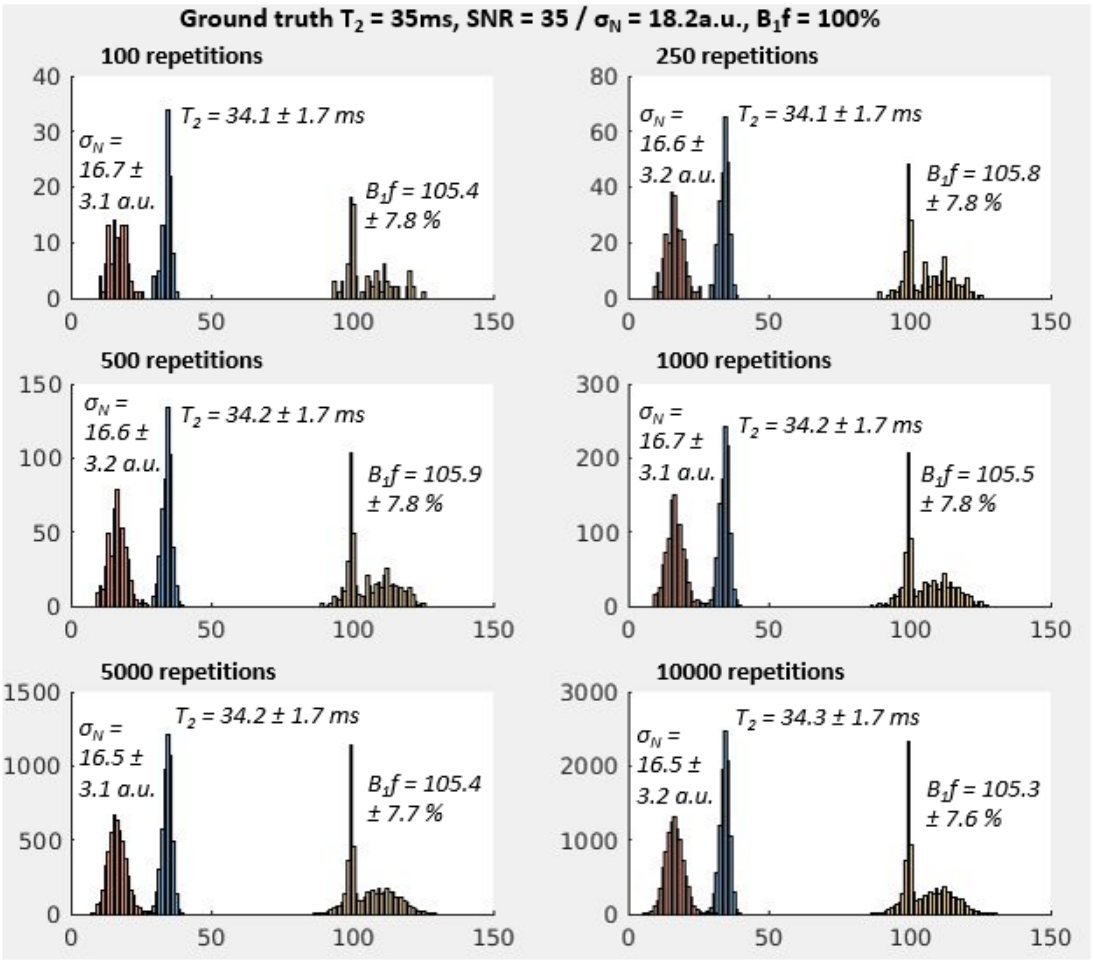


Figure 3.3: Dependence of the mean and standard deviation of the estimated parameters on the number of repetitions. For T_2 - the main parameter of interest – convergence is observed from 500 repetitions within a range of 0.1ms (compared to the value yielded by very large numbers of repetitions), as predicted by the statistical analysis. A single component sEPG model was used (section 3.2.5) with ground truth parameters B_1f=100%, T_2=35ms, T_1=1400ms, amplitude α=4000a.u. and Rician noise σ_N (SNR≈35) in order to estimate B_1f, T_2, α and noise σ_N.

As shown in Figure 3.3 the T_2 estimate converges within an interval of 0.1ms (|34.2-34.3|=0.1ms) for 500 repetitions or above, consistent with the result predicted by the above statistical reasoning. The remaining parameter estimates behaved similarly as will be described and explained in later sections.

Therefore for this particular experiment, use of 500 or more repetitions provides a maximum divergence of 0.1ms (or 0.29% error in relation to the 35ms ground truth value) in the estimation of T_2. There remains an inherent bias of 0.7 ± 0.1ms ($T_{2,\text{estimated}}$-$T_{2,\text{nominal}}$=|34.3-35|=0.7ms) in the estimated T_2, characteristic of the estimation algorithm used. The latter is a known property of MLE methods, predicted in theory (Poot et al., 2013) and also seen in practice (McPhee and Wilman, 2017). It will also be further addressed in later sections (section 3.2.6 and onwards).

3.2.2 Exponential model and Gaussian noise

To thoroughly explore the issues involved in T_2 estimation the simplest possible model of signal behaviour was used as a starting point. This first step in the investigation of T_2 fitting considered a simple single component exponential decay model. The first set of simulations therefore examined the fitting of five data points created by such a model at different echo times (*TE*s) with the addition of randomly varying noise. This simulated realistically individually acquired physical single spin echo (SE) measurements with successively increasing *TE*. In the first examples the noise levels were rather low, such that we expect the signal distribution to be approximately Gaussian so that the LSQ and MLE methods are expected to return near identical results. The physics model in this case therefore was

$$s(TE) = \alpha \cdot e^{-\frac{TE}{T_2}}$$ Equation 13

where *s* is the signal at echo time *TE*, α the signal amplitude (intercept) and T_2 the decay constant, *TE*s and T_2 given in ms. Random noise was generated by the Matlab built-in function 'randn', producing random numbers drawn from a Gaussian distribution with a unit SD (MathWorks, Massachusetts, United States). Therefore by multiplying this number in each case by the desired noise σ_N and adding it to the spin-echo magnitudes calculated according to (Equation 13) signals with the desired SNR were synthesized. LSQ fitting was implemented with use of the 'lsqcurvefit' Matlab built-in function which solves nonlinear curve-fitting (data-fitting) problems in least-squares sense, and MLE was implemented using Equation 13 to model the signal in a Gaussian PDF calculation as the objective function to be minimised by 'fmincon' (MathWorks, Massachusetts, United States) as described in section 2.2.3 and Figure 2.31.

Thus we consider fitting the simulated signal from a single component exponential decay model with parameters amplitude (α) and time constant (T_2) of 1000 units and 30ms respectively, sampled at successive TEs of 30, 60, 90, 120 and 150ms. Random Gaussian distributed noise was added with σ_N varied systematically to create 5 SNR levels varying from 50 to 10, the signal magnitude used to calculate the SNR being here again defined as that of the highest (1st) echo. In Figure 3.4 an example set of simulations is shown for the intermediate SNR of 30.

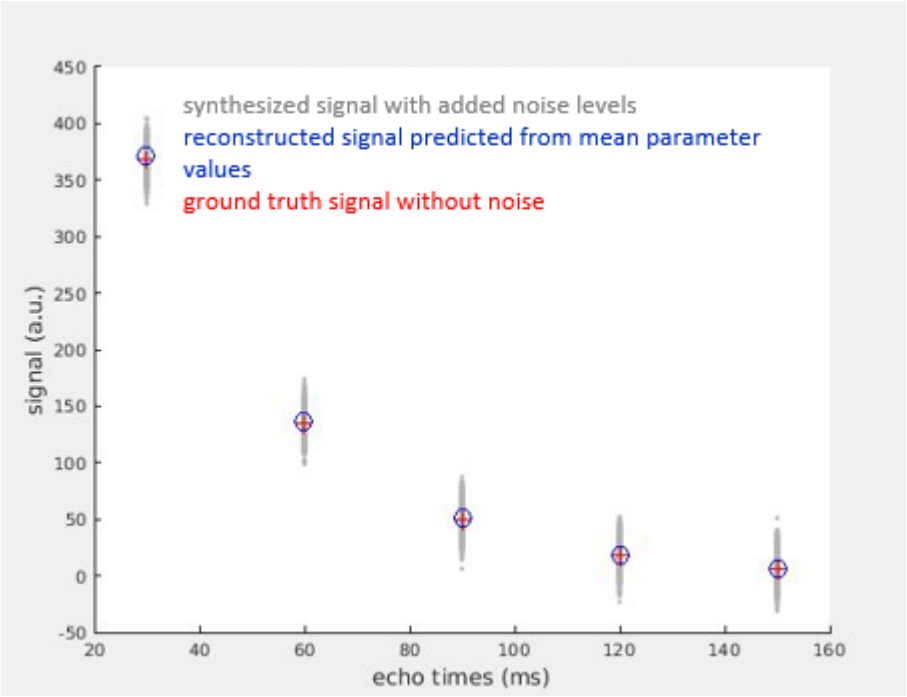


Figure 3.4: Example of 1000 simulations for SNR 30, exponential model with added Gaussian noise, ground truth parameters: amplitude α=1000 (arbitrary units), T_2=30ms. Grey dots represent the simulated signals, blue circles the signal predicted using mean values of the MLE parameter estimates and red crosses the ground truth signal without added noise. The mean parameter estimate-predicted signal closely overlaps the ground truth signal.

As can be seen the synthesized signals with the added noise (grey dots) are symmetrically distributed around the ground truth signal (red crosses) which in turn is closely approximated by the signal predicted by forward calculation using the mean parameter estimates (blue circles) obtained by MLE. Figure 3.5 compares MLE and LSQ minimisation T_2 estimates for this model and parameter set across the range of SNRs investigated.

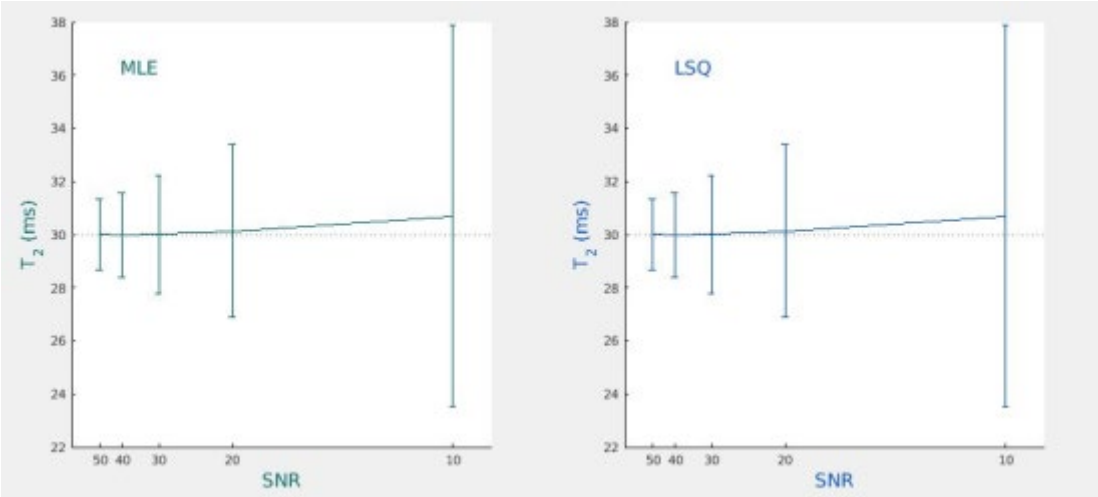


Figure 3.5: Comparison of T_2 estimates returned by LSQ and MLE fitting to 1000 simulations of an exponentially decaying signal (amplitude α=1000 units, time constant T_2=30ms) with added Gaussian noise SNR varying from 50 to 10. Under these conditions means and SDs of the T_2 estimates are virtually identical for ML estimation (left) and LSQ minimization (right).

Estimated T_2 mean, bias and σ_N were virtually identical using both methods; some minor differences were seen after the 3rd decimal digit in a few individual fits. Both the bias and the SD increased with decreasing SNR. This demonstrates the equivalence of the two approaches (LSQ and MLE) under appropriate conditions, satisfied in this case for a single component exponential model with normally distributed noise. It should be noted that this equivalence is predicated on the respective fit routines converging to the correct global minimum: as will be explored in later sections, this may depend on correct choice of the algorithm control parameters, especially any parameter bounds imposed, and the initial parameter estimates used to initialise the fitting process. In this comparison initial parameter values were chosen close to the true values (approximately from 10% to 50% deviation) and similarly their bounds were kept to physically meaningful ranges, for example only positive values for both α and T_2.

MLE, unlike the LSQ method, involves explicit estimation of the noise levels: noise σ_N is one of the parameters estimates returned by the routine and was seen to reasonably approximate the true values of added noise σ_N, although with some negative bias. As an example, for SNR 30 ground truth noise and the same α and T_2 as in Figure 3.5, the ground truth σ_N was 12.3 units while the mean estimated value was 8.8±3.8 (Figure 3.6).

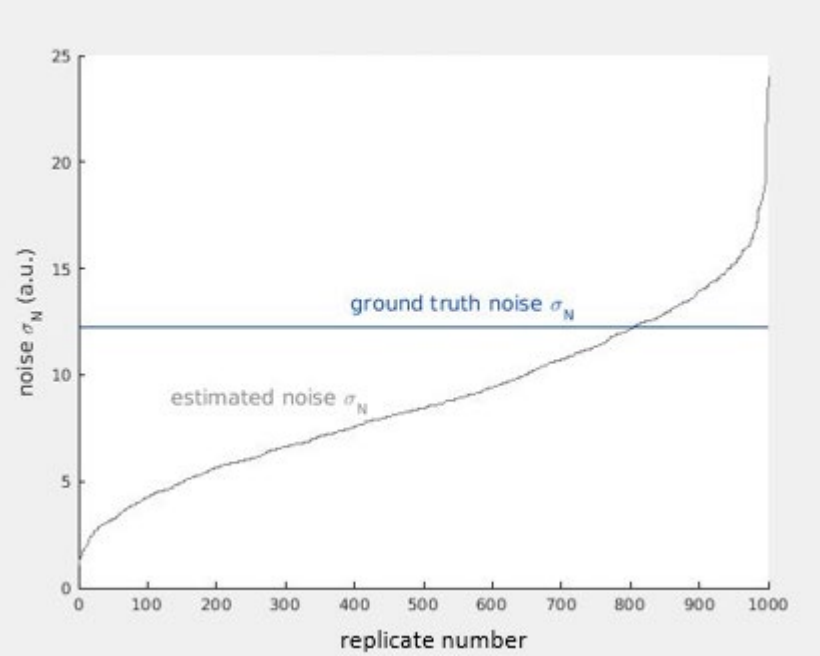


Figure 3.6: The MLE noise σ_N estimates for SNR 30 (Figure 3.4) arranged in ascending order for the entire set of simulations, showing overall a tendency to underestimation compared to the ground truth value.

The estimated noise σ_N was in general smaller than the ground truth value. This is likely to be correlated with the systematic positive bias in the T_2 estimate with increasing levels of noise due to an interaction between the proportion of signal explained by the model and the proportion of the signal assigned to noise. This concept is further explored in the following chapters and is particularly important when analysing physical *in vivo* data, since in that case the noise parameter must be

estimated as it is usually unknown, and its distribution with respect to the signal amplitude is governed by complicated factors such as non-uniform receive sensitivity and multi-coil acquisition schemes.

3.2.3 Exponential model and Rician noise

The next more complicated test case for which parameter estimate behaviour was investigated was the same mono-exponential decay model as in the previous section, but this time with Rician rather than Gaussian noise. First considering the forward model, Figure 3.7 illustrates, for a high SNR condition 10^5 replicates with added Gaussian (top panel) and Rician (lower panel) noise (equal noise σ_N in each PDF), for the same ground truth forward model used in the previous section, how the signal is distributed per simulated echo. Data shown are the 2nd to 5th *TE*s presented from right to left (decreasing SNR): the shortest *TE* has been omitted for clarity since the Rician distribution effect has least impact for this echo with highest local SNR. The graphs represent distributions of simulated signals generated as previously (section 3.2.2), with σ_Ns calibrated to achieve a SNR of 40 for the 1st echo. For this relatively high SNR the distributions appear approximately Gaussian with widths about 6 times σ_N (corresponding to 99.7% of the counts) as expected for normally distributed data. Figure 3.8 illustrates the simulated signal histograms with the ground truth noise σ_N calibrated to achieve a lower SNR of 20 for the 1st echo.

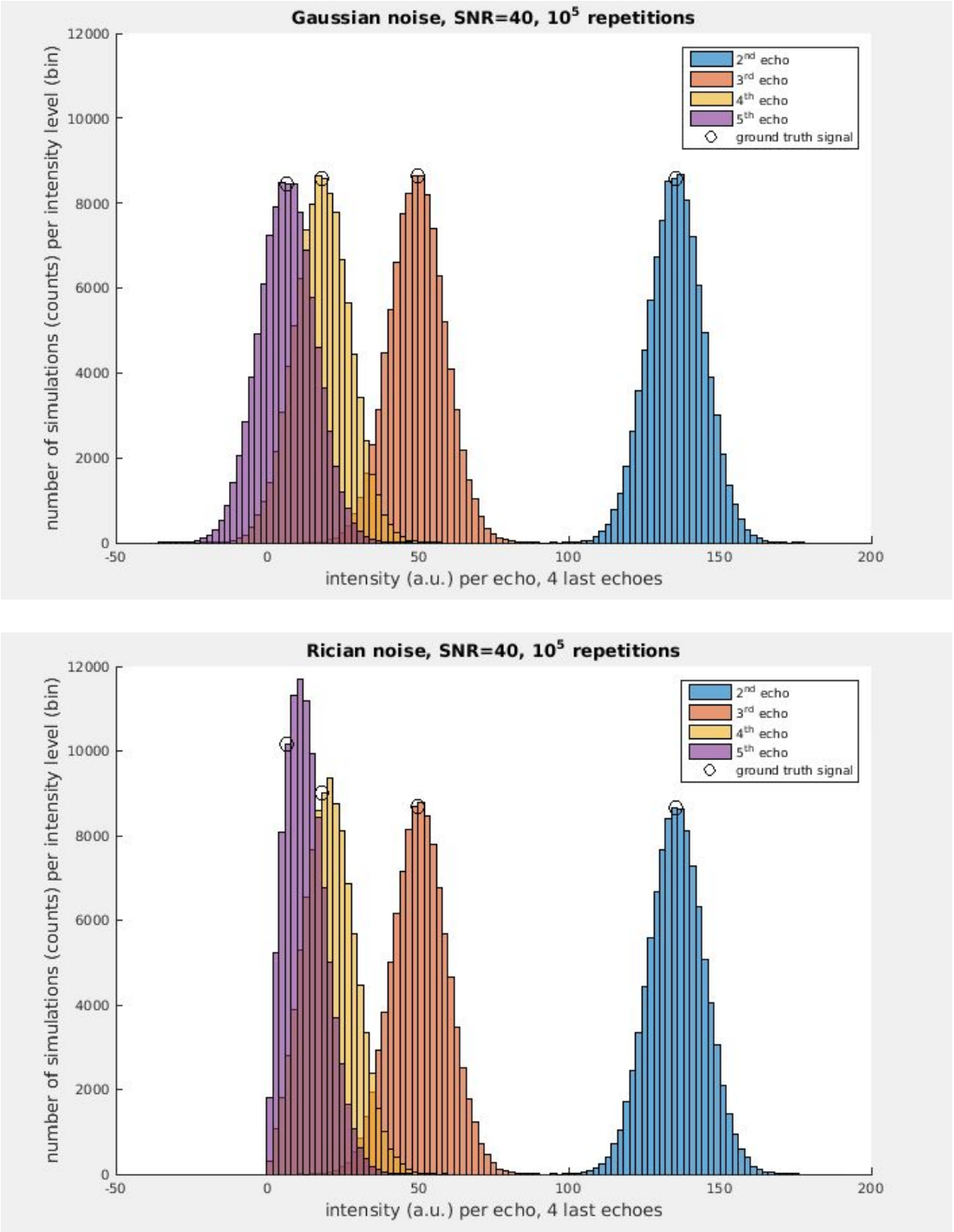


Figure 3.7: Histogram distributions of signal intensity for simulated spin echoes with successively longer *TE*s across 100000 repetitions with noise σ_N calibrated to give SNR = 40 for the shortest *TE*. Top: Gaussian noise distribution, bottom: Rician noise distribution. 2nd to 5th echo signals (highest to lowest intensity) are plotted from right to left (echo with shortest *TE* omitted for clarity).

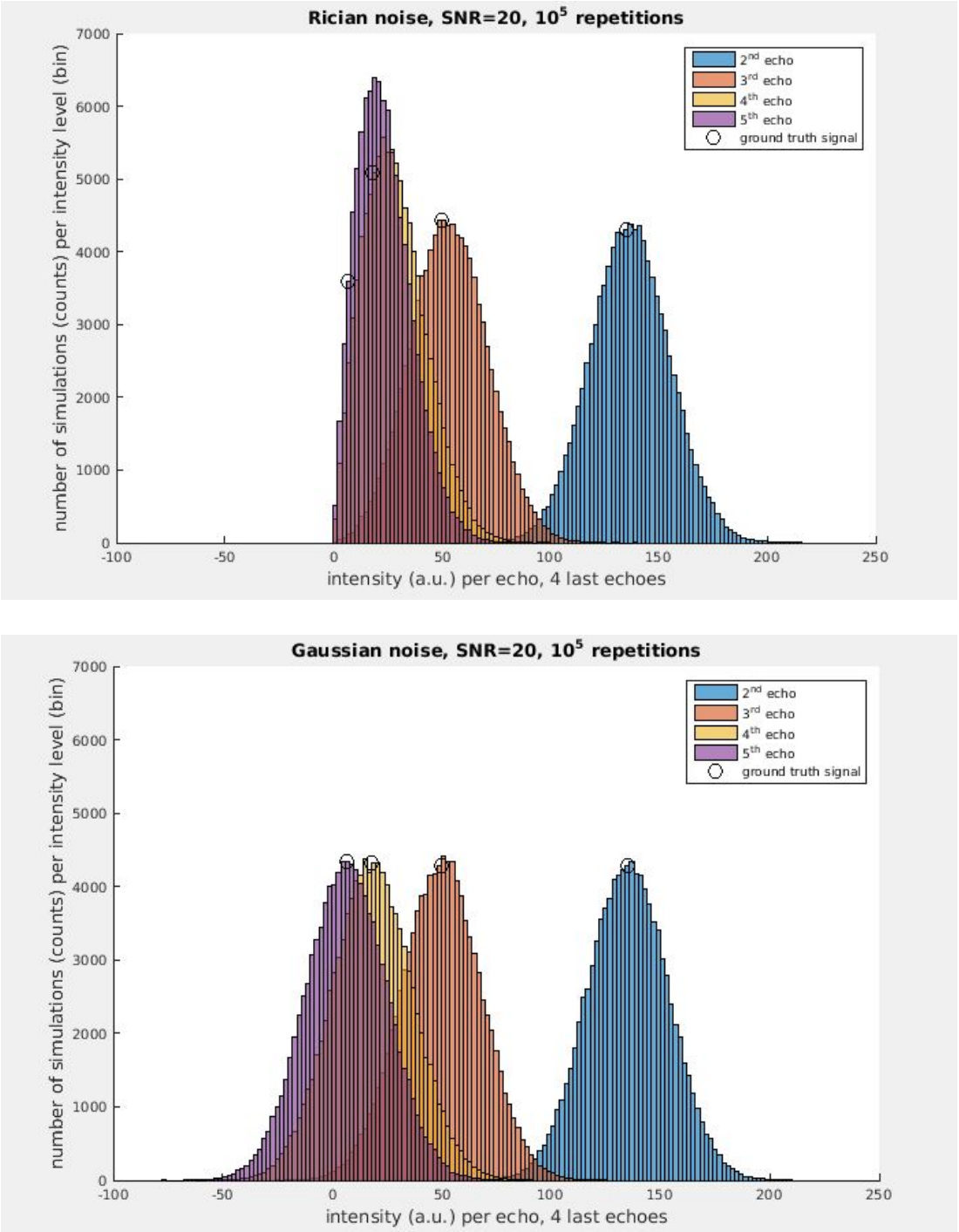


Figure 3.8: Histogram distributions of signal intensity for simulated spin echoes with successively longer *TE*s across 100000 repetitions with noise σ_N calibrated to give SNR = 20 for the shortest *TE*. Top: Rician noise distribution, bottom: Gaussian noise distribution. 2nd to 5th echo signals (highest to lowest intensity) are plotted from right to left (echo with shortest *TE* omitted for clarity).

The signal model with the Rician distribution is in practice obtained by taking the square root of the sum of the squared modelled (exponential) signal plus noise (1st

channel) and the squared noise (2nd channel). This mathematically equals the sum of true signal plus noise expected to be acquired in each channel for magnitude reconstructed images (McRobbie et al., 2006). In good agreement with the literature (Walker-Samuel et al., 2009) (Gudbjartsson and Patz, 1995) visual assessment of the histogram shapes in the lower panel shows that these vary from approximately Gaussian for high SNR in the first echoes to close to a Rayleigh distribution towards the last echo where the signal level and hence SNR is markedly lower, approaching zero. This also confirms the points made in sections 2.1.8 and 3.1 for theoretical and practical considerations with regards to noise in MRI.

Turning now to consider the behaviour of the parameter estimates with each fitting method, example results for LSQ and MLE fitting of 1000 simulated datasets for an exponential model with α=1000 and T_2=30ms with Rician giving a shortest *TE* SNR of 10 (the lowest SNR here investigated) is shown in Figure 3.9 - Figure 3.11. The minimisation parameter start values were the ground truth values for α and T_2 for both methods, and additionally 1 (a.u.) for the noise σ_N in the Rician PDF for MLE. The Levenberg-Marquardt method was used for LSQ which does not offer the option to specify bound constraints (sections 2.2.2 and 3.3.1). For MLE (section 2.2.3) bounds were highest echo amplitude < α < 2500a.u., 10ms < T_2 < 200ms and 0.1a.u. < σ_N < highest echo amplitude. Routine settings and convergence criteria are discussed in more detail during the stages of the final algorithm optimisation, from section 3.2.5 and onwards. In Figure 3.10 the parameter estimates' means and SDs yielded for all 5 SNRs between 50 and 10 from 1000 replicates at each noise level are compared for MLE fitting and LSQ minimization.

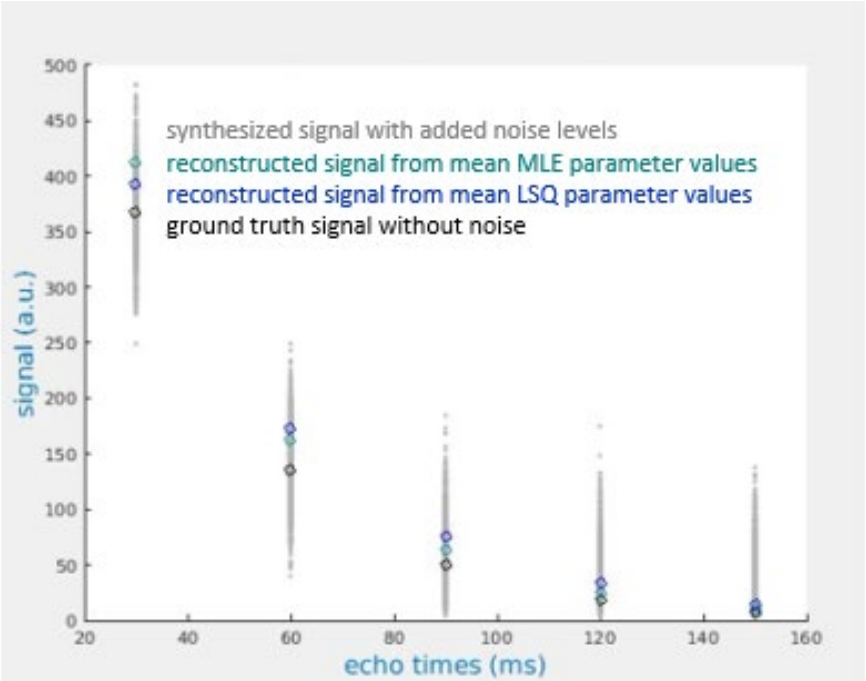


Figure 3.9: Example of a set of 1000 simulations for an exponential model with Rician noise SNR 10 (first echo), ground truth parameters: amplitude α=1000 (arbitrary units), T_2=30ms. Grey dots represent the synthesized signals, green circles the signal reconstructed using the mean MLE parameter estimates, blue circles the signal reconstructed using the mean LSQ parameter estimates, and black circles the ground truth signal without noise. MLE-derived signal amplitude estimates deviate from the nominal value for early echoes while T_2 estimates remain overall accurate (see subsequent figures).

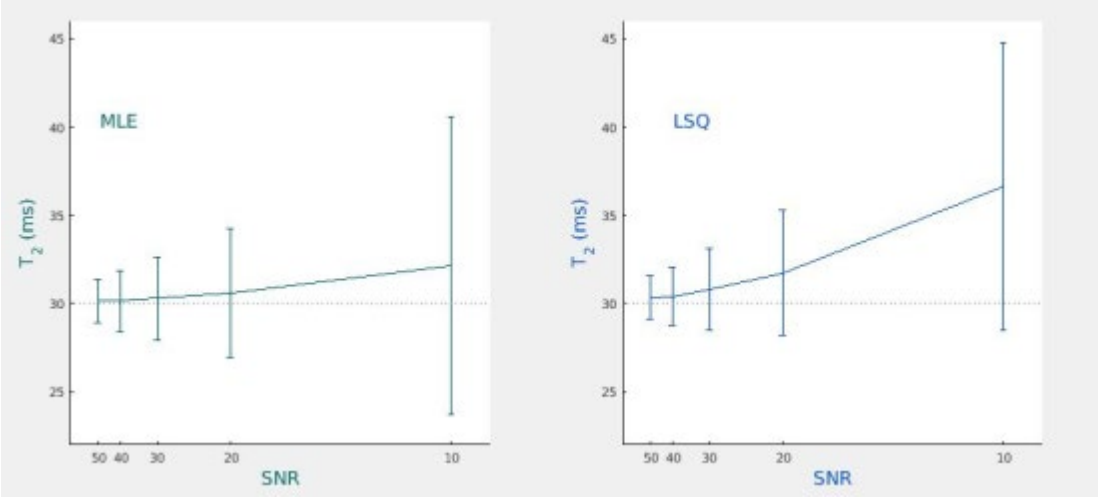


Figure 3.10: Comparison of T_2 estimates (means and SD) returned by LSQ and MLE fitting to simulations of exponentially decaying signals (amplitude α=1000 units, time constant T_2=30ms) with Rician noise yielding SNR from 50 to 10 (10^3 repetitions each). MLE T_2 estimates (left) are overall more accurate than LSQ T_2 estimates, which appear increasingly biased as noise increases (right).

Figure 3.11 shows the estimated parameter histograms for the lowest SNR of 10.

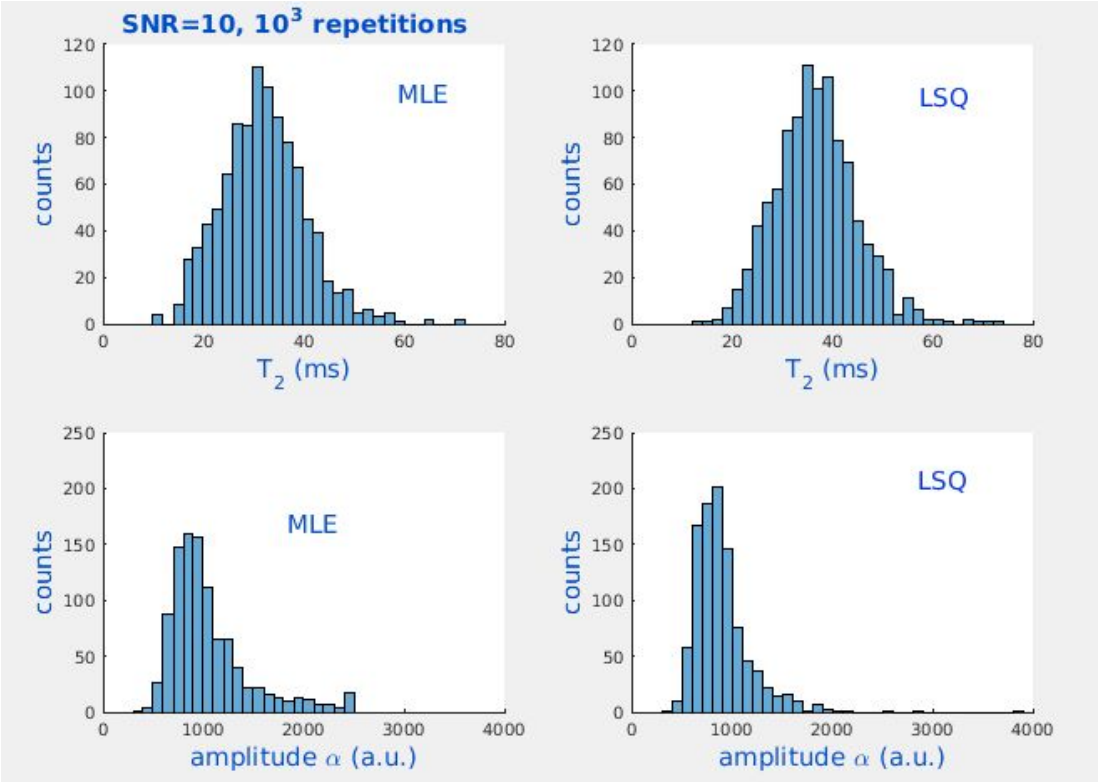


Figure 3.11: The distributions of the MLE parameter estimates for the exponential model with Rician noise; these appear to be non-symmetrical (skewed) in general.

The non-Gaussian, or more generally non-symmetrical shape of these histograms – such that the sample means differ from their modes, indicate that these parameter estimators are biased for both the MLE and LSQ cases. This is supported by Figure 3.12, showing that the parameter modes are overall closer to the ground truth value than the mean values shown in Figure 3.10. The 15.9 and 84.1 percentiles (or mode ± 34.1 percentiles) are presented in Figure 3.12 as an indication of the data spread approximately equivalent to a 1 SD error bar, which includes approximately 68.2% of the data for a normal (Gaussian) distribution. ML estimates are again less biased than those from LSQ.

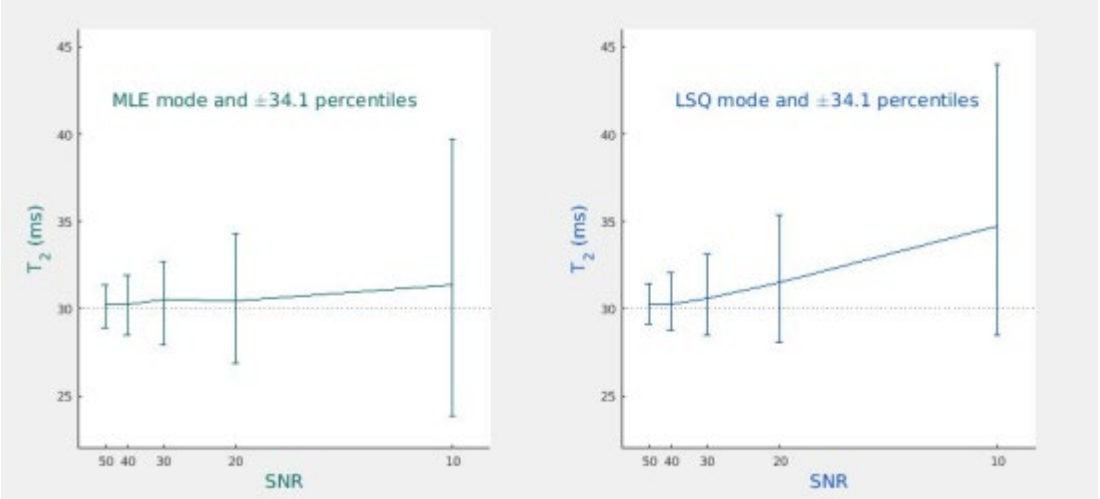


Figure 3.12: Comparison of T_2 estimates (modes and ±34.1 percentiles) returned by LSQ and MLE fitting to simulations of an exponential decay signal (amplitude α=1000 units, time constant T_2=30ms) with Rician noise yielding SNRs from 50 to 10 (10^3 repetitions each). Biasedness is in general decreased compared to means and SDs (Figure 3.10), while ML estimates still appear less biased than the respective LSQ ones.

For further clarity in Figure 3.13 T_2 estimate distributions comparing MLE and LSQ are presented with the use of boxplots and whiskers – reporting medians, quartiles and considering extreme values as outliers.

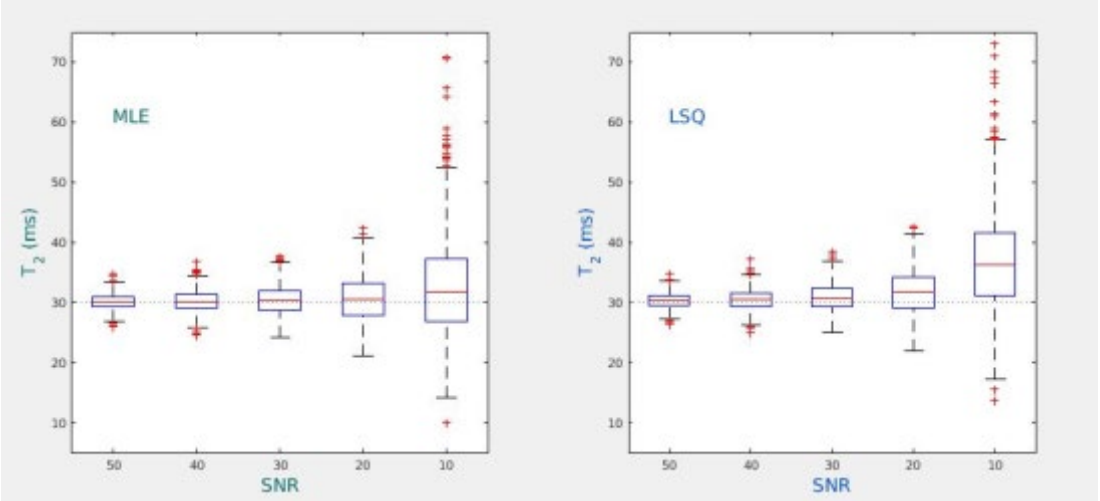


Figure 3.13: Comparison of T_2 estimate distributions with boxplots (medians and percentiles), returned by LSQ and MLE fitting to simulations of exponential decay signals (amplitude α=1000 units, time constant T_2=30ms) with Rician noise yielding SNR from 50 to 10 (10^3 repetitions each). Behaviour in terms of bias is for LSQ closer to the results expressed by means and σ_Ns (Figure 3.10) and for MLE closer to results expressed by modes and percentiles (Figure 3.12), making the apparent differences between the two methods more conspicuous due to their conceptually different PDF structure.

While medians are closer to modes for MLE (showing relatively less bias), they are closer to means for LSQ (more bias), revealing more clearly the differences between the two methods.

It was previously noted that for the lowest SNR of 10, the signal reconstructed using the mean of the estimated parameters across all of the simulated replicates significantly deviated from the ground truth signal, as shown in Figure 3.9. This was not least due to many cases (repetitions) for which the ML estimation for amplitude α substantially departed from the true values for the initial echoes, thus raising the mean returned values (1^{st} and 2^{nd} echoes). This was mainly an effect of echo trains which by chance were created with a markedly low second echo due to the added noise (SNR = 10) – markedly negative for these cases, plausibly also emphasised by the relatively wide bound constraints for the amplitude. This adds steepness to the echo succession (i.e. between the two first) and combined with the additive Rician noise effect for the later echoes leads to a combination of increased α estimate and reduced T_2 estimate for MLE compared to LSQ. A typical example is shown in Figure 3.14. This is a phenomenon that applies also to EPG models as will be seen in later sections.

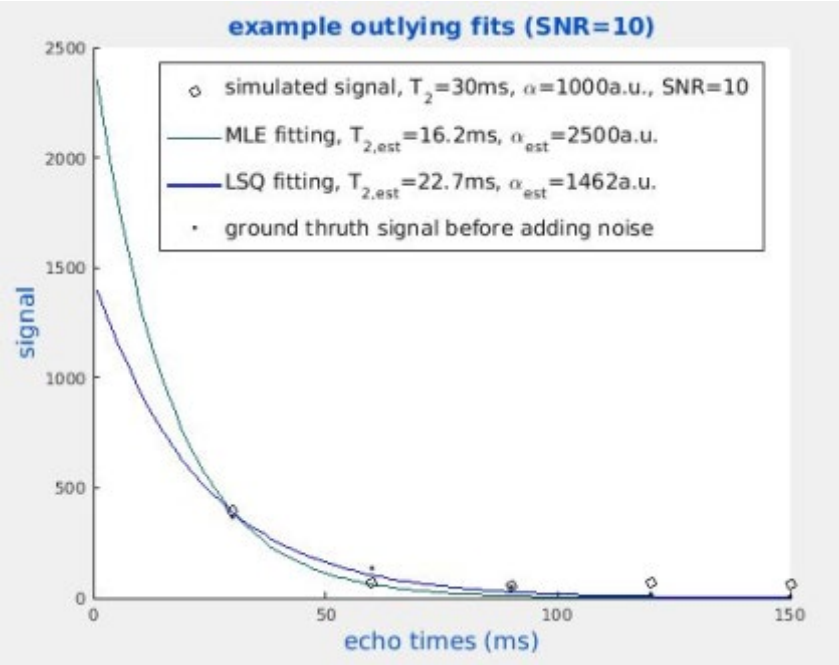


Figure 3.14: An example fit (simulation repetition) returning spuriously high amplitude (intercept) and relatively low T_2 estimates, due to the (high) added noise creating a markedly low second echo. This was seen in roughly 10-20% of that set of simulations.

This apparent deviation of the expected values in Figure 3.9 is however related also to the way that results are reported. Re-plotting the signal re-composed from the estimates' modes, namely the values used in Figure 3.12 for SNR = 10, we obtain a more representative characterization of the fitting process, as shown in Figure 3.15.

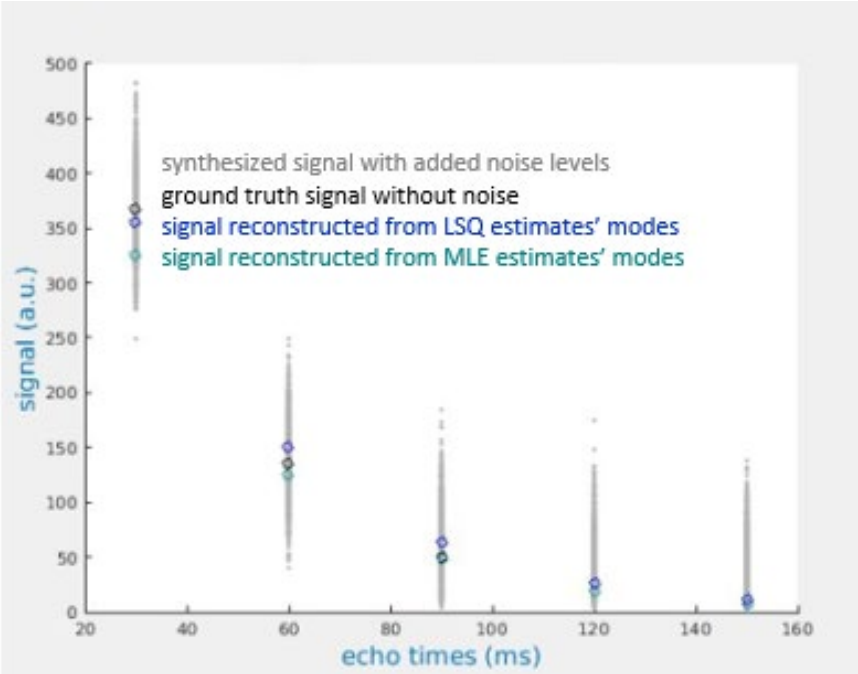


Figure 3.15: The example set of 1000 simulations for SNR 10, from Figure 3.9, here with the signal reconstructed using the estimated parameters modes instead of means. In this manner MLE results express more accurately both the Rician PDFs of the synthesized ('raw') data and any skewness in the estimated parameter distributions. On the other hand the LSQ method results tend to produce reconstructed signals that pass closer to the middle of data distributions as these are considered Gaussian.

The signal reconstructed using the MLE method parameter estimates is expected in general to pass under the hypothesized Gaussian peak of the synthesized signal for two reasons, as is apparent in this graph. First, because MLE accounts correctly for the Rician noise; this is particularly apparent where SNR is low, typically for high noise images and in any case for later echoes. Second, because of the estimator bias that is inherent in the MLE method and is dependent upon the specific experimental conditions, as exemplified in the results presented above. The skewed distributions of the parameter estimates are related to this inherent estimator bias. Conversely, the signal re-composed using the LSQ estimates is expected in general to pass through the peak of the approximate Gaussian distribution of data points, according to this method's assumptions, and this is partly apparent in the above graph. There

is however the additional tendency for the first and particularly for the last LSQ estimate-mode-reconstructed echoes, to pass under the ground truth data points. This is due to the inability of the LSQ method to account for the additive nature of the Rician noise most apparent in the latter, low signal echoes; this Rician noise adds some extra 'curvature' to the echo trains, which the mono-exponential function cannot adequately describe. This effect is well known and reported in the literature, and can be partly mitigated by the addition of a constant to the mono-exponential model (Milford et al., 2015).

The examples above confirm that MLE T_2 estimation in general appears to perform better than LSQ minimization estimation in terms of accuracy: bias when using MLE reduces to only a fraction of that when using LSQ minimization. The use of exponential type models is an historically established way to describe and extract signal parameters (e.g. the T_2 time constant) for multi-echo spin echo acquisitions (as the CPMG sequence), however more advanced and accurate models also exist. These are now discussed in the following sections.

3.2.4 Single component EPG model and Rician noise

The purpose of this section is to extend the methods developed for exponential model parameter estimation presented above by application to estimating EPG model parameters, a formalism which more accurately predicts real-world MRI CPMG echo train amplitudes (section 2.1.7). Taking first an idealised single-component EPG model in which slice profile effects are not considered, the

parameters to be estimated are the flip angle ϕ – assuming the refocusing angle to be twice the excitation angle, T_2, amplitude α and noise SD (σ_N). Within the EPG model T_1 is also a parameter, however variation of T_1 over a reasonable range has been shown to have negligible effect upon the accuracy of T_2 estimation (Marty et al.) and for that reason T_1 will be fixed in the following work at 1400ms, a T_1 value for muscle commonly reported in the literature (Azzabou et al., 2015).

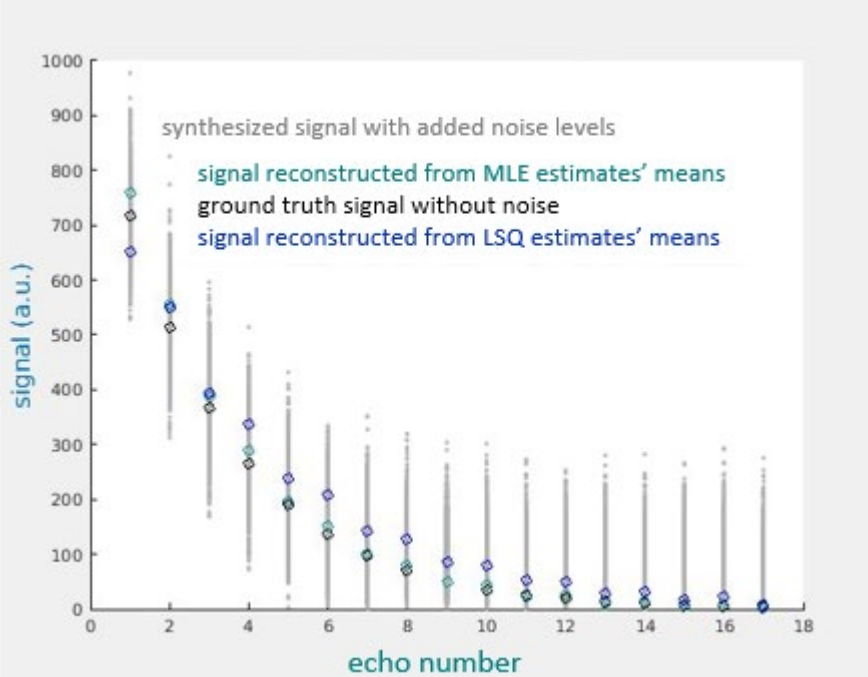


Figure 3.16: 1000 simulations for a single component example EPG model with Rician noise giving an SNR of 10, ground truth parameters: amplitude α=1000 (arbitrary units), T_2=30ms, flip angle φ=90˚ (180˚ for refocusing). Grey dots represent the synthesized signal, green circles the signal reconstructed using the mean of the MLE parameter estimates, blue circles the signal reconstructed using the mean of the LSQ parameter estimates and black circles the ground truth signal values. Reconstructed signal values based on the MLE parameter estimates deviates significantly less from the ground truth signal than from the equivalent LSQ estimated parameters.

LSQ and MLE parameter estimation behaviour for an EPG model simulated dataset with Rician noise is exemplified in Figure 3.16, for the SNR 10 case (the lowest SNR

here tested), with T_2 30ms, amplitude (α) 1000 arbitrary units (a.u.) and excitation and refocusing angles 90˚ and 180˚ respectively (no B_1 field errors).

The echo train length (ETL) of the synthesized signals was 17 with 9.9ms echo spacing. For fitting, parameter start values and bounds were once again kept within physically meaningful limits (e.g. no negative or extreme values) – a more detailed investigation of their impact on fitting behaviour will be presented later during the evaluation of the final versions of the finally deployed algorithm. In Figure 3.17 the means and σ_Ns of the estimated T_2, modes (and percentiles) and medians (and quartiles) yielded for 5 SNRs between 50 and 10, fitting to 1000 simulations for each are compared for the MLE and LSQ methods. In Figure 3.18 the parameter estimates' histograms are presented for the 5th, lowest SNR of 10. Reconstruction of the signal based on the returned MLE-estimated parameter means is systematically much closer to the ground truth signal than the signal reconstructed from the LSQ estimates.

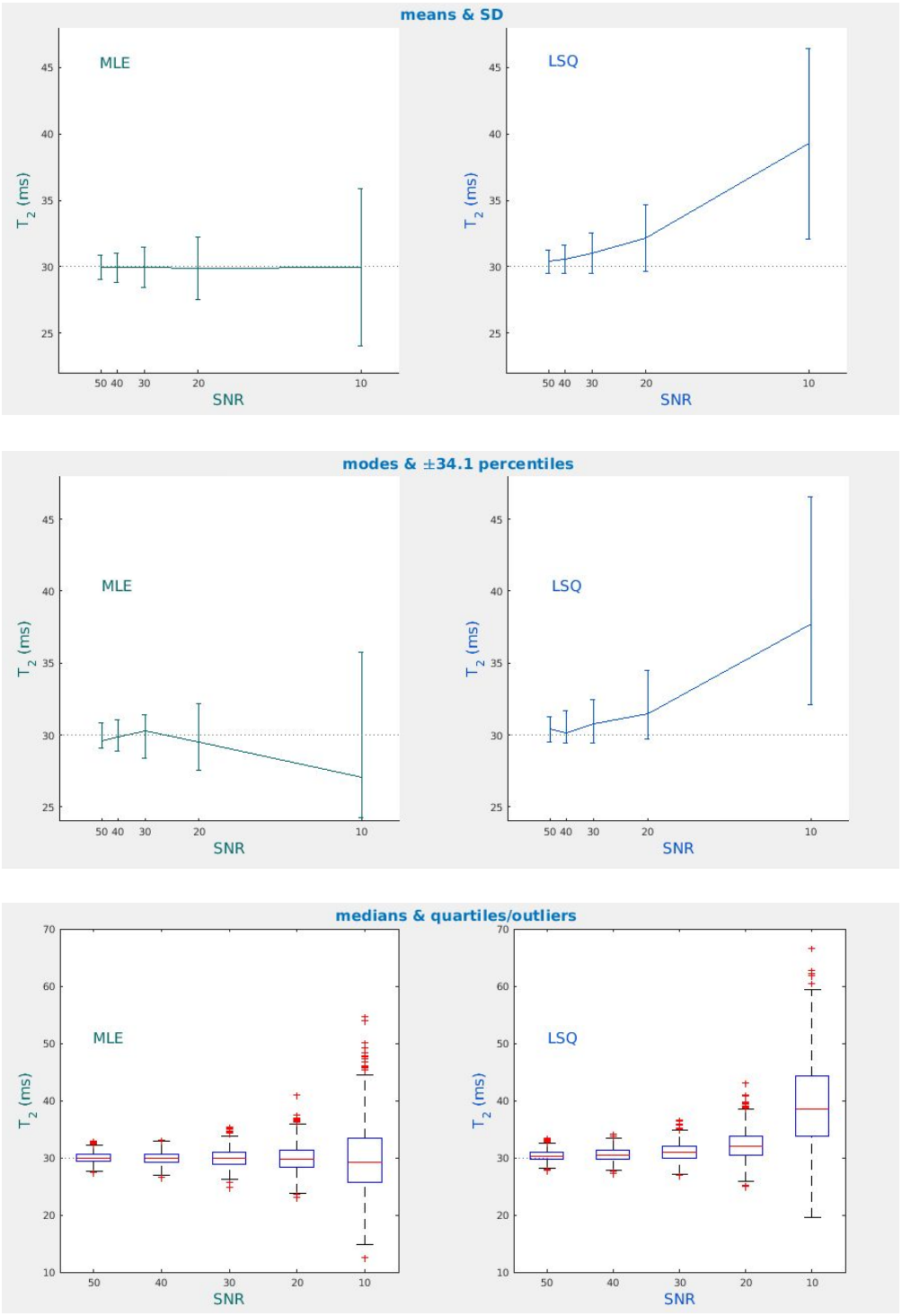


Figure 3.17: The dependence on SNR of MLE and LSQ T_2 estimates presented in terms of means and SD (top row), modes and percentiles (middle row) and medians and quartiles (bottom row) over 1000 replicates of modelled EPG signals (amplitude α=1000 units, time constant T_2=30ms, refocusing angle = 2 x excitation angle = 90˚); Rician noise was added to yield SNR from 50 to 10 (10^3 repetitions each). There is systematically less bias in ML estimated T_2 in comparison with LSQ T_2 estimation as well as a marked effect of the distribution shapes upon the mode values, for the given histogram binwidths and number of samples (counts).

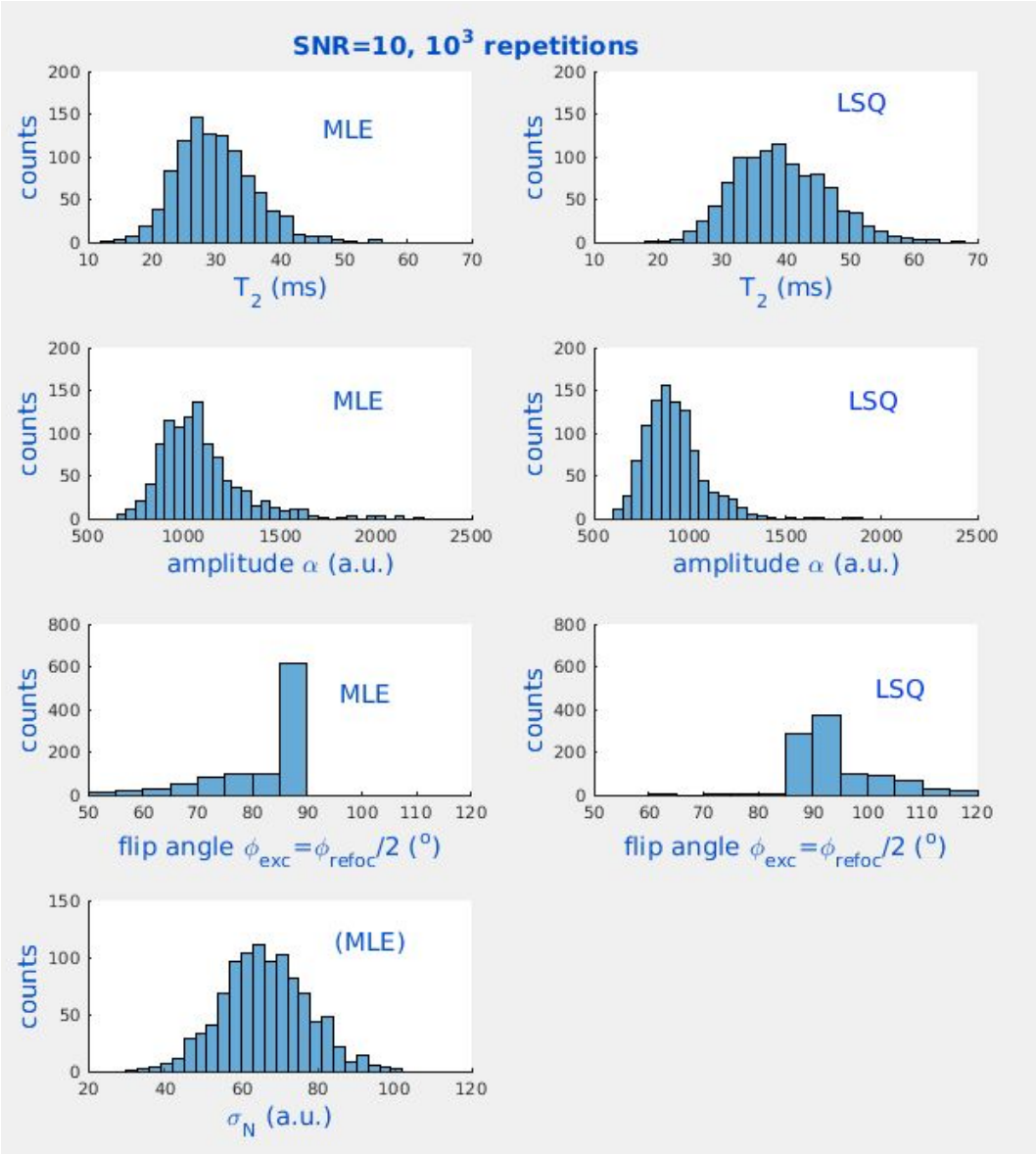


Figure 3.18: The distributions of the parameter estimates for the EPG model with added Rician noise (highest level σ_N or lowest SNR = 10), appearing non symmetrical (skewed) in general (extreme values/outliers were omitted for clarity).

The predominantly positive deviation in the α estimates is consistent with the skewness of the amplitude estimates' distributions shown in Figure 3.18 (2nd row, 1st graph) while MLE estimated T_2 remains broadly more accurate, i.e. closer to the ground truth value of 30ms (1st row, 1st graph). However the shape of the estimated T_2 distribution explains the larger departure of its mode (27.1ms with ± 34.1 percentiles -2.8 and +8.7ms) from the nominal value (30ms) in comparison with its

mean (29.9 ± 5.9ms) and median (29.2ms with quartiles -3.5 and +4.2ms). The one-sided shape of the excitation angle MLE estimate distribution is due to the symmetry in the EPG function for equal flip angle deviation above and below 90˚. An upper bound of 90˚ was therefore used as this was not seen to have significant effect on the rest of the parameters. The role of the excitation and refocusing angles in the fitting process will be examined in more detail via the slice profile effects introduced in following sections.

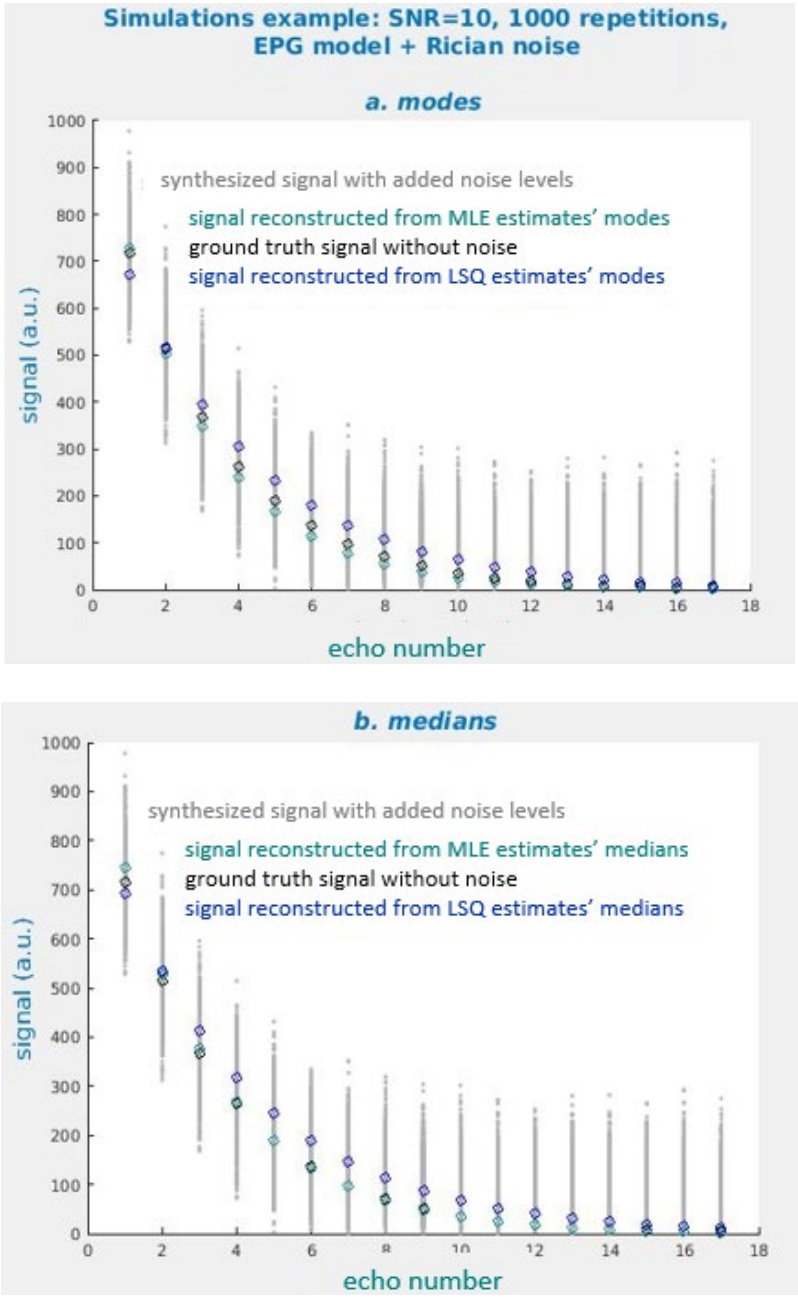


Figure 3.19: Example of a set of 1000 simulations for an EPG model with Rician noise, SNR of 10, ground truth parameters: amplitude α=1000 (arbitrary units), T_2=30ms, flip angle φ=90˚ (180˚ for refocusing). Grey dots represent the synthesized signal and black circles the ground truth signal values. Top panel: green circles represent the signal reconstructed using the modes of the MLE parameter estimates, blue circles the signal reconstructed using the modes of the LSQ parameter estimates. Bottom panel: green circles represent the signal reconstructed using the medians of the MLE parameter estimates, blue circles the signal reconstructed using the medians of the LSQ parameter estimates. Once again re-composed signal based on the MLE parameter estimates deviates significantly less from the ground truth than that reconstructed from the equivalent LSQ estimated parameters.

The graphs in Figure 3.19 show how the re-composed signal based on the modes and medians of the returned parameter estimates' histograms behave, in the case of MLE and LSQ fitting, for the case of SNR = 10. Less bias is seen than with use of the exponential model as presented in the previous section (3.2.3) which is likely to be due to the longer and more densely spaced echo train used in this case. Further exploration of this topic is presented in Chapter 6.

It is informative now to examine the behaviour of MLE fitting in the case that the noise σ_N in the Rician PDF is not considered a parameter to be estimated but is instead fixed. This would correspond to the practical circumstances where an independent estimate of σ_N was available, e.g. from the image background. There is extensive literature on both noise estimation and denoising methods in MRI (Henkelman, 1985, Mohan et al., 2014) and this is particularly relevant for MLE methods since the question of whether *a priori* or simultaneous with MLE noise estimation is optimal may be posed (Sijbers and den Dekker, 2004). In the following examples results from MLE fitting performed with the noise σ_N fixed at the ground truth values, that is at the noise σ_N levels added to the EPG signal when synthesizing the signal replicates to be fitted during simulation, were compared with results (as previously presented) in the case that noise σ_N was one of the estimated parameters (Figure 3.20). Fixing the noise σ_N was not seen to offer any improvement in the fitting experiment, rather there was a slight overall deterioration of the parameter estimates as noise levels increased. This is likely to be related to both the non linear nature of the EPG model and the not necessarily normalised PDFs (section 2.2.3) participating in the log-likelihood minimization.

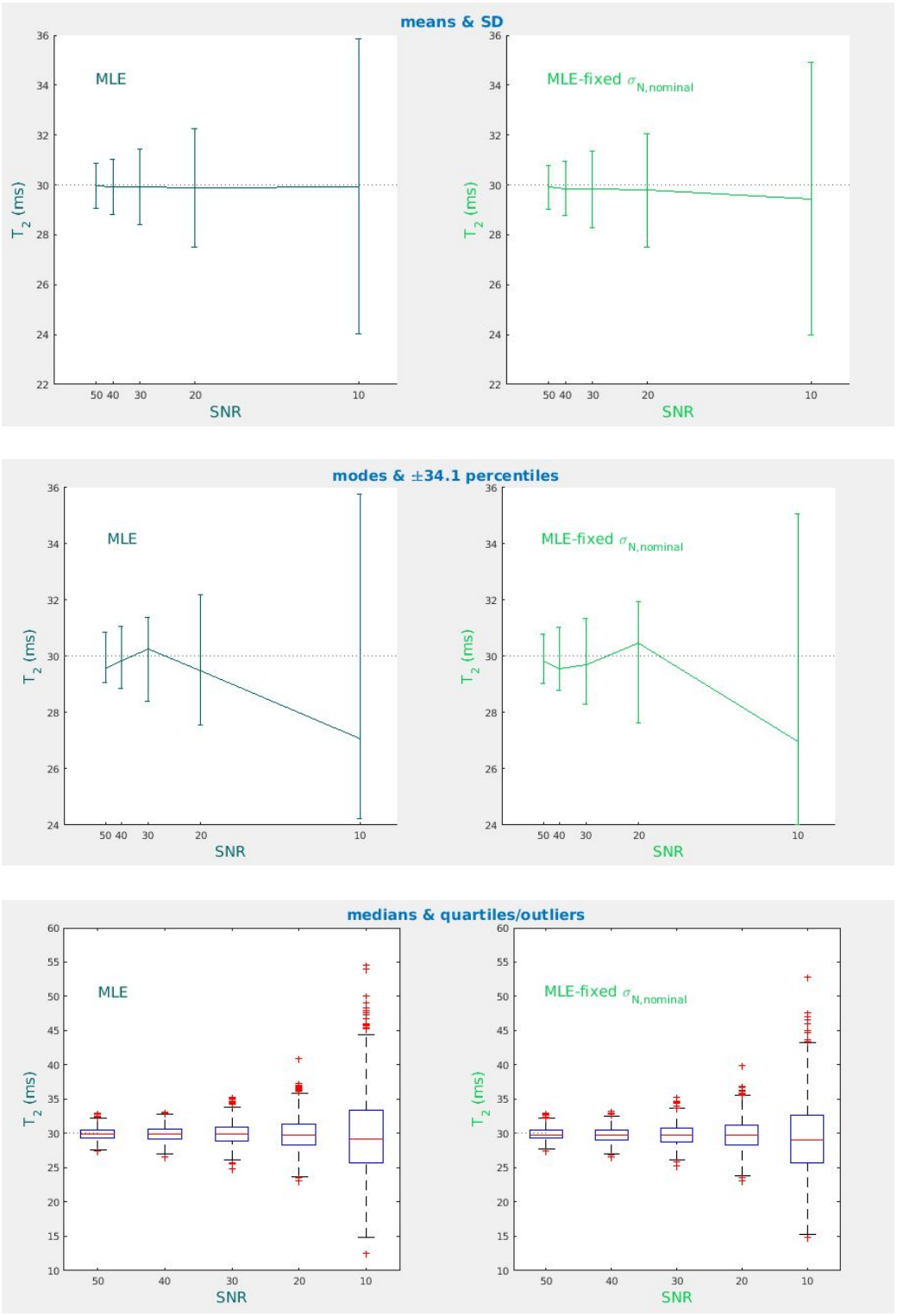


Figure 3.20: Comparison of T_2 estimates with the use of means and SD (top row), modes and percentiles (middle row) and medians and quartiles (bottom row) on results returned by MLE fitting to simulations of EPG modelled signal (amplitude α=1000 units, time constant T_2=30ms, refocusing angle = 2 x excitation angle = 90˚) with Rician noise yielding SNR from 50 to 10 (10^3 repetitions each). In the left column noise σ_N was an estimated parameter while on the right column it was fixed at its nominal value during the fitting process. Using fixed noise σ_N at the nominal (correct) levels was not seen to improve results.

Nevertheless this is a useful validation step for the method development, demonstrating that the proposed MLE method does not require input from separate noise σ_N determination methods.

3.2.5 Inclusion of slice profile effects in the EPG model

In the analysis so far presented slice profile effects have been ignored in modelling of CPMG signal. However these effects have been shown to have significant impact upon parameter estimation (McPhee and Wilman, 2017). As will be described below the effective flip angles vary across the slice profile so that the above EPG model applying at each slice position is calculated and then summed and averaged to be thus transformed to the slice selective EPG model (sEPG). The effect of this correction is demonstrated in a simple example (Figure 3.21) where a signal calculated using the sEPG model, with no B_1 calibration imperfections and no added noise, was fitted with both a simple EPG-MLE model as presented in the previous section and with the sEPG-MLE method introduced here.

For selective excitation and refocusing, excitation and refocusing angles vary at different locations across the slice profile due to the combined effect of the non-ideal RF pulse shapes available in practice and the applied gradients (section 2.1.8); the signal finally detected is the vector sum (resultant) of the different signals arising at successive locations across the slice profile. This was performed for 40 uniformly spaced divisions across the slice profile and for each division the flip angles were calculated by approximating the RF pulse time domain shape using the hard pulse method, successively applying the appropriate rotation matrices (pseudocode in

Figure 2.29, section 2.1.8). Based on these effective flip angles the signal at each slice division is calculated ('slice profile EPGs') and then averaged over the slice width to produce the final apparent echo train ('sEPG'). Consequently – even if there are no B_1 field imperfections as in this example (i.e. the nominal excitation and refocusing angles are 90˚ and 180˚ respectively), the resultant signal (black circles) deviates substantially from a pure exponential decay.

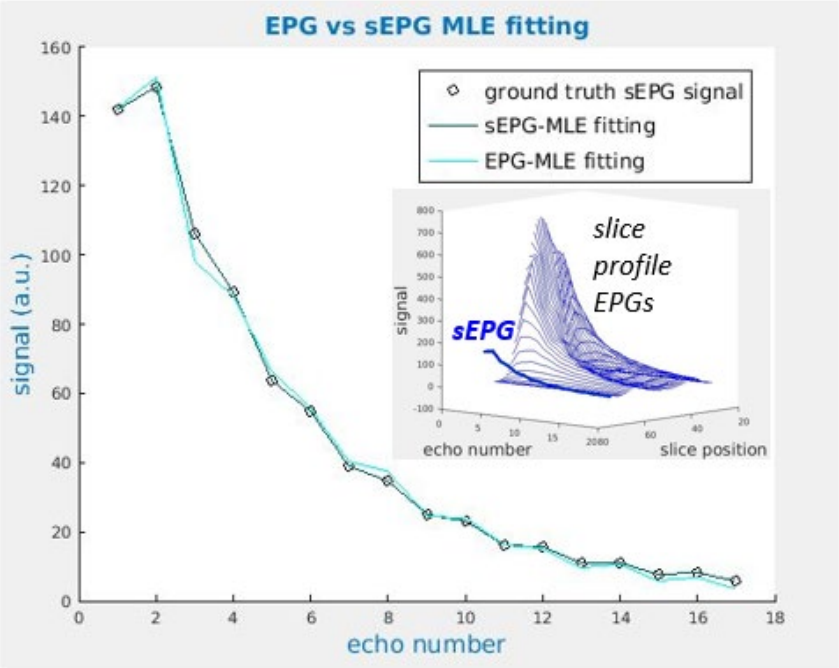


Figure 3.21: Ground truth simulated noiseless sEPG signal with 100% B_1 field factor (scaling), fitted with EPG-MLE (cyan line) and with sEPG-MLE (dark green line) models, showing how the first, non-slice selective EPG version, cannot match the ground truth signal, the unaccounted for slice profile effects also creating flip-angle and T_2 estimate errors.

The figure demonstrates that even in the case of zero noise, the best fit of the non-slice selective EPG model to the simulated signal (cyan line) cannot match the ground truth signal, and even this best fit returns an incorrect overall flip angle (or equivalently B_1) value and T_2 estimate error of about 30%; the effective flip angles are a distribution from the nominal value (in this case) in the middle of slice to

gradually reduced towards its edges. In contrast the sEPG model accounts for this distribution (figure inset) and the reconstructed signal parametrized with its estimated values (dark green line) are virtually identical to the ground truth values. The differences between the two models are further illustrated in Figure 3.22. In the case of a uniform and correctly calibrated B_1 field, the sum of the EPG signals across the slice will include perfectly excited and refocused signals from the middle of the slice as well as a variety of coherences from Hahn and stimulated echoes due to suboptimal flip angles towards its edges (Section 2.1.8). Typical features of the resulting echo train signals are the relative elevation of all echoes relative to the 1st, with signal elevation remaining apparent towards the end of the echo train, and the alternate echoes amplitude modulation effect. This alters significantly the shape and curvature of the signal evolution with time, compared with the exponential decay resulting from the simple EPG model with perfect excitation and refocusing flip angles. The contribution of Rician noise in amplitude reconstructed images further alters the signal's apparent evolution due to its additive behaviour particularly evident towards the end of the echo train as SNR decreases. When it comes to model parameter estimation, there is a potential interaction between the sEPG signal and the noise contribution. In the simulation here this was seen to contribute variably due to random noise contributions occasionally mimicking features of a plausible EPG signal. In addition in the sEPG case the SNR of the later echoes relatively to the first is substantially higher than the of the EPG case (Figure 3.22, lower graph, inset), which is of note since we have adopted the convention of defining the overall SNR as the ratio of the highest (generally first or second) echo amplitude to the noise level (σ_N) added for each set of simulations.

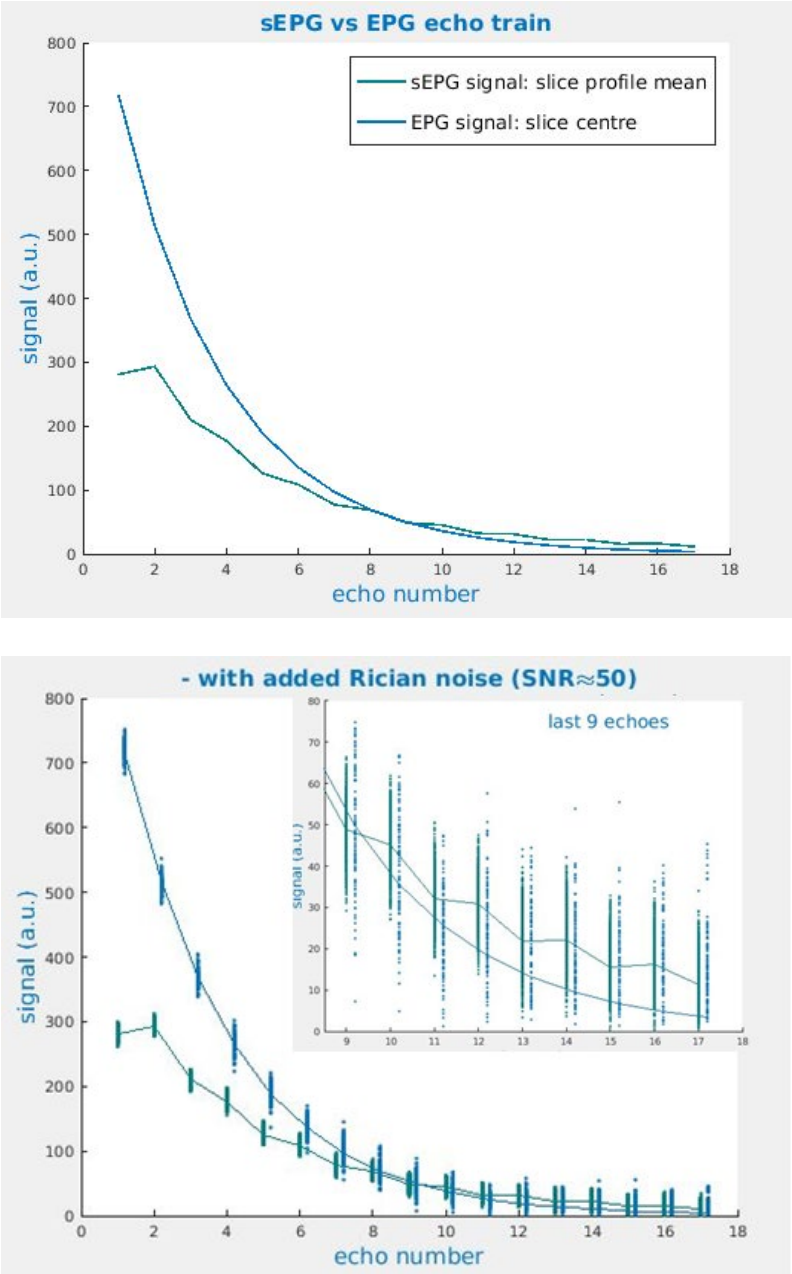


Figure 3.22: Behaviour of the sEPG model: the slice profile effects on the EPG signal, without (top panel) and with (bottom panel) added Rician noise. The slice profile effects clearly increase SNR markedly towards the end of the echo train.

Therefore sEPG signal data sets compared to EPG signals of the same nominal SNR will exhibit significantly higher SNR at their later echoes.

In order to more comprehensively characterize sEPG parameter estimation, the simulation experiments conducted at this stage were expanded to sets of 10000

repetitions for 10 different noise levels (SNR from 50 to 3). The workflow algorithm for signal synthesis and MLE fitting is shown in Figure 3.23. For comparison, LSQ parameter estimation was additionally performed: because the sEPG formalism is in essence a complex number problem (section 2.1.7), it was necessary to use the Levenberg-Marquardt algorithm which cannot handle bound constraints (Matlab, MathWorks, Massachusetts, United States). Initialization parameter values for the LSQ routine were identical to those used for the MLE calculations.

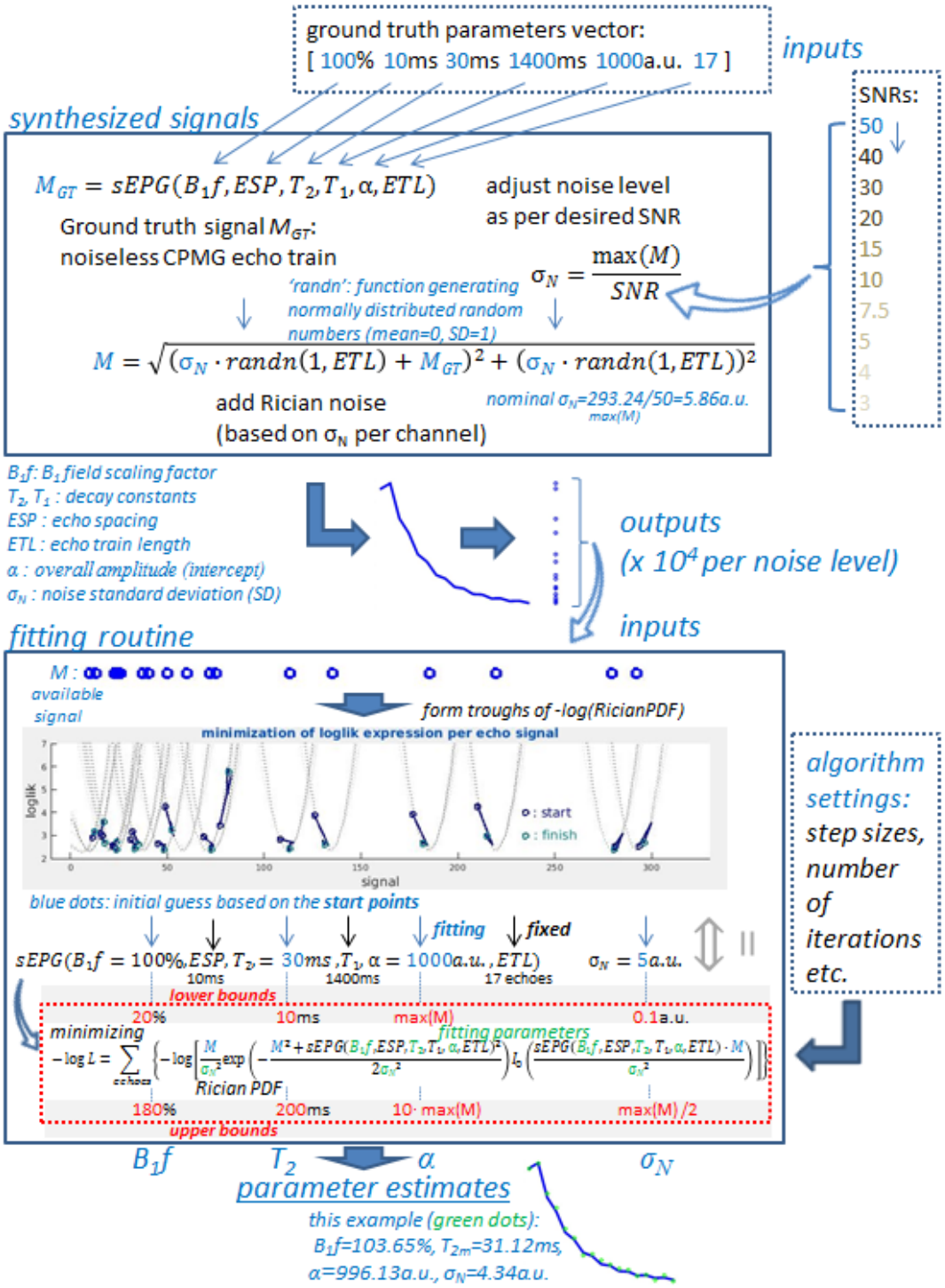


Figure 3.23: A schematic illustrating the implementation of the simulations parameter estimation tests at this stage. Signals are synthesized by adding Rician noise to the sEPG-produced echo train (upper box), and subsequently fed into the fitting routine (lower box) where their joint negative log-likelihood is minimized in order to yield the (MLE) parameter estimates, subject to specified start values, bound constraints and algorithm settings.

LSQ and MLE fitting to sEPG model simulated data with Rician noise are compared for the case of SNR 10 in Figure 3.24. The marked elevation, relative to the ground truth values, of the reconstructed signal based on both LSQ and MLE estimated parameter means (upper graph) is due in part to the estimates' skewed distribution. This is mitigated to an extent by using estimate distribution modes or medians (middle and lower graph) to reconstruct the signal. The parameter estimate distributions for this SNR = 10 case are shown in Figure 3.25, their general asymmetry supporting these observations. The highest estimate skewness is observed for the amplitude α, showing a prolonged upper tail, presumably explaining in a large part the reconstructed signal departures seen in Figure 3.24(a). This amplitude distribution upper tail feature is correlated with the opposite skewness in the estimated T_2 distributions. In Figure 3.26, it is shown that α estimates above 1500 a.u. are associated with low T_2 estimates, and account almost entirely for T_2 estimates less than 20ms in the complete dataset. Removing those amplitude estimates effectively removes the lower tail in the estimated T_2 histogram. This observation is consistent with effects seen in fitting the simple exponential model discussed in section 3.2.3 (Figure 3.14) where random noise can elevate initial echoes and/or diminish later echoes relative to their corresponding ground truth values in a way that causes the decay to appear markedly faster than it would in the absence of noise. Comparison with the respective LSQ estimates (excluding the same subset of α estimates) reveals that MLE T_2 distributions are tighter and closer to the ground truth value (bottom graphs).

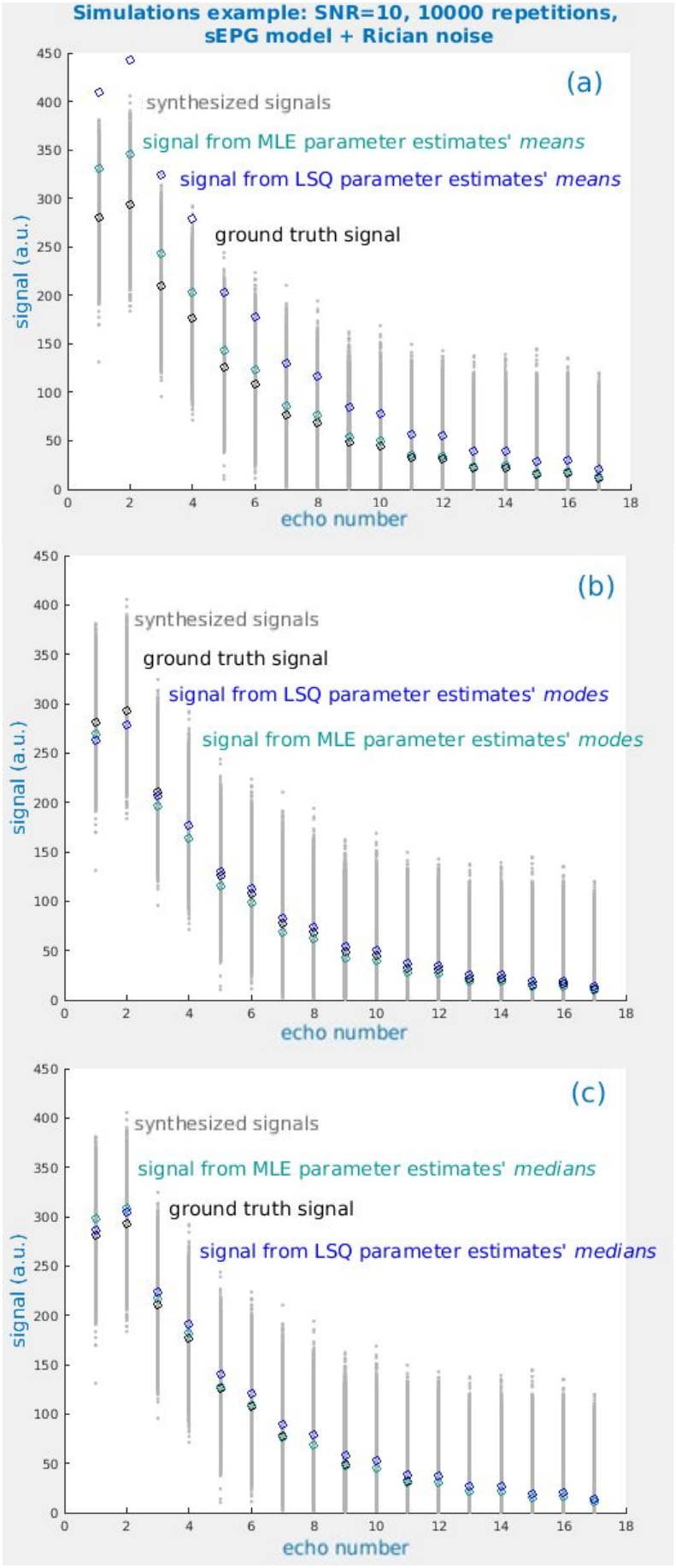

Simulations example: SNR=10, 10000 repetitions,
sEPG model + Rician noise
(a)
synthesized signals
signal from MLE parameter estimates' means
signal from LSQ parameter estimates' means
ground truth signal
signal (a.u.)
echo number
(b)
synthesized signals
ground truth signal
signal from LSQ parameter estimates' modes
signal from MLE parameter estimates' modes
signal (a.u.)
echo number
(c)
synthesized signals
signal from MLE parameter estimates' medians
ground truth signal
signal from LSQ parameter estimates' medians
signal (a.u.)
echo number

Figure 3.24: Example of a set of 10000 simulations for an sEPG model with Rician noise yielding SNR of 10, ground truth parameters: amplitude α=1000 (arbitrary units), T_2=30ms, 100% B_1 field factor). Grey dots represent the synthesized signals and black circles the ground truth signal values. Top (a): green circles represent a signal reconstructed using the means of the MLE parameter estimates, blue circles the signal from the means of the LSQ parameter estimates. Middle (b): green circles represent a signal reconstructed using the modes of the MLE parameter estimates, blue circles a signal reconstructed using the modes of the LSQ parameter estimates. Bottom (c): green circles represent a signal reconstructed using the medians of the MLE parameter estimates, blue circles a signal reconstructed using the medians of the LSQ parameter estimates. Re-constructed signals based on the MLE parameter estimates deviate in general markedly less than from the equivalent LSQ estimated parameters.

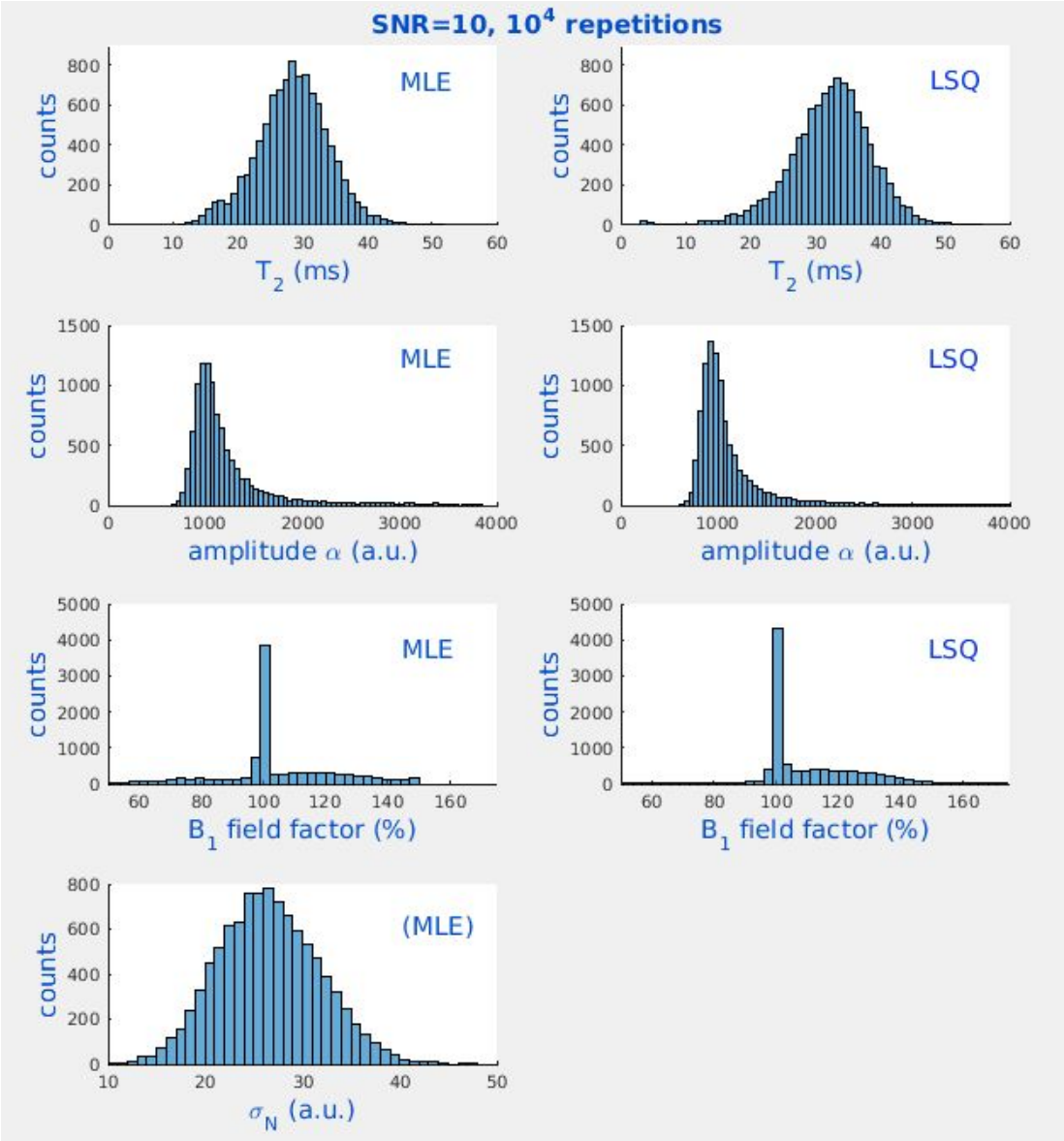


Figure 3.25: Histograms showing distributions of the parameter estimates for the sEPG model of Figure 3.23 and Figure 3.24 with added Rician noise (SNR = 10), which appear non symmetrical (skewed) in general (extreme values/outliers were omitted for clarity).

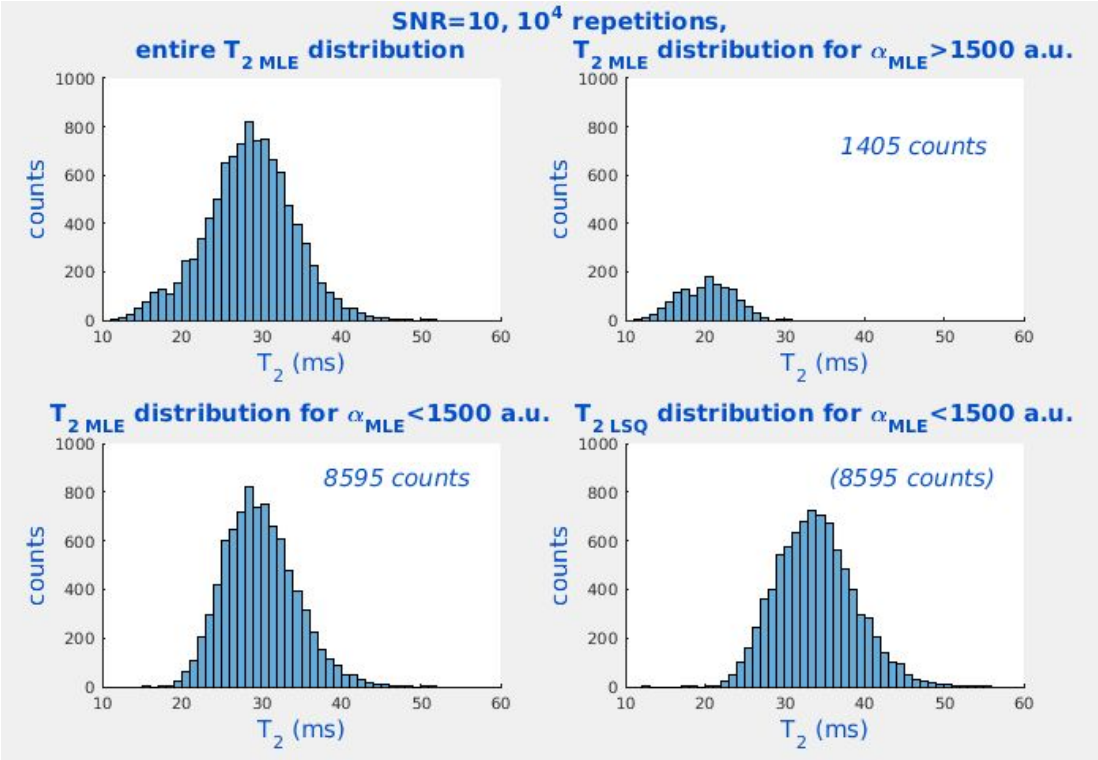


Figure 3.26: T_2 estimate histogram distributions corresponding to all 10^4 SNR 10 replicate EPG signals (top left), to the subset of replicate signals yielding α estimates above 1500a.u. (top right) and the remaining subset yielding estimated α below 1500a.u. (bottom left) and the latter's comparison with LSQ T_2 estimates (bottom right) as compared to the entire estimated T_2 distribution (top left). The fits yielding amplitude estimates above 1500a.u. are almost entirely those yielding T_2 estimates below 20ms.

In Figure 3.27 an example sEPG model simulation demonstrating the above effect is shown. Due to the random added noise the initial two echoes were in this case higher than the ground truth signal while the next three (3rd, 4th end 5th) conversely by chance were markedly lower, and this combination causes a significant T_2 underestimation with both MLE and LSQ methods.

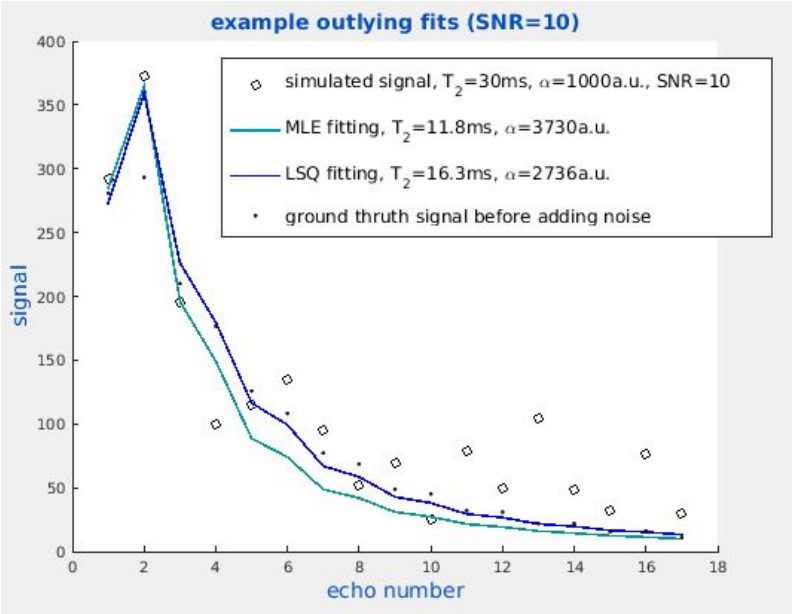


Figure 3.27: An example fit (simulation repetition) returning spuriously high amplitude (intercept) and relatively low T_2 estimates, due to the added noise distorting echo heights in a manner that causes acceleration of the apparent signal decay.

Interestingly in this case LSQ actually returned parameter estimates closer to the ground truth due to the relatively high added noise: this additive Rician noise is, under the LSQ assumptions (Gaussian homoscedastic noise), incorrectly attributed in part as elevated signal, and this partly compensates for the apparent accelerated decay early in the echo train. MLE properly accounts for the Rician noise and is affected only by the intrinsic bias described here related to the steepness of specific echo train formations as in the above paradigm.

A sub-population of such fits manifesting low T_2 and simultaneously high α values existed for each noise level. The value of 1500a.u. can be chosen (as above) as an approximate lower bound indicator for this pool of fits because in general amplitude estimates did not fall below 500a.u. at the opposite end (left tail) of the distribution – only a very small percentage (up to 1.5%) for the last three highest noise levels. As is shown in Table 3.1 for MLE, the number of such fits approaches half of the cases

for the lowest SNR, while for LSQ it also increases for SNR down to 7.5 almost in parallel with MLE, largely comprising of the same synthesized signals.

Table 3.1: The sub-populations of simulations (out of a total of 10^4 signal replicates) returning amplitude estimates above 1500a.u. grow larger with increasing noise for both the MLE and LSQ methods, while largely overlapping. However the noise-distribution-related bias for LSQ effectively starts compensating this type of fitting error and eventually dominates results for lower SNR (below 10).

SNR	***50***	***40***	***30***	***20***	***15***	***10***	***7.5***	***5***	***4***	***3***
α_{MLE}***>1500a.u.***	0	1	8	149	548	1405	2233	3459	4210	4835
α_{LSQ}***>1500a.u.***	0	1	10	171	517	1071	1300	1135	905	651
α_{MLE}***>1500a.u.*** $\cap$ α_{LSQ}***>1500a.u.***	0	1	7	126	444	1022	1277	1107	850	532

Then the opposite bias introduced by the increasing Rician noise floor – as it appears for the LSQ minimization – reduces the effect and dominates the LSQ fitting process as SNR further decreases.

The above trends can be also seen in the overall results presented in Figure 3.28 and Figure 3.29, as a relatively stable small negative bias for the first 3 or 4 high SNRs which continues in a similar fashion for MLE but eventually reverses towards the opposite direction (overestimation) for LSQ. Distribution medians (including quartiles and outliers) appear to more accurate characterise the results as means (and SD) can take extreme values, e.g. top right graph for LSQ, and modes can collapse for extremely skewed distributions, e.g. middle left graph for MLE where the mode becomes the lower T_2 bound of 10ms (Figure 3.23) for very high noise. For those reasons and because the SNR range of interest mostly seen in our *in vivo* data sets

was predominantly between 50 and 10 (section 3.1) the results for that range (50>SNR>10) are also shown in closer detail in Figure 3.29 – consistent also with the type of figures in previous sections for completeness. In this case there are no extreme values or estimate values restricted by bounds imposed in the fitting process. While trends can be seen to be the same an additional feature more clearly seen in this figure is that down to SNR=15 LSQ appears to be superior overall in terms of T_2 estimate bias. This is visible in the graph for means, and also true for medians (Table 3.2).

Table 3.2: The MLE and LSQ estimated T_2 medians for 50<SNR<15 with LSQ estimates being closer to the ground truth value of 30ms.

SNR	***50***	***40***	***30***	***20***	***15***
$T_{2\ MLE\ (}ms)$	29.698	29.633	29.504	29.263	29.055
$T_{2\ LSQ}\ (ms)$	29.790	29.794	29.830	30.067	30.607

The initial negative bias similar to that seen with MLE becomes gradually compensated due to the LSQ Rician noise-related opposite bias, such that all metrics (means, modes and medians) cross the ground truth value at SNRs between 30 and 20, suggesting an erroneous apparent superiority for LSQ over MLE for this range.

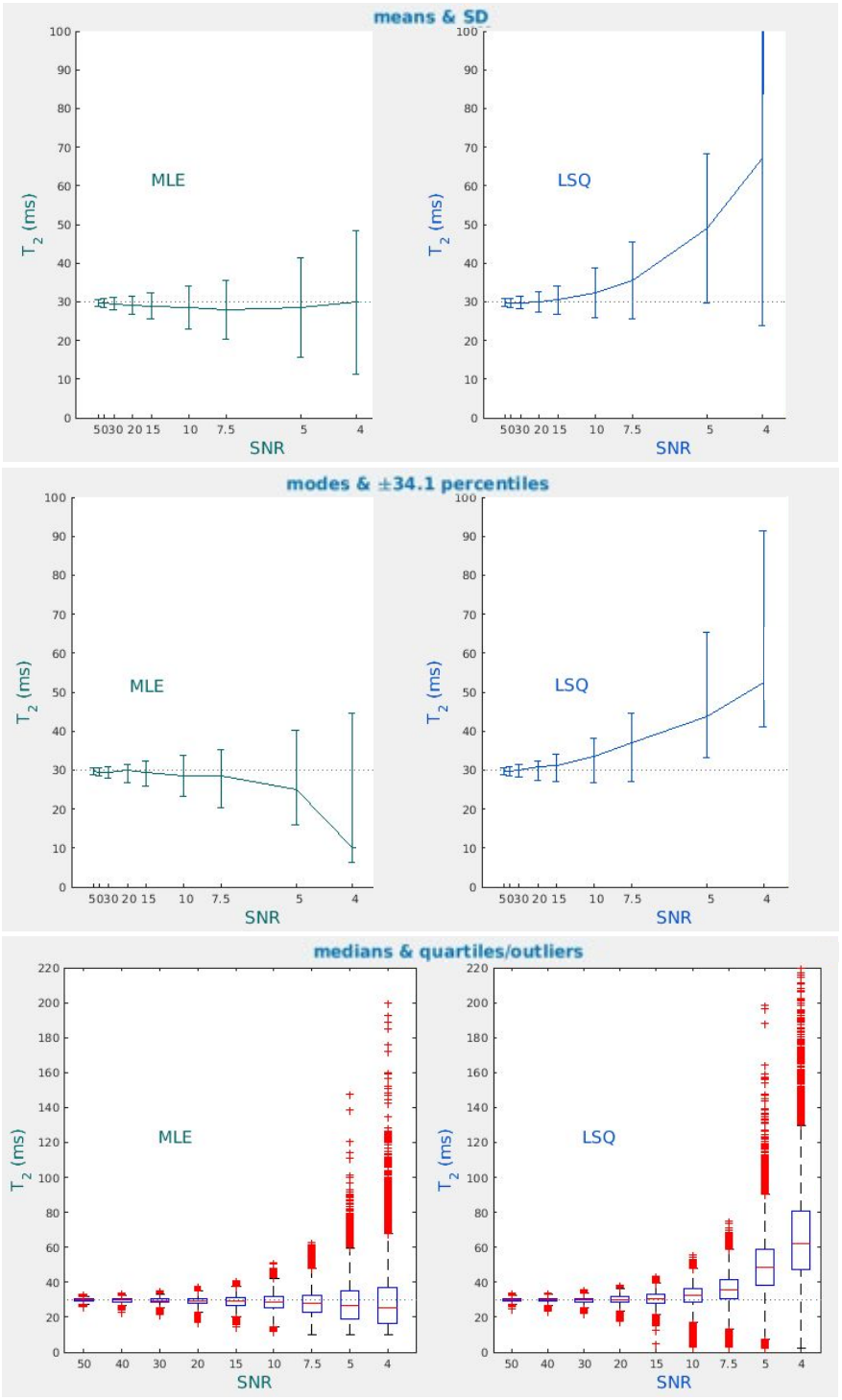


Figure 3.28: Comparison of T_2 estimates from 10^4 simulations presented in terms of means and SD (top row), modes and percentiles (middle row) and medians and quartiles (bottom row) of results returned by LSQ and MLE fitting to simulations of sEPG modelled signal (amplitude α=1000 units, time constant T_2=30ms, B_1 field factor = 100%) with Rician noise producing varying SNR from 50 to 4. There is in general less bias in MLE estimated T_2 in comparison with LSQ T_2 estimation while choice of the specific statistical metrics to be used (i.e. means and SD or otherwise) markedly affects the way that results are presented.

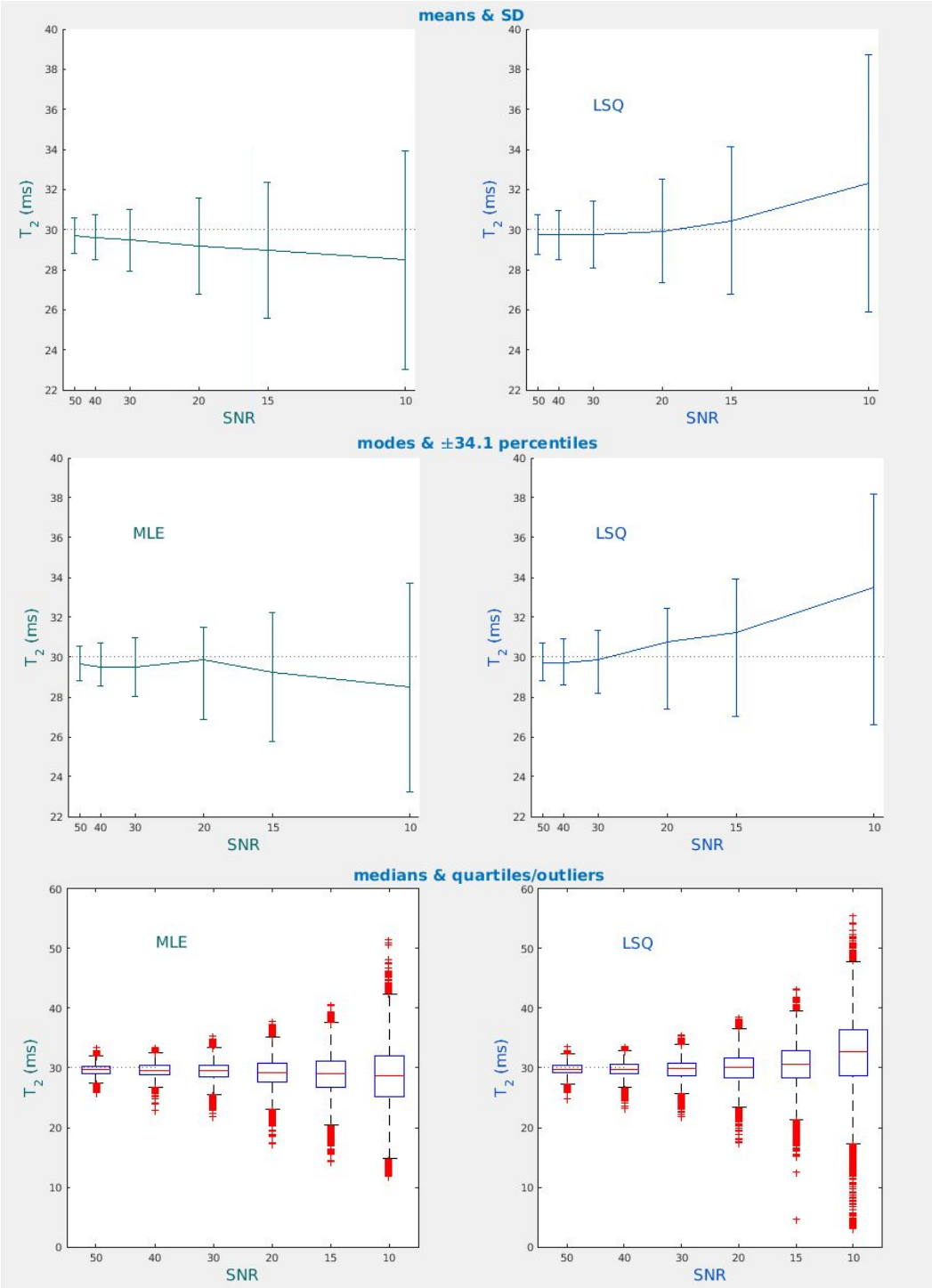


Figure 3.29: Comparison of T_2 estimates with the use of means and SD (top row), modes and percentiles (middle row) and medians and quartiles (bottom row) on results returned by LSQ and MLE fitting on simulations of sEPG modelled signal (amplitude α=1000 units, time constant T_2=30ms, B_1 field factor = 100%) with Rician noise producing varying SNR from 50 to 10 (10^4 repetitions each). Both methods exhibit similar negative model-related bias for initial high SNR, however with increasing SNR for LSQ fitting the effective positive bias in estimated T_2s due to the rising apparent Rician noise floor in that case reverses this trend.

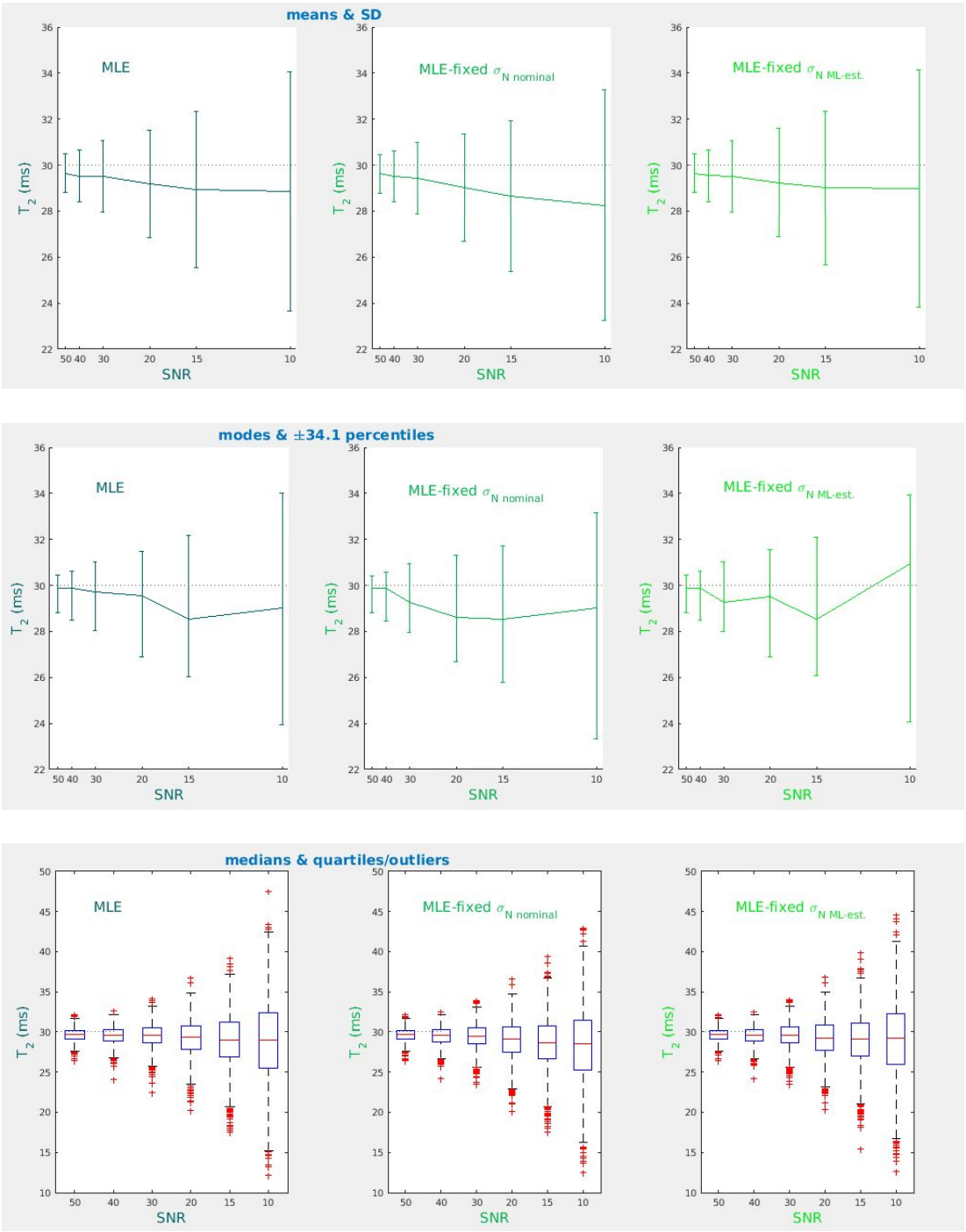


Figure 3.30: Comparison of T_2 estimates with the use of means and SD (top row), modes and percentiles (middle row) and medians and quartiles (bottom row) on results returned by MLE fitting on simulations of sEPG modelled signal (amplitude α=1000 units, time constant T_2=30ms, B_1 field factor = 100%) with added Rician noise of varying SNR from 50 to 10 (10^3 repetitions each). On the left column noise σ_N was an estimated parameter, on the middle column it was fixed at its nominal value during the fitting process and on the right column it was fixed on the previous mean MLE σ_N estimate for each SNR. In both cases using fixed noise σ_N was not seen to improve results.

The behaviour of MLE fitting in the case that the noise σ_N in the Rician pdf is not a parameter to be estimated but rather is known *a priori* and fixed is shown in Figure 3.30. The first 1000 synthesized signals from the previous experiment were fitted while fixing σ_N at both its nominal ground truth value of the noise added to each channel (middle column) and the mean value yielded from the previous simulations when it was fitted as a free parameter by the MLE method (right column) – for those first 1000 counts. The mean of the MLE estimate was used as its distributions were seen to approach a close to normal (symmetric) shape in general (last graph of Figure 3.25). In both cases fixing the noise was not seen to offer any significant improvement overall in bias and precision when compared to the results from the case where it was also estimated from the data (left column), once again demonstrating that the proposed MLE method does not require input from separate noise σ_N determination methods (Walker-Samuel et al., 2010).

3.2.6 Behaviour and fine tuning of the single component sEPG-MLE algorithm

Having established in principle a working algorithm to fit CPMG signals with the use of the sEPG formalism accounting for the presence of Rician noise, we then have to investigate its effectiveness and reliability under varying real-world conditions. A first step in that direction is to examine the dependence of the parameter estimate bias, if any, upon the parameter values across the practical range of interest for each. We can next consider strategies to implement the method, and potentially ways for further optimisation in the context of real-world data.

The previous experimental work suggested the existence of bias affecting the set of model parameter estimates with different underlying causes which are likely to dynamically interact. In this section a more detailed investigation of the behaviour and limitations of the fitting routine will be presented, leading to further improvements in the method. For this section, results are presented in terms of medians and percentiles via box and whisker plots only, for the main range of interest of SNR between 50 and 10.

Parameter start points (initial values) and bound constraints, imposed when initialising the fitting routine, directly influence the algorithm speed and ability to converge to an accurate parameter estimate vector, in particular with regards to a global *vs.* local minimum. The general approach with regards to setting parameter constraints was empirical such that the fitting proceeded reliably producing physically meaningful results. There are 4 possible types of fitting outcomes, returning parameter estimates corresponding to: i) a global minimum, ii) a local minimum, iii) the bound values or iv) failure to converge. The aim of the work now presented was to develop strategies favouring convergence to global minima, avoiding as far as possible the remaining suboptimal outcomes. Estimates returned at the bound values can serve as alerts for cases where the fitting fails or is suboptimal. This may occur for example when parameter start points are far away from the true values or data are very noisy and/or for physical reasons ill-posed e.g. very low B_1 field. For those cases bounds also had to be kept reasonably wide, in order to retain the possibility of converging to the correct values. The $B_1 f$ parameter

estimation is particularly sensitive to these factors, and it is also strongly affected by the existence of local minima as will now be explained.

Since the loglik space is multidimensional (with number of dimensions equalling the number of estimated parameters) this work is based on studying variations along one dimension with the other parameters at fixed or narrowly confined, representative physically realistic points, mainly focused on the range of interest for T_2, and for multi-component models (section 3.3) also for *ff*. A full study would require a multidimensional analysis, however the findings presented next are sufficient to characterise the key functionality of the proposed method and physical meaningfulness of the yielded results. The existence of a physical model predicting the shape of the loglik surface with respect to B_1f, based on prior knowledge from MR theory, is particularly relevant in this analysis.

3.2.7 The role of B_1f in the fitting process

Effects of start points (parameter initialization)

The B_1f parameter variations appear to have a major impact upon both the signal construction (in simulation) and its fitting via the behaviour of the sEPG model in the minimization algorithm. As a percentage scale factor accounting for B_1 transmit inhomogeneity it scales the magnitude of the RF pulse, taking a value of 100% for a slice profile with 180° effective refocusing angle at its centre. Depending on the specific scanner software and settings there might an additional scale factor between the excitation and refocusing pulse amplitudes. B_1f thus combines with the RF pulse waveshape to determine the precise excitation and refocusing flip angles across the

slice profile. Previous work fitting slice profile corrected models constrained B_1f as 0 < B_1f < 100% (Lebel and Wilman, 2010, Petrovic et al., 2015); i.e. it was assumed that for example a nominal 170° refocusing pulse produces the same echo train as a 190° pulse. Later work showed that this assumption fails for slice-selective pulses (McPhee and Wilman, 2017). The existence of closely similar solutions for B_1f values symmetric about 100% causes, particularly in the presence of noise, redundancy and hence instability in the parameter estimates. Since B_1f equally above or below the correct value produces similar but not identical signal models (Figure 3.31) this may lead to errors in the final parameter estimates. This type of B_1f variation has therefore been considered a potential confound in relevant published work (Lebel and Wilman, 2010, Azzabou et al., 2015). Hence characterizing the role of B_1f is a key point in optimising the fitting process.

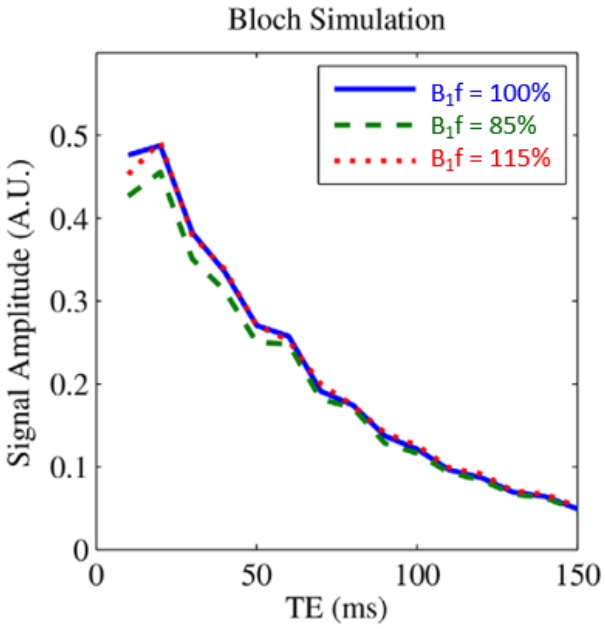


Figure 3.31: T_2 decay curves with T_2=50ms and B_1f 85, 100 and 115% computed using a slice selective Bloch equations based simulation showing asymmetry in the signals produced for B_1 amplitude symmetric about the ideal value (McPhee and Wilman, 2017).

It is informative for that reason to first examine the case of a noiseless signal and investigate the effect that changing the B_1f initial value for the iterative fitting routine has on the final parameter estimates. In Figure 3.32 this behaviour is shown for the case of a noiseless single sEPG signal with an 80% B_1f ground truth value. This corresponds to experimental cases where a B_1 field deviation is present (e.g. RF field spatial inhomogeneity) but at the same time SNR is overall very high.

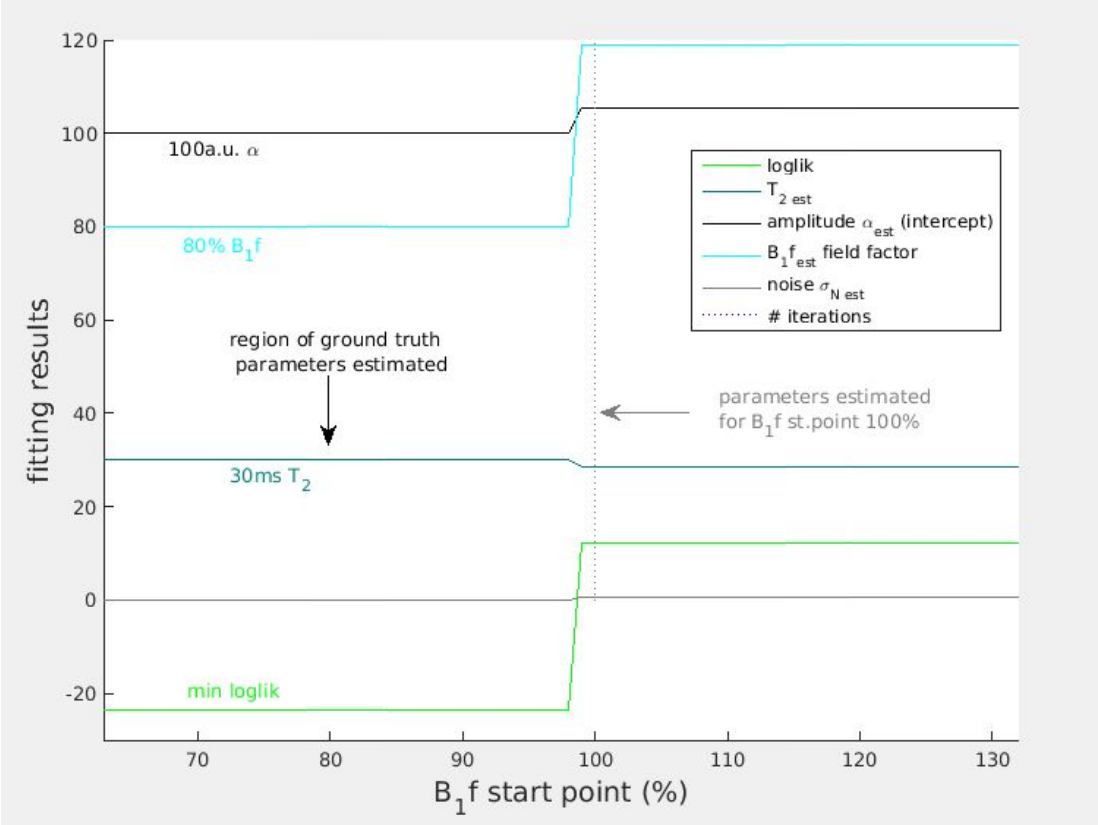


Figure 3.32: The effect of B_1 field factor start point (initial guess) on the sEPG-MLE minimization, with all other start point and parameter bounds kept equal. If, as in this noiseless signal case, the ground truth B_1f was 80% and we start with an initial value of 100% (or higher), spurious incorrect parameter estimates will be returned (indicated by the grey dotted line).

In this example it can be seen that in the B_1f starting value region around 80% at convergence the global minimum log-likelihood (loglik) is reached and the ground truth parameters are returned. However it can also be seen that there exists a region

for B_1f start values above 100%, centred around 120%, where the algorithm converged such that loglik reaches a secondary local minimum, and here the resulting parameter estimates deviate from the ground truth. Using for example a B_1f 100% initial value in the algorithm (as a natural choice if a single common value is required for all pixels across an entire image) would return these incorrect parameter estimates – even if the estimation precision is very high due to the lack of noise. The algorithm returns in that case an erroneous B_1f estimate of about 118.9%, corresponding to a B_1 deviation from 100% of opposite sign to the ground truth, i.e. one with similar absolute error but above rather than below 100% (in this example close to 120% rather than the correct value of 80%). Due to the non-symmetrical slice profile behaviour – as also explained in section 3.2.5 – converging to this opposite polarity (but not equivalent) B_1f error value results in erroneous values also for the remaining parameters, e.g. significantly underestimated T_2 of 28.5ms *vs.* the 30ms true value. Notably, solutions converging to the non-global minimum yielded non-zero values for the model noise σ_N, contrasting with the zero noise ground truth in this case.

In Figure 3.33, for the range of B_1f initial values, the corresponding steps of the log-likelihood iterative minimization, and final number of iterations needed, are shown. The minimization process is tracked by plotting the entire loglik curve as a function of B_1f initial value (top graph) as it proceeds through the minimization: each curve presents the current loglik for a specific iteration step, for minimisations initialised with B_1f start values along the x-axis.

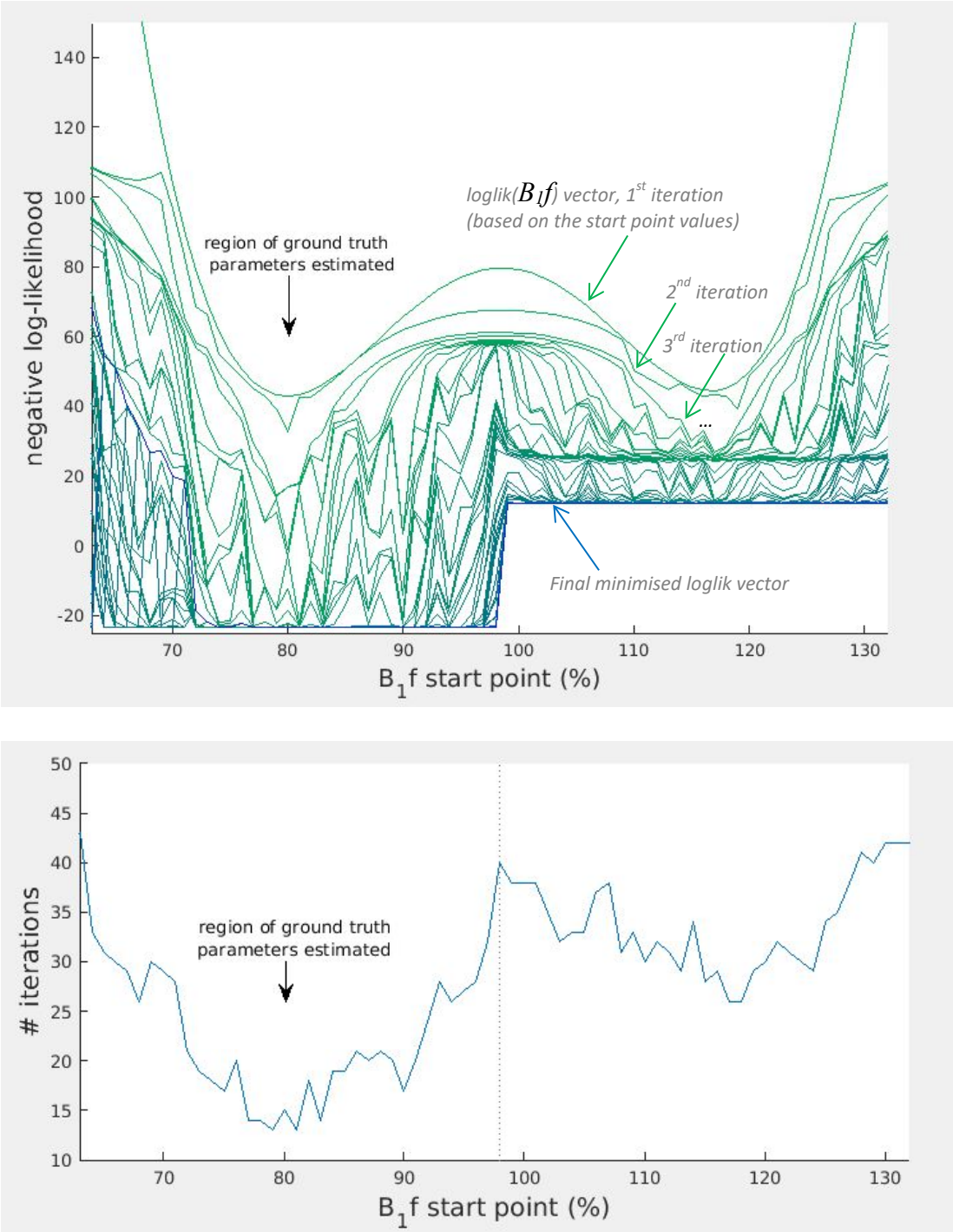


Figure 3.33. Above: behaviour of the loglik minimization from its initial loglik(B_1f) vector (green curve) for a range of B_1f values as start points, to the final minima showing a two-stepped behaviour: the global minimum range around the true (80%) value and a local minimum around the opposite region of B_1f around 120% (bottom blue line). Below: the corresponding number of iterations. It can be seen that the most rapid convergence is achieved at the global minimum region and close to the ground truth value of 80%.

While convergence was most rapid in the global minimum region (bottom graph),

where initial values were close to the ground truth B_1f value (80%), within the non-

global local loglik minimum region for B_1f > 100%, minimization was also relatively fast close to the near-symmetric opposite final estimate of 118.9% - the fastest for the upper B_1f region. This is a useful piece of information to be used next to further improve the fitting routine.

The above experiment demonstrates that model-related systematic parameter estimate bias, due to the RF pulse non-symmetric behaviour for scaling imperfections above and below the optimal (close to 100%) B_1f may exist. This is due to the non global, local minimum that the log-likelihood may fall into as a consequence of the chosen start values and bound constraints. A common way to address the existence of local minima in a test statistic minimization process is to adequately grid (scan) initial values across the associated parameter space (Bevan, 2013), in this case the entire range of possible B_1f initial values.

However, in the present case, we have *a priori* from MRI theory that there will exist both the global minimum and a secondary non-global minimum in the loglik parameter space, with these approximately symmetrically positioned about B_1f = 100%, and the apparent two, upper and lower B_1f regions are clearly separated. The correct parameter estimates are those corresponding to the lowest final loglik in turn closest to the ground truth values. Taking together the above ideas – including the minimum iterations information – a more efficient fitting strategy than scanning across the entire range of B_1f initial values for every pixel to be analysed can be constructed by making use of the initial loglik turning points. In Figure 3.34 it is shown that the position of both the loglik central turning point and the two minima with respect to initial B_1f depend on the choice of the other parameter start point values.

In addition, the central turning point may not necessarily be a central local minimum as it can coincide with a single minimum, or a wider plateau-shaped minimum-region between the local minima may exist (right panel).

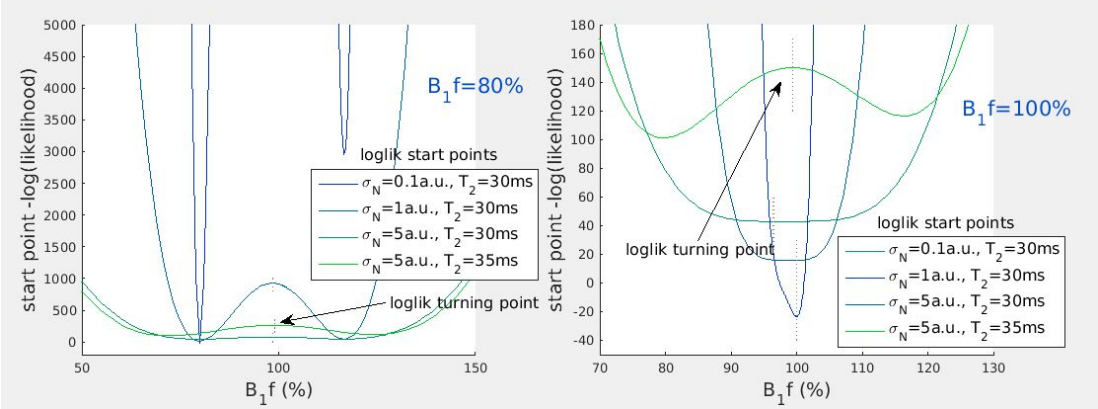


Figure 3.34: The shape of the initial loglik vs B_1f for different sets of the remaining parameter start point values, for a noiseless signal of 30ms T_2 and B_1f 80% (left) or 100% (right). The separation of minima, the existence of a secondary minimum and the position of the provisional minima and turning points is influenced by the choice of parameter start points.

Therefore to avoid converging into the incorrect non-global local loglik minimum with respect to B_1f, the conventional approach of repeating the entire fitting routine for a comprehensive range of plausible B_1f initial values can be reduced to testing the two most probable candidates, the B_1f initial values corresponding to the two minima (global and local) if these are separated by a loglik central turning point. An additional step must be taken otherwise, for the case that a single minimum emerges for the initial loglik, because it can be truly single – as is the case for the 100% B_1f shown at the right graph of Figure 3.34 – or only apparently single, driven by parameter start values substantially different from the ground truth. This is manifested by one or both minima coinciding with the turning point. In that case updating the provisional loglik with use of the previous step's parameter estimates resolves the ambiguity: the recalculated loglik will subsequently lead to resolution to either two minima (global

and local) or the global minimum and the turning point value again (updated); when the latter happens, because the loglik minimisation has already reached the true single global minimum, the local minimum/turning point value is also very close to that and slightly higher. In both cases the global minimum can be identified through its lower loglik. This strategy can be seen as a special, more efficient (condensed version) of the well-known parameter grid (or parameter scan) technique previously mentioned. A schematic representation of the process is shown in Figure 3.35. Examples of implementing this strategy are shown in Figure 3.36.

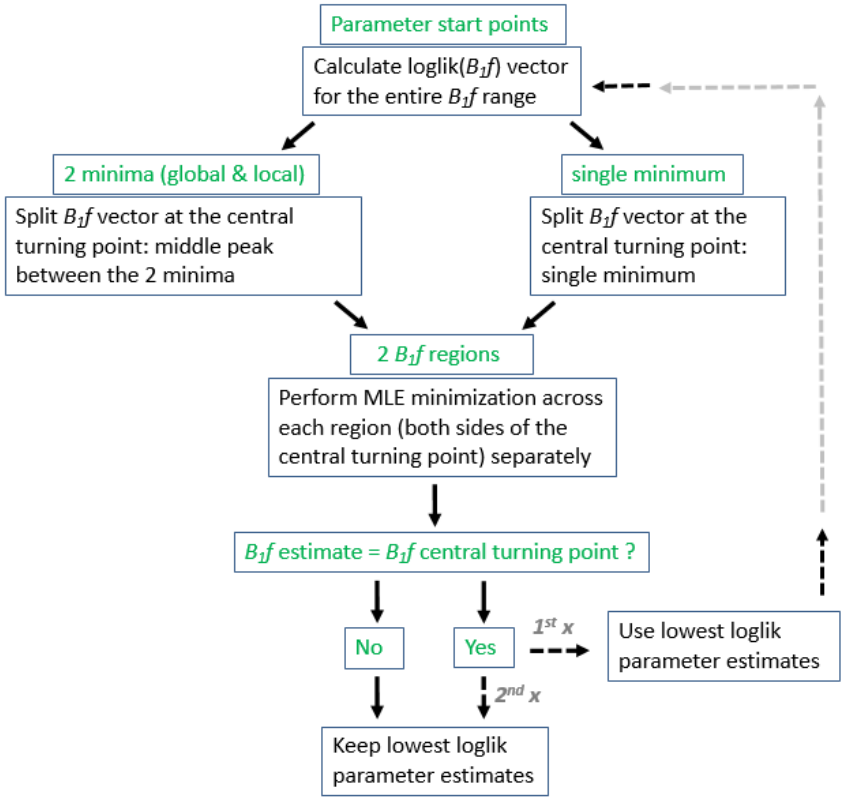


Figure 3.35: Flow diagram of the B_1f gridding strategy to efficiently determine the correct global minimum

In all cases the global minimum corresponding to the ground truth is reached, however if the initial loglik manifests a single minimum the extra step described above is required to clarify whether this is truly so also for the final (minimized) loglik.

Finally in the case of a (true) single minimum this can again coincide with the central turning point for both sides (*B₁f* regions), with the minimization producing that same result (example not shown here).

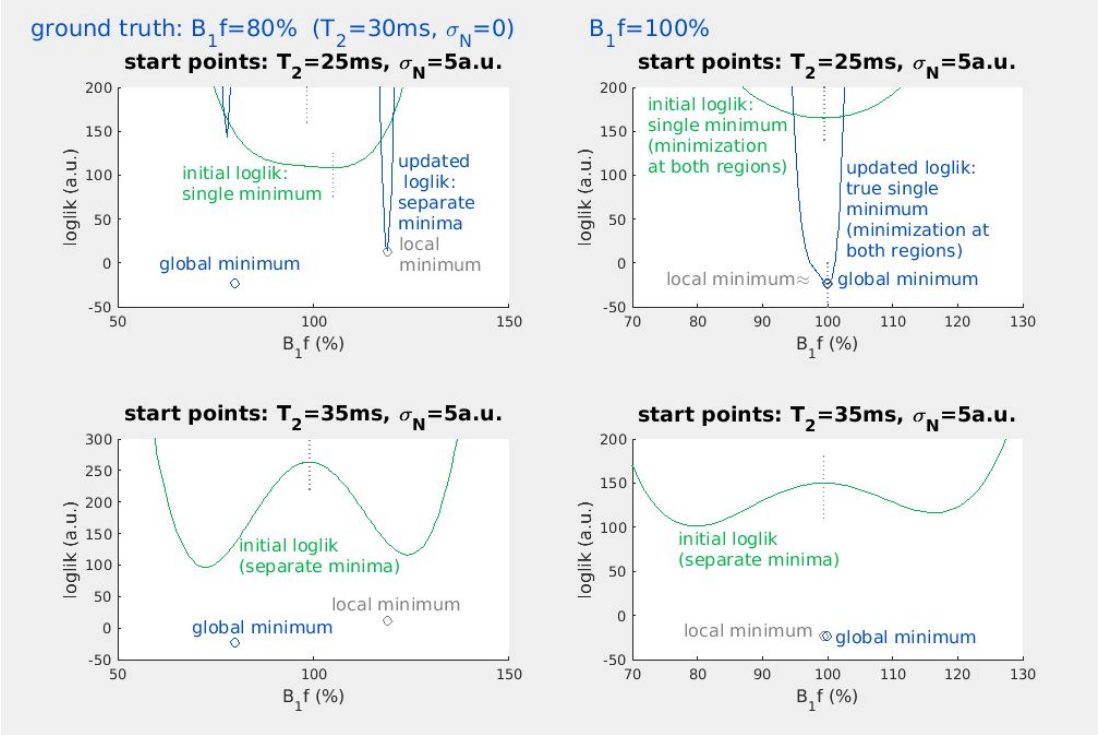


Figure 3.36: Examples of the fitting strategy for different initialization conditions for a ground truth noiseless signal of 30ms T_2 and B_1f either 80% (left) or 100% (right), similarly to Figure 3.34. The upper graphs depict the case of 25ms T_2 start point where the initial loglik showed a single minimum and by recalculating it with the results of this loglik minimization it revealed the true existence (left) or not (right) of a secondary local minimum. The lower graphs depict the case of 35ms T_2 start point where two minima where shown at the initial loglik and by minimizing it, either a true secondary local minimum was revealed (left) or the minimization for that region finished at the turning point bound, producing an apparent local minimum which similarly to a true local minimum was rejected because of its higher loglik (right).

The presence of noise in more realistic datasets can complicate the clearly separated bimodal loglik behaviour with respect to B_1f expected theoretically. Noise introduces additional uncertainty even in the case of a B_1 field without error, i.e. for ideally scaled RF pulses. This can be demonstrated with the following simulation experiment: first a synthesized image is produced using a forward sEPG model with ground truth

B_1f varying from 50% to 150%, increasing in steps of one unit per pixel along the y-axis, and T_2=30ms and amplitude α=1000a.u. To each pixel the same level of random Rician distributed noise is added for 100 repetitions along the x-axis, to produce 6 SNRs between 1000 and 100 for the B_1f = 100% pixels. This would translate to real SNR – using the previous more common definition of highest echo to noise σ_N ratio – up to 4 times lower (due to the sEPG model), with SNR further diminishing up to 3 times from the middle of the images to their edges (due to the increasing B_1f deviations).

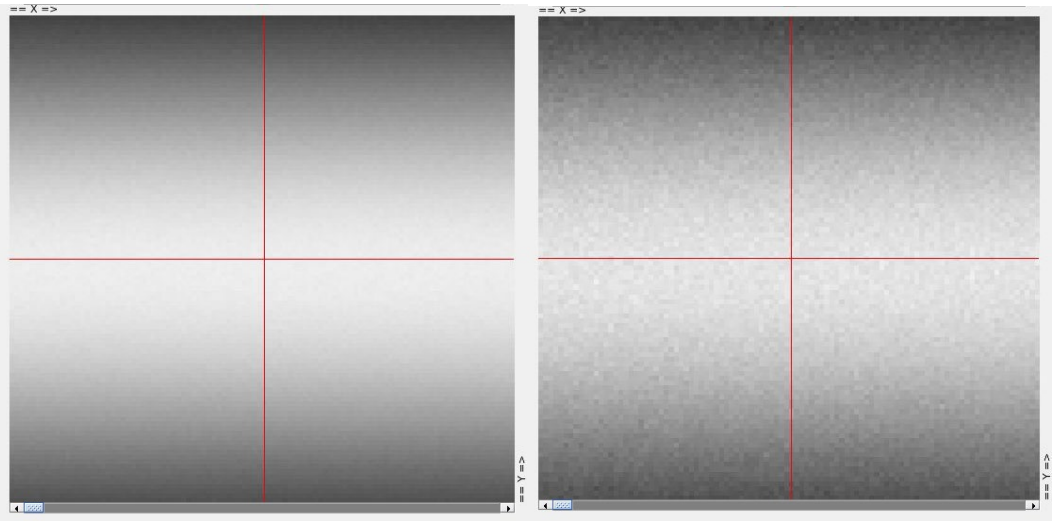


Figure 3.37: Examples of simulated images with varying B_1 field (from 50 to 150% from bottom to top) for nominal SNR 800 (left) and 200 (right).

These values were chosen to simulate real experimental conditions where the physical processes of randomly added thermal noise and B_1 field deviations are both present. Examples of two of those images are shown in Figure 3.37.

Then the fitting process described above was performed for these images. According to this strategy, for the lower part of the images (up to the middle horizontal red line representing uncompromised, 100% B_1f field) the loglik corresponding to the lower

B_1f range minimization – below the B_1f turning point and with utilization of the lower B_1f start point – would be expected to be lower, returning the correct parameter estimates. Similarly for the upper parts (above the red line) of the images, the loglik corresponding to the higher B_1f region minimization would be expected to be lower and therefore to prevail. However the presence of noise alters this behaviour and a significant proportion of simulations returned the minimum loglik for the opposite polarity B_1f error region than the ground truth. This increasing cross-B_1f-region contamination with increasing noise levels is shown in Figure 3.38.

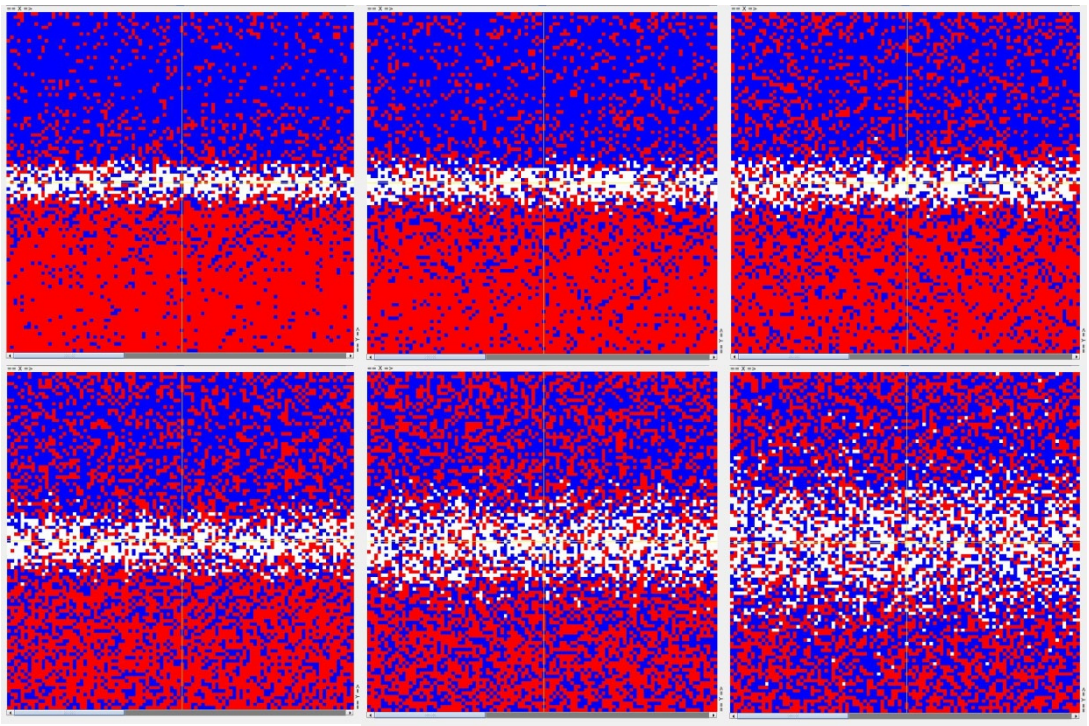


Figure 3.38: Effect of noise on the loglik minimization. In red the points for which the lower B_1f region loglik was minimum, in blue the points for which the upper B_1f region loglik was minimum and in white the points for which a single (global) minimum was produced. Nominal (intercept defined) SNR decreasing from 1000 (upper left image) to 100 (lower right image). This cross-B_1f-region contamination appears to increase with increasing noise.

The result of this noise-driven contamination of correctly estimated parameters in a region with results corresponding to the opposite B_1f polarity region is erroneous variation in the estimated parameter values. The middle bottom graph in Figure 3.38 matches well with images from real datasets, where the receive coil sensitivity is low. Its nominal 200 SNR would translate to a real – due to B_1 field inhomogeneity and noise – observed SNR of 60 to 20, as mentioned above.

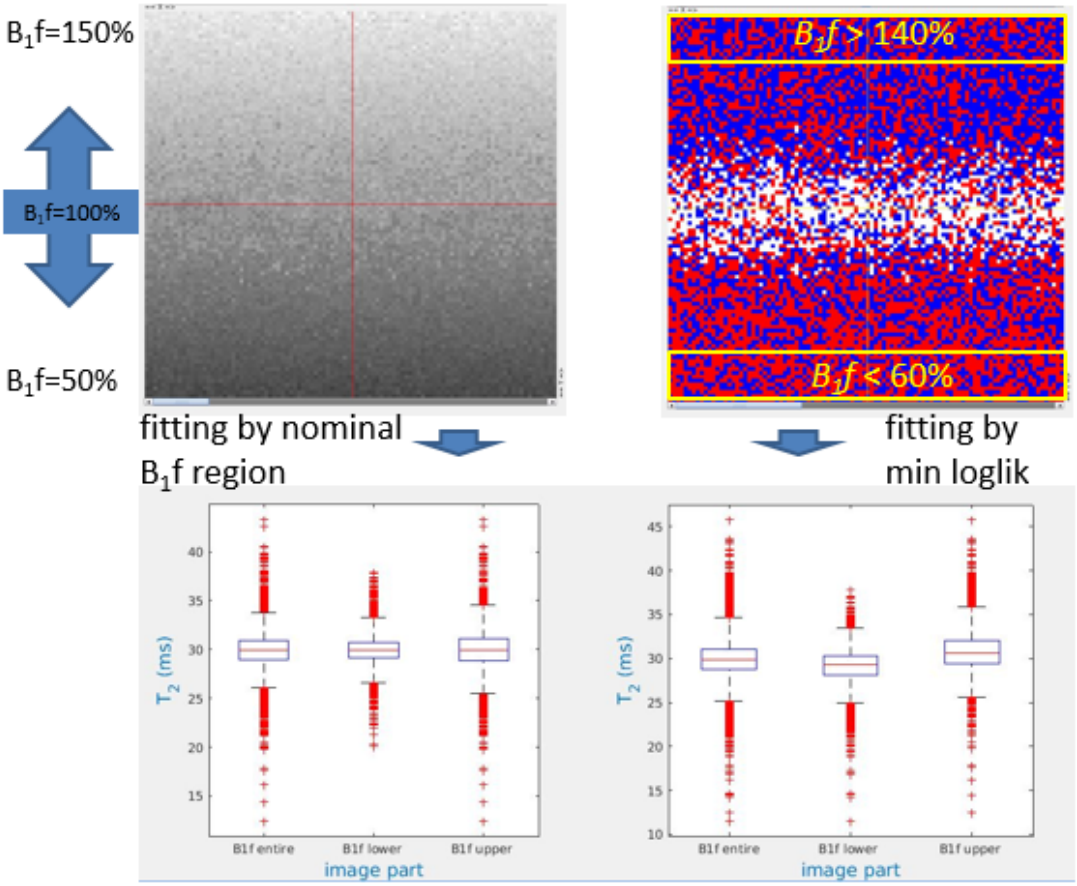


Figure 3.39: The effect of noise to the parameter estimation when this is done with use of the minimum loglik (right) compared to the estimation using the *a priori* knowledge of the correct B_1 image region (left). While the entire image's B_1f distributions are broadly similar (left boxplots for each image), there are significant differences seen when constraining B_1 factors only lower or only higher than the central turning point (approximately 100%) with use of the loglik minimization (middle and right boxplots for each image). For clarity T_2 estimates for lower B_1f < 60% and upper B_1f > 140% (areas in yellow boxes) are shown in Figure 3.40.

A comparison between the T_2 estimates obtained first with B_1f constrained to lie within the correct B_1f region versus the estimates obtained by minimizing the loglik (without B_1f constraints) is shown in Figure 3.39.

In each case all pixels across the entire image show a T_2 similar distribution centred around 30ms, the ground truth value. However in the case of using minimum loglik-based convergence with B_1f unconstrained there are systematic differences in median T_2 seen for the two constituent parts of the image, with B_1f above and below 100% respectively.

Overall these systematic median T_2 differences appear to be of the order of 1ms, however the phenomenon is more marked near the image edges rather than towards the centre. As shown in Figure 3.40 the differences seen in the loglik-estimated T_2 near this synthesized image edges, where the B_1 field factor falls below 60% or above 140% simulating a markedly compromised RF field, can reach or exceed 2ms.

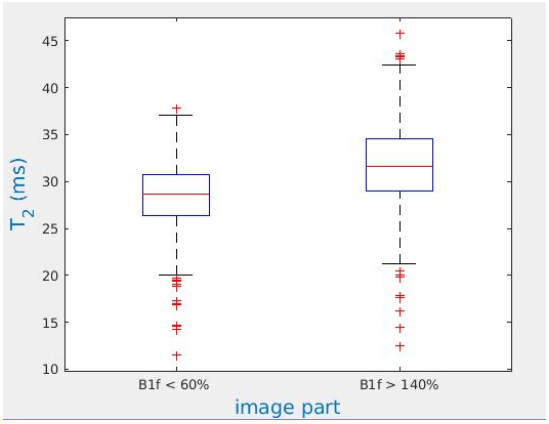


Figure 3.40: The departures in T_2 due to minimization owed to the opposite B_1f region loglik are more prolonged towards the image edges where the RF field is more compromised: T_2 results for lower B_1f < 60% and upper B_1f > 140% (areas in yellow boxes) from Figure 3.39.

In order to better understand this behaviour, it is beneficial to examine the way that the added noise interacts with the ground truth signal echo train. For that reason a similar simulation as for the previously studied image (Figure 3.39) is created, for ground truth signals of B_1f either 80 or 120% and the same levels of added noise in order to create an SNR of approximately 50. The signals returning the lowest loglik for the correct B_1f (global minimum) in each case were compared with those returning the lowest loglik for the opposite region B_1f (local minimum), for each of the 2 B_1f cases (80 or 120%) (Figure 3.41).

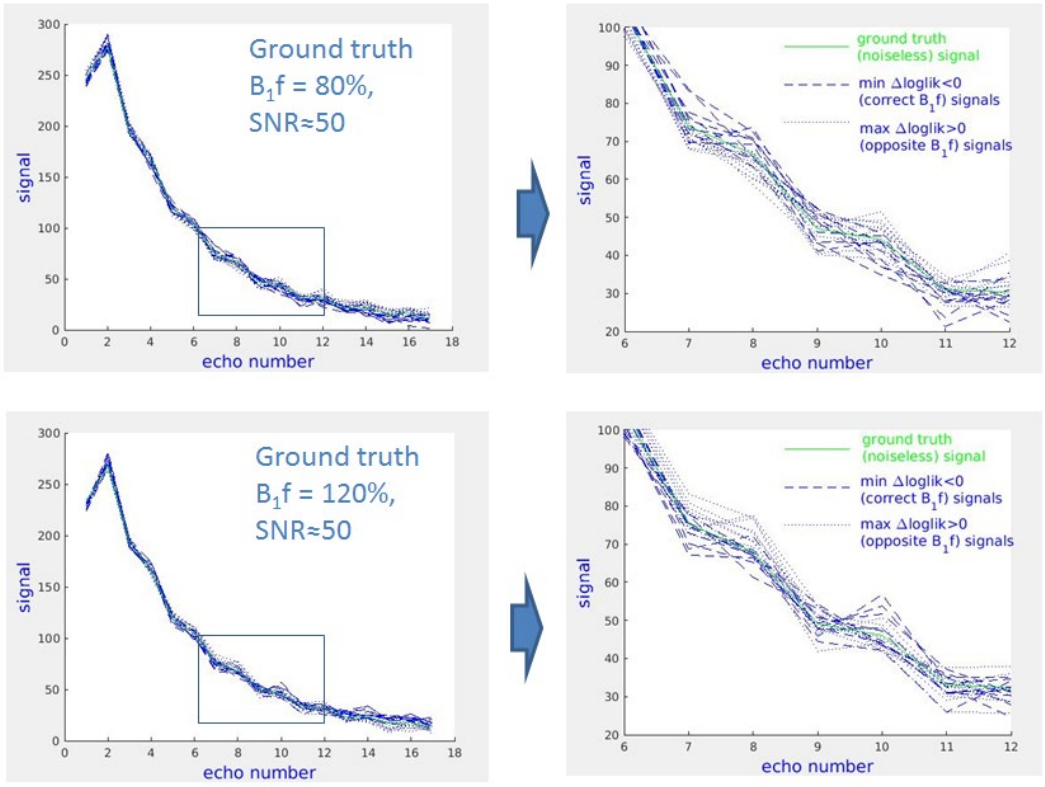


Figure 3.41: Behaviour of the fitting strategy for opposite B_1f regions, with the dashed lines denoting minimum loglik emerging for the correct B_1f region and the dotted lines minimum loglik emerging for the opposite B_1f region. It can be seen that this is related to the way that the noise gets distributed along the echo train.

It is evident that random noise can preferentially amplify or diminish certain echoes in combinations that can switch the minimum loglik to the opposite B_1f region. For the ground truth B_1f = 80% case there is an obvious trend for signals that cross around the middle of the echo train from below the ground truth noiseless signal to above, to minimize their loglik for the opposite B_1f, around 120%. Contrariwise, for the ground truth B_1f = 120% case, signals that cross around the middle of the echo train from above the ground truth noiseless signal to below, to minimize their loglik for the opposite B_1f, around 80%. This demonstrates that there are cases where the differences between the two opposite B_1f signals can be masked with the addition of noise to the extent that the original (ground truth) parameters cannot be recovered by the estimation process.

It is therefore informative at this point to examine the differences between closely matching noiseless signals from opposite but nearly equivalent B_1f regions, as shown in Figure 3.42.

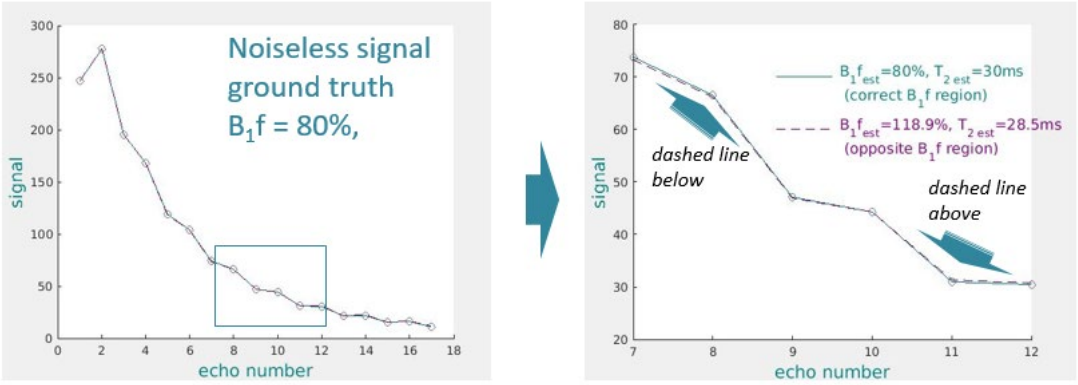


Figure 3.42: Fitting of the noiseless ground truth 80% B_1f signal at the two opposite B_1f regions. The returned signal from the upper (spurious) B_1f parameter estimates (dashed line) crosses from lower to higher values of the ground truth signal (continuous line) at echo 10.

Similarly to the previous experiment's trend, the fit corresponding to the incorrect upper B_1f region crosses from lower than the ground truth signal values to higher ones at echo 10. This is due to the structure of this noiseless signal; its fit showed that from the upper B_1f region it is the 118.9% B_1f-valued signal that best matches the 80% ground truth valued signal, and it exceeds the latter in amplitude at the latest echoes: the dashed line corresponding to the B_1f = 118.9% signal is below the solid line corresponding to the B_1f = 80% signal until the 10th echo, and above it afterwards. This behaviour is shown in more detail in Figure 3.43 where the echo train projections are expanded across the B_1f range. Dashed vertical lines indicate the previous example values for B_1f = 80% and its closer match B_1f = 118.9% (left) or their subtraction (right).

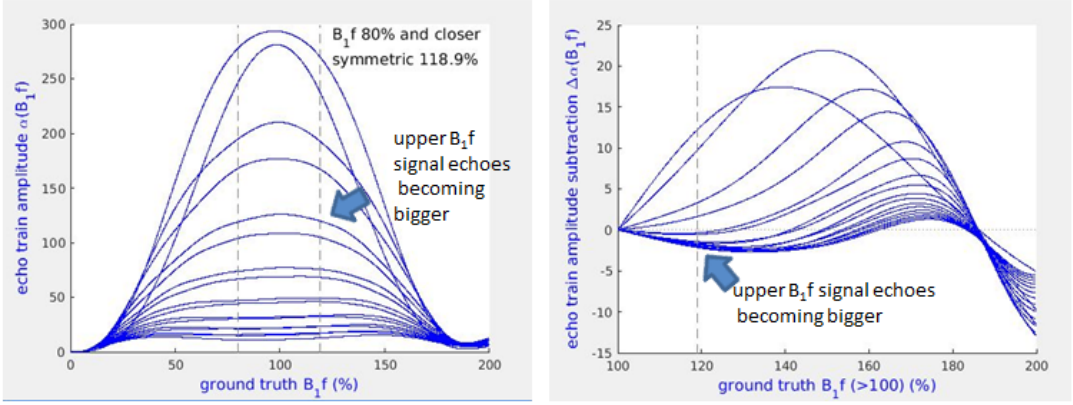


Figure 3.43: Noiseless 30ms T_2 signal along the echo train (17 echoes) and the B_1f range (left), with each curve respresenting the behaviour of one (e.g. the nth, n=1, 2, …, 17) echo in the series. In general signals for the lower B_1f ground truth region are larger than the closer equivalent upper B_1f region signals (that best match them), but not for all echoes and B_1f values (right). $\Delta\alpha(B_1f)$ = echo train amplitude (lower B_1f) - echo train amplitude (best matching upper B_1f). Dashed vertical lines show the echo heights for B_1f = 80% and its closer equivalent B_1f = 118.9% signal (left) or their subtraction (right).

It appears therefore that random noise can modify specific echoes along the echo train in a way that alters the emerging loglik, causing it to switch to a region of the

parameter space associated with spurious parameter estimates. As an example, modifying the above noiseless signal by 10 (arbitrary) units at the 8th and 10th echo only, can switch its minimum loglik to the opposite B_1f region, causing a considerable difference (of 1.7ms) in the T_2 estimation (Figure 3.44).

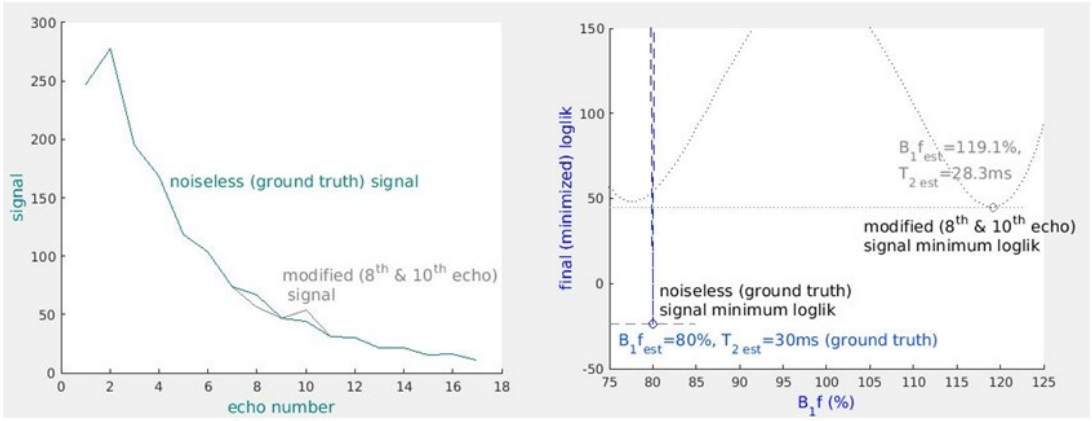


Figure 3.44: Altering two echoes of the ground truth noiseless signal (simulating the added noise effect, left) switches the minimum loglik to the opposite B_1f region with concomitant spurious parameter estimates (right).

These findings are linked to both the slice profile effects described in section 3.2.5 and the echo train behaviour described in section 3.2.3: the fast T_2 and high amplitude sub-pool of simulations seen there, were in a similar fashion due to the effects of noise seen along the echo train, and lead to deviations for the parameter estimates and overall bias for the method. As it has been demonstrated in this section B_1f effects play a key role in the echo train formation and can be classified via the loglik. This will next be combined with the spatial distribution properties of the B_1f behaviour to further improve the method by means of spatial regularization.

3.2.8 Spatial regularisation of B_1f error polarity

In MRI acquisitions B_1f typically varies smoothly, with few or no oscillations below or above optimal calibration levels through images; relevant examples from the literature are shown in Figure 3.45 (Sinclair et al., 2012).

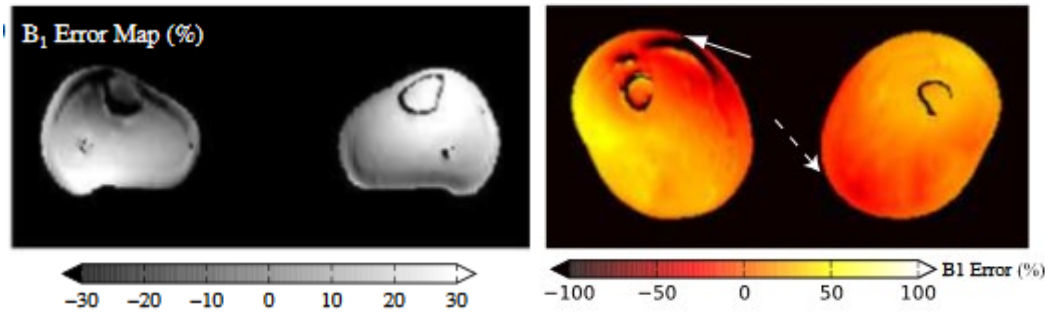


Figure 3.45: B_1f (or "error") maps presented in published work from our group. Left, for calf and right for thigh images (Sinclair et al., 2012).

These examples were produced by an established method using 2 different excitation flip angles (the “double angle method”, DAM) with B_1f determined from the relative amplitudes of the detected signals (Stollberger and Wach, 1996). For the purposes of this work, we are mainly concerned with the polarity of the B_1f error, which may be assigned to one of 2 conditions, i.e. below or above optimality as explored in the above sections. B_1f magnitude maps, such as those in Figure 3.45 may therefore usefully be reduced to simpler binary maps representing the B_1f polarity error condition for each pixel. In Figure 3.46 the basic idea of such a map is demonstrated, in the case of the established DAM method B_1f mapping.

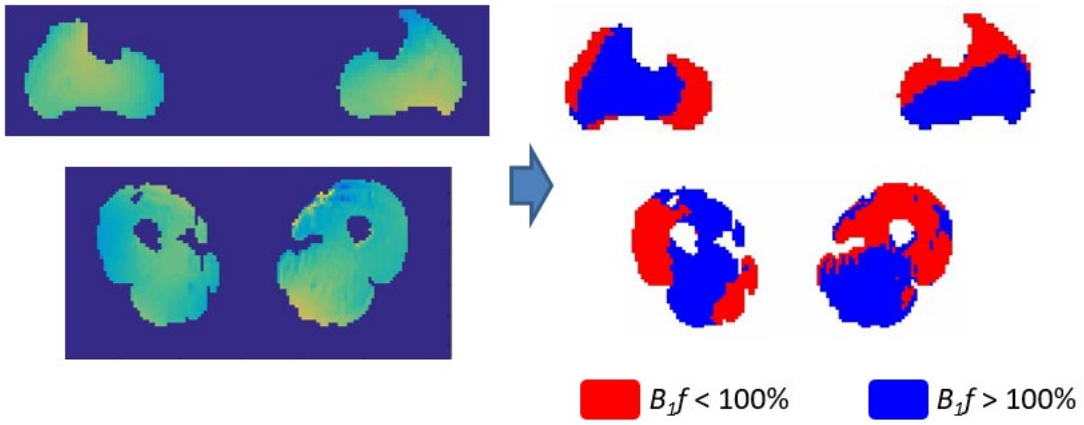


Figure 3.46: Transformation of conventional DAM method B_1f maps (left) to B_1f error polarity maps (right) from data available in our department from calf (above) and thigh (below).

These maps are informative of the overall behaviour of B_1f; they corroborate that it varies smoothly as mentioned above, and takes relatively large image areas before it changes from above to below the nominal correct value (B_1f =100%) and vice versa. Knowledge of these patterns suggests a simple method to partially regularise the parameter estimation procedure by enforcing local consistency in the B_1f error polarity relative to 100%. In this way a simple spatial filter can be created to mitigate the problem of ambiguous fitting due to the close-to but imperfectly symmetric behaviour of the echo trains with respect to B_1 amplitudes either side of the nominal 100% value.

The B_1f polarity maps corresponding to the B_1f maps initially produced using the sEPG-MLE routine without constraining B_1f, have the mottled appearance shown at the left of Figure 3.47, similar with the simulation experiment results previously shown in Figure 3.38.

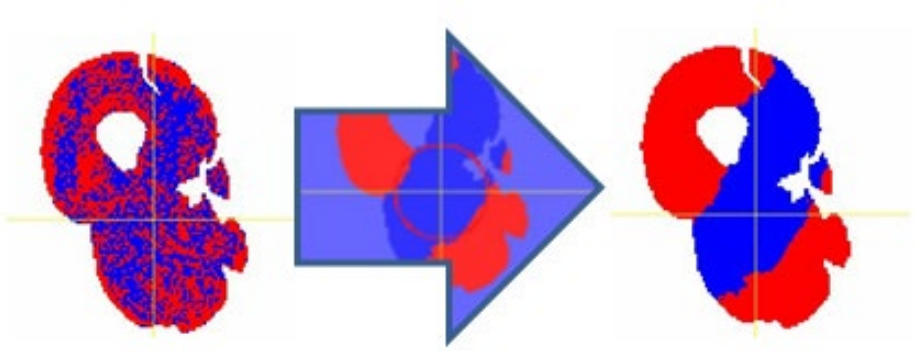


Figure 3.47: Applying a 20 pixel radius disc-shaped median filter in the loglik-B_1f error polarity map obtained for a healthy volunteer with a preliminary estimation procedure with B_1f unconstrained, produced a more credibly physically consistent B_1f polarity map similar to those obtained from the conventional DAM B_1f-mapping in Figure 3.45 and Figure 3.46.

In principle, a local median filter can be used at each pixel to correct its B_1f (error) polarity, based on the majority vote of its neighbouring pixels. By assigning 1 for lower and -1 for upper B_1f region pixels as in the initial B_1f-loglik dependent maps presented in Figure 3.38, this majority vote is simply the median of the B_1f polarity sign of the pixel in question and its neighbouring pixels within a defined ROI: if positive the pixel in question is assigned as a lower B_1f region pixel (whether it was already or not) and if negative as an upper B_1f region pixel. In this work the neighbourhood was defined as a circular ROI centred on the pixel in question, which was itself included in the neighbourhood. To avoid edge distortion or discontinuity issues, empty or single minimum (assigned a zero value) pixels in the neighbourhood are not included. Empirically it was found that a ROI with radius 20 pixels produces smooth B_1f spatial variation while avoiding merging equal sign but clearly separated regions (over-smoothing and loss of contrast) (right part of Figure 3.47).

In practice from the preliminary estimation procedure both parameter estimates' vectors are kept for the cases that 2 minima exist, and the spatial regularisation scheme is used to choose the one corresponding to the correct B_1f polarity in the region of the pixel in question. For comparison purposes and for clarity – for this particular example only – any single minimum B_1f-loglik pixels were also included, assigned a value according to the B_1f polarity that this minimum had with respect to the turning point bound, as shown in Figure 3.36 – whereas after the minimisation is concluded the true turning point is effectively the minimum. It can be seen that the resulting B_1f polarity map is similar to the conventional DAM maps from similar acquisitions, without evidence of issues like over-smoothing, edge distortion etc. However in general it seems favourable to keep single minimum B_1f-loglik fits neutral (zero valued in terms of B_1f polarity). This is because the location of their minimum typically is around 100% B_1f, very close to the loglik turning point bound, and on either side. This can be misleading for the B_1f polarity map since such fits would be given equal representation in the majority vote (median filter) as significantly compromised B_1f pixels when in reality their B_1f is optimal or very close. Application of the filter to the simulated data set analysed and presented in Figure 3.39, including this amendment is shown in Figure 3.48.

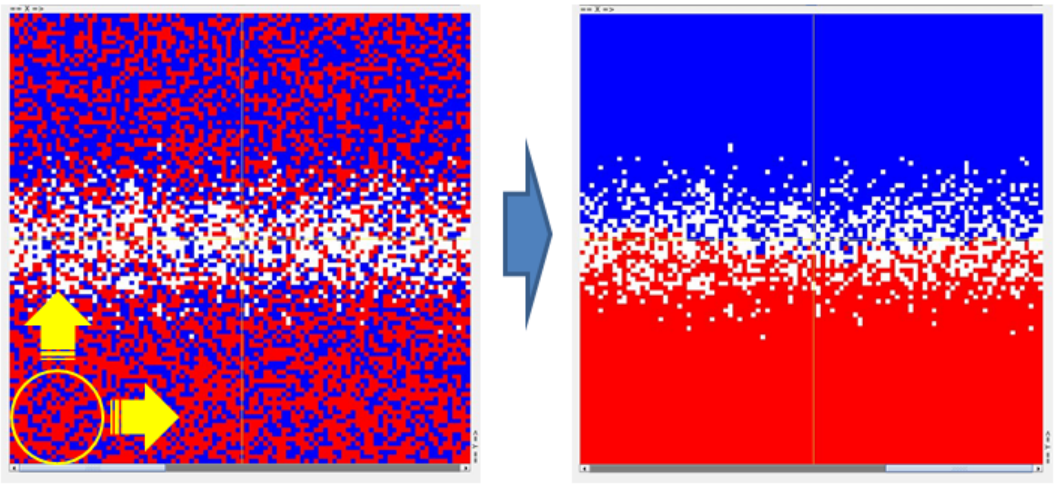


Figure 3.48: Applying a 20 pixel radius disc-shaped median filter in the initial B_1f polarity map produced for the simulated image analysed and shown previously in Figure 3.39.

As it can be seen (right panel) the B_1f polarity map after the spatial regularisation process is almost entirely consistent with the ground truth B_1f values, i.e. 1 (red) for $B_1f < 100\%$, -1 (blue) for $B_1f > 100\%$ region pixels and 0 (white) for the single minimum (and B_1f) in the central part of the image. In that region (centre) very few pixels retained the opposite (spurious) B_1f polarity which has negligible effect on the estimated parameters due to the approximately optimal B_1f scaling: the associated numerical results for the parameter estimates here emerge virtually identical with the ones already presented at the bottom left boxplots of Figure 3.39 since application of the median filter largely collected the parameter estimates corresponding to the ground truth, *a priori* known B_1f polarity and respective loglik.

The above experiments while exploring the effect of choice of B_1f start points have also led to further structuring (fine tuning) of the proposed method. This discussion is continued next for the bound constraints applied for B_1f.

Bound constraints for B_1f

To investigate the behaviour of the fitting process with respect to the B_1f bounds a simple simulation experiment was carried out in which ground truth B_1f was varied across its entire range i.e. $0 < B_1f < 200\%$, and the remaining parameters were kept constant: amplitude α=1000a.u, T_2=30ms and noise σ_N=0a.u. (noiseless signal). The returned B_1f estimate shows departures from the ground truth values for the extreme ranges lower than 20% and higher than 180% (Figure 3.49). At those B_1 levels very low signals are available (as shown in Figure 3.43) with echo heights up to 10^6 times smaller than the nominal amplitude, irregular evolution of the echo train (not the expected mainly decreasing behaviour) and therefore little signal contrast available along the echo train. On the other hand these conditions are generally out of the range of interest in real datasets (and practically rarely seen), as it has been seen in real data, presented later in this thesis.

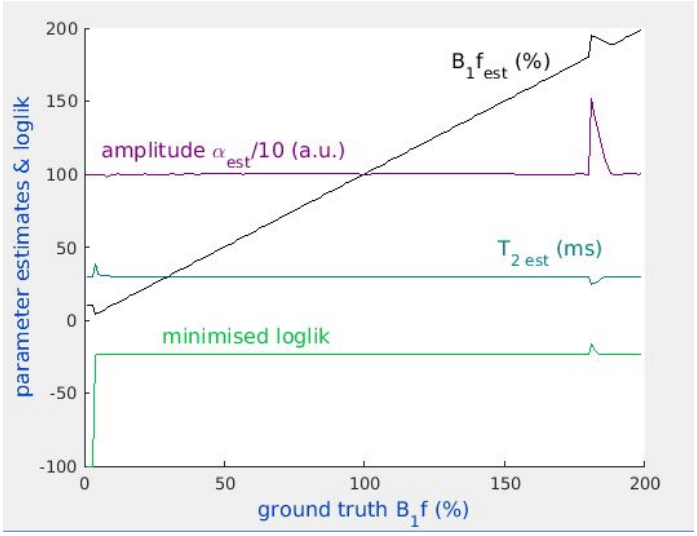


Figure 3.49: Parameter estimates and loglik for noiseless signals (T_2=30ms, α=1000a.u.) of the entire B_1f range of 1 to 199%. Significant departures from the ground truth are seen for very low (below 20%) or very high (above 180%) B_1f values.

For those reasons B_1f bounds were set to 20% and 180% - as in fact they were also used in generating all the results presented so far and for all further work in this thesis.

3.2.9 Remaining parameters' initial values and bound constraints

T_2: Decay constant.

Initial values

The fitting routine as presented and developed so far appears to be broadly insensitive to T_2 starting value variations in terms of the numerical results for the parameter estimates. This was tested for different levels of noise (from noiseless signal to low SNR) and different ground truth T_2 and B_1 values (optimal and suboptimal B_1f) (Figure 3.50, Figure 3.51). However the T_2 initial value choice influences the evolution of the loglik behaviour during the fitting process, and this is worth examining to further characterise the method. The T_2 initial value choice's main effect is to influence the shape of the loglik *vs.* B_1f curve (initial or updated version) in the first iterations as explained in the beginning of the previous section regarding the dependence of the fitting procedure on the B_1f start values (Figure 3.32 through Figure 3.36). The exact position or existence of the provisional loglik minima (i.e. if there appear to be two minima or only one) and B_1f polarity central turning point depend also on the T_2 start point. As it can be seen in Figure 3.50 for the case of noiseless signals, higher T_2 start values lead more often to 2 apparent loglik minima (instead of a single minimum) and to increased distance between them.

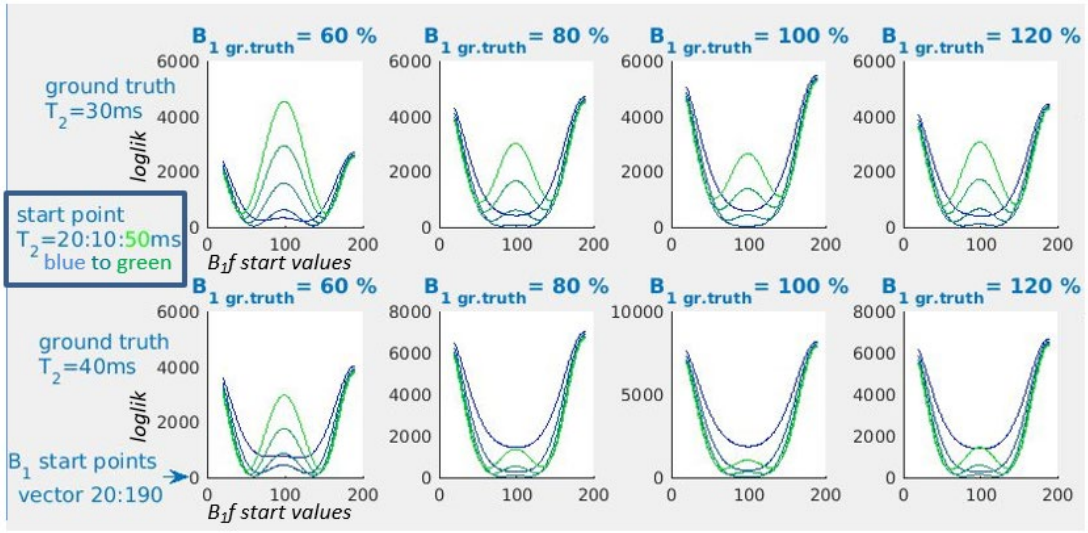


Figure 3.50: Dependence of the initial loglik (vertical axes) *vs.* B_1f start value on the T_2 start point for ground truth T_2 30 (above) or 40ms (below), T_2 start points from 20 to 50ms in steps of 10ms (shaded blue to green curves correspondingly) and for different ground truth B_1f (RF field) values (horizontal axes). Higher T_2 start point values tend to produce separate initial (apparent) loglik minima and place them further away from each other.

This does not seem to affect the final convergence of the minimization process (i.e. the returned parameter estimates and related metrics) but only the intermediate path. This can be demonstrated by a set of simulations of 100 repetitions fitting a noisy (SNR ≈ 20) 30ms T_2 signal, for 10 different T_2 start point values at each repetition, from 20 to 65ms in steps of 5ms. The medians of the parameter estimates and related metrics for each set of 100 repetitions are shown in Figure 3.51.

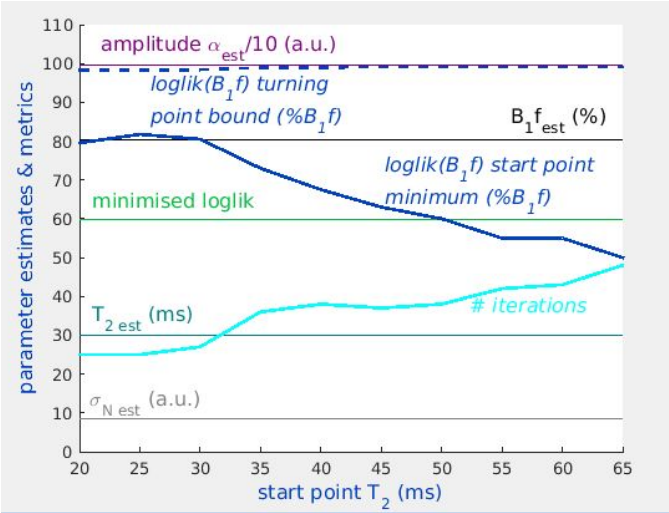


Figure 3.51: Medians of the parameter estimates (B_1f, T_2, amplitude α,noise σ_N) and related metrics (min loglik, B_1f-loglik minimum and central turning point start values) of 100 repetitions of fitting a 30ms T_2, 80% B_1f, SNR 20 signal with varying T_2 start point. Parameter estimates and min loglik are stable while the path of minimisation can vary as manifested from the B_1f-loglik initialization minimum and turning point respective values and the number of iterations. These are the only curves that are non constant, reflecting the behaviour seen in Figure 3.50.

The parameter estimates are stable, even though slightly different to the ground truth values due to the relatively low SNR of 20 (estimates: B_1f=80.2%, T_2=30.1ms, α=997a.u., σ_N=8.6a.u. – ground truth: B_1f=80%, T_2=30.0ms, α=1000a.u., σ_N=10a.u.). The final minimum loglik (the minimised objective function) is also stable. However the shape of the initialization loglik vs B_1f curve changes substantially for different T_2 start point values as denoted from the location of its central turning point (dashed thick blue curve) and more notably its initial minimum (thick blue curve). This different path through the parameter space for different T_2 start points is reflected in the number of iterations required for convergence, which increases for increasing start point T_2. The latter is in turn related to the initial minimum loglik position, which as seen above for higher T_2 start points is further away from the central loglik(B_1f) turning point, and necessitates more iterations or loglik updates to reach convergence. These findings validate the structure and corroborate the reliability of

the fitting strategy described previously (as demonstrated in Figure 3.36) in terms of the yielded parameter estimates, while at the same time provide useful insight to further optimise the method.

T_2 Bound constraints

Similarly to the choice of T_2 initial values, the choice of T_2 bounds did not seem to have a major effect on the fitting process. While not necessarily biologically realistic in this specific context, low T_2 estimates, down to 5ms, occasionally served as markers of suboptimal fitting, or signals artefacted in terms of noise, RF field etc. For T_2 values below that, fitting is unfeasible because the algebraic solutions of the sEPG model are very close to zero and some (late) echoes may become negative. Signals with T_2 > 150ms are out of the range of interest in this work, since muscle and fat typically have T_2 below 150ms. Therefore the lower and upper bound T_2 values used in the fitting algorithm were – at this initial stage – set to 5 and 150ms respectively.

α: amplitude (intercept).

Start points

The fitting routine appears to behave very stably with regards to α (amplitude) start point variation, in a fashion similar to that for T_2 start point values described above. The dependence of the position or existence of the provisional loglik minima (i.e. if there appear to be two minima or only one) and $B_1 f$ polarity central turning point upon the initial value of α is shown in Figure 3.52 for different ground truth T_2 and B_1 conditions.

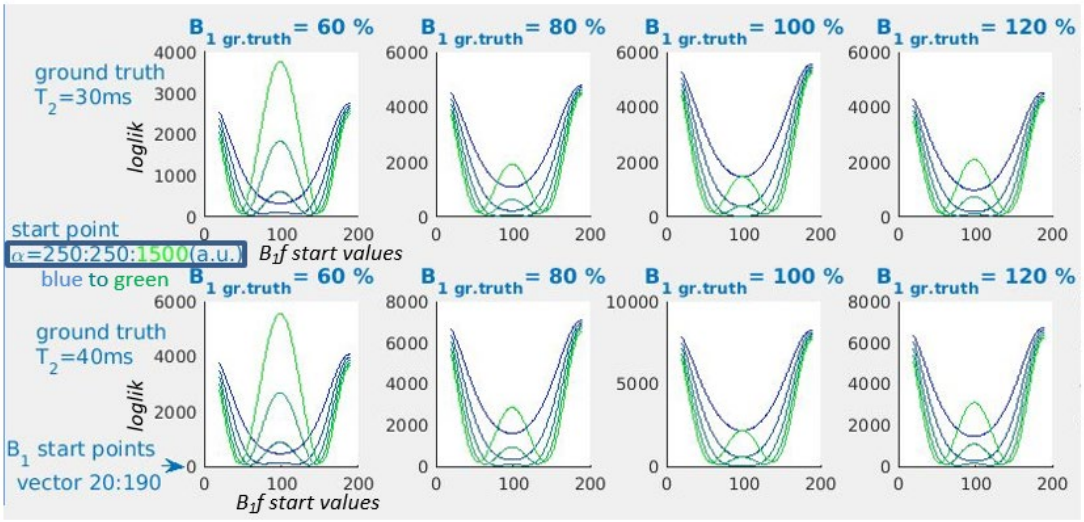


Figure 3.52 Dependence of the initial loglik on the amplitude α start point for ground truth T_2 30 (above) or 40ms (below), α start points from 250 to 1500a.u. in steps of 250a.u. (blue to green curves respectively) and for different ground truth B_1f (RF field) values. Higher α start point values tend to produce separate initial (apparent) loglik minima and place them further away from each other.

Higher α start points lead more often to two apparent loglik minima (instead of a single minimum), with their separation increasing with initial α. A simulation of 100 repetitions was performed, fitting a SNR ≈ 20, 30ms T_2 sEPG signal for 10 different α start point values at each repetition, from 500 to 1500a.u. in steps of 250a.u. The medians of the parameter estimates and related metrics for each set of 100 repetitions are shown in Figure 3.53.

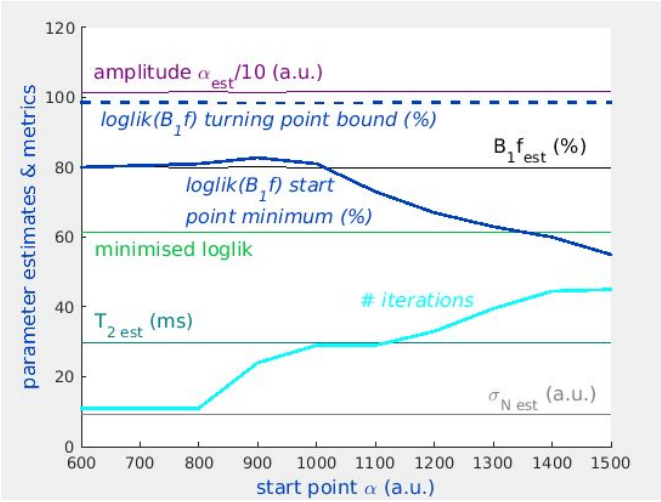


Figure 3.53: Medians of the parameter estimates (B_1f, T_2, amplitude α ,noise σ_N) and related metrics (min loglik, B_1f-loglik minimum and central turning point start values) of 100 repetitions of fitting a 30ms T_2, 80% B_1f, α = 1000a.u., SNR = 20 signal with varying α start point. Parameter estimates and min loglik are stable while the path towards minimisation can vary as evidenced by the B_1f-loglik initialization minimum and turning point respective values and the number of iterations.

The median parameter estimates returned are close to the ground truth values (median estimates: B_1f=79.7%, T_2=29.7ms, α=1017a.u., σ_N=9.2a.u. – ground truth: B_1f=80%, T_2=30.0ms, α=1000a.u., σ_N=10a.u.), and the minimised loglik is also independent of the initial value of α. Consistently with the results for noiseless signals shown in Figure 3.52, the local minima separation and the number of iterations taken increase in general with higher α start points, suggesting that choice of the α initial value does within the range investigated not affect the final convergence.

α bound constraints

The fitting routine returns the amplitude estimate by means of the initial nominal value of α before any decay, slice profile or phase history effects. Even in the case of an infinite T_2 value (constant echo amplitude with increasing TE) and perfect slice profile and flip angle conditions, α cannot be smaller than the maximum echo signal,

therefore the latter was set as the lower bound for α. No upper bound was specified as this was not found to be necessary. These rather wide amplitude bounds have been useful in the investigation of the returned parameter distributions seen in previous sections (e.g. Figure 3.18, Figure 3.25, Figure 3.26).

σ_N: underlying Rician noise standard deviation

Start points

The overall behaviour of the final parameter estimates and the *B_1f*-loglik related metrics with varying σ_N start points are shown for an example ground truth signal in Figure 3.54.

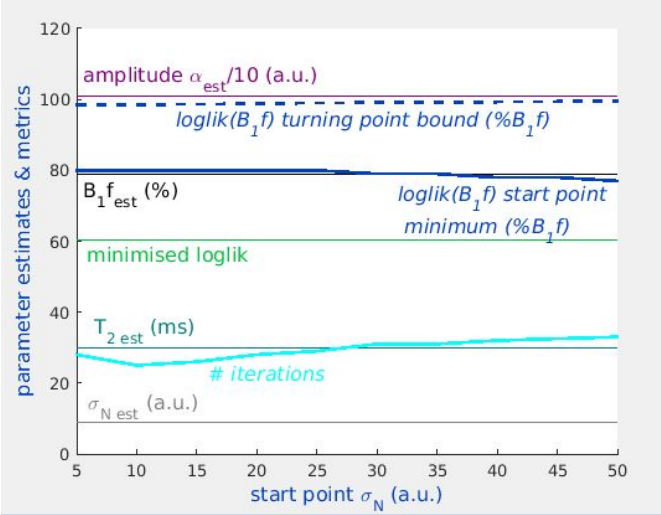


Figure 3.54: Medians of the parameter estimates (*B_1f*, *T_2*, amplitude α, noise σ_N) and related metrics (min loglik, *B_1f*-loglik minimum and central turning point start values) of 100 repetitions of fitting a 30ms *T_2*, 80% *B_1f*, SNR 20 signal with varying σ_N start point. Parameter estimates and min loglik are stable while there are subtle differences overall in the path of minimisation with regards to the *B_1f*-loglik initialization minimum and turning point values and the number of iterations.

Once again the final parameter estimates were constant and independent of initial values, while there were minor differences apparent for the loglik(*B_1f*) minimum and

central turning point location and hence for the minimization path. The number of iterations shows a clear minimum at the ground truth value (10a.u.) – unlike the roughly monotonic behaviour seen for the T_2 and α start points. However it seems favourable to initiate the fitting with higher σ_N start point values rather than lower. This can be illustrated by an example of how the routine proceeds for two symmetric σ_N start points – 5 and 15a.u. – for one example of the simulated signals, as shown in Figure 3.55.

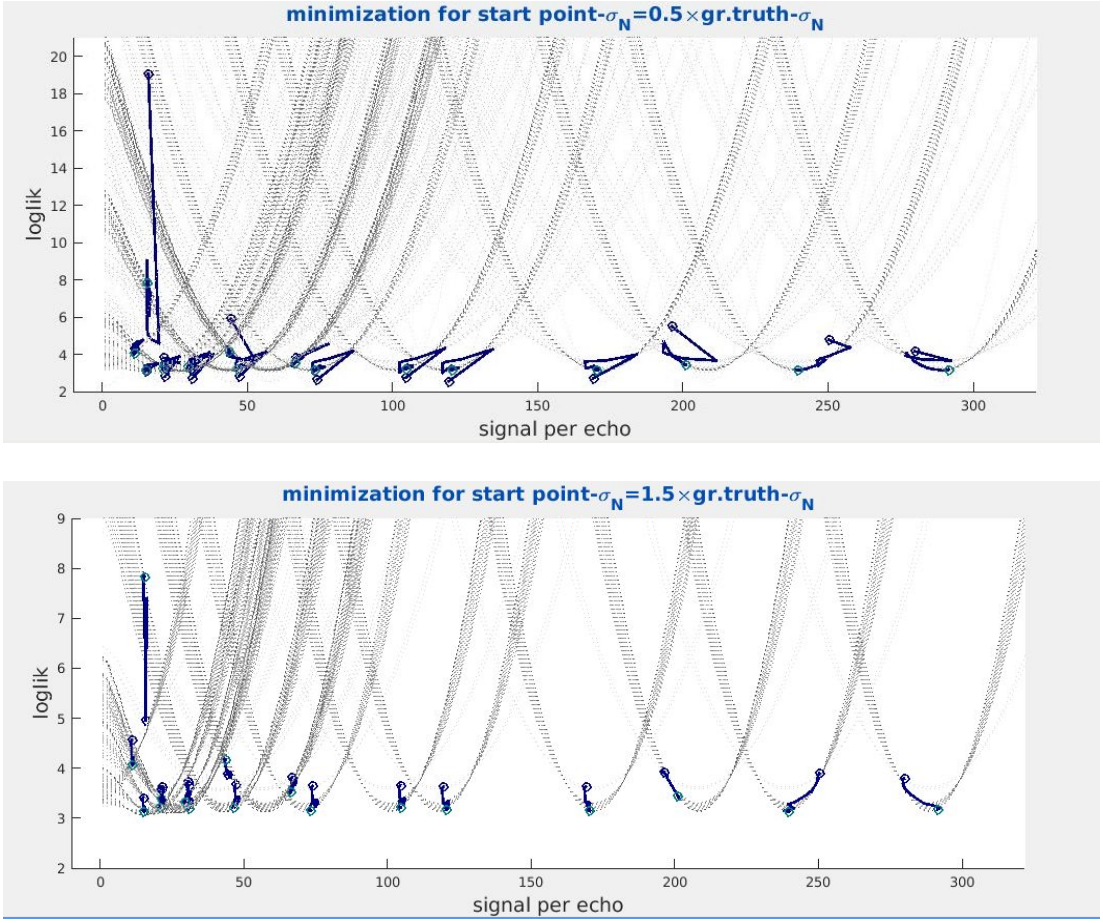


Figure 3.55: The path of minimization for a σ_N start point lower than the ground truth value (above) as compared to higher (below). During the fitting process the negative logarithm of the Rician pdf along the echo train (dotted curves) is being adjusted through successive iterations and its smallest value (the loglik) compatible with the applying sEPG model is the final convergence point. Even though the number of iterations is almost the same the process appears more efficient in the second case.

Trajectories are more direct in the case of the higher start point, while in the case of the lower start point for many echoes the local initial loglik was lower than its final value and was readjusted via a more complex path. This can be seen also in the parameter space (Figure 3.56) – in this case in the loglik, T_2 and α dimensions – where a shorter and smoother trajectory can be observed for a higher σ_N start value, reflecting a smaller number of iterations as also shown in Figure 3.54. The rapid vertical motion of the loglik is predominantly due to σ_N as the scaling factor (denominator) in the Rician pdf equation, while the influence of the other parameters increases in the horizontal directions (Equation 7, Figure 3.23) (Bates, 1988).

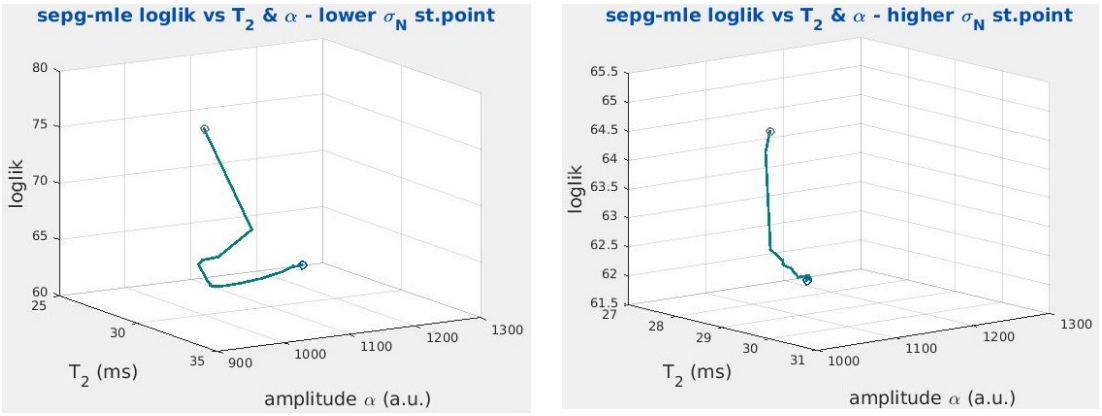


Figure 3.56: The path of minimization for a σ_N start point lower than the ground truth value (left) as compared to higher (right) in the T_2 and α parameter space. For the case of the higher σ_N start point, changes of the parameters are much smoother and the loglik range also smaller.

σ_N bound constraints

When fitting to very noisy and low quality data, σ_N can interact strongly with the α estimation and lead to suboptimal fitting: typically spuriously high estimated noise can mask the other parameters true behaviour, in an analogous way to the role of the noise floor in the LSQ simulations results shown in previous sections. The

potential impact of this may be reduced by defining an upper bound for σ_N. Setting the σ_N upper bound to half the maximum echo signal was empirically seen to maintain stable fitting for the vast majority of cases and at the same time, return of this bound noise value in the final parameter estimate vector served as an indicator for potential outliers from the expected parameter estimate behaviour. More concrete examples on this point are given in section 3.3.2. A similar effect was seen at the opposite end of the noise σ_N distributions, if the noise lower bound was set to zero: even for high SNR signals there is a tendency within the algorithm for the noise estimate to often reach zero, as its σ_N is in the Rician pdf denominator and can rapidly increase it (the pdf) – or equivalently decrease the loglik as previously seen. However for exactly zero noise σ_N loglik is algebraically always also zero and the minimisation cannot continue (the noise σ_N cannot be negative), returning spurious parameter estimates. In both simulations and real data noise σ_N typically shows rather tight distributions, with a slight underestimation (about 10%) and well above zero, e.g. for a 1000a.u. signal intercept minimum noise will be at least 2-3a.u. (Figure 3.18, Figure 3.25). For that reason a very small value (just above zero) was set as the noise σ_N estimate lower bound: for both the simulations and real data examined here 0.1a.u. or grey scale units respectively, in order to prevent it from becoming zero.

The above investigation refers to the role of start points and bounds in MLE parameter estimation in limited parameter space regions, relatively close to the ground truth values, and where the data accurately represent a single component sEPG signal, i.e. without multi-component or confounding physical effects occurring. More challenging cases closer to real experimental conditions, as for example

suboptimal image quality and parameter interdependence in multi-component models, will be addressed in the following sections, after a consideration of phantom validation.

3.2.10 Phantom data validation

The sEPG-MLE fitting routine tool development described above was performed principally with simulations, and subsequently will be further optimised by tailoring it to the characteristics of real *in vivo* image data. At this stage the validity of the method, at least for single component models, was also tested on phantom data since the ground truth parameters (T_1, T_2, proton density) are known and given by the manufacturer. The ISMRM/NIST System Phantom and MESE sequence described in section 3.1 were used (TR= 3000ms, 17 TEs from 9.9-168.3ms with 9.9ms interval, a single 2 mm slice at the T_2 spheres level, matrix 210x210, in-plane resolution 0.8x0.8 mm), using a simple 2-channel quadrature transmit/receive ('TxRx') head coil. For this experiment 5 ROIs corresponding to T_2 from 32 to 133ms were segmented as shown in Figure 3.57. This is the range of interest for T_2 for this work, since as seen before these are the values associated with musculoskeletal anatomy from lean muscle to pure fat.

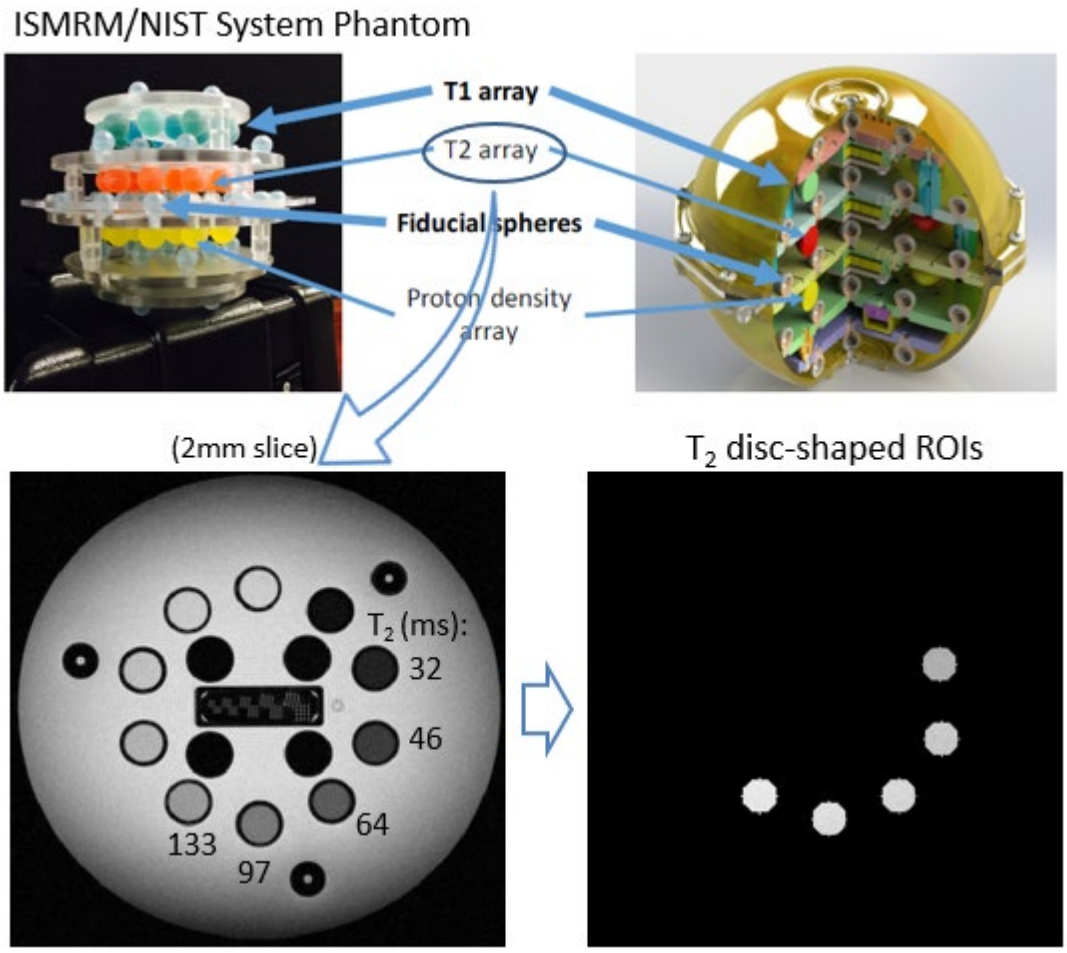


Figure 3.57: The phantom experiment. From the T_2 array of the ISMRM/NIST System Phantom, 5 disc-shaped ROIs were segmented out corresponding to the T_2 spheres in the range of interest between 32 and 133ms.

For this experiment a single value of T_1 of 1000ms was used and the start point for T_2 (single component sEPG model) was set at 80ms, with all other algorithm settings kept as introduced in sections 3.2.6 and 3.2.8. This was seen to better fit the higher T_2 of 133ms sphere which is near the algorithm upper T_2 limit (150ms) – this step removed one outlying fit (single pixel) that was seen with the original settings. Ground truth T_1 was ranging from 458ms (T_2 32ms sphere) to 1332ms (T_2 133ms sphere), reversely than real data where for muscle (T_2 30-40ms) T_1 is around 1400ms and for fat (T_2 130-140ms) T_1 is around 370ms. The sEPG-MLE results are compared to the ground truth values in Table 3.3.

Table 3.3: Comparison of the proposed sEPG-MLE method results with the ground truth (per manufacturer) values for the phantom ROIs chosen.

Ground truth T_2 and SD (ms)	sEPG-MLE method parameter estimates T_2 and interquartile ranges (ms)
31.97 (0.083)	32.13 (30.78 33.45)
46.42 (0.014)	46.29 (45.12 47.39)
64.07 (0.034)	66.24 (64.98 67.64)
96.89 (0.049)	96.80 (93.98 99.040)
133.27 (0.073)	142.00 (138.20 145.69)

Good agreement was seen in general. Estimated noise σ_N were varying from 25 units (T_2 32ms sphere) to 50 units (T_2 133ms sphere), broadly consistent with the experimentally estimated noise level of 29 units seen in section 3.1. This corresponds to an observed SNR from 50 to 30 respectively, which might explain the larger departure seen for the highest T_2 sphere of 133ms. Noise σ_N may also be dependent upon the system calibration differences within the raw images (Figure 3.1). The T_2 ground truth SD similarly increases for the higher T_2 spheres – with the exception of the first – which might be related to the manufacturing specifications (e.g. solution characteristics) or with slower decays (longer T_2 and more signal) intrinsically associated with bigger uncertainty.

3.3 Transition to real-world data and optimisation

3.3.1 2-component sEPG model and Rician noise

A single component sEPG model is not adequate to describe the transverse magnetisation decay in muscle affected by neuromuscular diseases. A multi-component sEPG model is now developed to account for the compartmentalization of tissue in neuromuscular diseases. This will involve a relatively fast decaying component describing the signal emanating from intra- and extracellular tissue water (Foster and Hutchison, 1987) – often referred to as "structured" water (McRobbie et al., 2006) – corresponding to lean muscle, a second slower decaying component corresponding to infiltrating fat, together with a fat fraction (*ff*) parameter describing the amplitude ratio between these two components. To reduce the number of fitted parameters and improve convergence, commonly agreed fixed values have been adopted for both the fat and the muscle-water compartments' T_1s – as these values in general have little effect on the remaining parameters' estimation (Marty et al., 2016). The fat component T_2 may also plausibly be fixed, based on the assumption that it is not expected to change with disease, and that fat and water molecules may be considered physically separated with little or no exchange occurring between them. The objective here is to facilitate the fitting process by limiting the number of free parameters compared to the number of data points available per pixel – defined by the echo train length and number of averages (i.e. separately acquired repeat acquisitions – usually one) – thus maximizing the available degrees of freedom and therefore statistical power (Azzabou et al., 2015). Typical values for T_1 and T_2 for the fat compartment (T_{2f}) at 3 Tesla are 400 and 130ms respectively (Han et al., 2003); at

this stage of the algorithm development T_{2f} was set at 138ms as this was the average value produced from single sEPG component fitting in several subcutaneous fat regions from our datasets. LSQ and MLE yielded parameter estimates for a representative 2 component simulated sEPG dataset with different SNRs were firstly compared, as now described, with results presented in Figure 3.58.

All simulations were performed in Matlab. A forward sEPG model was implemented incorporating Bloch equation calculations of the excitation and refocusing angles across the slice profile, using known RF pulse shapes in use on our 3T scanner (Siemens Prisma). This model was used to generate 1000 replicates at each of 3 levels of randomly generated additive Rician noise, of an echo-train with ground truth parameter values representing a typical *in vivo* situation, and acquisition parameters matching in vivo protocols in use in our centre: water T_2=30ms, fixed T_1=1400ms, B_1f =90%, ff = 8%, fixed fat T_2=138ms and T_1=400ms, α = 1000a.u., inter-echo spacing = 10ms, no of echoes = 17. Three fitting approaches were investigated: 1) MLE for a 2-component sEPG signal model with 5 free parameters: muscle-water T_2, B_1f, ff, α and noise σ_N 2) Standard LSQ for the same sEPG model with free parameters muscle-water T_2, B_1f, ff and α, 3) Standard LSQ for the same parameters with the same free sEPG parameters as in 2) estimated for the same sEPG model with an added constant baseline to partially account for the effects of rectified noise. Start values were set close to the ground truth values and identical for both MLE and LSQ: 100% B_1f, 35ms (water) T_2, 5% ff, 1000a.u. α and 5a.u. for the noise σ_N (case 1, MLE) or the baseline constant (case 3, LSQ).

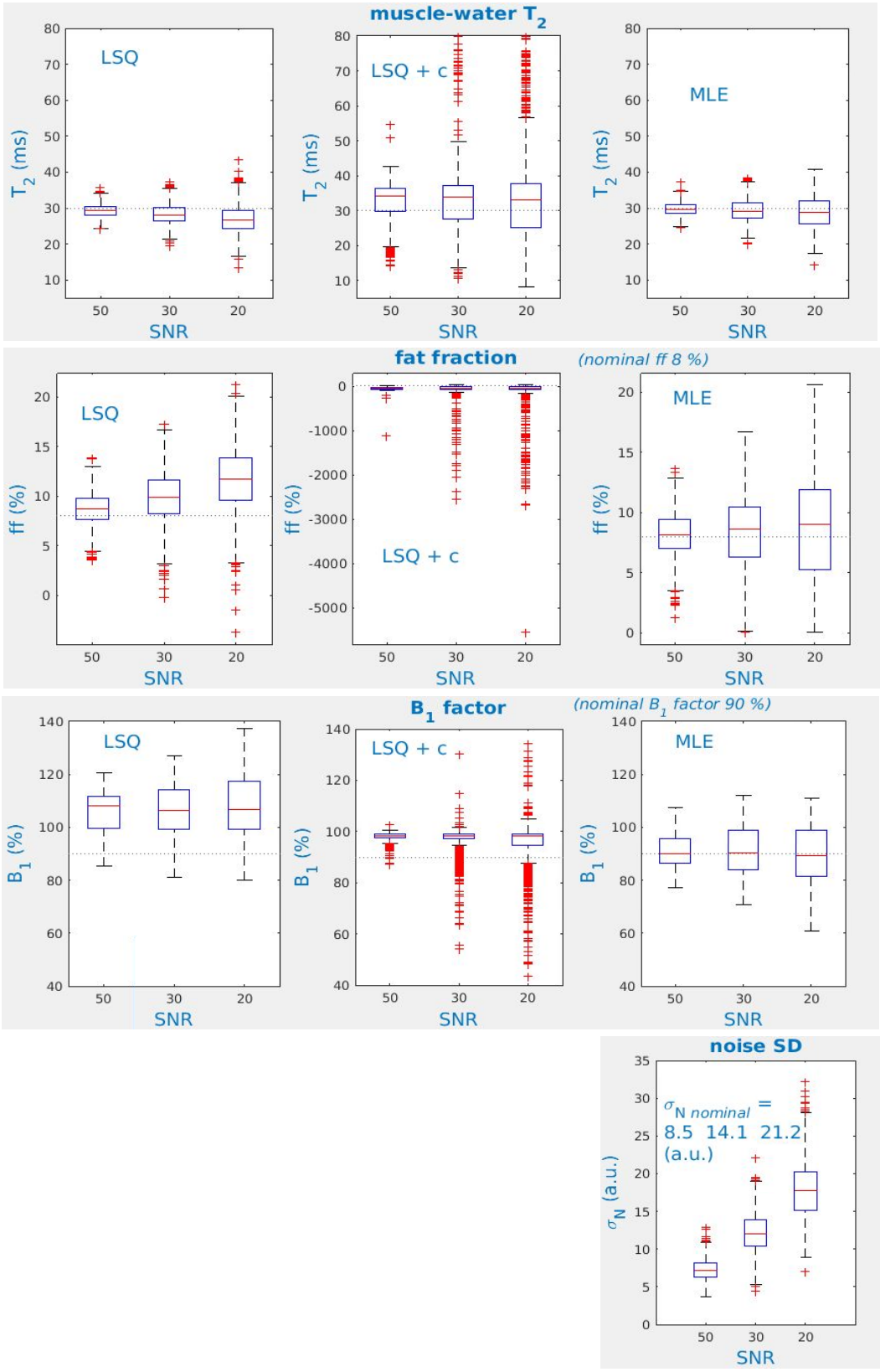


Figure 3.58: Two component (fat &water) sEPG model T_2 (T_{2m}), fat fraction, B_1 factor and underlying noise SD estimates for simulated data, comparing LSQ, LSQ+baseline and MLE methods, for a ground truth model with T_2=30ms, *ff*=8% $B_1 f$=90% and SNR=50, 30 and 20 Rician noise signals. MLE showed superior accuracy (less bias) compared to LSQ.

Bounds for MLE were as described in sections 3.2.7 to 3.2.9, plus 0.1 and 100% for *ff* (lower and upper respectively), while no bounds could be used for LSQ as explained in section 3.2.5.

The addition of the constant to the model for LSQ estimation for the most part was not helpful resulting for these ground truth parameters in frequent spurious *ff* estimation, with the parameter mainly taking non-physical negative values. This is consequential to the inability to specify bound constrains for LSQ due to the algebraic implementation of the method in this case (the Levenberg-Marquardt algorithm). The previously observed trend for LSQ single component sEPG model parameter estimation of increasing T_2 bias with decreasing SNR was in this case reversed, partly due to the now possible interaction between the effects of the *ff* and noise parameters at long *TE*, and partly due to B_1f overestimation biasing the other parameter estimates; relatively short water T_2 (T_{2m}) estimates were returned due to the LSQ estimated B_1f being mainly above 100%, related to slice profile asymmetry effects (section 3.2.7) which cannot be accounted for with LSQ (a very small proportion of B_1f estimates were at the correct region below 100%).

For the MLE approach, B_1f estimation was largely accurate and less overall bias was seen in the rest of the parameters. The method of majority vote, the principle used in the spatial regularization with respect to B_1f estimation described in the previous section, can help towards that direction. In this case it is based on the B_1f polarity that showed majority in the total number of repetitions for each set of simulations: about 60% of the non-single minimum fits proved to be from the lower (below 100%) B_1f region – as expected due to the 90% ground truth B_1f – thus enabling switching

to that region and therefore to return the associated correct parameter estimates. This is a similar but more simplified version of the paradigms shown in Figure 3.38, Figure 3.39 and Figure 3.48.

3.3.2 Behaviour and robustness of the 2-component sEPG – MLE fitting algorithm

In contrast to the single-component sEPG model previously investigated, the 2-component sEPG model here includes three additional parameters, the fat fraction (*ff*) and the T_1 and T_2 decay rates associated with the fat compartment (T_{1f}, T_{2f}). In Figure 3.59 a ground truth noiseless signal, calculated with the example parameter values used in the previous section is shown, simulating the signal from a fat-containing muscle tissue.

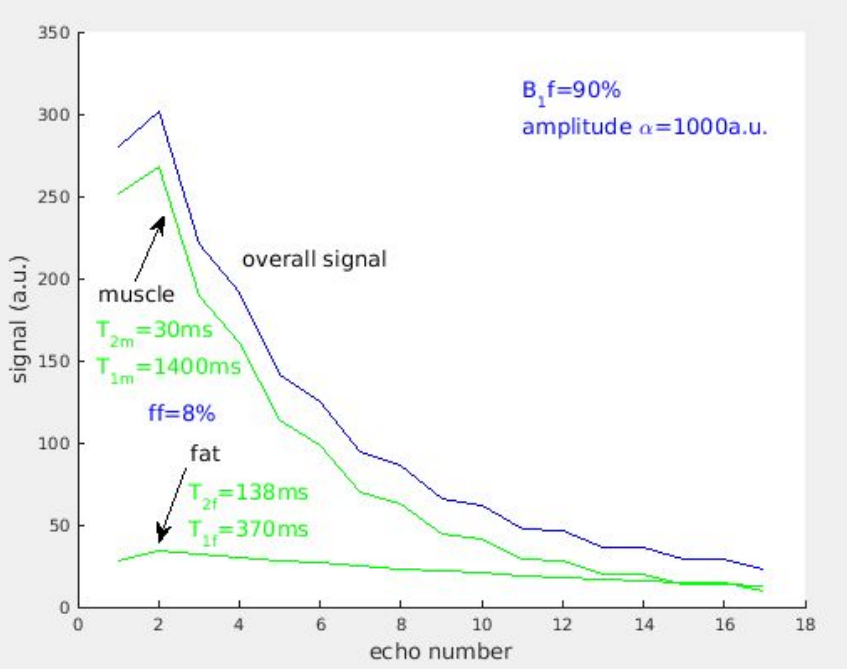


Figure 3.59: Simulation of a decay of a slice-selective CPMG signal and its two sEPG fat and water components with 8% *ff*.

Since in this work T_{1m}, T_{1f} and T_{2f} decay rates are fixed at specific values, in the 2-component sEPG-MLE fitting routine, compared with the single component case, the only additional estimated parameter is *ff*. For the remaining free parameters (B_1f, T_{2m}, amplitude and noise σ_N), testing the effects for each parameter upon the fitting process of its start point and bound constraints showed behaviour very similar to that described in sections 3.2.6 through 3.2.9 for the single component model. However additional considerations arise due to the role of the *ff* parameter which introduces complex interplay between the two signal components and the remaining free parameters. Therefore first the *ff* generic behaviour is presented here, followed by features of the overall 2-component model behaviour owing to parameter interaction.

ff: fat fraction.

Start points

Fat fraction can show great variability within typical imaging datasets. For healthy volunteers *ff* in muscle is usually below 10% and for patients mostly up to 50% but it is not uncommon for *ff* to be anywhere up to 100% for completely fat replaced areas. Dixon methods typically return by design *ff* values mostly up to 50% (Fischmann et al., 2012, Morrow et al., 2016, Ricotti et al., 2016), although precise values depend also on the particular implementation (2 or 3 point technique, field strength, corrections etc.). Other modalities have yielded higher values up to 85%, as for example MR spectroscopy (Forbes et al., 2014) and IDEAL-CPMG (Sinclair et al., 2016), again depending on the particular application and the disease examined. In

more recent work published on *post hoc* multicomponent modelling methods (similar to the sEPG-MLE method presented here) *ff* upper range values reach about 70% (Azzabou et al., 2015, Marty et al., 2016) while in earlier work higher values up to 90% or above have been reported (Kan et al., 2009). While intramuscular fat infiltration is generally associated with disease, there are also additional inter-muscular fat regions, fat around nerves, and adipose tissue that may contaminate analyses due to imperfect image segmentation. For all of the above reasons, the entire range of possible *ff* is of interest, potentially from 0 to 100%. The effect of *ff* start points on the estimation process is examined here for a wide range of *ff* ground truth values (Figure 3.60). For this example set of simulations, start points for T_{2m} and amplitude were set at the forward model ground truth values of 30ms and 1000a.u. respectively without adding any noise. Initial B_1 values were variable, determined according to the local loglik minima detection algorithm described before. Start values for *ff* were varied systematically between 0 and 100% as the parameter under investigation. Noise σ_N start point was 5 units at this stage in accordance with its fitting behaviour seen in section 3.2.6. Interestingly a higher T_{2m} start point of 50ms produced more stable results, as shown in Figure 3.61.

The routine mostly returns the expected ground truth values, especially for mid-range and upper *ff* start point values. For that reason a 50% *ff* start point was chosen for the general case moving forward. However there are regions of *ff* start point values that even for this case of noiseless signal, returned spurious parameter estimates. This was typically due to suboptimal loglik minimisation to a local minimum choosing an erroneous $B_1 f$ estimate of the opposite polarity to the true

value, as seen previously, introducing bias to the remaining parameter estimates.

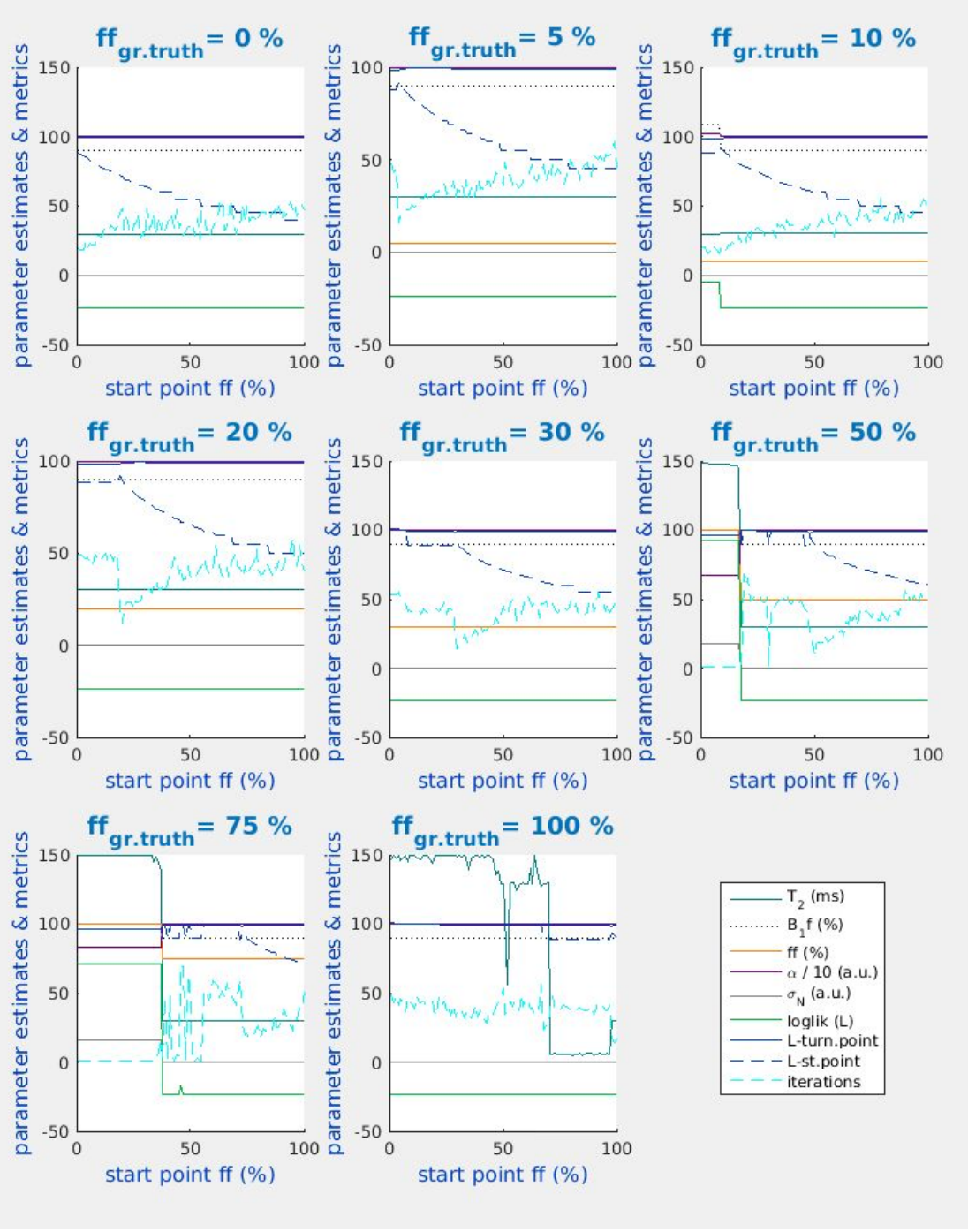


Figure 3.60: Parameter estimates (B_1f, T_{2m}, *ff*, amplitude α,noise σ_N) and related metrics (min loglik, B_1f-loglik minimum start point and turning point) from fitting a 30ms T_{2m}, varying *ff* (between 0 and 100%), 90% B_1f, noiseless signal vs varying *ff* start point. Parameter estimates and min loglik are relatively stable however regions of *ff* start points exist where spurious parameter estimates are returned, due to erroneous loglik minimisation associated with B_1f polarity. The path of minimisation again can vary as manifested from the respective B_1f-loglik initialization minimum and turning point values and the number of iterations.

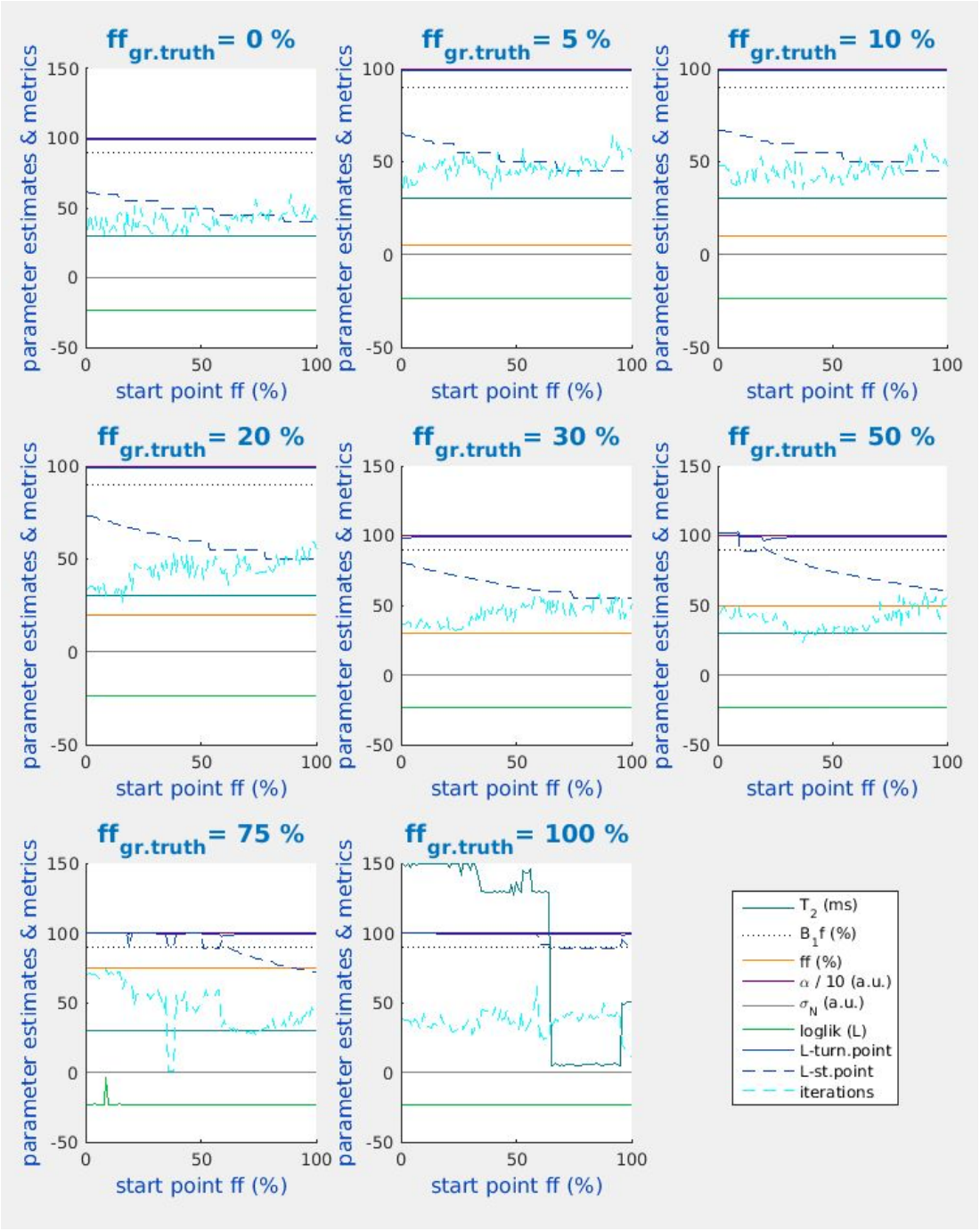


Figure 3.61: The same experiment as presented in Figure 3.53, in this case for a 50ms T_2 start point instead of 30ms – all other start points and bounds equal. A more stable fitting behaviour is seen with ground truth valued parameter estimates and minimum loglik returned in almost all cases.

This is related to an apparent effect of the suboptimal B_1 start value on the turning points for the initialisation loglik seen in several cases of low *ff* start points, as for instance for 10, 50 and 75% ground truth *ff* (Figure 3.60). This is likely to be associated

with the tendency for increased separation of the local B_1f-loglik minima with higher T_2 start value, as seen in section 3.2.6. However this in essence is a matter of parameter interdependence which is further examined below.

ff bound constraints

Bound constraints for *ff* were not seen to have any impact in the fitting process and were kept within the physically meaningful limits of 0 and 100%.

Combined parameter attributes (interdependence).

In the following examples simulated signals are examined in order to link the simulations with the validation stage of this work.

A suboptimal fitting example seen in Figure 3.60 that is an interesting case of 3 parameter interaction, is the low *ff* start point region for ground truth *ff* 10% (3rd graph of the top row). Here, even for an 8% *ff* start point (close to the ground truth *ff* 10%) and a noiseless signal, initial values of T_{2m} (30ms) and α (1000a.u.) at the ground truth, and a 5a.u. σ_N start point, the first-pass loglik curve had a single minimum for B_1f 104%. The subsequent final loglik minimisation (applying the upper and lower B_1f range algorithm described in section 3.2.6, Figure 3.35) produced the parameter estimates:

B_1f (%)	T_2 (ms)	ff (%)	α (a.u.)	σ_N (a.u.)	loglik	iterations	
96.1	148.8	100	428.5	15.2	0.1886	190	(lower B_1f region)
108.8	29.4	9.6	1022.2	0.2	-0.0050	16	(upper B_1f region)

Since the lowest loglik appears for the upper B_1f region, in this simulation experiment the corresponding final parameter estimates would be recorded. However in real-world imaging applications the spatial regularization / majority vote strategy would be also possible to retain the lower B_1f region results which are clearly spurious since 3 out of 5 final parameter estimates were at or very close to bound values: T_{2m} (close to upper bound of 150ms), ff (upper bound 100%) and noise σ_N (upper bound 15.2a.u., 1/20 of the maximum echo intensity). The issue could be resolved by increasing the T_{2m} start point to 50ms, which in turn changed the lower region B_1f start point and bound to 65 and 99.1% respectively and led the minimisation reach its true minimum (Figure 3.61). Due to this more efficient fitting behaviour T_2 of 50ms has been used as the (T_2) minimisation start point from hereafter.

Role of the σ_N upper bound

While a σ_N start point value somewhat higher than the ground truth was seen to aid efficient minimisation, an excessively high upper bound can inhibit the process. In the previous numerical example this effect was obvious because the σ_N upper bound was returned as the final estimate, thus revealing the problem. However there can be situations where such a high bound can permit the minimisation to follow a problematic path with erratic outcomes. In Figure 3.62 examples of this are shown,

for the same noiseless ground truth signal experiment, but in this case with the σ_N estimate upper bound raised ten-fold.

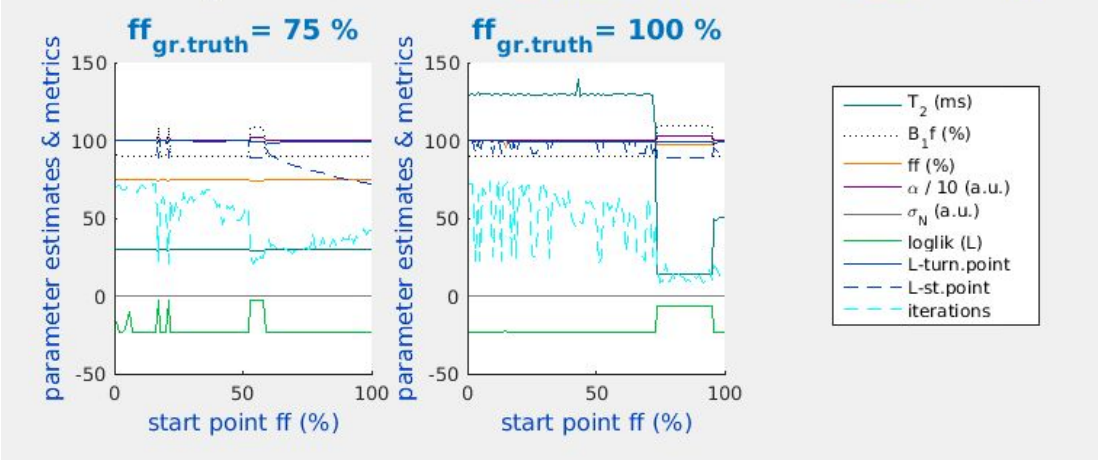


Figure 3.62: The experiment of Figure 3.60 repeated for 75 and 100% ground truth *ff* noiseless signals, in this case for the noise σ_N upper bound set to 1/2 the maximum echo intensity – instead of 1/20 – all other start points and bounds as before. Several ranges of *ff* start points are seen where loglik minimisation suboptimally produced spurious parameter estimates. Notably an erroneous B_1f polarity estimate (above 100%, dotted black lines) is returned for those fits (ground truth B_1f was 90%).

In those cases the high noise σ_N upper bound was reached after only a few iterations keeping the minimisation away from the true global loglik minimum, producing similar results as in the previous example when parameter estimates reached bound constraints. Choosing a much lower noise σ_N upper bound (1/20 of the maximum echo intensity) resolves this issue, in this case of a noiseless signal. However for real data an initial σ_N of ½ of the maximum echo signal produced stable fits when combined with the T_{2m} and *ff* start points given above (50ms and 50% respectively), without reaching bound constraints (for any of B_1f, T_{2m}, *ff* and σ_N) or other manifestations of problematic fitting.

3.3.3 Fat compartment signal model and calibration

Due to the multi-peak proton resonance spectrum of fat and the possibility of tissue compartmentalisation (sections 2.1.2 and 2.1.4), fat tissue CPMG signals are not necessarily expected to be adequately described by a single T_2 component model. In this section the topic of optimising the fat signal model is explored. The overall aim is to establish a fixed parameter 2-component fat signal model which combined with a single water component will enable accurate estimation of T_{2m} by fixing as many parameters as possible, while maintaining an empirically accurate model of the fat signal.

2-component behaviour of fat signals

The simulations experiment described in Figure 3.62 above revealed an important behaviour occurring when fitting a 2-component sEPG model to ground truth 2-component signals with *ff* at, or close to, 100%: there can occur a numerical degeneracy whereby, even for a noiseless signal, a set of estimated parameters different to the ground truth values yields a lower global loglik minimum than the true values, and predicts a signal virtually identical to that parameterized by the true values.

As it can be seen from the right graph in Figure 3.62, for simulated ground truth *ff*=100% tissue there is a considerable range of *ff* start points close to 100% (about 75 to 95%) for which the final minimisation was suboptimal, returning a lower *ff* estimate of 97.2%, a corresponding small amplitude short T_2=13.9ms, and an

erroneous 109.1% B_1f. The actual ground truth signal and both the correct and erroneous parameter predicted signals are shown in Figure 3.63.

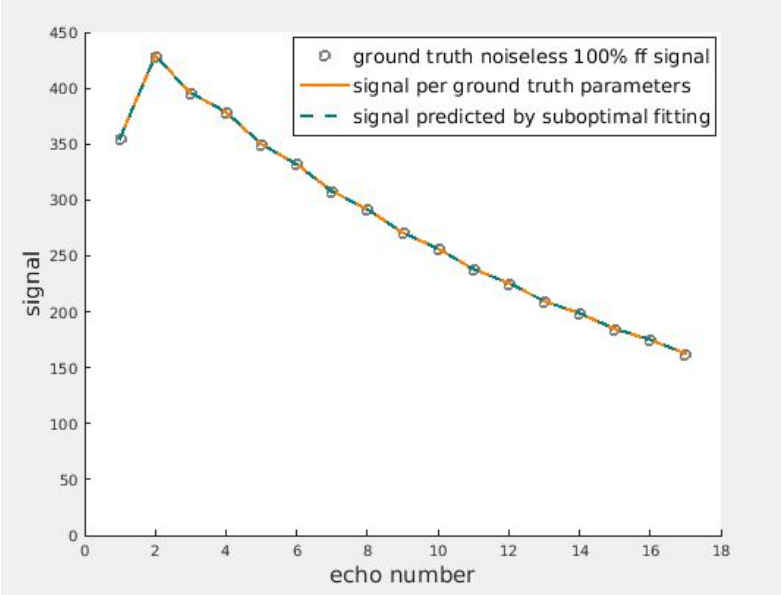


Figure 3.63: Comparison of the correctly (yellow line) *vs.* the erroneously (green dashed line) predicted signal with reference to the ground truth signal (grey dots). Even for noiseless signals there are regions of the parameter space for which, as in this case, an erroneous estimation can emerge favourable (lower loglik).

The predicted signals are virtually identical and since the actually incorrect parameter estimates produced a lower loglik, these will be the preferred values recorded for parameter estimation. While this behaviour was subsequently generally avoided by more careful algorithm initialisation (i.e. increasing the T_{2m} start point to 50ms) it points to a probability that may arise in real data with high *ff*. While this in essence is a numerical issue – since it emerged here for theoretically simulated noiseless signals – for high *ff* regions similar effects may co-occur with true multi-component sEPG signal behaviour due to physical compartmentalization.

The multi-component behaviour of very high *ff* tissue is demonstrated by conducting an analogous experiment using *in vivo* data. In Figure 3.64 an example fit is shown for

signal from adipose tissue (subcutaneous fat) from the lower limb of a healthy volunteer. To be comparable with the noiseless simulation examples, a pixel with minimal visible echo train alterations due to noise or artefacts was selected from a good quality dataset as a representative case. The multi-echo imaging CPMG dataset used is taken from the motor neuron disease study explored in Chapter 4, with acquisition methods as described in section 4.2. A longer echo train (22 echoes) was available with these data, a difference which at this high SNR would have negligible impact in the parameter estimates, as explained in section 6.2.

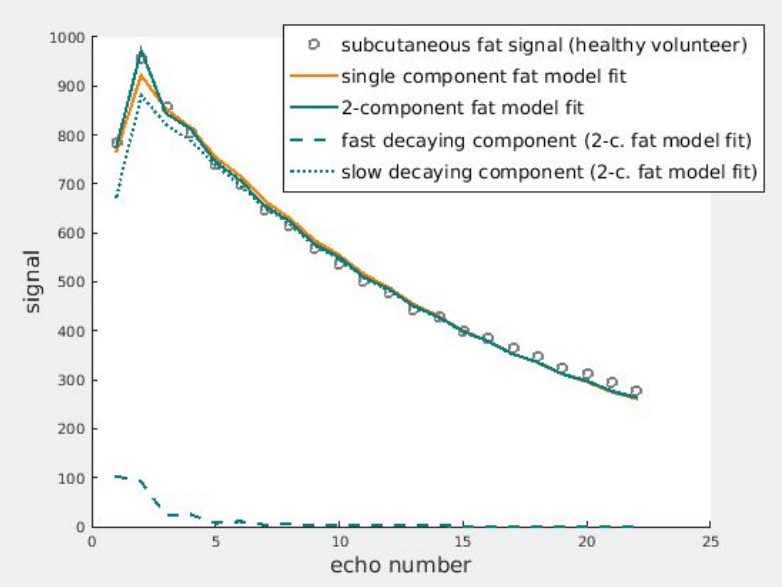


Figure 3.64: Comparison of fitting a single pixel adipose tissue (subcutaneous fat) signal with a single-component and a 2-component sEPG model. The latter is seen to more adequately explain the signal, both visually and via its lower loglik.

A single component model was tested (yellow line fit) which consistently with observations reported by other authors (Burakiewicz et al., 2017) returned a relatively long T_2 for fat of 138.7ms. However fitting to a 2-component sEPG model as above significantly improved the fit quality, both visually (green solid line) and through its reduced loglik and other fitting metrics (reduced noise estimate and

higher R^2). For a fixed long component (T_{2fl}) at 137ms, the estimated short component T_2 (T_{2fs}) was 7ms and the amplitudes ratio ($\alpha_l/(\alpha_s+\alpha_l)$, where α_l and α_s are the long and short component amplitudes respectively) was 62.8%. Both models returned the same, credible value of B_1f after spatial regularisation. Since in this case the fitting routine has already been optimised (with regards to initialisation and bounds) this finding suggests a true multi-component behaviour of fat signals, as also reported elsewhere (Azzabou et al., 2015). Therefore, in order to determine a better, optimal model for the fat compartment generally applicable to our *in vivo* data, investigations were extended from the single pixel case here described to consider multiple subcutaneous adipose ROIs.

Biased B_1f estimation for high ffs

An issue arising with 2-component fitting of signals from regions of high *ff* is that under these conditions there is a tendency to overestimate B_1f, shifting the majority of (the non-single minimum) fits predominantly to the upper (>100%) B_1f polarity region. This is demonstrated in the following simulation. A synthesized image is produced using a forward sEPG model with B_1f varying from 50% to 150% in the vertical direction, increasing in steps of one unit, and a simulated ground truth 2-component fat signal – as above hypothesised – with a fast T_{2fs}=5ms component and a slow T_{2fl}=137ms component in equal (50%/50%) proportion and overall amplitude α=1000a.u., and adding random Rician noise distributed with σ_N = 5a.u., with 100 repetitions along the x-axis. This σ_N would in practice result in observed SNRs in the highest amplitude echo-image of between 50 and 20, imitating conditions affecting real data.

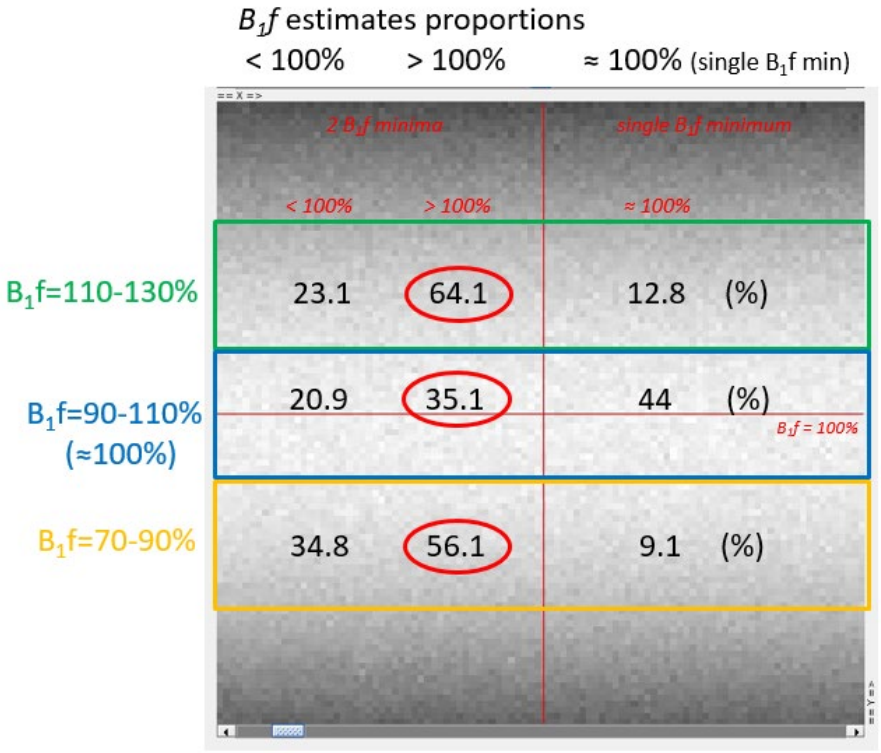


Figure 3.65: Behaviour of the 2-component fat model fitting in terms of B_1f estimation, for a synthetic image of varying B_1f and SNR between 50 and 20. The 2-component model tends to preferentially lead the minimisation to the upper B_1f region (B_1f > 100%, circled values).

The simulated data was fitted with a 2-component sEPG model for fat, as described previously (Figure 3.64); Figure 3.65 presents, for 3 ranges of ground truth B_1f, the proportion of pixels in each for which the final minimization converged giving final B_1f estimates above, below or approximately at 100%. The loglik(B_1f) minimum gridding strategy (section 3.2.6, Figure 3.35) was used here without spatial B_1f polarity regularization. The 2-component model fitting to data generated from a forward model with ground truth parameters close to those obtained in the example described above (Figure 3.64), shows a clear tendency to converge to a non-local minimum located in the upper B_1f region when in reality B_1f is equal or less to 100% (middle and lower delineated zones respectively in the simulated image). If the spatial regularisation majority vote strategy presented earlier was applied for example to these 2 regions this would lead to incorrect final parameter estimates as

can be seen from the leftmost 2 columns of B_1f proportion values – the rightmost column percentages being neutral with regards to the regularisation scheme because in those cases the minimisation has led to a single minimum. For instance for the lower 70-90% B_1f zone the 56.1% majority of (erroneous) B_1f > 100% pixels would according to the spatial regularisation algorithm lead to also switching the 34.8% of (correct) B_1f to the upper region; this was seen to actually return a median loglik lower than the correct B_1f (<100%) did. This is likely to be linked to the combination of the added noise and the shape of the particular echo trains, as seen for the B_1f effects in section 3.2.7.

The above findings suggest that for the purposes of fitting experiments to determine the optimal signal model in very high *ff* adipose regions, the B_1f polarity spatial regularisation scheme as previously implemented should be suitably modified. This can be done by using prior information regarding local B_1f polarity obtained by fitting first to the muscle compartments only, using the spatial regularisation as previously implemented, since this was seen to adequately address the discrepancies due to B_1f deviations (sections 3.2.8 and 3.3.1) – these being relatively low *ff* regions for which the above overestimation problem does not occur. If thereafter relatively small subcutaneous fat ROIs are chosen near muscle areas of homogenuous B_1f, fitting in these ROIs can then be performed imposing the B_1f polarity condition existing in the neighbouring muscle pixels (Figure 3.66). If it is in the lower, positive polarity region (broadly under 100%), perform the minimisation regularising towards the lower B_1f-loglik minimum, and conversely if the muscle B_1f is of the opposite polarity.

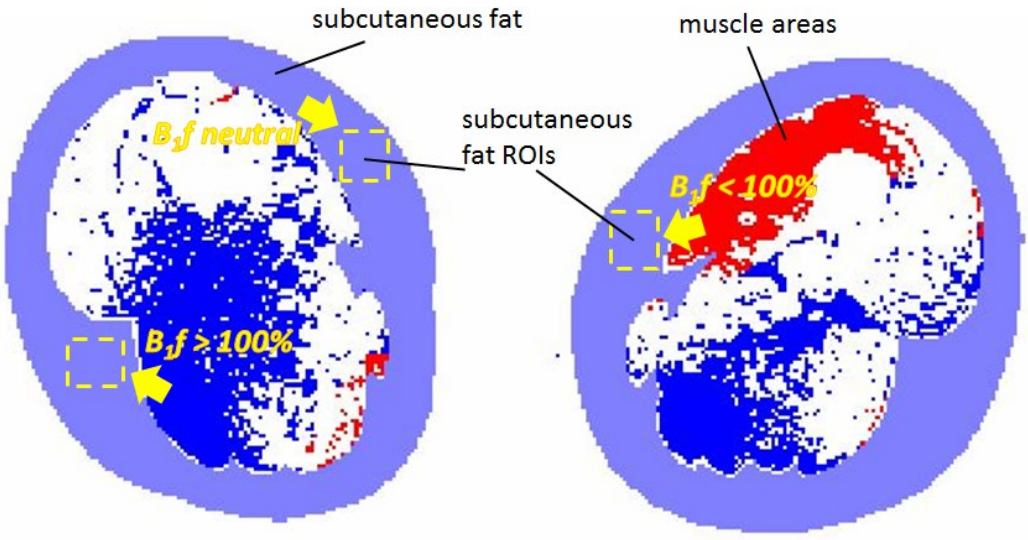


Figure 3.66: A schematic representation of the spatial regularisation strategy used for the fat model determination in subcutaneous adipose ROIs. Relatively small adipose tissue ROIs were used in order to allow the local B_1f polarity condition to be obtained as the adjacent tissue (muscle areas) B_1f polarity classification.

In the case that the minimisation in the muscle area has shown a single minimum (white areas in Figure 3.65) the fat ROI pixels will in the general case be a mix of different B_1f types (lower, upper and neutral) and no regularisation can be performed; the direct parameter estimates for each pixel can simply be recorded.

Fine tuning of T_2s for the 2-component fat model

On fitting the 2-component simulated signals with a single component in the previous simulated image experiment, the returned single component T_{2f} estimate was in general lower than the ground truth value (137ms) for the 2-component forward model long component T_2 (T_{2fl}) signal. This suggests that the optimal T_{2fl} estimate in the 2-component fat model is likely to be higher than the typical 130-140ms values found when fat signals are fitted with a single component model. The behaviour of

the 2-component fitting for fat signals introduced above (Figure 3.64) suggests that there is commonly a fast T_2 (T_{2fs}) component – so far seen to be below 20ms, a long T_2 component (T_{2fl}) above 120ms and their ratio is typically around 40-60% or higher. Following the reverse process and testing the single component fitting of such signals reveals that T_{2fl} should be at the range of 200ms if T_{2f} (the ground truth value for the single component fat model) is around 135-140ms respectively – the value commonly seen in real data and consistent with published work (Burakiewicz et al.). Examples of this relationship are shown in Figure 3.67. An important outcome of this process is that the longest T_{2fl} (198ms) provides the best matching ground truth 2-component model signals to the single component model being fitted, keeping $137 < T_{2f} < 150$ms, for the $50 < B_1f < 150\%$ range.

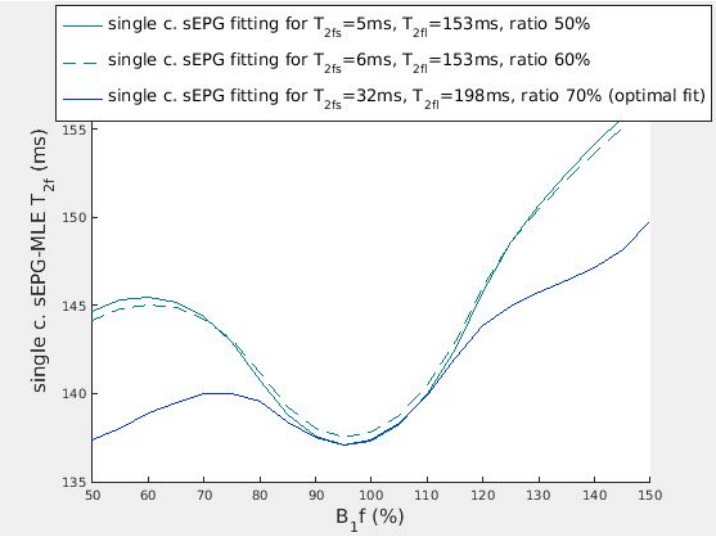


Figure 3.67: T_{2f} (single component) estimate dependence on B_1f for a 2-component fat model ground truth noiseless signal of T_{2fs} = 5, 6 and 32ms, T_{2fl} = 153 or 198ms and inter-component ratio of 50, 60 and 70%. For a broad range of B_1f values the single component fitting returns values similar to the values commonly seen in real datasets.

Further experimentation with fitting to *in vivo* subcutaneous adipose ROIs data also revealed that setting T_{2fl} at a fixed value at 198ms instead of 137ms was seen to

significantly improve fitting, as for example with the signal used in Figure 3.64, now re-fitted as shown in Figure 3.68.

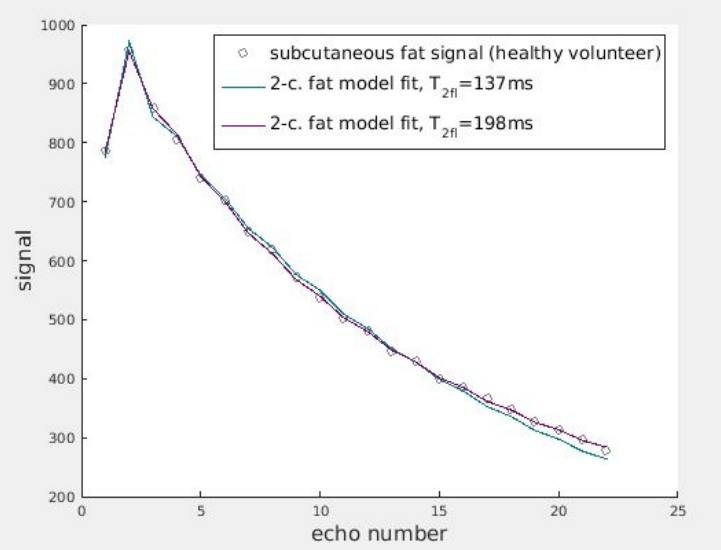


Figure 3.68: Optimising the 2-component fat model fitting: setting the long component T_{2fl} at 198ms significantly improves the quality of fit. The same single pixel adipose tissue (subcutaneous fat) signal as in Figure 3.64 was used.

This value was seen to best match the 2-component fitting to the single component fitting estimates (of T_2f 135-140ms) for B_1f close to 100% (as demonstrated from the noiseless signals fitting examples in Figure 3.67) which in turn correspond usually to the majority of pixels in typical images – as for instance the white areas in Figure 3.66. The way that T_{2fl} and the rest of this 2-component fat model parameters were determined was by repeating the subcutaneous fat ROIs investigation using prior spatial regularisation information as suggested above (Figure 3.66) for three different sEPG-MLE method versions for fat: the single component, the 2-component when fitting for both T_{2fs} and T_{2fl} and the 2-component when fitting for T_{2fs} while T_{2fl} takes fixed values within a chosen range. Using again good quality datasets from the previously mentioned motor neuron disease study (Chapter 4) including amyotrophic lateral sclerosis (ALS) patients, Kennedy's disease (KD) patients and healthy

volunteers, 8 subjects' images were chosen, showing the highest quality in terms of both SNR and of relatively mild or absent artefacts (motion, flow, chemical shift etc.): 4 healthy volunteers, 2 ALS and 2 KD patients; from each image 4 fat ROIs of 56 pixels each were defined. In Table 3.4 the overall results are shown as medians plus interquartile ranges for the sets of 8 by 4 by 56 individual pixel (echo train) fits. The 2-component sEPG-MLE fitting using a variable but fixed T_{2fl} yielded the most precise (smaller interquartile ranges) and accurate (lower loglik and σ_N and higher pseudo-R^2) results. The pseudo-R^2 introduced here is a variation of the conventional R^2 statistical metric: $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$ where N is the number of observations in the model, y is the dependent variable, $\bar{y}$ is the mean of the y values, and $\hat{y}$ is the value predicted by the model. The numerator of the ratio is the sum of the squared differences between the actual y values and the predicted y values, or the unexplained variance. The denominator of the ratio is the sum of squared differences between the actual y values and their mean, or the explained variance. The model used in the calculation of the unexplained variance here was defined as the squared root of the summed squares of the modelled via the estimated parameters sEPG signal, plus the estimated noise σ_N (1st channel), and the estimated noise σ_N alone (2nd channel): $\hat{y} = \sqrt{(sEPG(p, TE) + \sigma_N)^2 + \sigma_N^2}$, where p are the estimated sEPG parameters and TE the echo times. This takes into account the theoretical and practical aspects of Rician noise in MR images previously discussed (sections 2.1.9, 3.1, 3.2.3), since as the MLE method applied here estimates both the sEPG model parameters (p), and the underlying Rician noise levels (σ_N), and these may interact in the model fitting process, noise is in essence part of the model. It should however be

noted that its additive effect only has major impact in the low amplitude final echoes where any negative noise contribution becomes positive through the magnitude image reconstruction. This implies that this pseudo-R^2 is actually positively biased, since it always increases the model-explained signal via the above σ_N additive effect, while in reality Rician noise can also decrease signal, especially at the initial high echoes. Nevertheless this pseudo-R^2 provides an operational, approximate but pragmatically useful goodness of fit metric: empirically this approach returned higher, more credible pseudo-R^2 values (closer to unity), than when using the sEPG function alone to model the predicted y values ($\hat{y} = sEPG(p, TE)$), in an approach that may be considered a simplified version of Efron's pseudo-R^2 (UCLA, 2011). While estimated noise levels (Rician σ_N) and loglik (the main MLE statistic) are quantitatively specific to the individual signal fitted, pseudo-R^2 as a normalised quality of fit metric offers complementary information. The Akaike information (AIC) and Bayesian (BIC) criteria (Box et al., 2015) were also calculated as conventional meaures of statistical model quality.

Table 3.4: Overall results for estimates of the fat model parameters: T_{2fs}, T_{2fl}, the 2-component ratio, B_1f, σ_N and the related metrics loglik and pseudo-R^2 (see main text for its use and definition), Akaike information (AIC) and Bayesian information (BIC) criteria, for 3 variations of the sEPG model within the sEPG-MLE algorithm. The models were fitted to pixel data from a total of 32 subcutaneous adipose ROIs from 8 participants (see text). Median estimate values are stated, with interquartile ranges are given in brackets. A clear superiority both in accuracy and precision was seen with the use of a 2-component model with fixed T_{2fl} (last column; Here fitting was repeated with fixed T_{2fl} tested across a range between 100 and 500ms in steps of 1ms, with the value yielding the minimum overall loglik being reported).

	single component sEPG-MLE for T_{2f}	**2-component sEPG-MLE fitting T_{2fs} & T_{2fl}**	**2-component sEPG-MLE with varying fixed T_{2fl} ('grid')**
T_{2fs} (ms)	140 (136 145)	19 (12 30)	40 (38 42)
T_{2fl} (ms)		165 (155 181)	198 (tested range: 100 – 500ms)
fat components' ratio (%)	-	69 (65 73)	67 (64 69)
B_1f (%)	95 (86 96)	115 (75 119)	97 (77 110)
σ_N (a.u.)	18.2 (13.8 25.3)	5.8 (3.9 10.0)	3.5 (2.9 4.4)
loglik (a.u.)	95 (89 102)	70 (61 82)	59 (55 64)
AIC / BIC	198 / 220	152 / 185	128 / 155
pseudo-R^2	0.986 (0.983 0.988)	0.998 (0.997 0.999)	0.999 (0.999 1.000)

In Figure 3.69 the medians and interquartile ranges for the single- and 2-component sEPG-MLE fitting per each fat ROI across all subjects for loglik, σ_N and pseudo-$R^2$are shown.

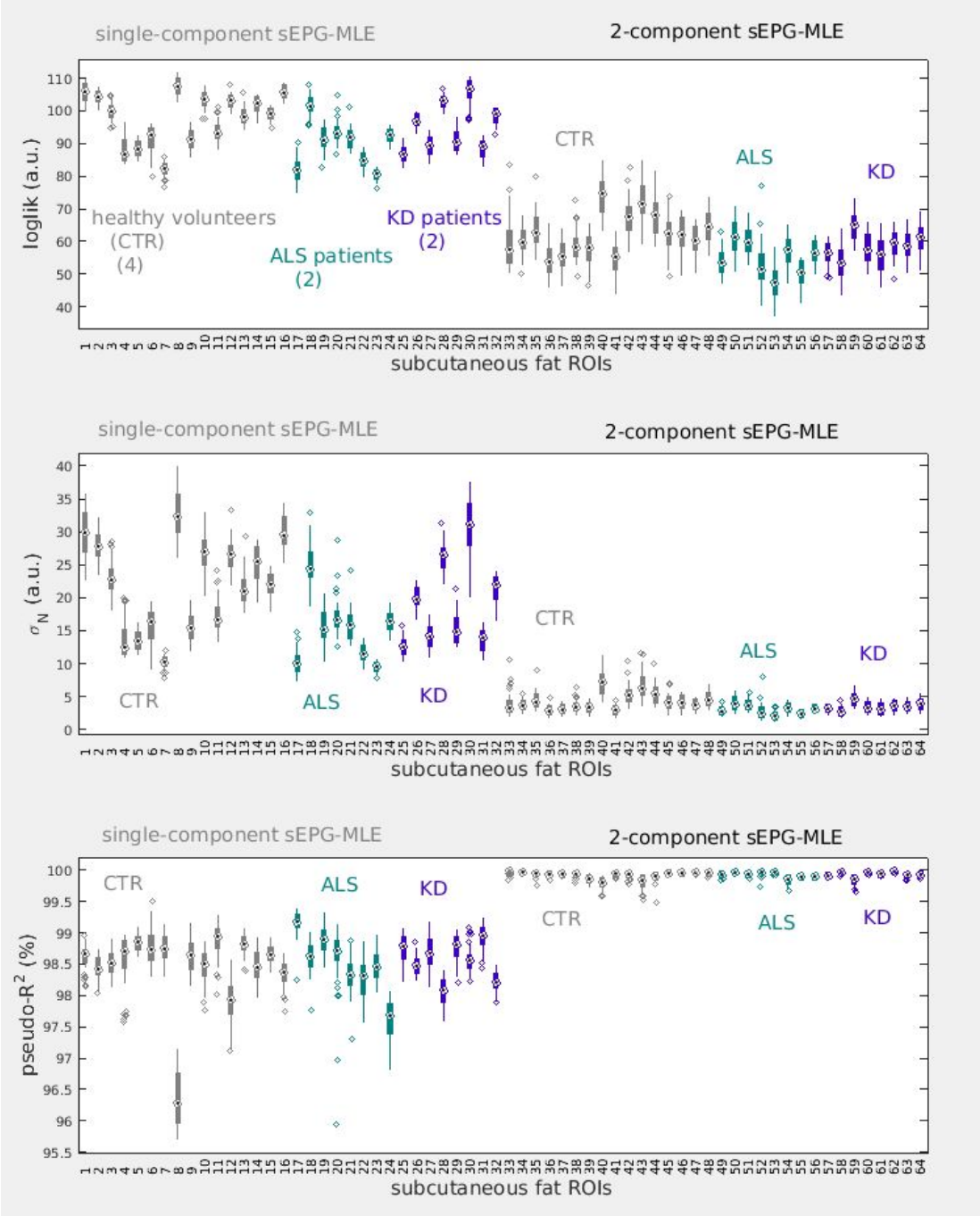


Figure 3.69: Comparison of the single- and 2-component sEPG-MLE fitting per fat ROI for all 8 subjects tested via the loglik, noise σ_N and pseudo-R^2 distributions as quality of fit metrics. The 2-component model results appear superior than the single component while healthy controls and the two types of patient groups produced similar results.

The 2-component model consistently returned lower loglik, lower, more physically consistent σ_N, and higher pseudo-R^2, close to one.

An additional feature seen here is that there was no significant difference in the yielded fat parameters between the three groups of subjects (controls, ALS and KD patients), for both T_{2fs} (p=0.137) and the fat components' amplitudes ratio (p=0.363), thus further validating the results of Table 3.4. In Figure 3.70 the T_{2fs}, 2-fat-component ratio (long over short fat component amplitudes ratio as a percentage) and T_{2f} (single component fitting) distributions are shown for completeness, per subcutaneous adipose ROI and across all subjects included.

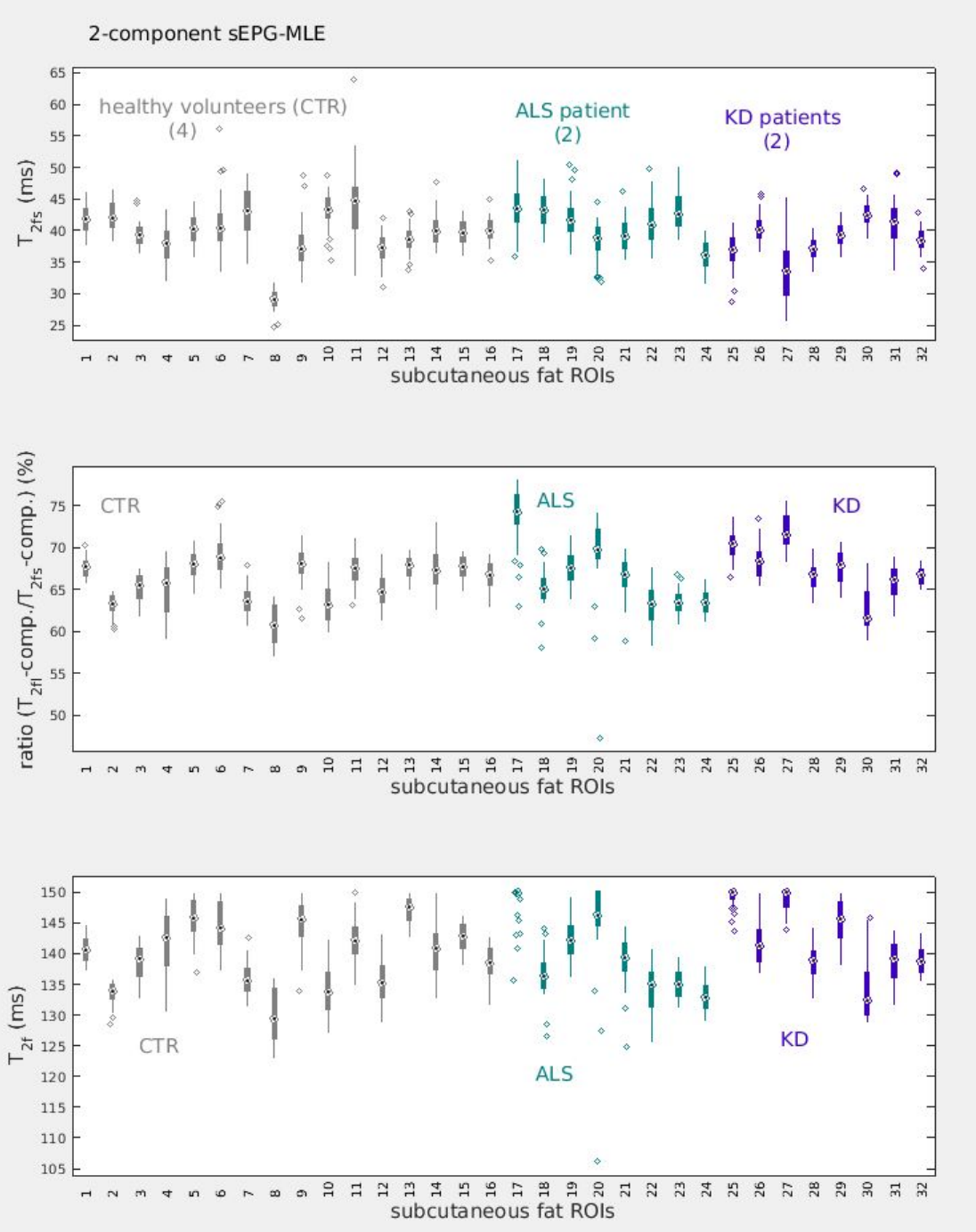


Figure 3.70: The yielded T_{2fs}, 2-fat-component ratio and T_{2f} (single component) distributions per fat ROI and across all 8 tested subjects.

Behaviour and robustness of the 2-component fat model

As can be seen in Table 3.4 the optimal model and parameter combination, i.e. the 2-component fat model with fixed T_{2fl}, produced also the $B_1 f$ estimate closest to 100%

compared to the other methods. From the 32 ROIs examined – and based on the adjacent muscle tissue B_1f behaviour (Figure 3.66), 14 were in neutral B_1f regions close to 100%, 10 in lower B_1f (<100%) regions and 8 in upper B_1f regions (>100%). This is a meaningful outcome consistent with the variations across the B_1f spectrum (above and below optimality) expected in real conditions. It also follows that the B_1f estimation errors seen with the other two methods, most notably with the alternative 2-component method (when fitting both T_{2fs} and T_{2fl}), is accompanied by suboptimal estimation performance for the remaining parameters.

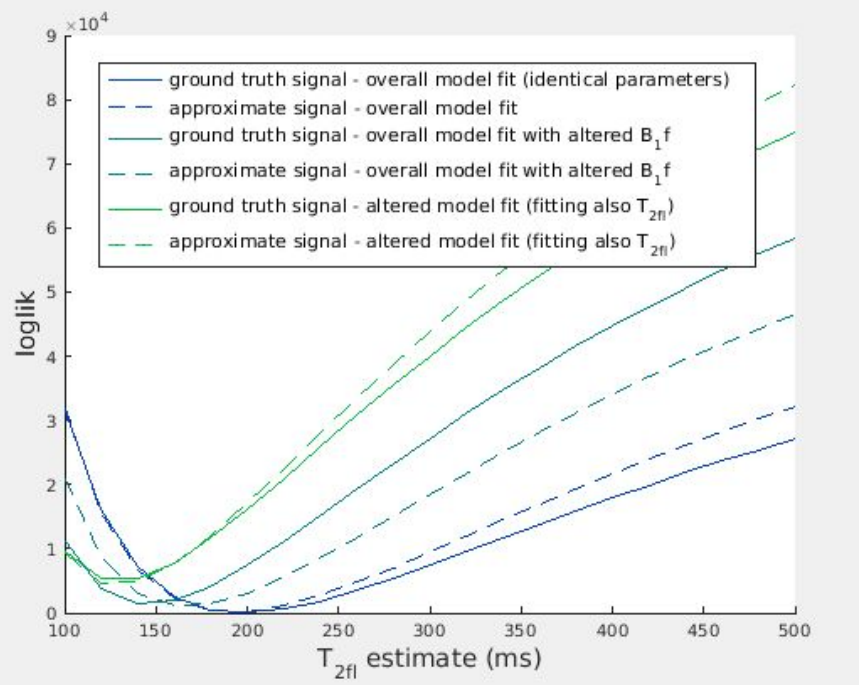


Figure 3.71: The effect of varied parameters within the fat model on the loglik determining T_{2fl}, for a ground truth signal with parameters equal to the final 2-component fat model parameters (T_{2fs}=40ms, T_{2fl}=198ms and 2-fat-componant ratio=67%) and an approximate signal (T_{2fs}=35ms, T_{2fl}=198ms and 2-fat-componant ratio=70%) (blue lines). It can be seen that increased B_1f (dark green lines) and varied T_{2fs} and the 2-fat-component ratio in the model (light green lines) are associated with decrease of the T_{2fl} and increase of the loglik estimates

In Figure 3.71 a few theoretical examples (noiseless signals) are presented showing the impact that erroneous parameter values in the fat model used for fitting have on the T_{2fl} determination.

In the first case the final 2-component model best describing fat (B_1f=97%, T_{2fs}=40ms, T_{2fl}=198ms and 2-fat-componant ratio=67%) – as extracted above – is examined, for the identical (ground truth) signal and an approximate signal (T_{2fs}=35ms and 2-fat-componant ratio=70%), to simulate the variations expected in real samples, e.g. anatomical ROIs. The lower blue and dashed blue lines show their respective loglik behaviour, demonstrating that the parameter departures in the approximate signal do not affect the T_{2fl} estimate since the minimum loglik still appears at 198ms. In the second case B_1f in the fat model is changed to 115% - the median estimate for the alternative 2-componant model (Table 3.4) – and the T_{2fl} estimate substantially decreases and loglik increases, as depicted by the middle dark green and dashed dark green lines. Finally in the last case, when changing also T_{2fs} and the 2-fat-component ratio in the model to the respective alternative model values (approximately T_{2fs}=20ms and 2-fat-componant ratio=70%), estimated T_{2fl} further decreases and loglik increases – depicted by the lighter green and lighter green dashed lines. This is likely to be related to the dependence of loglik on the echo train shape as well as parameter interdependence, as seen in the preceding paragraphs and in section 3.2.7. It seems in general that fitting data with a model within which both T_{2fs} and T_{2fl} are free (Table 3.4, 2nd column) tends to emphasize the slow fat component (the fast component is reduced in terms of both T_{2fs} and ratio) and this in turn is associated with incorrect B_1f estimates and higher loglik.

The fat ROI presented in Figure 3.72 below is a typical example of the effectiveness of the spatial regularization strategy described above.

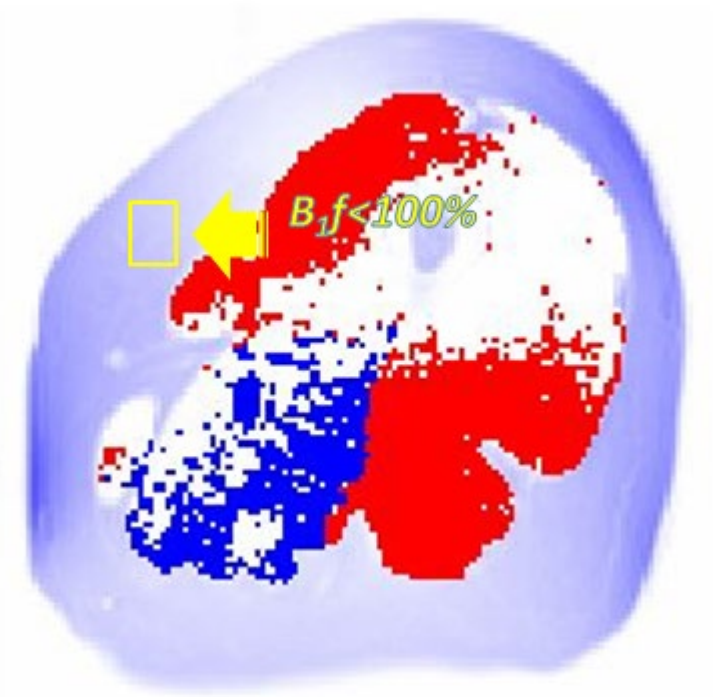


Figure 3.72: An example fat ROI (yellow rectangle) where according to the adjacent muscle-area-based spatial regularization, B_1f was switched to the lower region. The minimization performed under those conditions was seen to produce parameter estimates much closer to the average values across all subjects and ROIs.

These anatomical regions (anterior medial) of the lower limbs are common areas of B_1 transmit field deviation (Figure 3.45). Without correction, of the 56 pixels in the fat ROI, 33 produced lower loglik for B_1f in the upper region (>100%) and 23 in the lower region (<100%). This, according to the basic majority voting spatial regularization introduced in section 3.2.8 – would lead to switching of B_1f for all pixels in the ROI to the upper region; this would return the following median results:

T_{2fs}=35.0ms, 2-fat-component ratio=62.6%, B_1f=126.6%, σ_N=4.1a.u., loglik=62.5a.u. and pseudo-R^2=0.999

However, to account for the B_1f bias effects seen both in real adipose tissue data (Table 3.4) and simulation (Figure 3.65), our improved scheme dictates that the spatial regularization will be driven by the B_1f polarity condition pertaining in the adjacent muscle tissue, rather that intra-ROI majority voting, since this was seen to be more robust in general (Figure 3.48, Figure 3.66, Figure 3.72). In this case B_1f for all of the ROI pixels will be switched to the lower region, with the minimization then producing the median results:

T_{2fs}=39.3ms, 2-fat-component ratio=67.6%, B_1f=71.9%, σ_N=4.2a.u., loglik=62.7a.u. and pseudo-R^2=0.999

These results are markedly closer to the average values extracted from all ROIs (Figure 3.70) thus corroborating the validity of this approach.

Water content within fat - choice of T_{2m} lower bounds

While it might be assumed that adipose regions have effectively 100% *ff*, histological results have shown that the proportion of pure fat in subcutaneous fat tissue is in fact broadly below 90% (Thomas et al., 1962). This is now explored by fitting the previously chosen fat ROIs this time with an updated multi-component model, which consists of the 2-sEPG-component model for fat established above (T_{2fs}=40ms, T_{2fl}=198ms and 2-fat-component ratio=67%), plus a single sEPG component representing water, linked through the *ff* parameter:

$$s(TE) = (1-ff)\cdot sEPG(B_1f, T_{2m}, \alpha, \sigma_N, TE) + ff\cdot[0.33\cdot sEPG(B_1f, T_2 = 40ms, \alpha, \sigma_N, TE) + 0.67\cdot\ sEPG(B_1f, T_2 = 198ms, \alpha, \sigma_N, TE)] \quad \text{Equation 14}$$

where *s(TE)* is the signal at echo time *TE* and the model is MLE fitted to the measured data to estimate $B_1 f$, T_{2m}, *ff*, α (amplitude) and σ_N (Rician noise SD).

Note: In this case T_{2m} represents signals from water components within the fat compartment rather than muscle.

This model was used in three variations of the fitting routine, respectively setting the T_{2m} lower bound (lb-T_{2m}) to 5, 10 and 15ms. Constraining T_{2m} in this way was seen to affect the quality of fitting and the final parameter estimate distributions – in a somewhat similar way that assigning different discrete values to a fixed T_{2fl} in the model did previously. Figure 3.73 shows that the returned loglik for lb-T_{2m}=5ms is higher, and both T_{2m} and *ff* reduced, compared to the results for lb-T_{2m}=10 or 15ms, resembling the trends in the results for the alternative 2-component fat model seen previously (Table 3.4). This suggests a lower quality of fit and at the same time loss of physical meaning of the results since it appears that part of the otherwise very high fat content is substituted by a very fast decaying component combined with somewhat lower *ff*. As it has been seen before (for instance the paradigm shown in Figure 3.64) fat signals under this type of fitting can commonly return mid-range *ff* estimates combined with a short T_{2m} introducing difficulty or error in classifying the type of tissue: fat regions can appear as two compartmental, muscle and fat infiltrated regions. Since – on the other hand – no substantial difference in the behaviour and quality of fit is observed between the results for lb-T_{2m}=10 and for lb-T_{2m}=15ms, 10ms can be set as a reliable working lower bound for T_{2m} in order to not further restrict results (i.e. as a lb-T_{2m}=15ms would do) in the muscle areas, the main anatomical regions of interest.

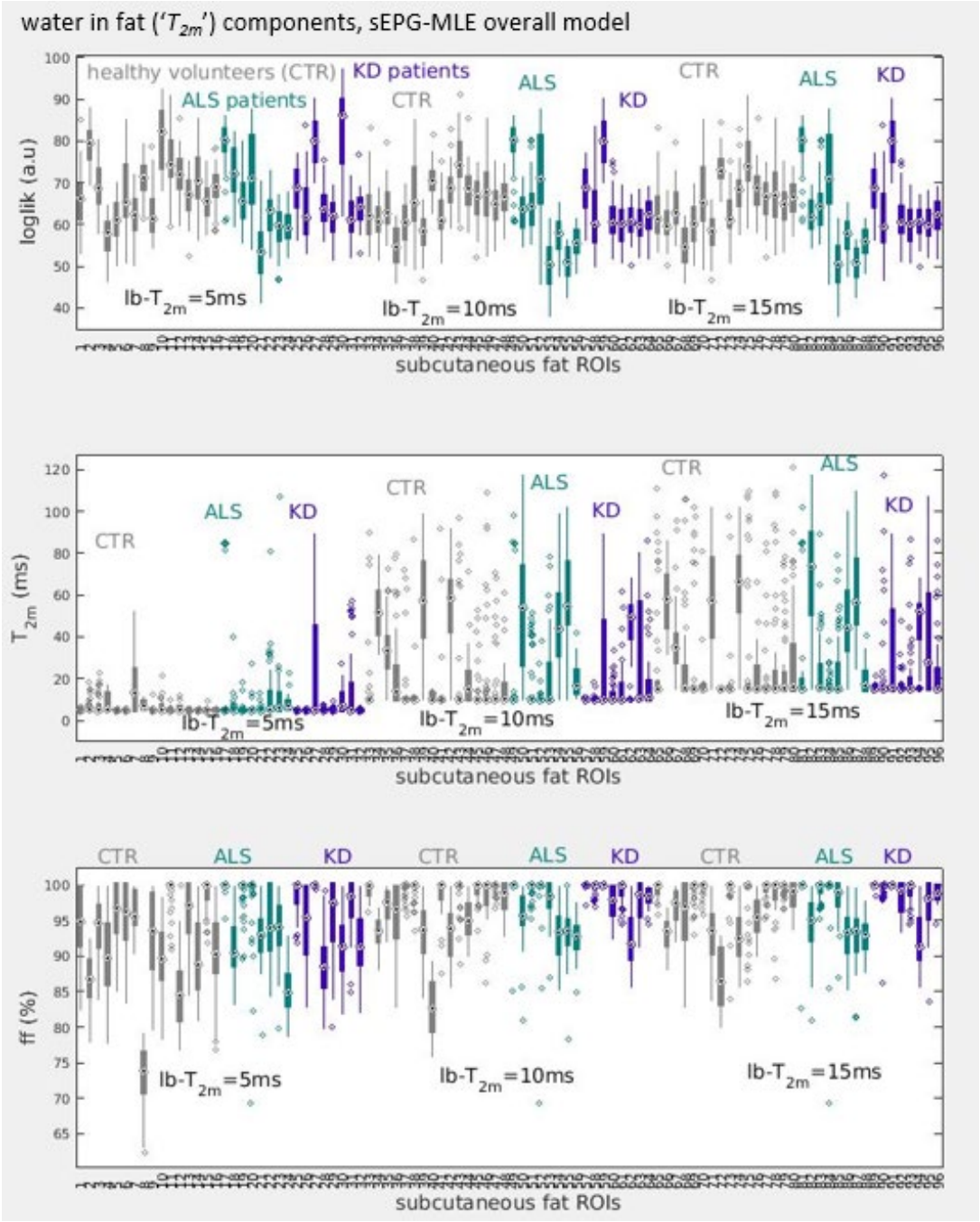


Figure 3.73: Comparison of the muscle-fat-component sEPG-MLE fitting applied to subcutaneous fat ROIs for all 8 subjects, for 3 different T_{2m} lower bounds (lb-T_{2m}) of 5, 10 and 15ms, with respect to the final loglik, T_{2m} and *ff*. The results for lb-T_{2m}=5ms return higher loglik and lower T_{2m} and *ff* compared to the results for lb-T_{2m}=10 or 15ms which are quite similar.

This on the one hand resolves the above issue of interpreting results and on the other hand can serve as means to identify pure fat or very high fat content regions. From the results in Figure 3.73, for lb-T_{2m}=10ms, 31 out of 32 ROIs have *ff* above 90% and

24 out of 32 T_{2m} below 20ms, and this parameter combination could form a criterion for characterizing fat tissue when using this model through an entire image.

The above criterion, based on the results seen so far regarding B_1f bias in high ff areas, can also be used to improve the spatial regularization process. By visual comparison of the initial B_1f binary maps before applying any regularization (as for example in Figure 3.47, left) and the raw T_2 weighted images it can be seen that high fat content areas or structures coincide with B_1f estimate in the upper region (negative polarity, coloured in blue) as shown in the examples in Figure 3.74 below.

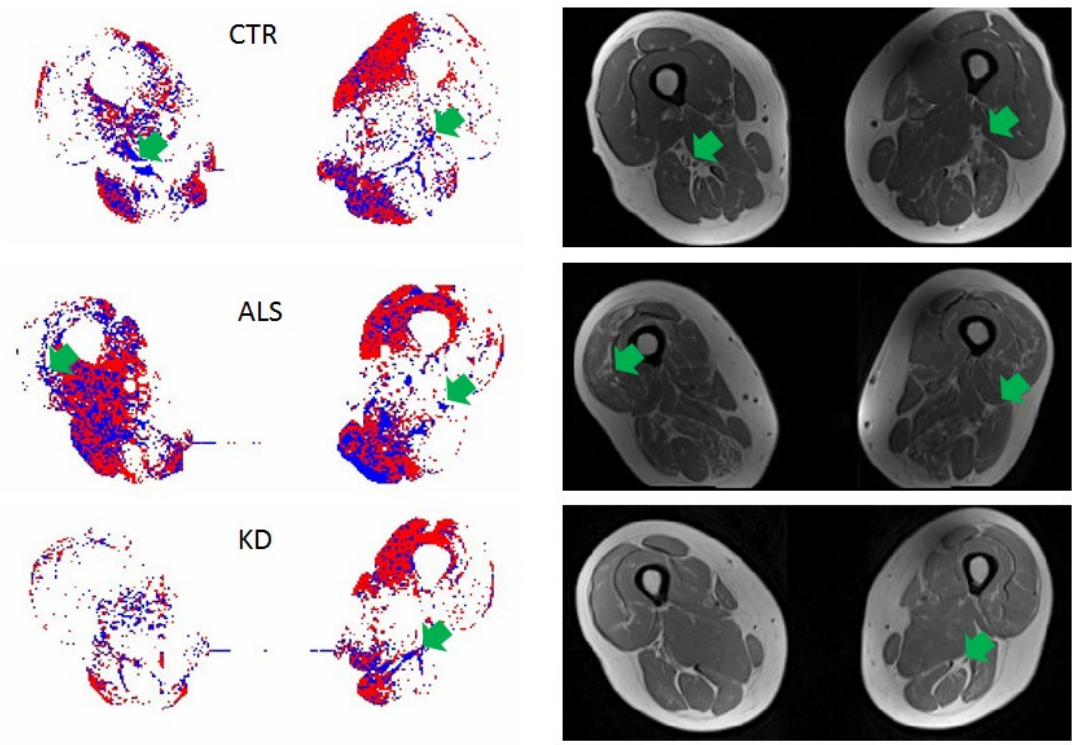


Figure 3.74: Examples of high fat content structures in one subject of each group (healthy volunteers, ALS and KD patients) raw images (right column) showing a correlation (green arrows) with the corresponding upper B_1f region estimation in areas showing otherwise neutral (single minimum, white areas) B_1f behaviour (left column)

Therefore non-participation of pixel fits manifesting this combined B_1f, T_{2m} and ff estimation behaviour to the spatial regularization part of the algorithm is likely to

increase its overall efficiency. More in particular, pixels for which estimated T_{2m} was less than 15ms and simultaneously estimated *ff* more than 90% - for which $B_1 f$ is typically overestimated – are not participating in the spatial regularization part of the algorithm.

3.3.4 Final multi-component model

Having established the overall form of the multi-component model for water and fat as expressed by Equation 14, it is informative to seek now more insight to its statistical behaviour. The tissue fat content (ground truth *ff*) seems to have an impact on the uncertainty introduced in the parameter estimates. Since previous studies limited *ff* mainly up to 50% (section 3.3.2), one of the main aims of this work was to explore the behaviour of the proposed methods across the full *ff* range. This was demonstrated in a simulation using the multi-component sEPG model (Equation 14) with the representative fixed values and initial values as defined in the previous section, for a range of ground truth *ff* values; majority voting regularisation was also included. The results are shown in Figure 3.75 where it can be seen that the dispersion of the T_{2m} and ff_a estimates (ff_a: the 'apparent' *ff*) around the ground truth value increases with increasing (ground truth) *ff*, is also restricted by respectively the lower and upper bounds, imposed in the particular MLE minimisation.

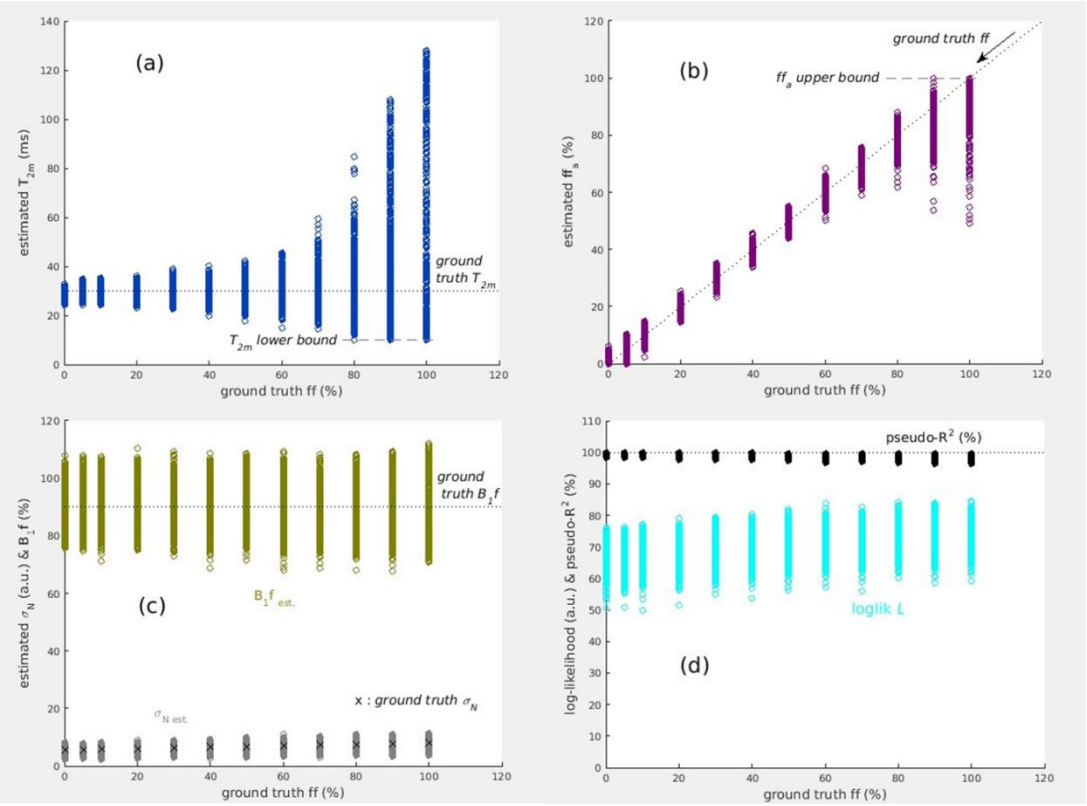


Figure 3.75: Dependence of the parameter estimation on the ground truth *ff* – for 12 discrete intervals between 0 and 100% *ff*, 1000 simulations at each, ground truth T_{2m}=30ms, B_1f 90%, SNR ≈ 50: (a) T_{2m}, (b) ff_a, (c) noise σ_N and B_1f estimates and (d) the respective log-likelihood and pseudo-R^2.

Bias (skewness) in the parameter estimate distributions of the T_{2m} and ff_a estimates is also evident, as shown in Figure 3.76 and Figure 3.77.

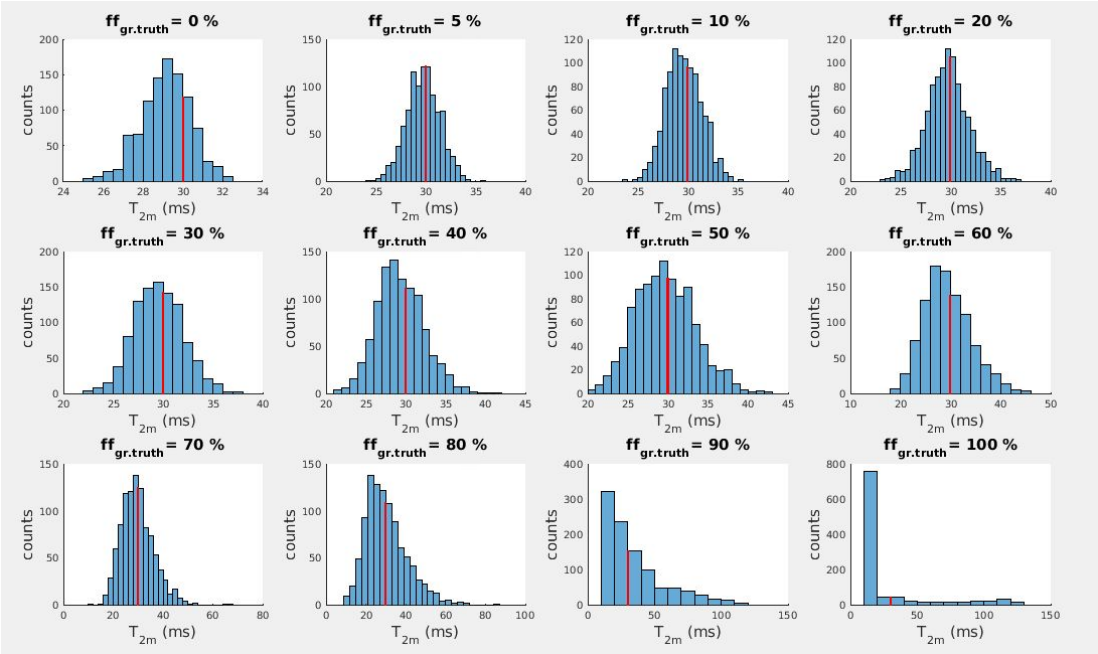


Figure 3.76: Distributions of the estimated T_{2m} for the 12 discrete values of ground truth *ff* from 0 to 100%. A general tendency for T_{2m} underestimation (bias) is seen. Red lines indicate the ground truth T_{2m} of 30ms.

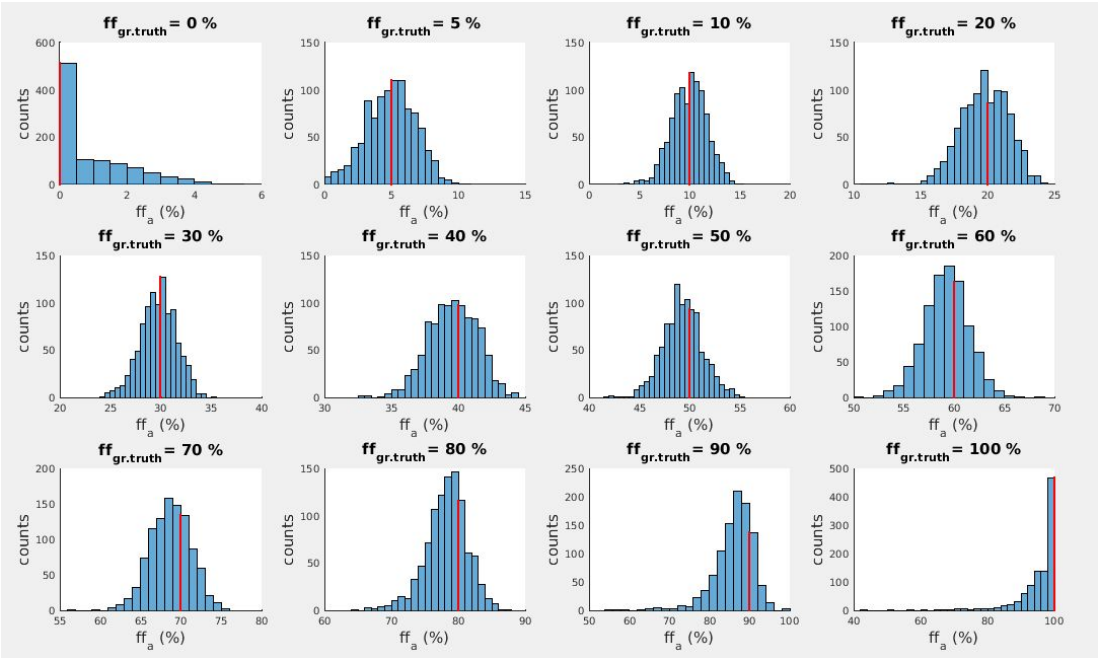


Figure 3.77: Distributions of the estimated ff_a for the 12 discrete values of ground truth *ff* from 0 to 100%. A tendency (bias) for ff_a underestimation is seen in general – despite the exception of the 3 first ground truth *ff* levels. Red lines indicate the ground truth *ff*.

It is important to emphasize that these observations as well as the overall conclusions with regards to the behaviour of the multi-component final model within the proposed sEPG-MLE algorithm might not be generalisable for the entire possible parameter space. Our main goal was to optimise the fitting routine for the estimated parameters' ranges of interest expected in the *in vivo* datasets studied in later chapters.

3.3.5 Corrections for acquisition conditions

To this point the development of the physical modelling has not taken into account additional physical factors affecting quantification relating to real world acquisition conditions. The computational steps described here aim to address these.

Magnetization transfer (MT) effects.

The example *in vivo* images presented in the above analysis appear as single slices. In reality however MESE/CPMG acquisitions are typically used producing multi-slice datasets, from which one or more slices are selected for analysis. Due to the high density of off-resonance RF pulses used to select neighbouring slices in such multi-slice acquisitions and for the relaxation rates seen in muscle (typically relatively short T_2 and long T_1) magnetization transfer (MT) effects might affect its signal intensity, while fat (relatively long T_2 and shorter T_1) is in general considered to be unaffected by MT (Marty et al., 2016, McRobbie et al., 2006). In order to quantify MT effects empirically the pixel signal intensities from a multi-slice acquisition were compared to a single slice acquisition for a healthy volunteer, with all other sequence settings equal to those used in the *in vivo* data examined so far. The ratio of the single slice

intensity and the intensity of the middle slice of the multi-slice acquisition was in average 1.20±0.01 for muscle (ROIs) and 1.02±0.01 for fat, corroborating that MT effects are negligible in fat. Therefore *ff* for this multi-slice multi-echo (MSME) type of acquisitions should be corrected as:

$$ff = \alpha_f / (\alpha_f + 1.2 \cdot \alpha_m) \quad \text{Equation 15}$$

where α_f and α_m are the respective amplitudes of the fat and the water component in the acquired signal. Since in this work calculation proceeds according to Equation 14 above, this can be re-written in terms of the initially estimated $ff_0 = \alpha_f / (\alpha_f + \alpha_m)$ (as a *ff* correction, ff_{CORR}):

$$ff_{CORR} = ff_0 / (1.2 - 0.2 \cdot ff_0) \quad \text{Equation 16}$$

in order to produce the final *ff* maps. This is an important correction in order to achieve more meaningful comparisons with results from other methods of fat content estimation, for instance with the Dixon method which is considered the golden standard in MRI.

Note: At the time of production of this report this was the only available experimental data with which to estimate MT effects, and was based on acquisition conditions similar to those pertaining for the ALS and KD data presented above and in detail in Chapter 4. This correction was applied throughout the results produced hereafter, with the reservation that it might be slightly inaccurate for different acquisition settings and datasets with different numbers of slices. The corrections used were

based on comparing a single *vs.* 7 slice acquisition whereas in most cases the patient study acquisitions consisted of 8 or 9 slices.

Incomplete T_1 recovery effects

This version of the sEPG-MLE algorithm described thus far assumes that each echo train starts from fully relaxed magnetization, i.e. fully recovered to M_0 along the z direction. However in general this might not always be the case, depending on the relationship between the tissues' T_1 and the repetition time (TR) used. The entire z-magnetization – accounting for the sequence (sEPG) history effects – available before each TR recovery period should be rescaled by the T_1 recovery term (1-exp(-TR/T_1)) where T_1 is tissue specific, because this will represent the z-magnetization available to participate in the next repetition cycle in the steady state (McRobbie et al., 2006). Therefore in a relaxometry experiment, the fat and water signal amplitudes and their relative proportion will effectively depend on this mechanism. In Table 3.5 typical values characterising this behaviour are shown.

Table 3.5: Examples of T_1 recovery ratios for 2 representative TRs – i.e. those used in the patient study protocols that produced the results presented in this thesis – for typical muscle and fat T_1s: for this calculation $T_{1\,muscle}$=1400ms and $T_{1\,fat}$=370ms.

TR (ms)	Fat T_1 recovery ratio at TR	Muscle T_1 recovery ratio at TR	Muscle proportional T_1 recovery ratio at TR
3500	0.9999	0.9179	0.9180
3000	0.9997	0.8827	0.8829

In the case of the muscle-fat component sEPG-MLE model, the proportional muscle T_1 recovery ratio at TR (last column) is the proportion of muscle-emanating signal

participating in the echo train with respect to the overall signal (including fat-emanating signal), and dividing by this value will correct the true muscle content in the *ff* equation (Table 3.6 below, 3rd column). Rearranging the equations in terms of the initial sEPG-MLE-yielded fat fraction $ff_0 = \alpha_f / (\alpha_f + \alpha_m)$ – which is not corrected for incomplete T_1 recovery – leads to the final *ff* correction expression (4th column).

Table 3.6: Examples of *ff* correction due to differential T_1 recovery rates for typical acquisition and tissue characteristics

TR (ms)	Muscle proportional T_1 recovery ratio at TR	Correction (reciprocal) for muscle amplitude α_m in *ff* (α_f: fat amplitude)	*ff* T_1 recovery correction $ffT_{1\text{-CORR}}$
3500	0.9180	$ff = \alpha_f / (\alpha_f + 1.0893 \cdot \alpha_m)$	$ff_0/(1.0893\text{-}0.0893 \cdot ff_0)$
3000	0.8829	$ff = \alpha_f / (\alpha_f + 1.1326 \cdot \alpha_m)$	$ff_0/(1.1326\text{-}0.1326 \cdot ff_0)$

Finally, in real world MESE/CPMG acquisitions the flip-angle profiles across the imaging slice for the protons in fat and those in water may not be in general aligned due to chemical-shift displacements. This is potentially most significant if for example, as is common in practice, the slice-selection gradient magnitude during the excitation pulse is different from the slice-selection gradient magnitude for the refocusing pulse, resulting in a different refocusing for the fat protons compared to the water protons and alterations in the overall signal echo trains, especially for high *ff*s and certain combinations of gradient amplitudes (Keene et al., 2020). Even though differently scaled excitation and refocusing gradients were also used in the data acquisitions analysed in this thesis (section 5.5.1), numerically simulating the above effect based on the sequences and scanner settings used, produced differences in the parameter estimates (e.g. T_{2m} and *ff*) only after the second decimal digit. For that reason no such corrections were included in the analyses presented here.

3.3.6 Quality control steps and presenting results

After pixel-wise fitting the appropriate multi-component sEPG model to *in vivo* CPMG data, before extracting pixel parameter estimates for statistical analysis, evaluation of the values obtained and the associated goodness-of-fit measures are necessary to identify remaining erroneous values. An example of this process was presented at the end of section 3.3.3, namely identifying fits which exhibit both low T_{2m} and high *ff* as 'pure fat' rather than muscle, as a form of quality control (QC) towards presenting only physically meaningful results. Such QC can be incorporated in the estimation algorithm pipeline itself, or performed as a form of post-processing, depending on the specific application. For example, the above high fat pixel identification and exclusion criterion has been used both to improve the spatial regulation phase of the algorithm, and before the final data presentation to exclude very high or pure fat regions from the final maps. However, its use may change if investigating adipose tissue is also of specific interest. Other post-fitting processing steps that were integrated in the software tool developed in this work were a minimum R^2 threshold, and 'zero signals' detection. For this R^2 threshold criterion the pseudo-R^2 described in the previous section was used, which in general for good quality signals takes high values near 100%. However, there were exceptions due to image artefacts, for instance due to flow, and segmentation imperfections, where pseudo-R^2 could take extreme negative values; therefore a threshold for pixel exclusion was set as pseudo-R^2 > zero. The second condition was useful where, due to similar acquisition related issues, the amplitude of several or more echoes in the echo train were virtually zero (below the noise) and the rest very low and random –

i.e. without following any meaningful pattern such as representing coherent decay or the noise floor (baseline); therefore any pixels with more than half their echo train being zero were discarded. Below (Figure 3.78) is presented an example raw image showing such features, and the resulting T_{2m} map after applying the above post-processing procedures. Fat areas mostly numbered in the range of a few tens of pixels (for instance for healthy controls) to a few hundred pixels (patients with severe fat infiltration); non-analysable signals (zero echoes, non-monotonic etc.) occurred much less frequently, about 5 to 10 pixels per image.

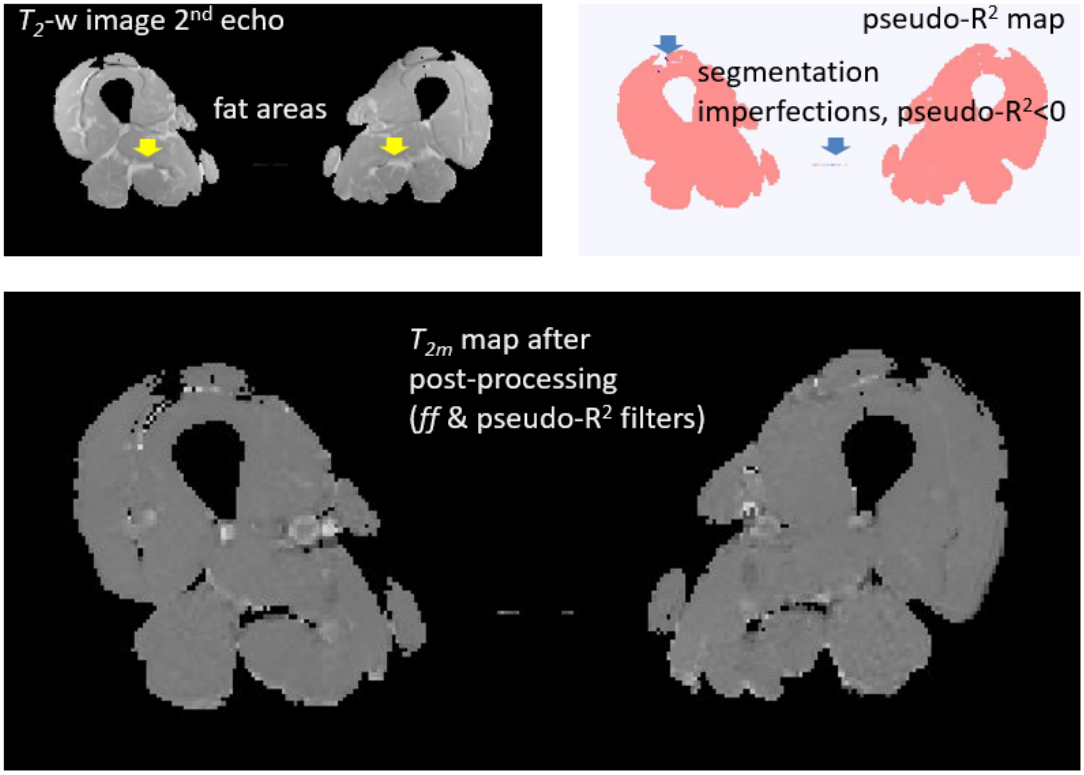


Figure 3.78: Example of a T_{2m} map for a healthy volunteer, after having excluded pixels of pure fat (T_{2m}<15ms & *ff*>90%) and of bad quality of fit (pseudo-R^2<0).

In the next section 3.3.7 examples of T_{2m} and *ff* maps are shown, with associated histograms and summary metrics.

3.3.7 *In-vivo* data examples

This section presents examples of the sEPG-MLE fitting methods applied to data sets of varying quality. Figure 3.79 presents example *in vivo* T_{2m} and *ff* maps, obtained using the sEPG-MLE muscle-fat-component model as optimised in sections 3.3.2 and 3.3.3 and after the basic post-processing steps described on section 3.3.4. Before fitting all datasets were segmented, yielding predefined anatomical ROIs as described in the specific patient group studies presented in Chapter 4 and Chapter 5. These examples all represent good quality data without obvious artefacts due to motion, flow, chemical shift etc. or large signal drop-outs due to B_1 or other inhomogeneities; also disease involvement for these patients appears to be relatively mild. Marked differences between patients and controls can be seen visually and via the histograms, most noticeably in patients the elevation in T_{2m} (about 10%) and widespread *ff* increases (up to more than 100%).

The examples in Figure 3.80 are from data exhibiting more obvious artefacts, predominantly from flow and motion (yellow and orange arrows) and other possible sources, sometimes in combination – for instance the B_1 inhomogeneity and edge/partial volume effects in subject (b). While B_1f plays a central role in the sEPG-MLE algorithm, it seems that even large departures from ideal conditions (100% B_1f) can be tolerated. Flow or motion effects can have a much bigger effect, especially when combined with B_1f departures – as in case (b) – since this both affects the accuracy of phase encoding and can severely distort the echo train. The minimization routine when applied blindly to signals from such artefactual areas often terminates at the parameter bound constraints, resulting for instance in the histogram spikes at

10ms for T_{2m} and 100% for *ff*. This can complicate the presentation and interpretation of results, in particular since opposite values of the distributions can cancel out – e.g. when results are reported as medians or means from a ROI or entire image, as seen for instance from the relatively lower T_{2m} of the compromised data compared to the previous good quality examples (for patients, e.g. ALS T_{2m} medians).

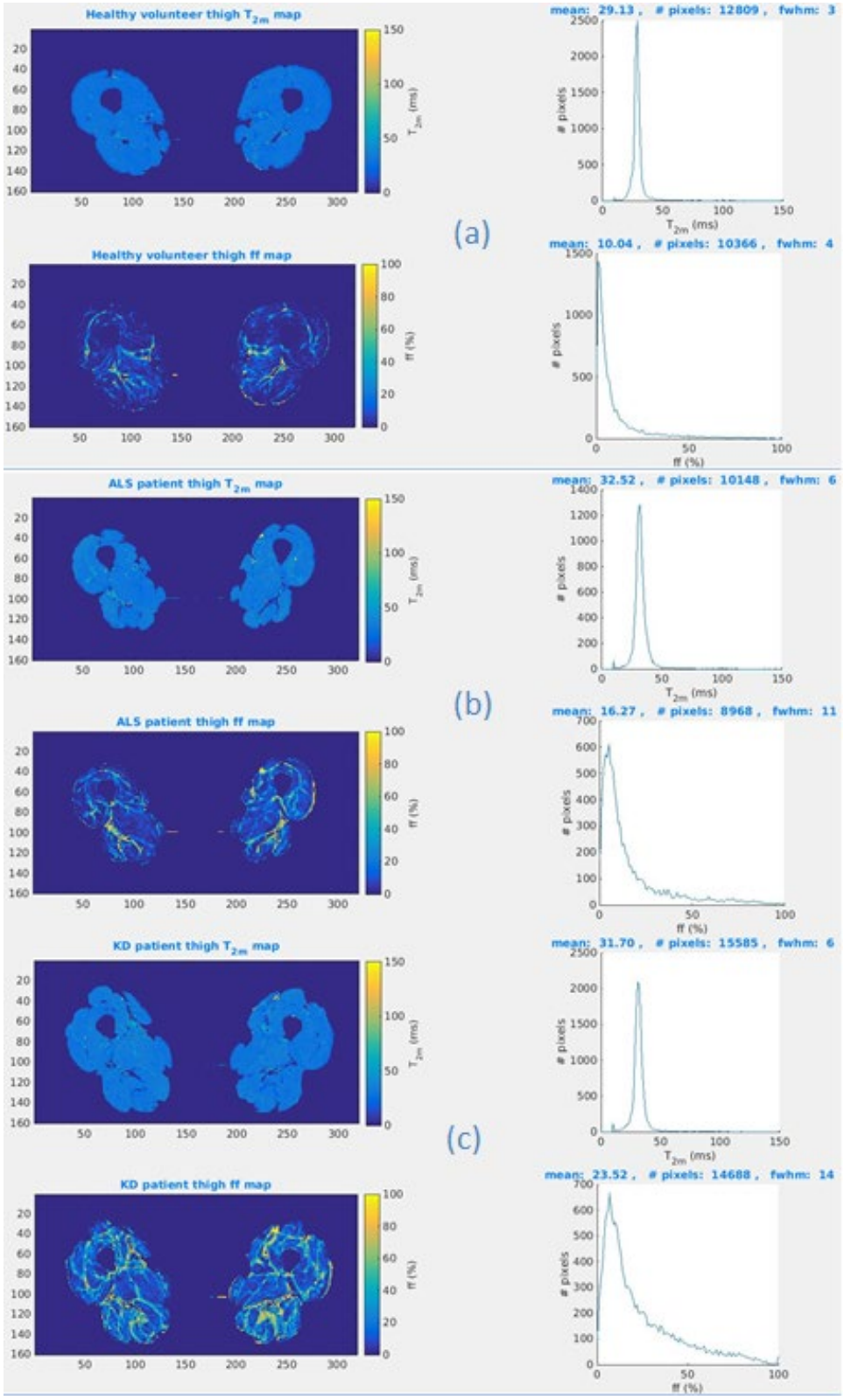


Figure 3.79: Examples of sEPG-MLE T_{2m} and *ff* maps and histograms for good quality data from a healthy volunteer (a), an ALS patient (b) and a KD patient (c).

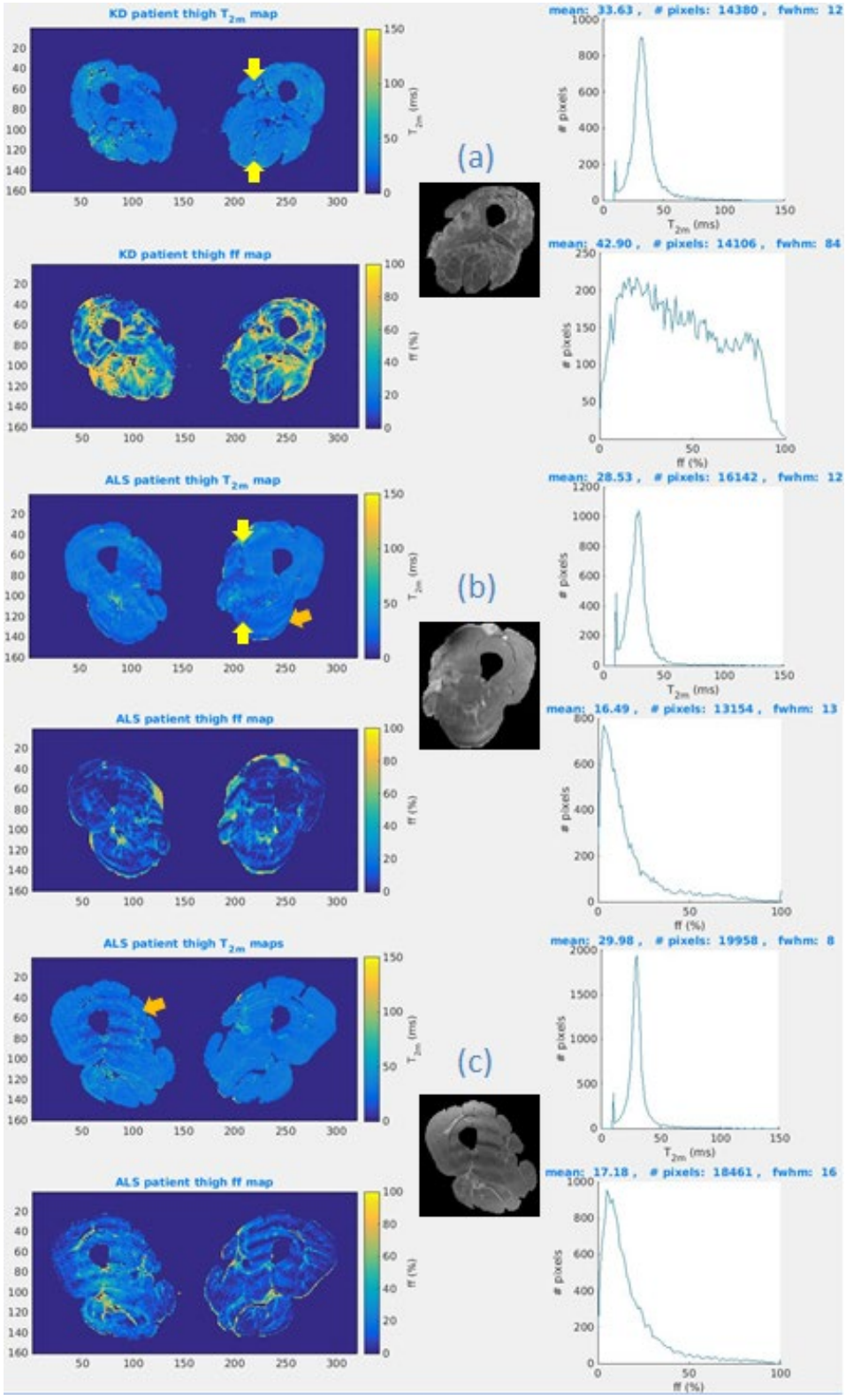


Figure 3.80: Examples of sEPG-MLE T_{2m} and *ff* maps and histograms for data with flow (yellow arrows), motion (orange arrows) and other apparent artefacts (RF field, partial volume effects etc.) from a KD patient (a) and two ALS patients (b-c). The grey scale panels are the raw T_2-weighted images (2nd highest amplitude echo).

Depending on the dataset – the proportion and severity of signal problems – additional post-processing and quality control steps can be necessary to objectively detect and remove non-meaningful or ambiguous parameter estimates from final analyses (Chapter 4, Chapter 5).

3.4 Summary and Conclusions

3.4.1 Conclusions

The aim of developing an optimised, fat-corrected T_2 (T_{2m}) fitting tool was accomplished in four main steps. Firstly, the sEPG formalism was used to achieve a model suitable to accurately describe neuromuscular CPMG image data. While the two main parameters of interest were initially T_2 and global signal amplitude (intercept) α, the B_1 field factor (B_1f) behaviour was also explored in modelling realistic RF pulse profile responses to avoid degenerate solutions; this led to the majority vote spatial regularisation scheme which was a key contribution towards improved accuracy. Second, the Rician distributed noise present in the images was addressed using MLE statistical modelling which incorporates the Rician noise σ_N as an additional parameter in the sEPG-MLE model. Third, the algorithm was extended to include a fixed-parameter 2-component approximation for the fat signal within a multi-component model, also estimating fat fraction (*ff*) as a parameter of the final model, relevant to *in vivo* data acquisitions in the patient studies. Finally, intermediate fit quality improvement and post-fitting QC steps, including spatial regularisation, principled setting of parameter bound constraints and start points, and parameter scanning strategies were implemented for robust and accurate fitting.

Final post-fitting QC procedures and means of presenting the quantitative analysis results are in general specific to the purpose of a particular study – examples will be presented for the neuromuscular disease natural history studies presented in Chapter 4 and Chapter 5.

3.4.2 Chapter Summary

Important methods developed and results presented in this chapter:

- Estimation of the Rician noise levels present in magnitude reconstructed MR images can be achieved with use of MLE methods which can explicitly account for non normal distributions.
- Combined with the sEPG model, MLE can provide accurate T_2 estimation from data acquired using MESE-CPMG sequences as was validated using phantom experiment data.
- Application of the current implementation of the sEPG formalism to MESE-CPMG data reveals non symmetry (non equivalence) of such signals with respect to B_1 field factor deviations below and above its nominal levels; this can be accounted for by integration of spatial regularisation into the log-likelihood-based sEPG-MLE fitting tool.
- A 2-component sEPG model with fixed parameters can adequately describe tissue-fat signals, under the assumption that inter- and intra-muscle fat produces signals equivalent to pure adipose tissue.
- This can be combined into an effective multi-component sEPG-MLE fitting tool for quantitative water and fat skeletal CPMG imaging data.

- Experimenting with parameter initial value grid techniques (parameter ‘scanning’) and bound constraints can be useful in optimising the fitting process and physical interpretation of the parameter estimates.
- The accuracy of fat-water quantification provided by the multi-component sEPG-MLE fitting tool can be further improved by correcting for additional phenomena affecting magnetization magnitudes in real-world acquisitions.

Chapter 4. sEPG-MLE T_{2m} estimation in motor neuron diseases: application in amyotrophic lateral sclerosis and Kennedy's disease

4.1 ALS and KD background

Amyotrophic lateral sclerosis (ALS) and Kennedy's disease (KD), also known as spinal bulbar muscular atrophy (SBMA), are two major motor neuron diseases. ALS is a rapidly progressive and fatal disorder characterised by severe impairment of motor function following the degeneration of the upper and lower motor neurons (UMN, LMN) and with a median time of 2-3 years from diagnosis to death (Keren et al., 2014). KD is a disabling disorder characterised by bulbar impairment and progressive weakness of the limbs, occurring at a slower rate than in ALS, and where life expectancy is not impaired (Grunseich et al., 2014). The cause of most ALS cases remains unknown, and the responsible genetic defects so far identified are heterogeneous, while KD always results from an expanded cytosine-adenine-guanine (CAG) repeat in the first exon of the androgen receptor (AR) gene inducing LMN degeneration leading to muscle denervation and weakness (La Spada and Wilson, 1991). At onset, ALS and KD may show similar symptoms, and distinguishing the two diseases is of clinical interest (Parboosingh et al., 1997).

Despite the efforts over recent decades, there are still no effective therapies available for both diseases. Promising targets for prospective therapeutics have been identified (Rinaldi et al., 2016), however a serious limitation for clinical trials is the lack of effective outcome measures for monitoring and assessing disease progression

(Mitsumoto et al., 2014). Together with these diseases' significant clinical heterogeneity, this limits the investigators' ability to detect effects of drugs on ALS progression, requiring clinical trials to be potentially extremely large and costly in order to be informative. There is therefore a strong need for biomarkers to refine disease stratification and identify more homogeneous patient groups and more sensitive outcome measures to better detect disease progression and treatment effects.

Skeletal muscle MRI can sensitively detect muscle involvement in neuromuscular diseases (Willis et al., 2013, Burakiewicz et al., 2017), and identify distinct pathological features, such as muscular fat infiltration or intra-muscular oedema (Straub et al., 2012). In our centre reproducible and responsive skeletal muscle MRI candidate outcome measures have been previously established, focused on quantification of muscle fat infiltration and replacement, and oedematous changes associated with inflammation. These measures have been validated in generally more slowly progressing neuropathies (CMT1A) and myopathies (IBM, LGMD2I) (Morrow et al., 2016) and are therefore likely to be highly responsive in this disease group.

MRI investigations of ALS and KD have so far focused predominantly upon the brain (Bede and Hardiman, 2014, Ferraro et al., 2016) while quantitative muscle MRI has been relatively overlooked (Bryan et al., 1998, Dahlqvist et al., 2019, Evans et al., 2014), and has mainly dealt with fat deposition (Dahlqvist et al., 2019, Hamano et al., 2004). In this chapter are described analyses by the author of multi-echo MRI data obtained as part of a large collaborative study performed by colleagues at the UCL Queen Square Institute of Neurology. A first report describing the study and cross-

sectional Dixon *ff* and STIR signal intensity findings has recently been published (Klickovic et al., 2019).

4.2 Tools and methods

For this study two patient groups were recruited, KD (N=21) and ALS (N=21), along with healthy controls, selected from two control sub-groups (N=16 each) with appropriate age and gender matching to the two patient groups. Mean age was 50.7y (SD 17) and 54.4y (SD 14.6; p=0.53) in the KD and KD-control group respectively. Mean age was 57.3y (SD 14.8) and 55.4y (SD 13.5; p=0.69) in the ALS and ALS-control group respectively. Height and weight did not significantly differ between the two patient groups and controls, except for ALS patients having a significantly lower body mass index (BMI) compared to their control group (p=0.009). Patients with genetically confirmed mutation of AR gene were included in the KD group while patients presented with a history of at least clinically possible disease according to revised El Escorial criteria were eligible for the ALS group. Additionally all participants underwent functionality tests according to established functional rating scales and detailed assessment with respect to their medical history and clinical and neurological status (Klickovic et al., 2019).

Several datasets were not available for analysis due to incomplete scans or participants dropping out from the study, or were excluded due to image or segmentation quality criteria, as described in the results section (4.4). The study was approved by the local ethics committee and all participants provided written informed consent. Healthy volunteers were scanned at two time points, from

November 2015 until January 2017 for the baseline examination and from November 2016 until January 2018 for the one year follow up; ALS patients were examined at three time points, from December 2015 until January 2017 at baseline, between June 2016 and May 2017 for the 6 months follow up and from November 2016 until November 2017 for the 1 year follow up examination; and KD patients at two time points, from December 2015 until November 2016 at baseline and from January until November 2017 for the one year follow up examination.

For the quantitative T_2 imaging the thighs of all participating subjects were examined at 3T (Siemens Magnetom Skyra) using a surface 'flex' body coil and a 12-element spine coil (integrated into the patient table) to cover proximal thighs, and a 'Peripheral Angio' (36-channel) lower limb coil, with a MESE sequence (TR= 3630/3500ms, 22 *TE*s from 10-220ms with 10ms interval, 9 x 6 mm slices, 12mm slice separation, matrix 320x160, in-plane resolution 1.3x1.3 mm). For all three groups twenty thigh muscles were manually segmented on one of the central slices – mostly the 5th (out of 9 or 8 slices in total). Individual muscles were outlined and separated from the surrounding subcutaneous fat, connective tissue and fascia etc. A single observer (U.K.) blinded to subject groups defined the muscle ROIs (excluding adipose or connective tissue, nerves, bones etc.) on images derived from the respective unprocessed shortest *TE* Dixon acquisition (*TE* = 3.45ms, see following paragraph) in all subjects using the itk-snap software (Yushkevich et al., 2006). This was done for the purposes of the collaborative study and publication mentioned above (section 4.1). Since the resulting segmentation masks were defined in the space of an acquisition different to the MESE sequence (i.e. the 3-point Dixon raw images), it was

necessary for these to be spatially adjusted by the current author using a custom-written software tool in order to match to the space of the T_2-weighted raw images. An additional tool was created to then apply these ROI segmentations to the raw T_2-weighted images. Selection of inter-subject comparable slices was assured by identification of specific anatomical landmark structures in the lower limb (lateral tibial condyle).

The main aim of the following analyses performed by the author were, following some further post-processing optimisation, to assess the effectiveness of the signal model and fitting method developed in Chapter 3 to extract useful information from this dataset, and compare these results with those obtained using previously existing methods, specifically the Dixon *ff* determination method and (STIR) fat suppressed imaging, described in the next paragraph and sections 2.1.2, 4.4 and 4.5. The proposed method uses the sEPG model as finally described in section 3.3.2, Equation 14; this is its generic mathematical expression (it cannot be written in closed form) describing the aggregated EPG (function) calculations across the slice profile for the combined muscle and fat compartments, which is subsequently fitted pixel-wise to the study data using MLE minimization in a custom-written Matlab tool (section 2.2.3). The parameters to be determined for each pixel are muscle water T_2 (T_{2m}), fat fraction (*ff*), B_1 field factor ($B_1 f$), overall amplitude (α) and noise standard deviation (σ_N). While T_1 for muscle and fat were fixed in accordance to existing literature values, and the fat compartment T_2 behaviour determined separately (section 3.3.2), the remaining sEPG model parameters (*TE*, ETL, RF excitation and refocusing profiles and duration vectors) were set according to sequence and scanner settings, as listed at

the top of this paragraph. A summary of the main algorithm settings used is listed in Table 4.1 below and also in Figure 4.2, section 4.3.1, via a comprehensive graph of its usage.

Table 4.1: sEPG-MLE fitting tool settings

Estimated parameters	Start points	Bounds lower	upper	**Fixed parameters**	
$\boldsymbol{B_1 f}$ (%)	Initial loglik(B_1f) vector troughs	20	180	$\boldsymbol{T_{1m}}$ (ms)	1400
$\boldsymbol{T_{2m}}$ (ms)	50	10	150	$\boldsymbol{T_{2fs}}$ (ms)	40
$\boldsymbol{ff}$ (%)	50	0	100	$\boldsymbol{T_{2fl}}$ (ms)	198
Amplitude $\boldsymbol{\alpha}$ (a.u.)	5·max(echo)	max(echo)	Inf	$\boldsymbol{T_{1f}}$ (ms)	370
Noise $\boldsymbol{\sigma_N}$ (a.u.)	20	0.1	max(echo)/2	$\boldsymbol{fat_{s/l}}$ ratio (%)	67
Convergence settings: Matlab 'fmincon' (constrained minimisation) function 'sqp' algorithm max # iterations: 200 step size tolerance: 10^{-6}					

To provide comparative data collected using the Dixon technique, three-point-Dixon (Glover and Schneider, 1991) images were analysed (2D gradient-echo TE1/TE2/TE3/TR= 3.45/4.60/5.75/102ms, flip angle 10°, 9x6mm axial slices, slice gap 12mm, FOV 42x21cm, matrix 320x160, pixel size 1.3x1.3mm, NEX=4). Images were post-processed offline with a Python programming language pipeline implemented by a colleague (Dr Christopher Sinclair), according to Glover and Schneider's algorithm (section 2.1.2) and separated fat (f) and water (w) images were used to calculate pixel-wise fat fraction (*ff*) maps according to *ff*(%)= 100*f/(w+f). This was performed for the same segmented muscles as above.

To compare the performance of the sEPG model with a more conventional exponential model, a multi-exponential function:

$$s(TE) = \alpha \cdot (1 - ff) \cdot e^{-\frac{TE}{T_{2m}}} + \alpha \cdot ff \cdot \left(0.3 \cdot e^{-\frac{TE}{45}} + 0.7 \cdot e^{-\frac{TE}{214}}\right)$$

where *s* is the signal at each *TE* and muscle water T_2 (T_{2m}), fat fraction (*ff*) and overall amplitude (α) were the parameters to be estimated, was fitted pixel-wise to the data using nonlinear least squares minimization (section 2.2.2) in a custom-written Matlab tool. The fat component parameters were determined separately as the mean values obtained when fitting to data from 4 subcutaneous fat ROIs in 8 representative subjects. In terms of quality control, values were excluded from the maps for pixel data which failed to meet empirically determined fit-quality criteria: R^2 (goodness of fit)>0.8, amplitude α lower confidence interval (CI)>0, amplitude α<10 times the 1st *TE* image amplitude, T_{2m}-CI-width/ T_{2m} <100%, *ff*<50% and T_{2m} <100ms.

Additionally transverse short tau inversion recovery (STIR) acquisitions (TR = 5200ms, TE = 39ms, TI = 220ms, NSA = 1, iPAT = 2, 31 slices, FOV = 420 mm, voxel size = 1.1 × 1.1 × 6.0 mm, slice gap = 6 mm) were performed separately at the level of the mid-thigh for both limbs.

4.3 Analysis optimisation

For this somewhat heterogeneous data set the post-fitting quality control steps introduced in section 3.3.3 needed further refinement to maximise the accuracy and precision of the parameter estimates. Even though in this study data were of high quality overall, there are common imperfections in many images, even in the good quality ones, that required better characterisation to avoid misleading interpretation of the results. Typical examples are motion and flow artefacts (even at mild levels),

B_1 field non-uniformity, and issues evidenced by parameter estimate histogram spikes. Importantly, these are common across different anatomical features in the acquired images, from healthy tissue to regions of severe pathology.

As seen in Chapter 3, iterative minimisation returning parameter estimates at or close to pre-defined bound values can be a manifestation of suboptimal fitting. An important example in the present context is T_{2m} estimates returned at or closely adjacent to the T_{2m} lower bound (10ms). The return of such values was characteristic of pixel subtypes with specific types of signal behaviour. A first subtype already discussed were those returning lower bound T_{2m} values in combination with high *ff*, identified previously as volumes containing effectively pure fat, already largely excluded by the previously established post-processing steps (section 3.3.6). Another pixel region subcategory returning lower-bound T_{2m} values occurred due to image artefacts as discussed previously (section 3.3.7). These were identified by significant spikes around 10ms in the respective T_{2m} histograms. It was found necessary in this study to further characterise these pixel subtypes – since both exhibit shortened T_{2m} – in order to distinguish intermediate to high fat content pixels from pixels returning non-meaningful spurious results due to artefacts. The former may ideally be retained for the analysis while the latter may be usefully discarded.

4.3.1 Assessment of preliminary fitting performance

Examples of obviously sub-optimal quality raw images found in the current dataset, in terms of artefacts, SNR and RF field homogeneity are presented in Figure 4.1, alongside images demonstrating severe neuromuscular pathology, which also

creates data fitting challenges. Good quality healthy volunteer raw images are also shown for comparison.

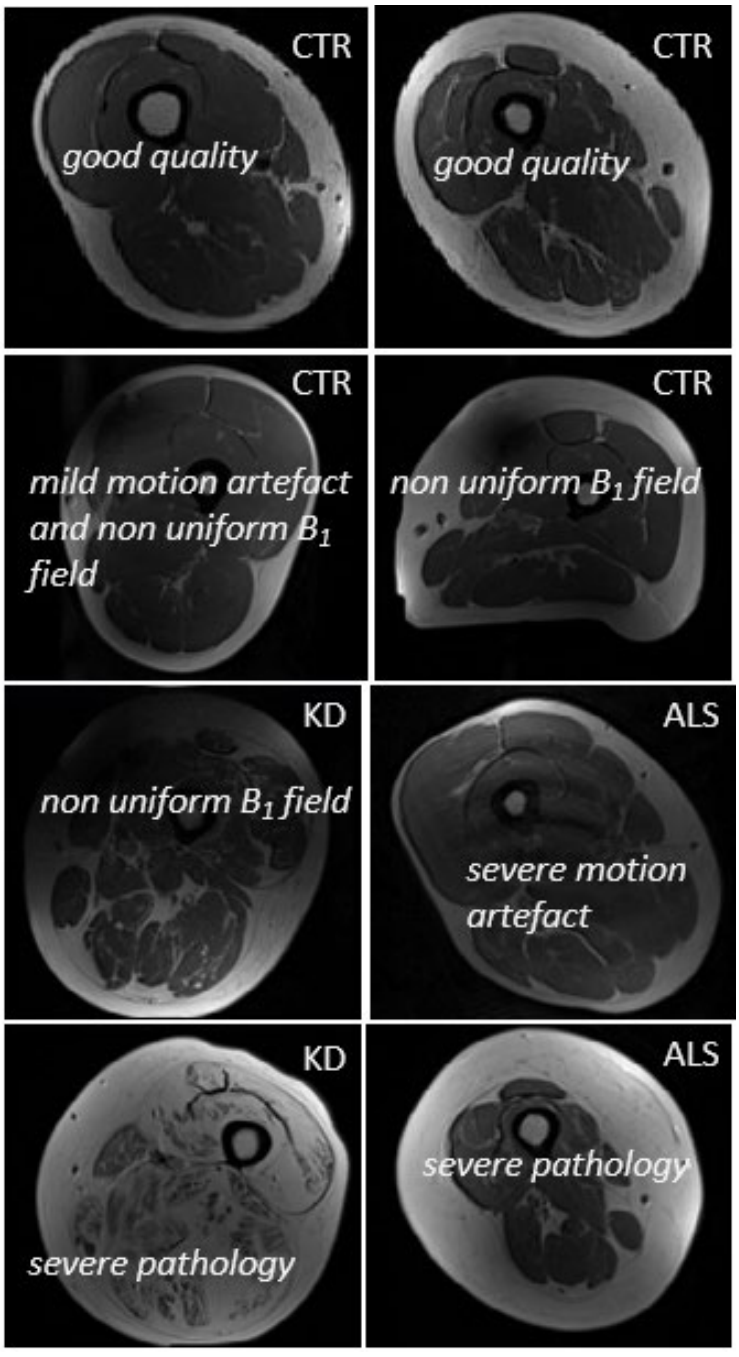


Figure 4.1: Examples of good quality healthy volunteer raw images (T_2-w, 2nd echo, *TE*=20ms) (CTR, top row) shown for comparison with those with the most severely suboptimal image quality (2nd and 3rd row), and those most severely affected by pathology within the study dataset (bottom row). Artefacts were mainly seen for patients (ALD/KD, 3rd row) and to a lesser extent for healthy volunteers (CTR, 2nd row).

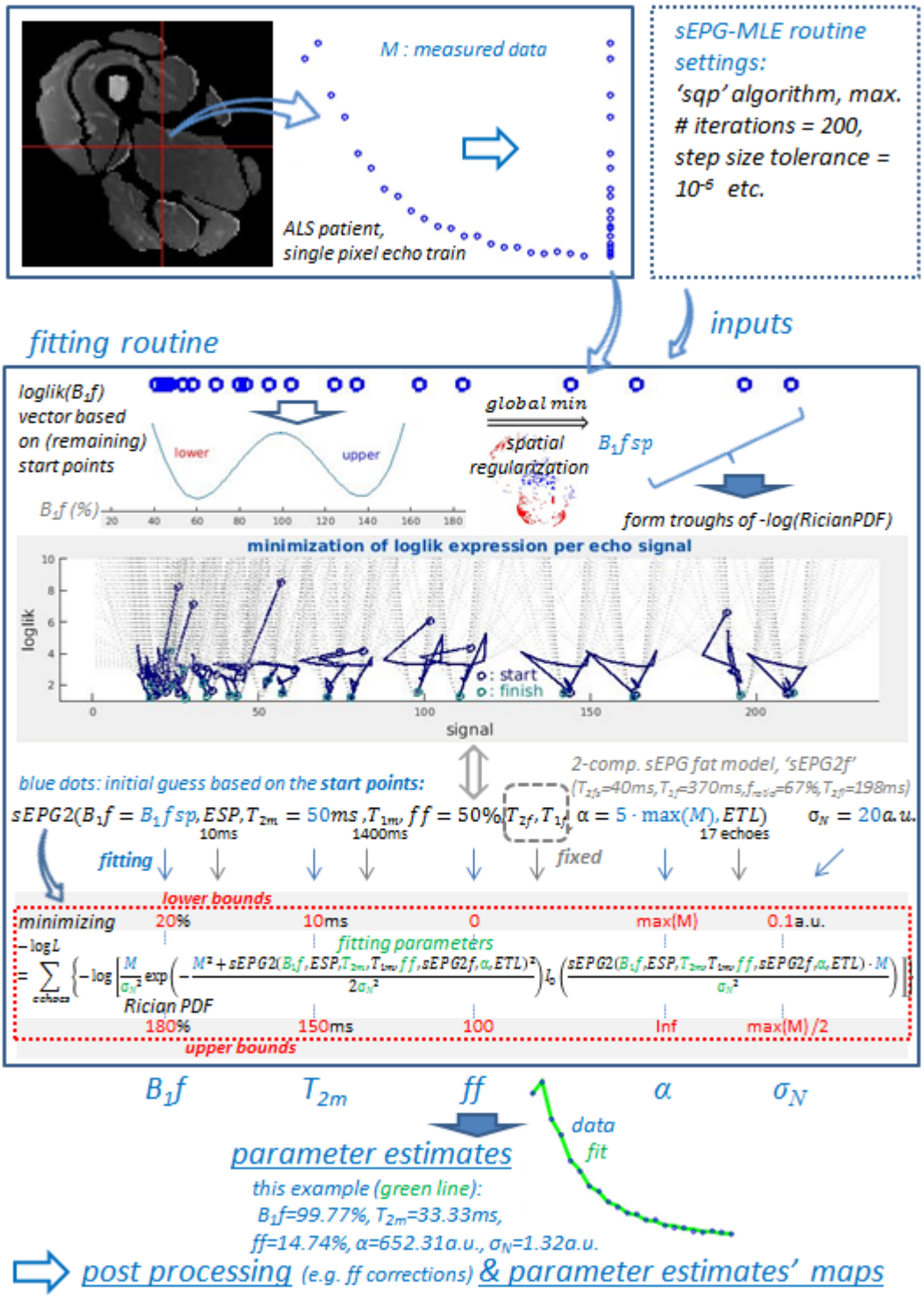


Figure 4.2: A scheme illustrating the implementation of the sEPG-MLE fitting algorithm for the *in vivo* data, including start point and bound constraint specifications and algorithm settings.

A preliminary data analysis using these exemplar CPMG data from the study was first performed to enable fine-tuning of the fit post-processing procedure before the final

parameter estimate mapping was performed for the full dataset. The workflow for the core part of the fitting procedure is shown in Figure 4.2 including the listed parameter estimate start and bound values.

A majority vote-based B_1f error polarity spatial regularisation strategy was incorporated in the pipeline to achieve physically consistent B_1f behaviour avoiding spurious estimates from non-local log-likelihood minima (section 3.2.6 and 3.3.2), and problematic (non-physical) or truncated-signal pixel data, as well as pure fat signal pixel data, were discarded (section 3.3.6), before the parameter estimate maps were stored. To more clearly explore various aspects of the fitting process, correlation graphs for the sEPG-MLE T_{2m} and *ff* estimates (after applying the above quality control steps) corresponding to the example datasets shown in Figure 4.1 are plotted in Figure 4.3: for each pixel from the thigh-level cross-sectional segmentation (as illustrated in Figure 4.2) the T_{2m} estimate is plotted against the respective *ff* value.

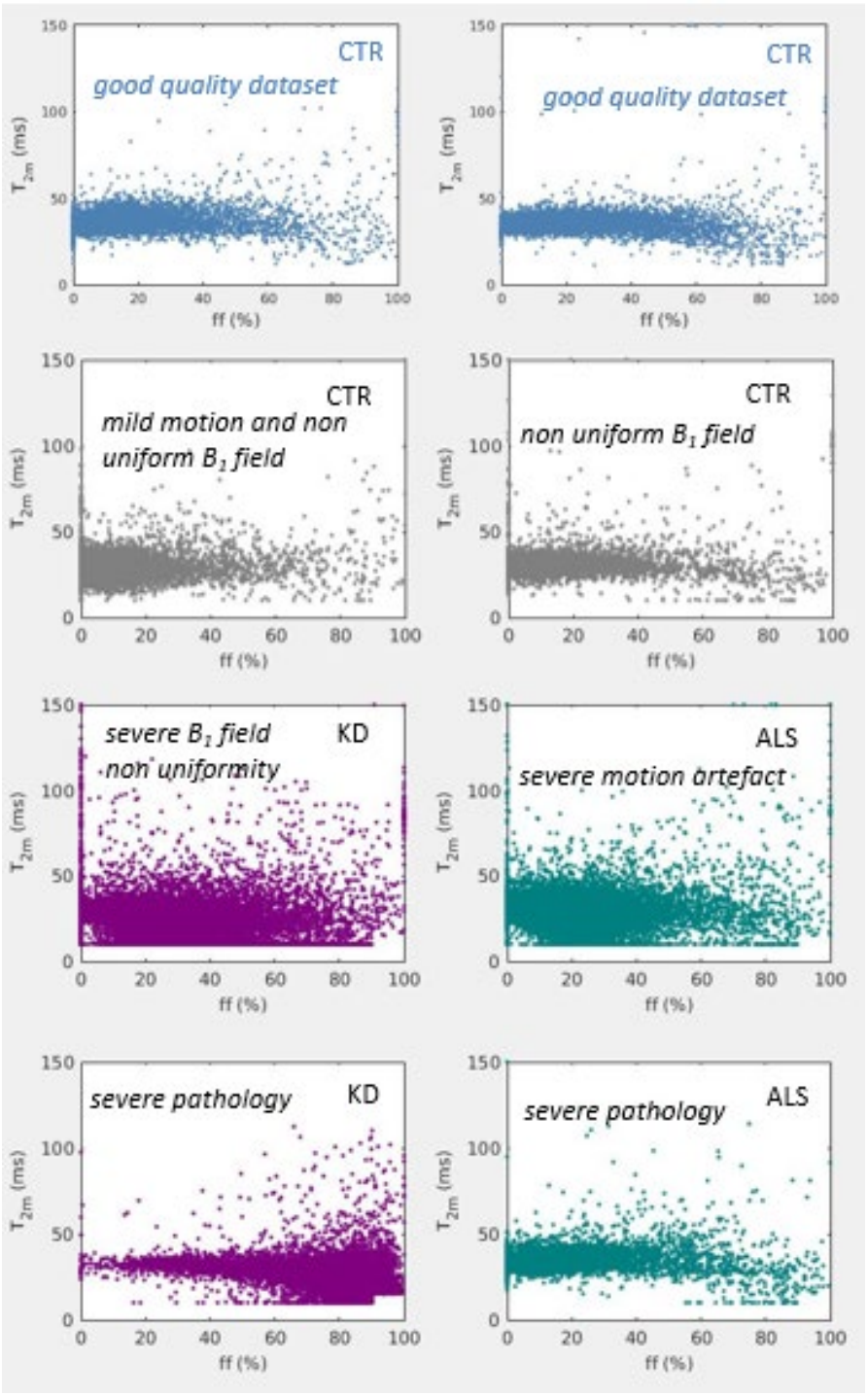


Figure 4.3: Pixel-wise T_{2m} and *ff* estimate correlation plots corresponding to the example data sets shown in Figure 4.1. T_{2m} is broadly uniform with increasing *ff* however on occasion the minimisation returned bound constraint values, mostly the 10ms lower T_{2m} bound. Normative T_{2m} from good quality healthy volunteer datasets (top row) appears constant and tightly clustered (around 30ms) across the majority of the *ff* range.

While T_{2m} appears in general uniform across the *ff* range, its lower bound (10ms) is returned in a number of cases, suggesting a failure of fitting. This is not seen in the

good quality healthy volunteers' datasets where T_{2m} appears clustered around a mean value of approximately 30ms for all values of *ff* up to approximately 60%, showing greater dispersion thereafter (top row). Bound T_{2m} values were returned for a small number of pixels for the relatively poor quality controls' acquisitions (2nd row), evidence as the sparse band of T_{2m} = 10ms data points at the upper *ff* range. This was seen more markedly in the poorer quality patient data sets, occurring e.g. across the entire *ff* range for the patient images in the 3rd row. For the bottom row of data from good quality images with severe fat infiltration this effect is confined to mostly higher *ff* – as expected from the multi-component sEPG-MLE algorithm behaviour (section 3.3.2 and 3.3.3). The excluded group of pixels with T_{2m} below 15ms combined with *ff* above 90% (i.e. those identified as 'pure' fat pixels, section 3.3.3) corresponds to the blank square at the bottom right corner of all graphs. As explained in section 3.3.2 these thresholds were set considering the average results – medians, percentiles and distributions – from the subcutaneous fat ROIs (Figure 3.70). The 10ms T_{2m} and mostly high *ff* pixel sub-group in the middle graphs can be considered a remaining category of fat pixels for which the model-fitting failed, in which the signal component assigned to T_{2m} actually represents an additional component emerging from fitting to the multi-component fat signal, rather than a true muscle water T_2 (T_{2m}) estimate (section 3.3.3). The continuous range of T_{2m} and *ff* values seen away from these cut-off values is expected to describe biologically meaningful T_{2m} and *ff* changes due to pathology. For this to be true, non-meaningful fits occurring within that distribution due to inadequate quality data should ideally be eliminated.

4.3.2 Quality control steps

Even though returned parameter estimates for a particular pixel may lie within biologically plausible ranges, as suggested above inadequate data quality may still affect the quality of fitting and resulting in unreliable estimated parameters. To identify such pixels pseudo-R^2 was next investigated as an additional quality control metric. Considering first the relationship between the returned T_{2m} estimates and pseudo-R^2 (Figure 4.4), a combination of returned T_{2m} at or close to the 10ms bound and low pseudo-R^2 was a prominent feature in quality-compromised patient data (3rd row), while being mostly absent in the healthy volunteer data (1st and 2nd row). It is also apparent that the short T_{2m} pixel regions ascribed above to meaningful features of high fat regions in the last row images exhibited the same high pseudo-R^2 as the remaining pixels, suggesting that that the model thus parameterized is actually a good fit to the data, supporting the interpretation of these results as physically meaningful.

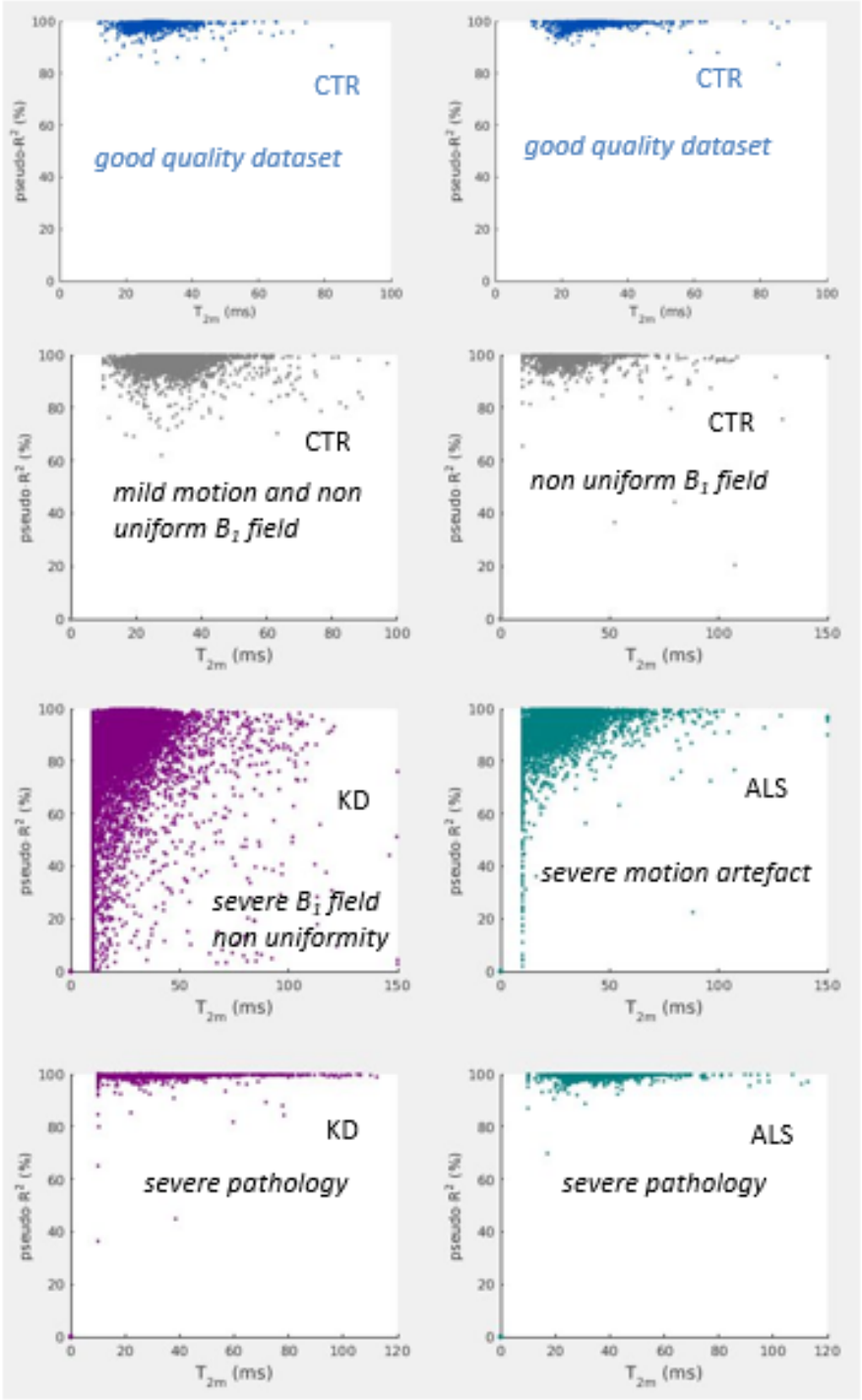


Figure 4.4: The correlation plots of T_{2m} estimates *vs.* the corresponding pseudo-R^2 metric corresponding to the Figure 4.1 raw images. It can be seen that pseudo-R^2 decreases with decreasing image quality (2nd and 3rd row, in order of increasing severity) but not with T_{2m} changes due to pathophysiological features (last row).

In other words, excluding all pixels returning a pseudo-R^2 ≤ 80 – 90% would preferentially remove unreliable fits due to poor data quality without significantly

impacting the pixel population returning meaningful T_{2m} values which might deviate from the normal range due to pathological processes.

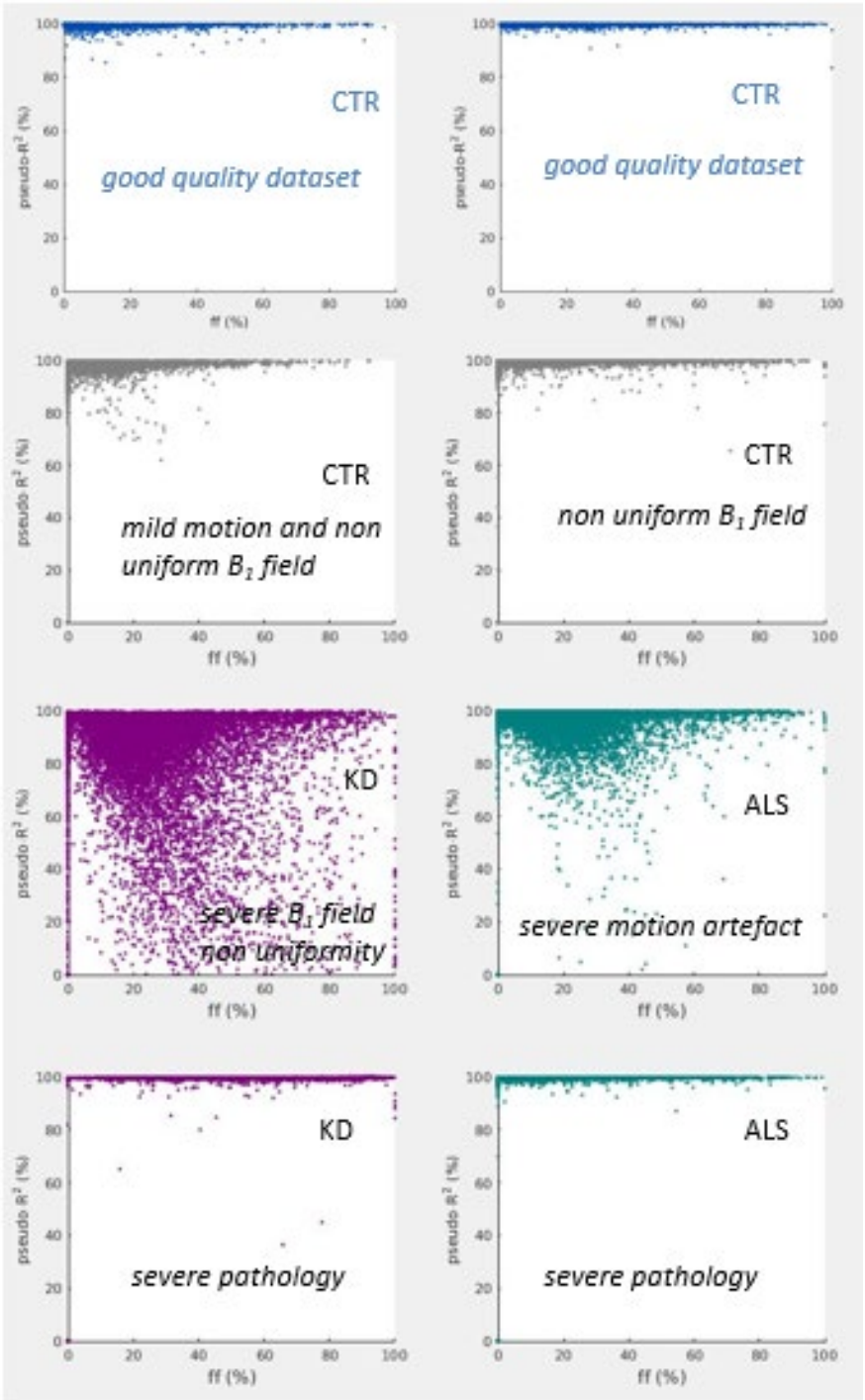


Figure 4.5: Correlation plots of *ff* estimates *vs.* the corresponding pseudo-R^2 values corresponding to the Figure 4.1 raw images. It can be seen that pseudo-R^2 is more commonly reduced with decreasing image quality (2nd and 3rd row, in order of increasing severity) uniformly along the *ff* range.

Next, considering the relationship between pseudo-R^2 and *ff* (Figure 4.5) it can be seen that as expected pseudo-R^2 tends to decrease with decreasing quality of data, but is otherwise independent of *ff* in each image.

Taking together the above, for this study data set, an additional post-processing step eliminating pixels returning pseudo-R^2 below 80% or T_{2m} below 11ms was considered appropriate. The effects of this are demonstrated in Figure 4.6 showing an example for each of the above cases.

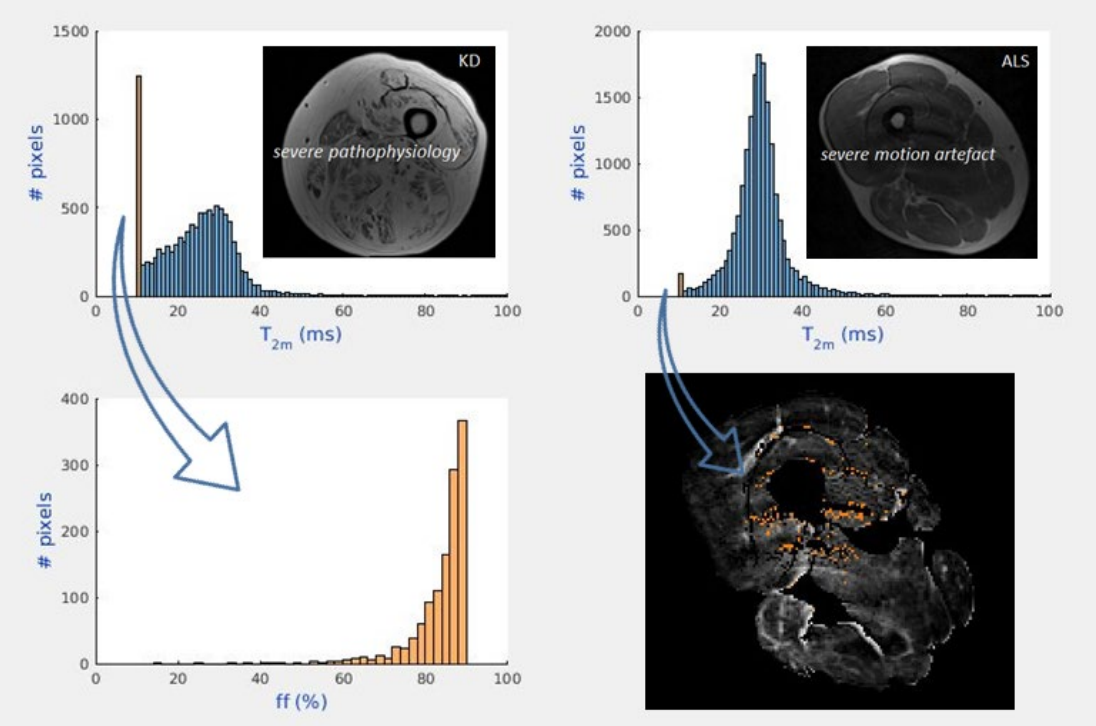


Figure 4.6: The effect of excluding pixels near the 10ms lower T_{2m} bound ($10<T_{2m}<11$ms) for two pixel subgroups of interest: left, from fat pixels where such values represent in reality an additional component within fat, and right, from pixels with data quality compromised by artefacts. In both cases the previous post processing steps of eliminating pixels for which $T_{2m}<15$ms and *ff*>90%, and pseudo-$R^2<80\%$ have already been applied.

For the left image manifesting severe pathology (fat infiltration), the pseudo-R^2 threshold had negligible effect (only 4 pixels eliminated) due to the good quality data and fitting, however several hundred pixels still contributed to an apparent 10ms T_{2m}

histogram peak. These were identified as high fat content pixels, as denoted by their *ff* distribution shown in the histogram below (extending the initial subgroup of T_{2m}<15ms and *ff*>90%), for which T_{2m} could not be reliably determined. For the right hand data set, manifesting severe motion artefacts, several hundred pixels were eliminated by both criteria, a significant proportion of which were identified by the T_{2m} lower band between 10 and 11ms. These are identified in colour in the bottom right *ff* map adjacent to pixels already missing due to the previously established quality control steps (section 3.3.6) (void pixels corresponding to the dark strips formed by motion artefacts), showing the complementary action of the different steps in the post-processing strategy.

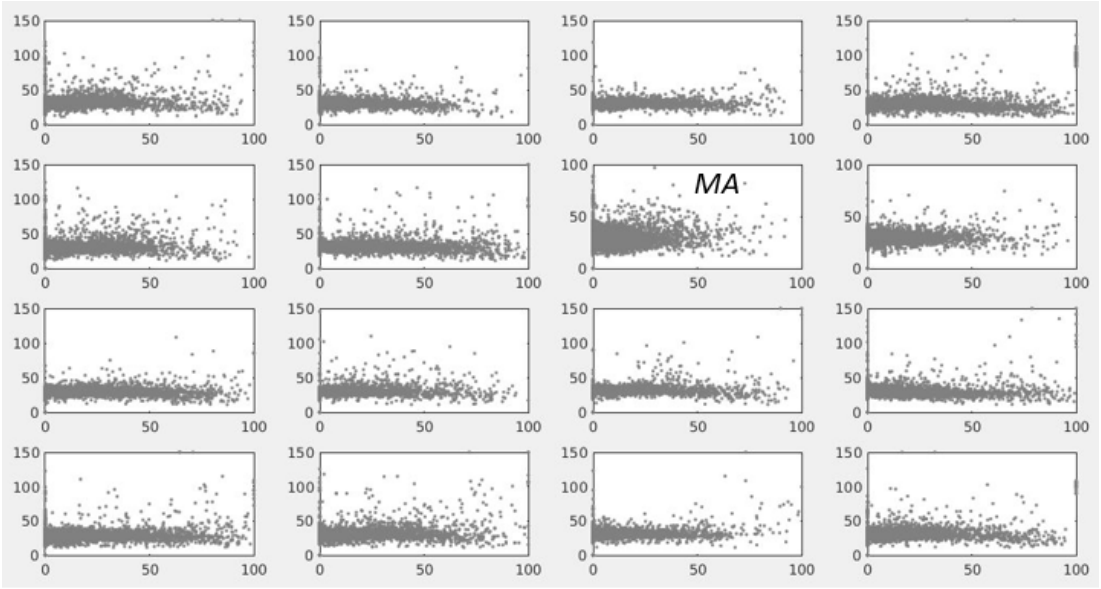
Once the above criteria are used to as far as possible eliminate data from pixels returning unreliable T_{2m} values, as a final step returned *ff* estimates should be corrected for the sequence-parameter specific inter-TR T_1 recovery difference between muscle and fat (section 3.3.2). Since the TR for the protocol used was in most cases 3500ms (section 4.2) *ff* should be corrected by the expression $ff/(1.0893-0.0893 \cdot ff)$ (Table 3.5). In addition an MT correction (section 3.3.2, Equation 16) was also applied. For the Dixon *ff* estimation a closed form correction similar to that proposed elsewhere (Azzabou et al., 2015) was used, namely by means of the expression $ff_{d\text{-}corrected}=ff_d/(0.45ff_d+0.55)$, to account for the multiple lipid resonance frequencies.

4.4 Results

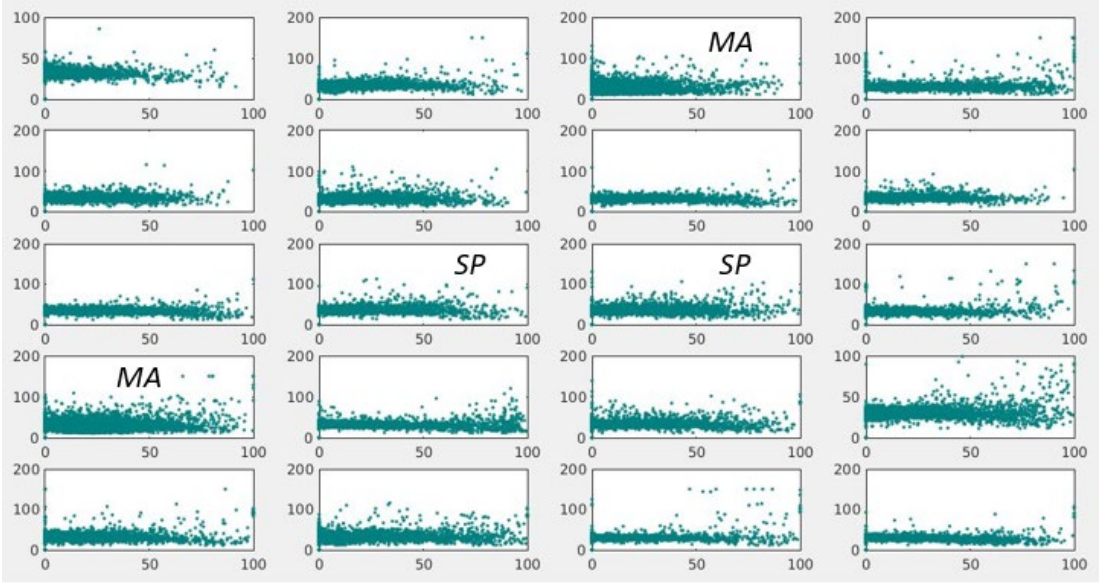
4.4.1 Data quality overview

At baseline there were no analysable data available for 3 participants: one ALS patient dropped out of the study, for one KD no thigh data were collected and for one more KD patient no compatible segmentation was available. The latter occurred because segmentations were performed on Dixon images and the patient in this case repositioned themselves during the scanning session. Data were collected at 6 months follow up only for the ALS group for 17 patients – 3 patients dropped out. At 1 year follow up 6 more ALS patients dropped out, 2 datasets were not analysable due to incompatible segmentation and one dataset lacked thigh images, leaving 8 ALS patient datasets at that stage. Also at the 1 year point, 1 volunteer and 1 KD patient also dropped out and 2 more KD datasets lacked thigh images, leaving 15 and 18 datasets respectively. Datasets were processed offline, using the sEPG-MLE algorithm (Chapter 3, Figure 4.2) and the quality control strategy described in section 4.3, with duration mostly between half and one and half hour per set of respective parameter maps. All results were computed on a 2.70-GHz Intel CPU and 8-GB RAM system.

To assess the overall data quality an overview of the parameter estimates at baseline by means of pixel-wise T_{2m} vs ff plots is shown in Figure 4.7 for all available segmented datasets and after the post-processing steps introduced so far.



T_{2m} (ms) vs ff (%) Healthy Volunteers



T_{2m} (ms) vs ff (%) ALS patients

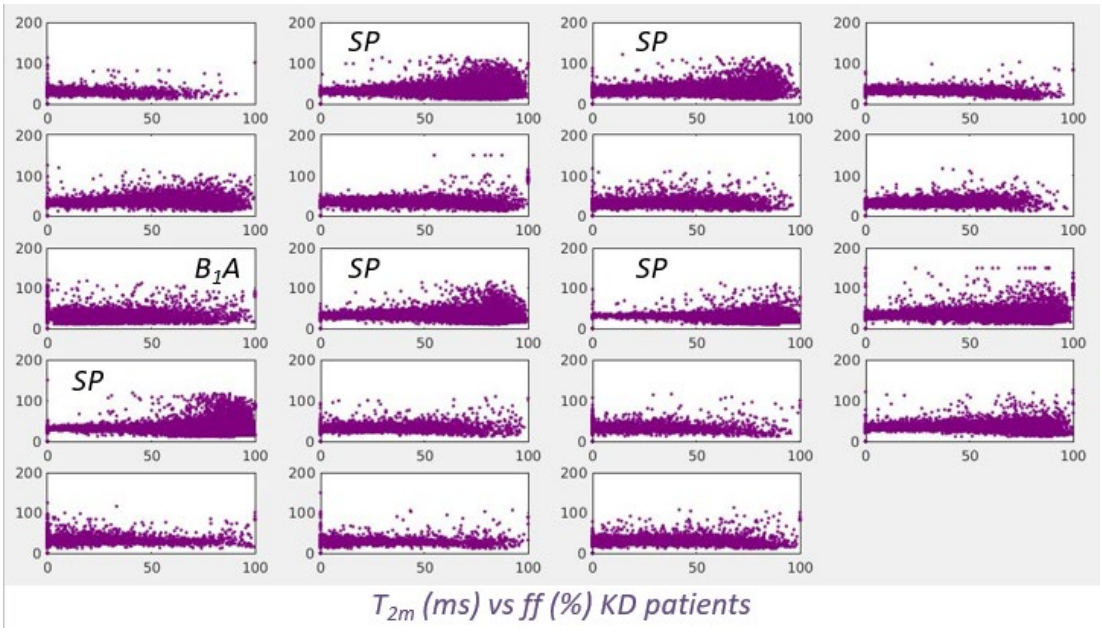


Figure 4.7: Overview of the estimated T_{2m} (0 – 150ms) vs *ff* (0 – 100%) for the thigh-level cross-sectional area maps for healthy volunteers (grey colour, top), ALS patients (green, middle) and KD patients (purple, bottom) at baseline. MA denotes motion artefacts, B_1A denotes compromised B_1 field / low SNR and SP denotes severe pathological involvement.

For healthy volunteers T_{2m} is clustered around 30ms, largely independent of *ff*, which itself was clustered at values mostly below 50%, with very few pixels having a *ff* exceeding that range. An exception was one data set (labelled "MA") which exhibited marked motion artefacts. For patients T_{2m} appears largely independent of *ff*, exhibiting larger dispersion for higher *ff* ranges associated with severe disease, and more for KD than for ALS patients. Mild dispersion was also seen for low to middle range *ff* in some cases with poor image quality. The latter cannot be determined by histogram evaluation alone, as can be seen in the examples in Figure 4.8, Figure 4.9, Figure 4.10 and Figure 4.11, where the T_{2m} and *ff* maps corresponding to the histograms are presented to allow visual assessment. Sub-optimal quality datasets are presented first, followed by good quality datasets.

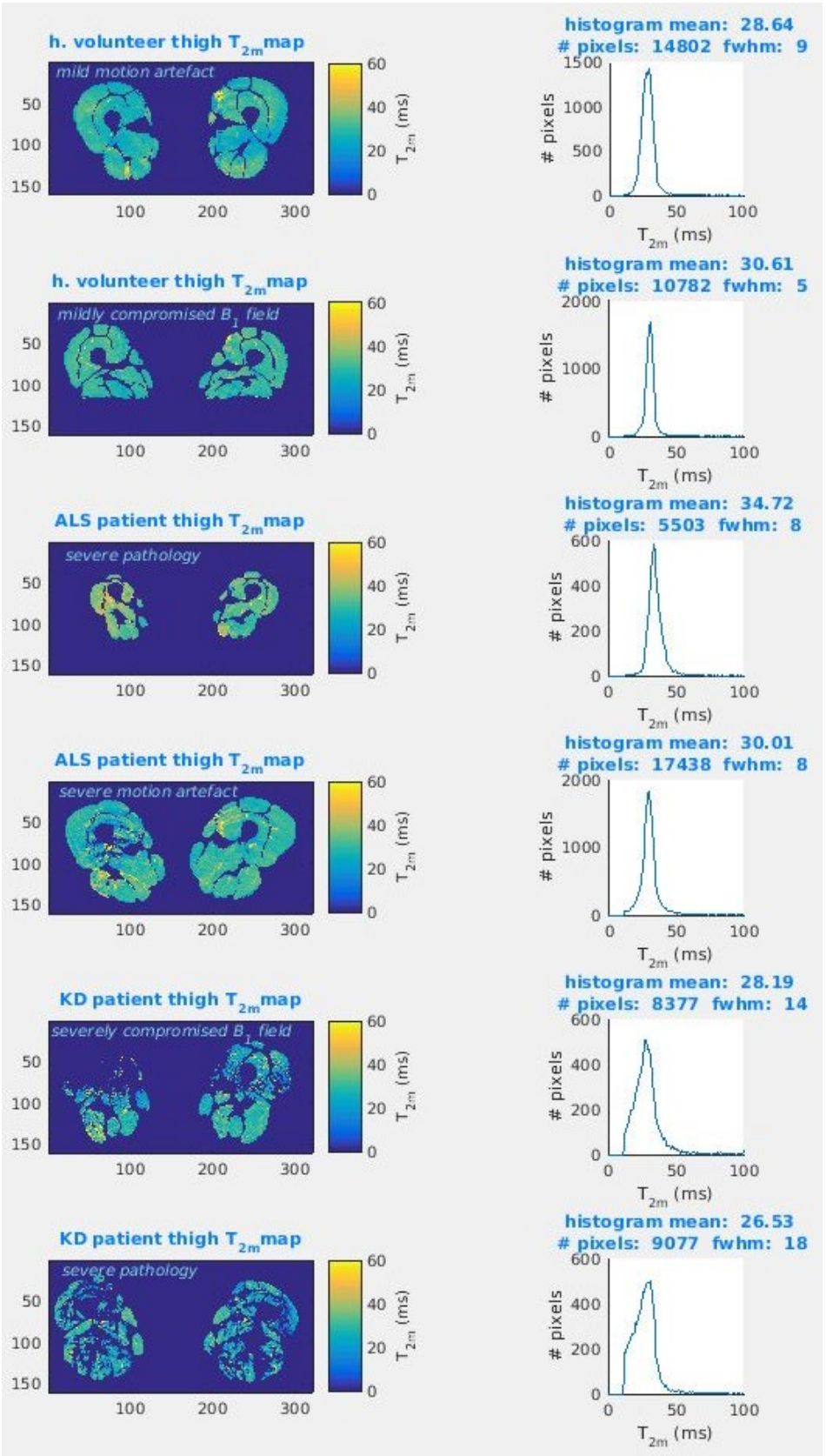


Figure 4.8: Examples of T_{2m} maps and histograms for representative (same as in Figure 4.1 through Figure 4.4) datasets with severely suboptimal data quality, and/or severe pathology

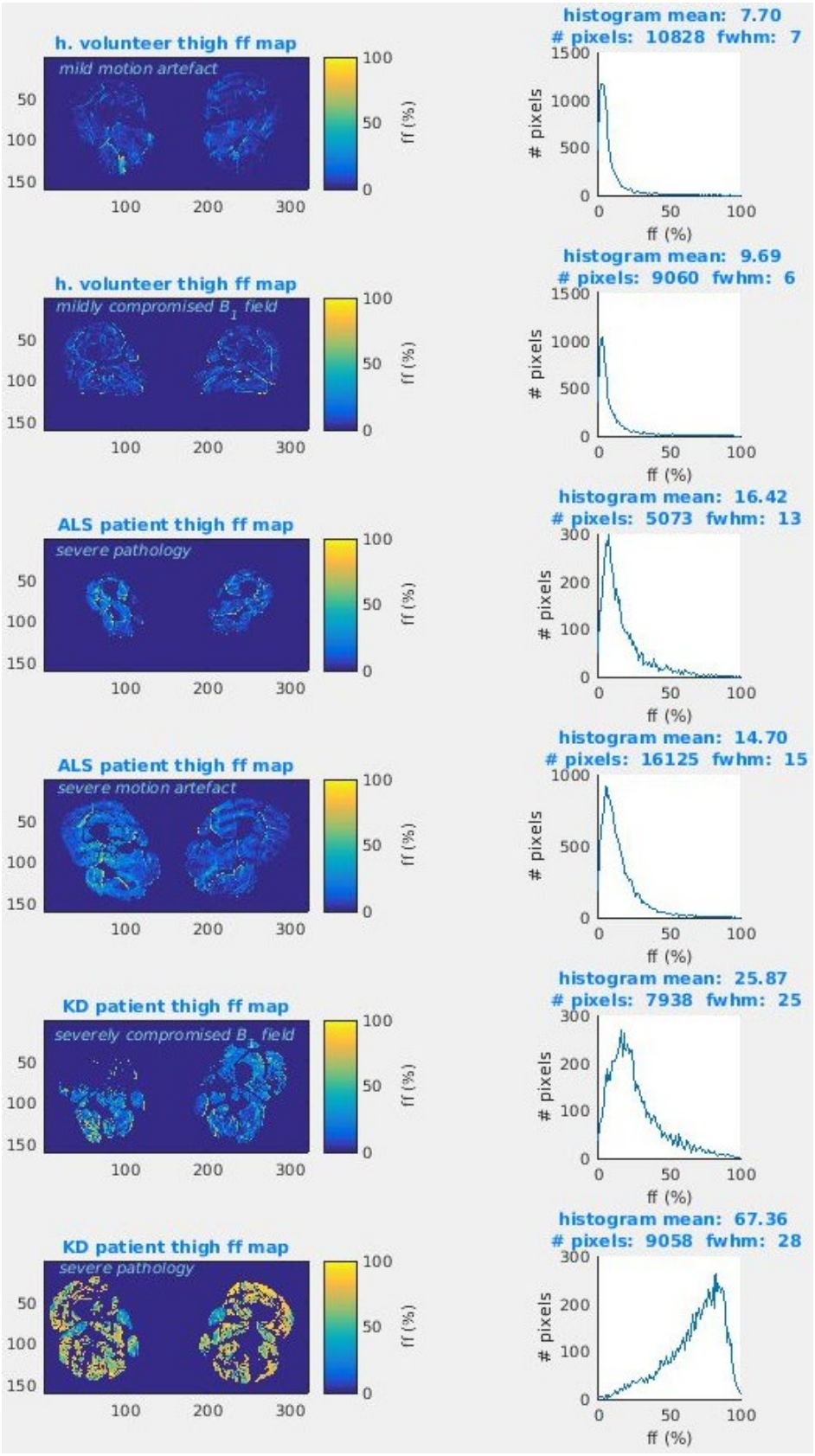


Figure 4.9 Examples of *ff* maps and histograms for representative (same as in Figure 4.1 through Figure 4.4) datasets with severely suboptimal data quality, and/or severe pathology

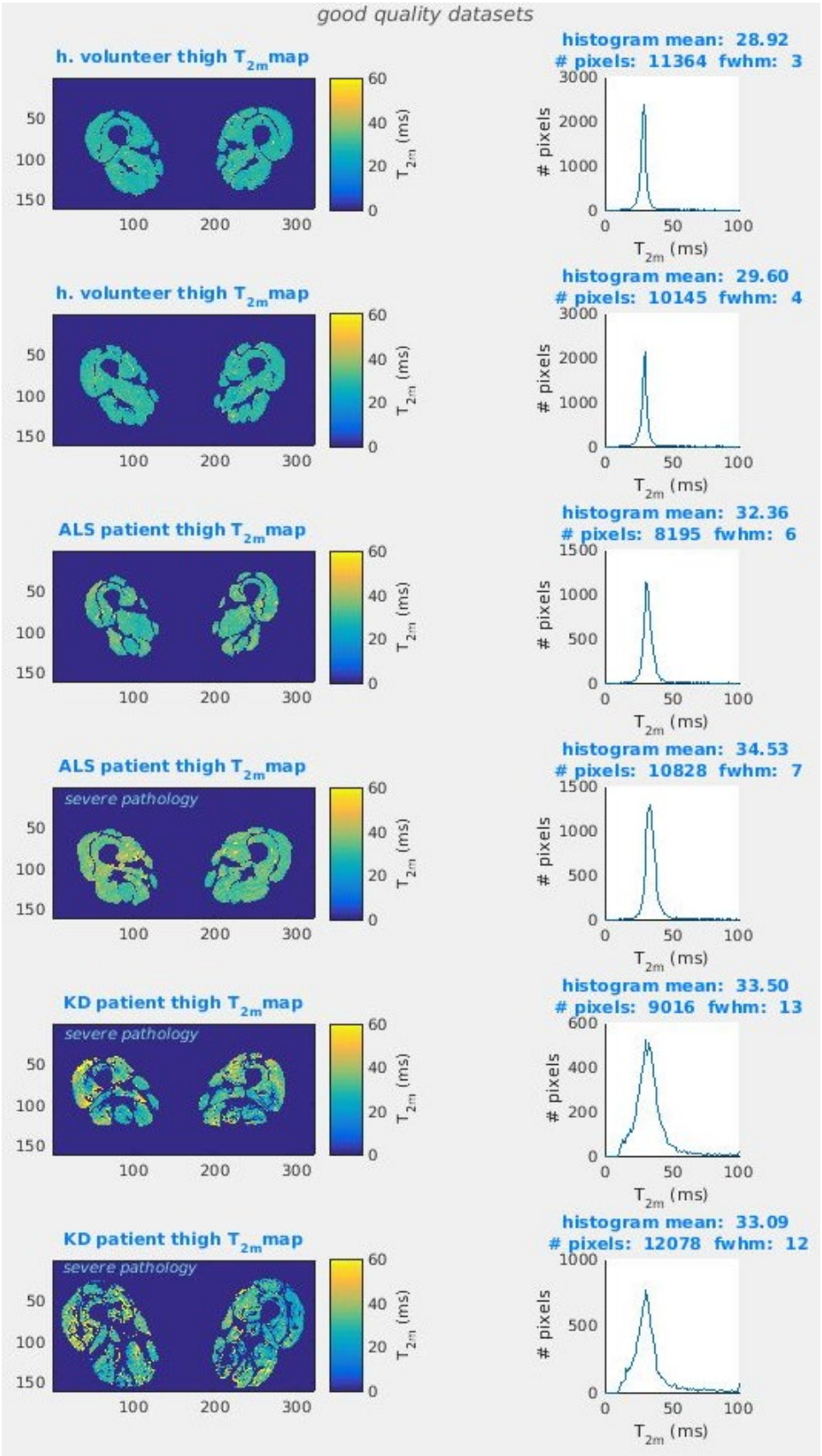


Figure 4.10: Examples of T_{2m} maps and histograms for representative good quality datasets including cases of severe pathology

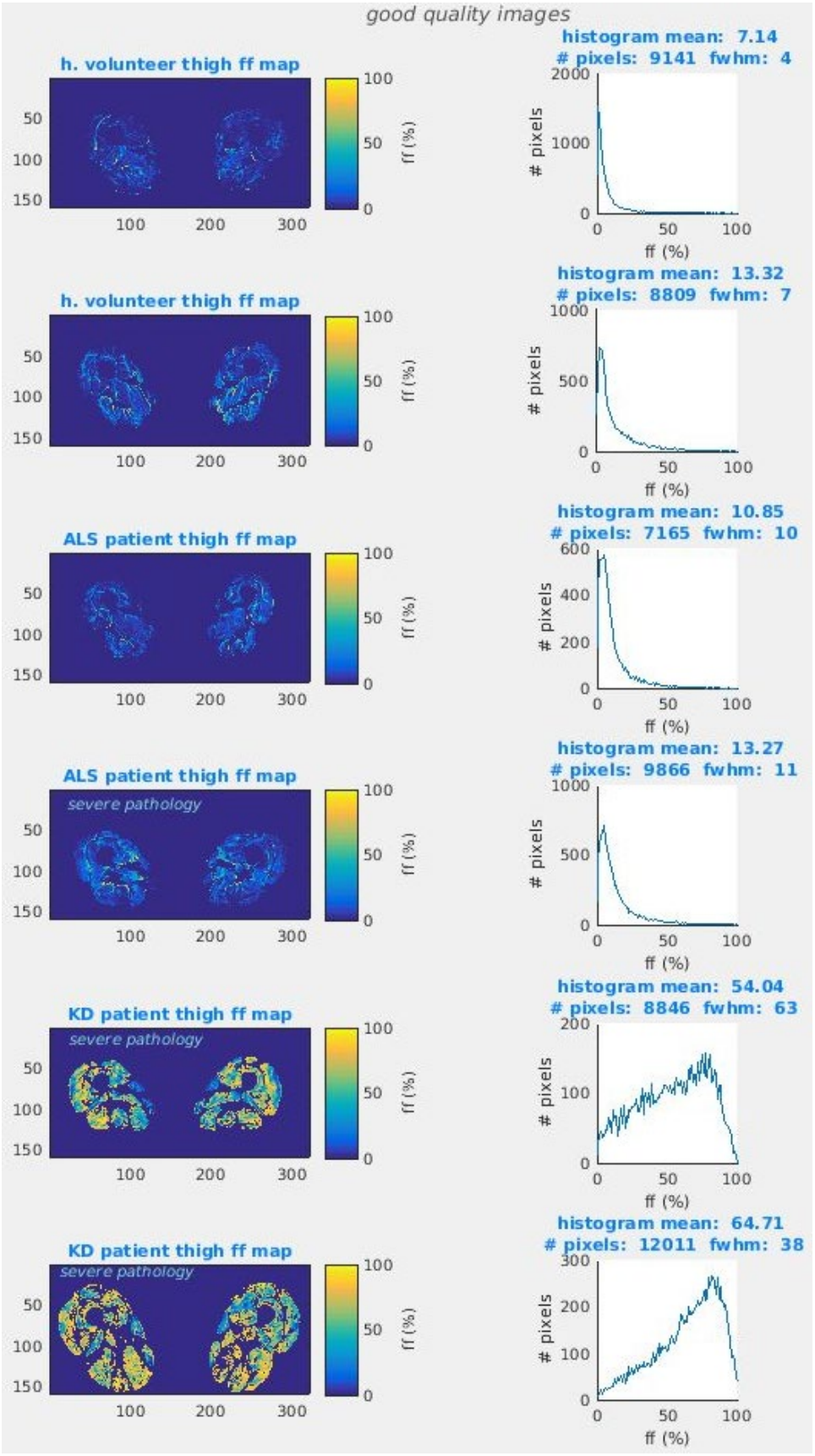


Figure 4.11: Examples of *ff* maps and histograms for representative good quality datasets including cases of severe pathology

This tends to validate the quality control strategy adopted in producing these results and shows also in more detail examples of the parameter estimates' behaviour described above.

Examples of parameter estimates per muscle ROI by means of boxplots are shown, in Figure 4.12 for T_{2m} and Figure 4.13 for *ff*, from the good quality datasets used also in Figure 4.10 and Figure 4.11, for the same 2 healthy volunteers, 2 ALS patients and 2 KD patients.

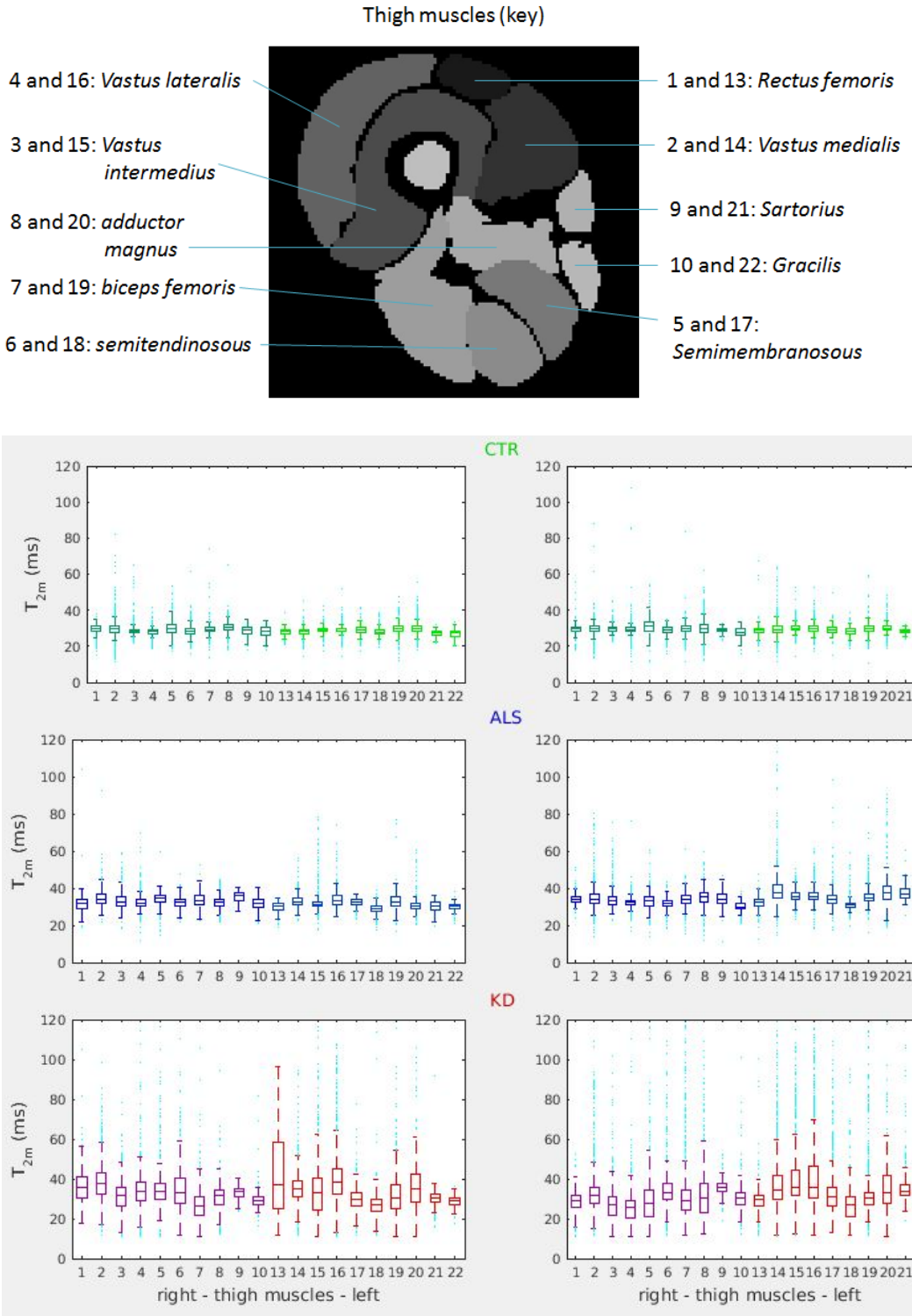


Figure 4.12: Example whisker-and-box plots of the estimated T_{2m} for the individual thigh-level muscle ROIs for 2 healthy volunteers (denoted CTR, green colour, top), 2 ALS patients (blue, middle) and 2 KD patients (purple/red, bottom) at baseline, from the good quality images shown in Figure 4.10 and Figure 4.11. 20 muscles ROIs per subject are shown, 10 for each limb (individual muscles are identified by number in the key above) excluding bone marrow. Note: the muscle ROI numbering used by the segmentation specialist (U.K., section 4.2), namely from 1 to 10 and from 13 to 22 for right and left limb respectively was also used here.

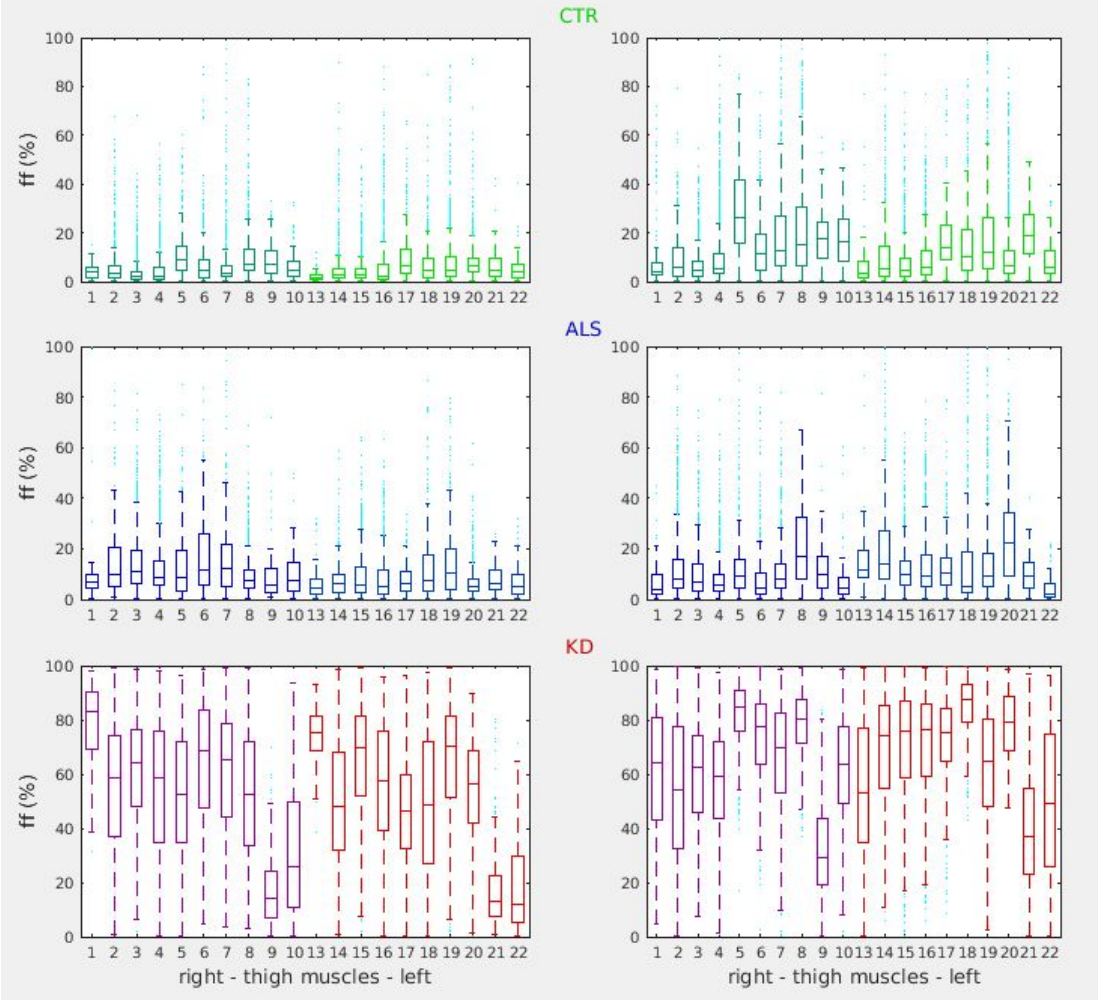
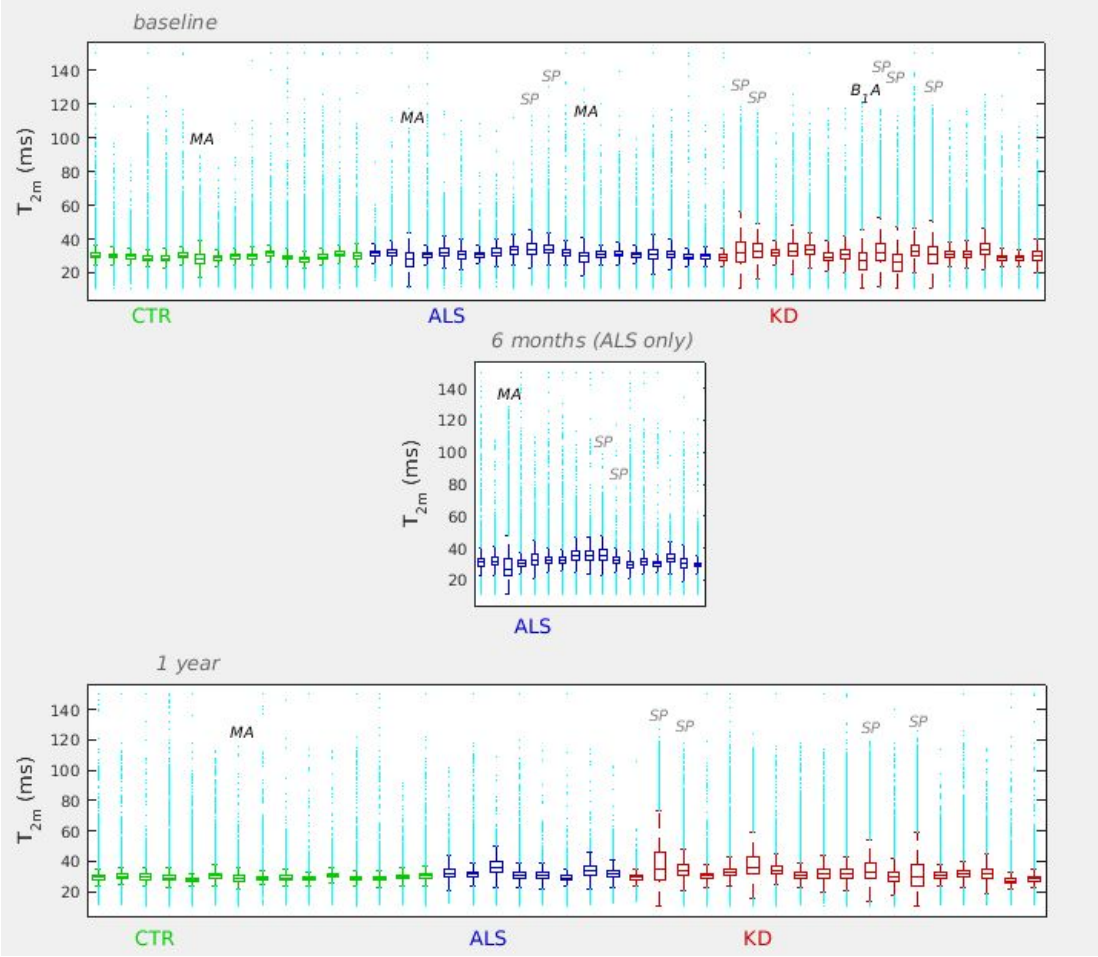


Figure 4.13: Examples whisker-and-box plots of the estimated *ff* for the individual muscle ROIs for 2 healthy volunteers (denoted CTR, green colour, top), 2 ALS patients (blue, middle) and 2 KD patients (purple/red, bottom) at baseline, from the good quality images shown in Figure 4.10 and Figure 4.11. 20 muscles ROIs per subject are shown, 10 for each limb (individual muscles are identified by number in the key above) excluding bone marrow.

The presence of muscle pathology appears to correlate in general with wider distributions (longer whisker-and-box plots), while T_{2m} distributions show departures from the normative values (around 30ms) both upwards and downwards and *ff* predominantly upwards (roughly above 10-15% *ff*). Individual muscle differences are also apparent, most prominently for the ROIs that were in some cases spared from severe disease, as for example the Sartorius and Gracilis inner thigh muscles, especially for severely affected KD subjects (bottom panels, last two muscles of each

limb); however no generalizable pattern was observed in disease severity differences between individual muscles in these example datasets.

Moving now to the patient and control dataset as a whole, Figure 4.14 shows an overview of the entire thigh cross-section T_{2m} and *ff* boxplot results, for all subjects (healthy volunteers, ALS patients and KD patients) longitudinally at baseline and at 6 months and 1 year follow-up. Each boxplot represents an individual subject, summarising the values from all pixels from all of the thigh cross-section muscles which, after the quality control steps, contributed to the final T_{2m} maps (top panel) and *ff* maps (bottom panel); outlier pixels are represented by cyan dots.



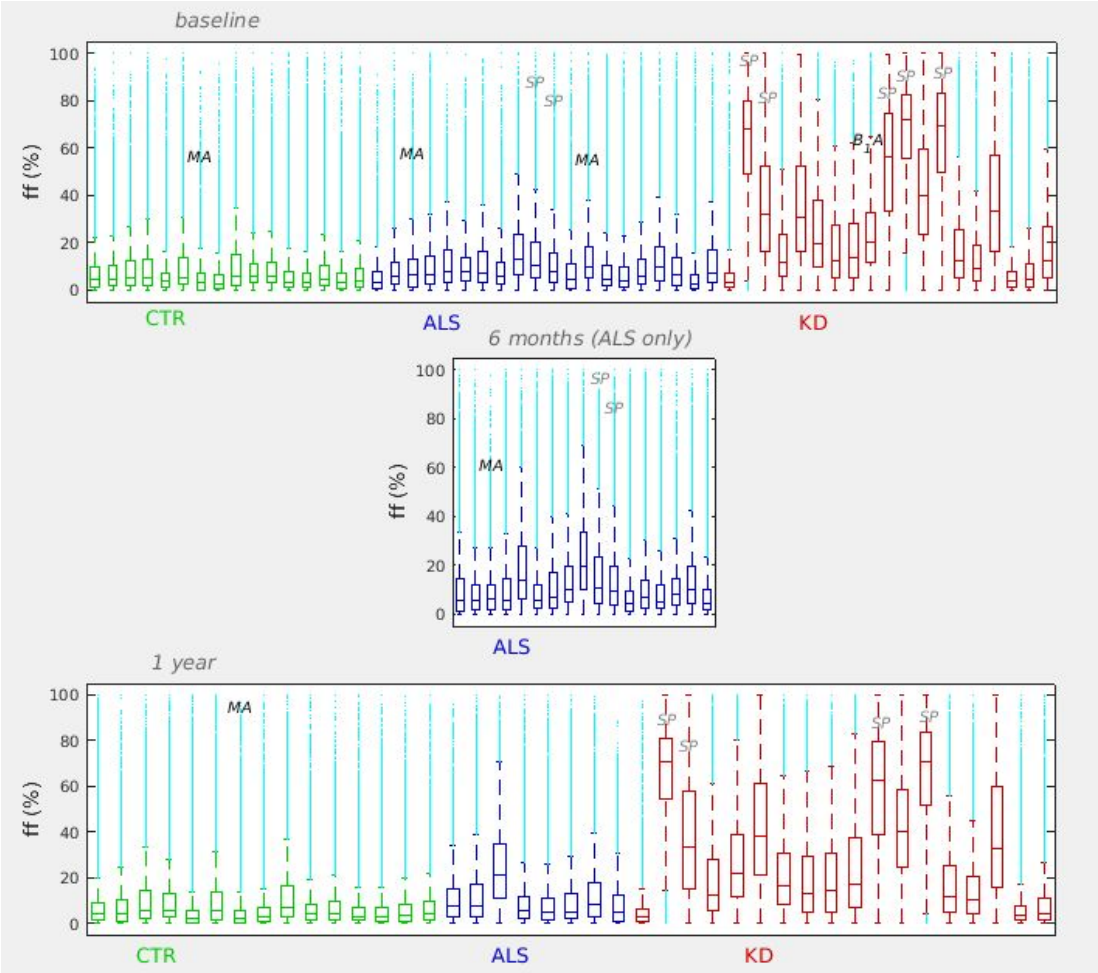


Figure 4.14: Overview of the estimated T_{2m} (0 – 150ms) and *ff* (0 – 100%) whisker-and-boxplots from the entire thigh-level cross-section muscles for healthy volunteers (CTR, green, n=16), ALS patients (blue, n=20) and KD patients (red, n=19) at baseline, 6 months and 1 year follow-up; outlier values (pixels) are represented by the cyan dots. MA denotes presence of motion artefacts, B_1A denotes compromised B_1 field / low SNR and SP denotes the presence of severe pathology. These results are overall consistent with those seen in the majority of muscles for each example subject shown in Figure 4.12 and Figure 4.13.

The trends in the parameter estimates followed by the majority of muscle ROIs for the representative subjects in Figure 4.12 and Figure 4.13, are in general also seen in these results, despite the variations observed between specific individual muscles. Wider T_{2m} distributions (longer whisker-and-box plots) and higher *ff* – especially in KD – are the predominant observations for patient data. Presence of artefacts (boxplots denoted as *MA* and B_1A) seems to correlate with parameter underestimation, and more prevalently for T_{2m} than for *ff*. The presence of pathology

seems to translate to broadening parameter distributions, predominantly upwards, and more so for *ff* than for T_{2m}.

4.4.2 Quantitative results – Histogram metrics

The results presented via Figure 4.6 and Figure 4.7 show that the T_{2m} and *ff* distribution plots can help distinguish between compromised data (and consequent suboptimal parameter estimation) and true biological changes, and this can be a useful tool to identify and discard inappropriate or confounding datasets. There were 6 severely compromised datasets, mainly due to motion artefacts (from 1 healthy volunteer at baseline and at 1 year follow up, and 2 ALS patients at baseline, and for one of these also at 6 months follow up) and 1 due to non-uniform B_1 field and locally low SNR (from a KD patient at baseline). These are denoted also in Figure 4.14. Other imperfections due to flow or edge (chemical shift, partial volume) artefacts or segmentation were also often present but usually were restricted to much smaller regions (strips under veins/arteries, ROI edges etc.). The above 6 datasets are excluded from the following group analysis. In Figure 4.15 the previous T_{2m} median values only (without whisker and box plots) are shown per muscle across all remaining subject datasets – including follow up scans, and in Figure 4.16 the respective *ff* median values.

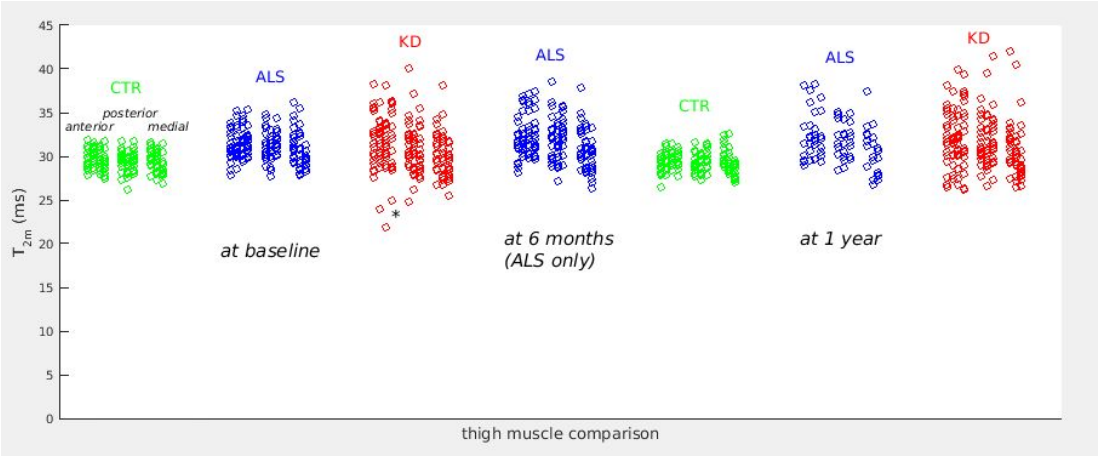


Figure 4.15: Left & right combined individual thigh muscle median T_{2m} across the 3 experimental groups. Note: In some highly fat infiltrated muscles, e.g. those indicated (*), where the muscle water signal was low, although the MLE parameter estimation performed well, the physical interpretation of the model parameters in terms of fat and water compartments appeared more complex than in the majority of the less affected muscles.

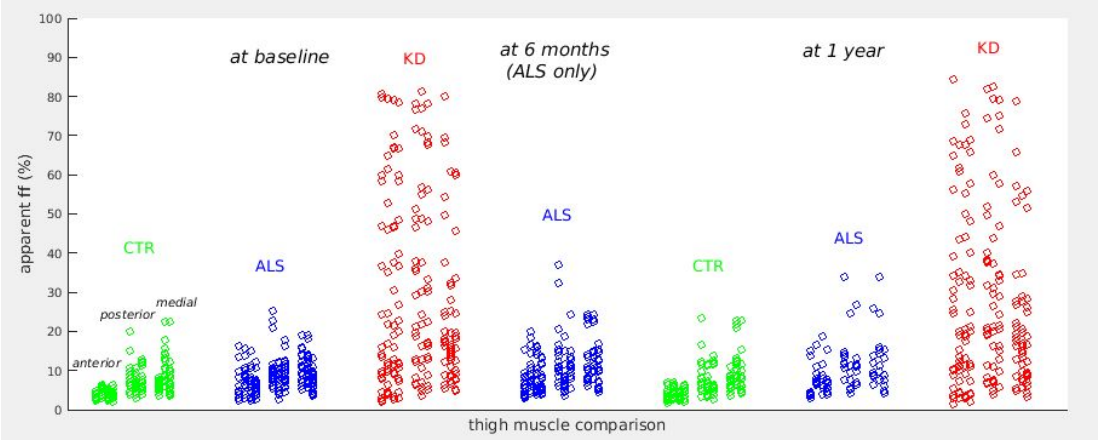


Figure 4.16: Left & right combined individual thigh muscle median apparent *ff* across the 3 experimental groups.

Visual inspection of these results also suggests that neuromuscular disease may cause both increase and decrease in the parameter estimates – compared with healthy controls – predominantly in T_{2m} and potentially also *ff*, while also affecting their distributions. Short T_{2m} appeared to be associated with high *ff* in many cases, which may pose challenges in the physical interpretation of the model parameters in terms of fat and water compartments (section 4.6, Chapter 7). It is therefore

instructive to examine and compare histogram metrics. In Figure 4.17 an illustrative example is given for a dataset with potentially the most severe pathological manifestations found in the entire study cohort.

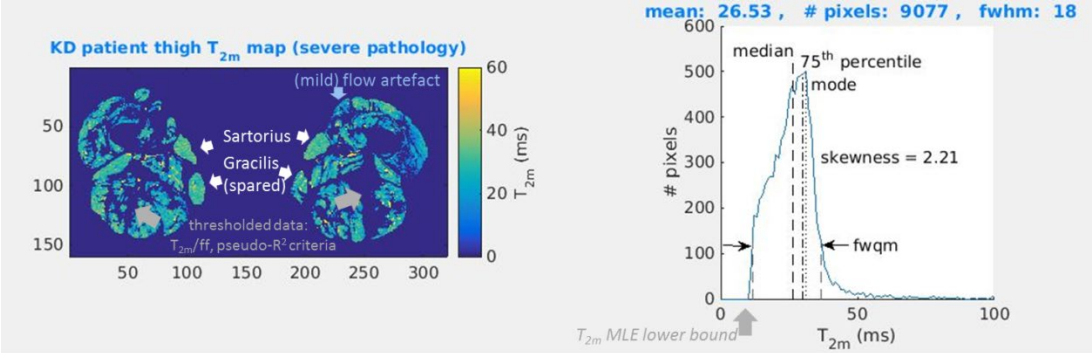


Figure 4.17: Example sEPG-MLE T_{2m} map from a KD patient with severe disease. Left: the Sartorius, Gracilis and other muscles return normative range T_{2m} (spared) while in several other regions many pixels did not survive the post-processing thresholds (dark grey arrows), in this case mainly due to severe pathology (high fat content) but potentially also due to mild artefacts, e.g. the flow artefact column of data in the phase direction (blue arrow). Right: the corresponding T_{2m} histogram, constraint by the MLE algorithm T_{2m} lower bound (10ms) and threshold (11ms) described in the text (dark grey arrow), and the related metrics: mean, mode, median, full width at quarter maximum (fwqm), 75th percentile and skewness of the distribution.

Some of the features mentioned previously are seen, such as the consistently spared muscle ROIs, artefacts, data post-processing and bound constraints effects, and conventional histogram metrics providing candidate outcome measures are illustrated in the right hand panel. The outlying T_{2m} results denoted with an asterisk in Figure 4.14 were from this same patient; due to the complications in the physical interpretation of these particular results this dataset was also excluded from the following quantitative analysis. In Figure 4.18 extracted T_{2m} histogram metrics are shown for all participants and scanning time points of the study – except the aforementioned datasets excluded according to quality of fit criteria and physical meaningfulness. Presented are the median, full width at quarter maximum (fwqm), 75th percentile and skewness from the estimated T_{2m} distributions, from all the

muscle ROIs (thigh cross-section) combined. In Figure 4.19 the corresponding *ff* histogram metrics are shown. In terms of T_{2m} ALS patients results appear more elevated *vs.* controls, and with consistently higher (medians, 75th percentiles) compared to their KD counterparts, and in terms of *ff* KD patient results appear to be higher overall and longitudinally more stable than those for ALS. At the same time distributions for both parameters are consistently broader and more varied (fwqm, skewness) for KD compared to ALS patients. These trends and the results overall are further discussed in section 4.6.

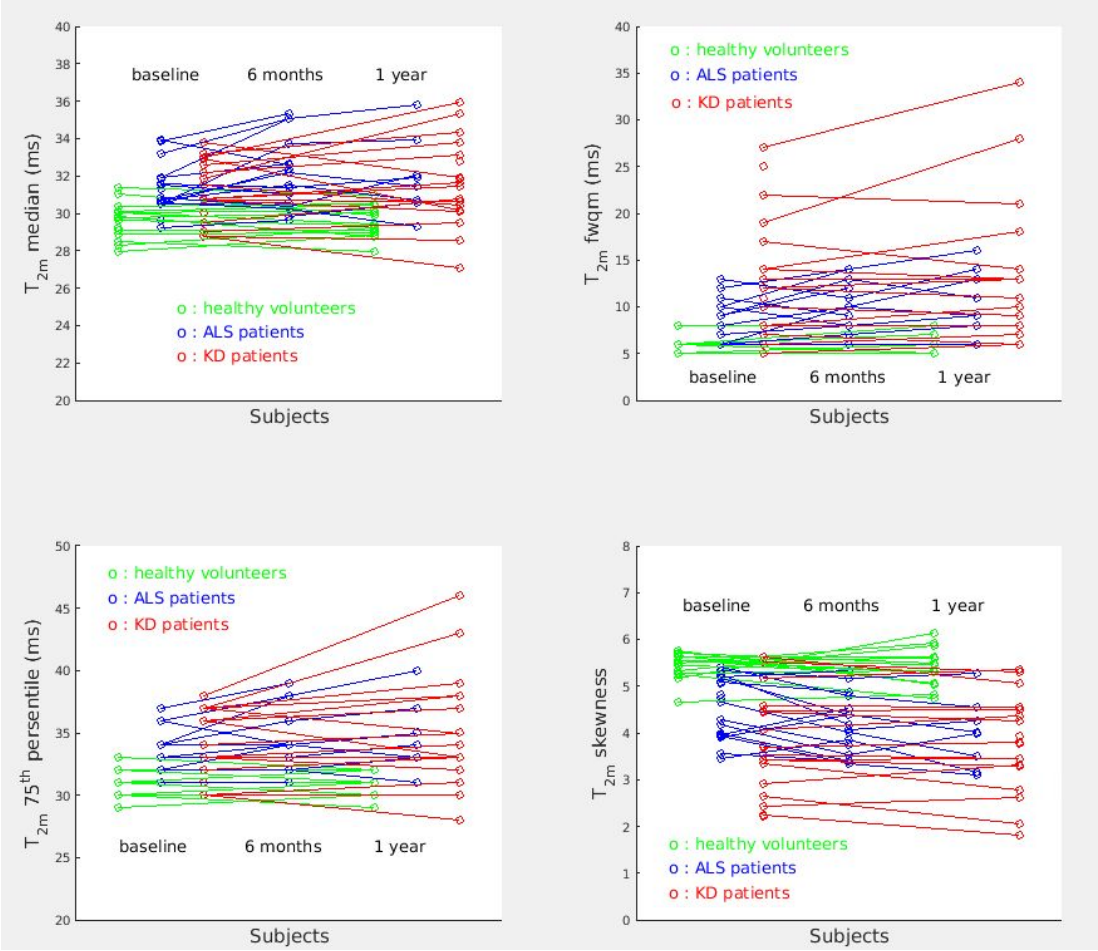


Figure 4.18: Histogram metrics comparison between healthy volunteers (in green), ALS patients (blue) and KD patients (red), at baseline, and at 6 months and 1 year follow-up scans, for median (top left), full width at quarter maximum (top right), 75th percentile (bottom left) and skewness (bottom right) of the T_{2m} estimates distributions.

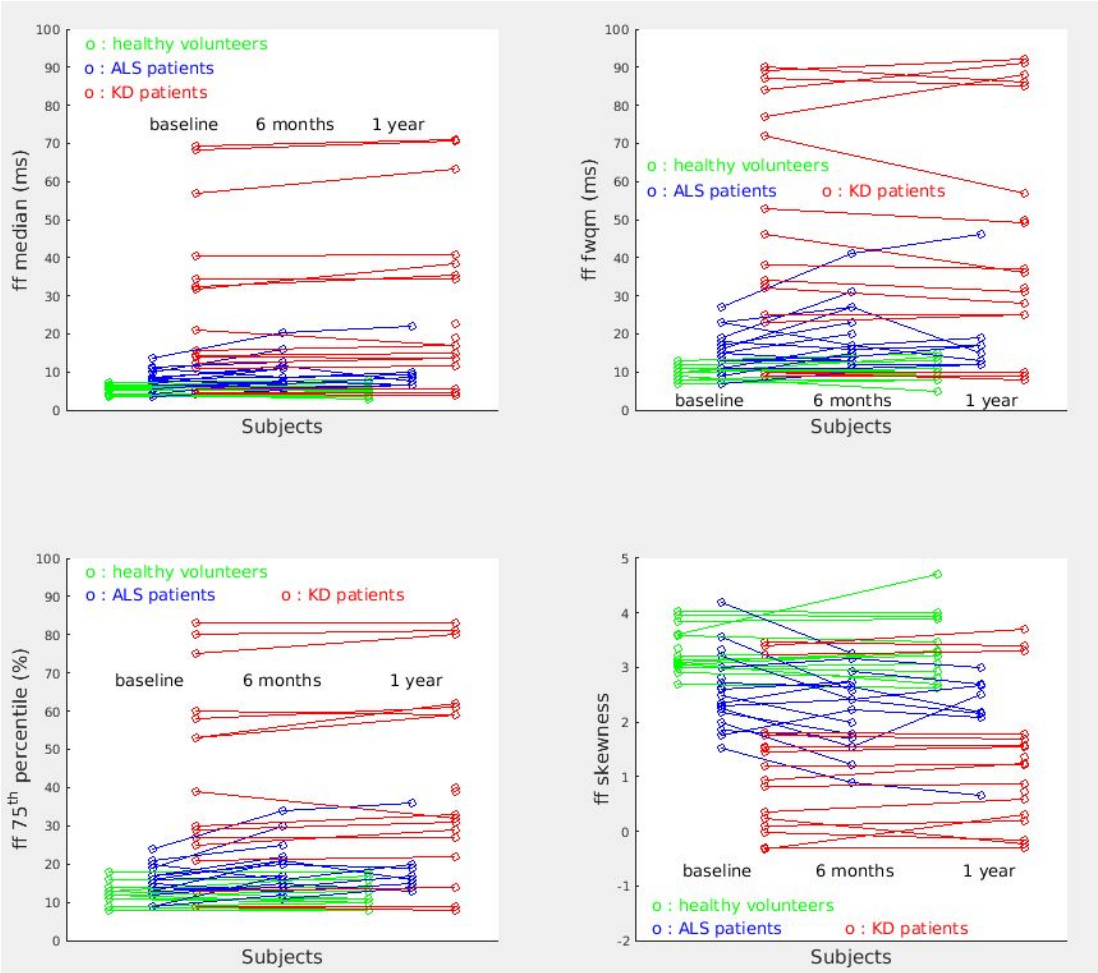


Figure 4.19: Histogram metrics comparison between healthy volunteers (in green), ALS patients (blue) and KD patients (red), at baseline, and at 6 months and 1 year follow-up scans, for median (top left), full width at quarter maximum (top right), 75th percentile (bottom left) and skewness (bottom right) of the *ff* estimates distributions.

Within this quantitative framework it might be worth noting that noise σ_N estimation across all subjects and time points converged to a mean value of 4.36±0.97a.u., calculated from the extracted medians from each thigh cross sectional sEPG-MLE estimated noise σ_N map. This shows remarkable stability across subjects and is consistent with the 3 to 5a.u. noise σ_N levels measured in a representative data set using the signal-background method in section 3.1. No marked differences were found in terms of this estimated noise σ_N between the three study groups (ALS and KD patients and healthy controls) or longitudinally. The compromised quality images

presented in Figure 4.8 and Figure 4.9 returned elevated noise level estimates (at 6.12a.u. on average), suggesting a correlation of increased (true or estimated) noise with artefactual data; the good quality images of Figure 4.10 and Figure 4.11 returned an average of 4.55a.u., close to the overall mean as expected.

To summarise, in both the ALS and KD patients there was evidence of varied patterns of T_{2m} changes at the level of individual muscles, predominantly elevation, and CTR median T_{2m} distributions were consistent between muscles. Whole thigh-level muscle group mean ± SD T_{2m} was 29.6 ± 1.0ms in controls, 31.3 ± 1.4ms in ALS, and 31.2 ± 1.6ms in KD (p=0.0004 and p=0.0007 *vs.* controls respectively) at baseline. Median *ff* across all thigh–level muscles was increased in ALS and more markedly in KD *vs.* controls. Statistical ANOVA comparisons are presented in the next section which considers also in comparison Dixon method results.

4.4.3 Comparisons with existing methods

The results obtained using the proposed sEPG-MLE algorithm are at this stage compared with more commonly used methods. The Dixon method is considered the *de facto* gold-standard MRI modality for fat quantification in muscle (Wary et al., 2015, Marty et al., 2016, McRobbie et al., 2006), while exponential model based LSQ fitting methods EXP-LSQ) have been widely implemented for single- or multi-component T_2 estimation (sections 2.2.2 and 3.2.2).

Dixon method

The equivalent of Figure 4.19 where the sEPG-MLE *ff* (henceforth ff_a, 'apparent' fat fraction) histogram metrics behaviour was illustrated, is shown in Figure 4.20 for the Dixon *ff* (ff_d) histogram metrics, for all participants and study time points. In the available segmented ff_d maps (section 4.2) spurious (pixel ff_d values) outside the expected (0-100%) range (i.e. not physically meaningful) were occasionally present, often reaching extreme values: these were excluded.

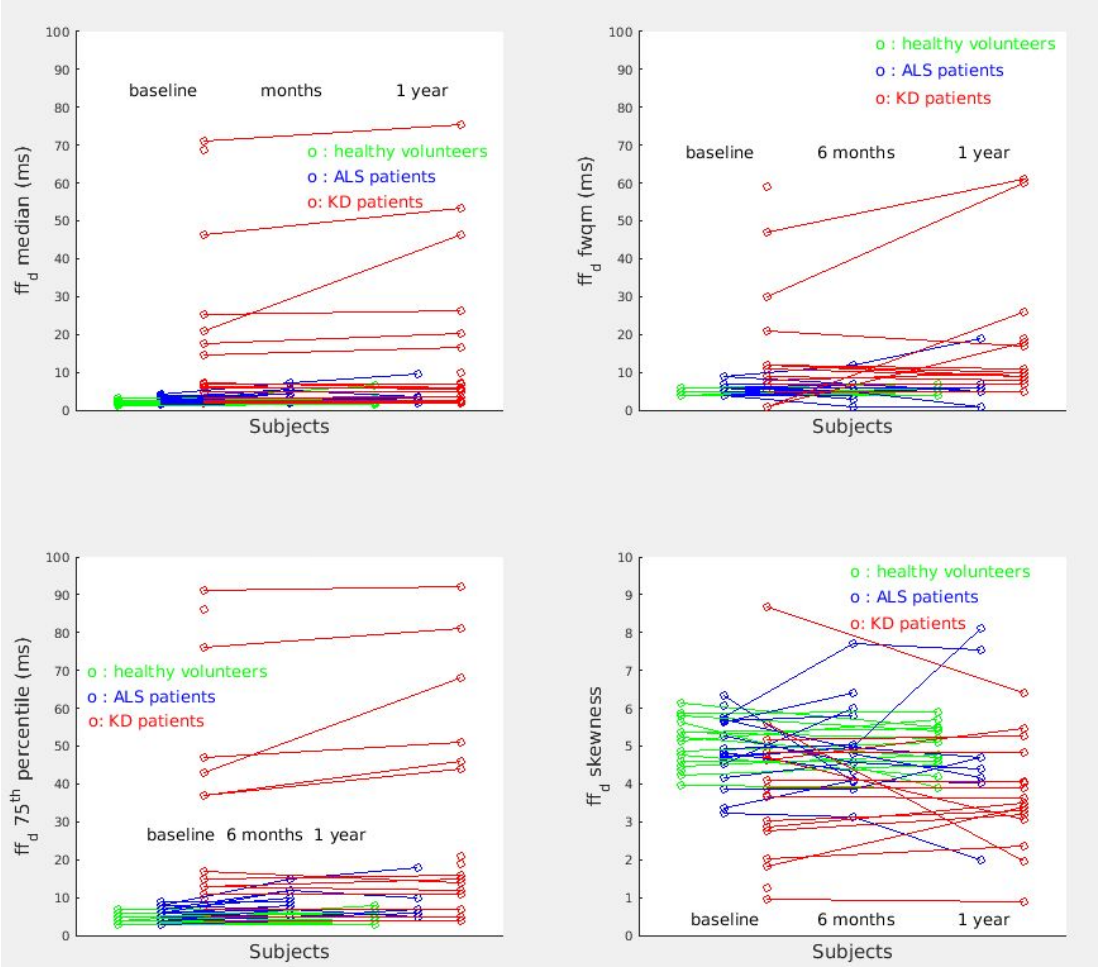


Figure 4.20: Histogram metrics for healthy volunteers (in green), ALS patients (blue) and KD patients (red), at baseline, and at 6 months and 1 year follow-up scans, for median (top left), full width at quarter maximum (top right), 75th percentile (bottom left) and skewness (bottom right) of the ff_d (Dixon *ff*) estimates distributions.

While both ff_d and ff_a medians are in the range from 0 to 75%, ff_d medians show the tendency to cluster more towards lower values, and more so for patients. This may explain the markedly lower ff_d than ff_a fwqm, and similar but milder differences in the 75th percentiles and skewness.

Initial one way ANOVA revealed significant differences in the sEPG-MLE estimated *ff* between male and female participants – with *ff* for females elevated – for both healthy controls (p=0.0069) and ALS patients (p=0.0021) at baseline (KD patients are only male). Similar trends were seen also for the follow up data, and also for T_{2m} which appeared visually about 0.5 to 0.8ms elevated for female participants compared to the males – without however reaching statistical significance. For the above reasons all subsequent ANOVA tests were conducted including also gender as a covariate.

In Table 4.2 one way ANOVA comparisons between the three groups (healthy volunteers, ALS patients and KD patients) per gender are shown, for the aforementioned T_{2m}, ff_a and ff_d histogram metrics at baseline. Overall differences between healthy controls (CTR) and patients (both ALS and KD) are observed with respect to T_{2m}. However group size appears to also affect this, with the p-value for the female CTR vs ALS group only marginally exceeding the 0.05 significance threshold, whereas the respective p-values for the male and the entire (both genres combined) groups (in lighter font) were much smaller. This is likely related to female groups being much smaller than the male groups (roughly half in size) for both the CTR and ALS cohorts. All three groups appear statistically separated from each other with respect to ff_a. For both T_{2m} and ff_a more clear differences were revealed for

fwqm and skewness than for medians and 75th percentiles. Conversely median and 75th percentiles were clearly statistically separable for the ff_d results between the three groups, but not its fwqm and skewness – except between KD and both the CTR and ALS groups for the latter.

Table 4.2: One way ANOVA comparisons between healthy volunteers, ALS patients and KD patients), for the T_{2m}, ff_a and ff_d histogram metrics per gender at baseline; p-values for statistically significant differences ($p<0.05$) are shown in blue colour, otherwise in red. The middle column of results in lighter colours is for the merged (mixed) male and female CTR and ALS groups comparison.

Histogram metrics ANOVA (p-values)		***CTR^{male} vs ALS^{male}***	***CTR^{female} vs ALS^{female}***	***CTR vs ALS (mixed)***	***CTR^{male} vs KD$^{(male)}$***	***ALS^{male} vs KD$^{(male)}$***
T_{2m}	median	0.0011	0.0538	0.0002	0.0034	0.8743
	fwqm	0.0153	0.0062	0.0004	0.0021	0.0097
	75th percentile	0.0007	0.0166	0 ($<10^{-4}$)	0.0009	0.1335
	skewness	0.0066	0.0004	0 ($<10^{-4}$)	0.0001	0.0054
ff_a	median	0.0114	0.0083	0.0023	0.0039	0.0072
	fwqm	0.0048	0.0066	0.0008	0.0004	0.0004
	75th percentile	0.0205	0.0153	0.0043	0.0007	0.0007
	skewness	0.0126	0.0016	0.0013	0 ($<10^{-4}$)	0.0002
ff_d	median	0.0180	0.0115	0.0013	0.0291	0.0228
	fwqm	0.0249	0.0079	0.0014	0.0131	0.0113
	75th percentile	0.0281	0.0491	0.0041	0.0081	0.0068
	skewness	0.0168	0.0049	0.0009	0 ($<10^{-4}$)	0.0004

In Table 4.3 one way ANOVA longitudinal comparisons for the T_{2m}, ff_a and ff_d histogram metrics are shown for each of the three groups (healthy volunteers, ALS patients and KD patients) per gender. For female ALS patients only, T_{2m} appeared to change (increase) significantly with time – from the baseline to the 6 months and 1 year follow up examinations. This can be seen as a further justification of the previously established separation by gender. As expected the CTR group results were

highly stable with time (between baseline and the 1 year follow up scan for this cohort), with p-values approaching unity in many cases.

Table 4.3: One way ANOVA comparisons for each group (CTR, ALS patients and KD patients) for the T_{2m}, ff_a and ff_d histogram metrics longitudinally and per gender; p-values for statistically significant differences (p<0.05) are shown in blue colour, otherwise in red.

Histogram metrics ANOVA (p-values)		***CTR^{male} baseline–1year***	***CTR^{female} baseline–1year***	***ALS^{male} baseline-6m–1y***	***ALS^{female} baseline-6m–1y***	***KD$^{(male)}$ baseline-1year***
T_{2m}	median	0.9635	0.6172	0.9808	0.0192	0.4982
	fwqm	0.4648	1	0.2196	0.0182	0.7123
	75th percentile	0.7902	0.7245	0.9217	0.0208	0.4112
	skewness	0.9368	0.6788	0.1301	0.0535	0.8309
ff_a	median	0.7584	0.833	0.3419	0.2294	0.8766
	fwqm	0.9232	0.7245	0.2861	0.2411	0.9093
	75th percentile	0.8401	0.9208	0.1791	0.1922	0.8137
	skewness	0.593	0.9779	0.2905	0.2838	0.9607
ff_d	median	0.8136	0.4299	0.4781	0.1414	0.8975
	fwqm	0.9659	0.7328	0.3722	0.1272	0.9952
	75th percentile	0.9521	0.6968	0.5445	0.1279	0.9718
	skewness	0.9986	0.9253	0.3424	0.2379	0.8777

An important observation from the above ANOVA investigations is that p-values for ff_a comparisons are markedly lower than for ff_d comparisons, and especially for the statistically significant differences between the 3 main groups (CTR, ALS and KD) presented in Table 4.1, which may be related to the previous observation of the apparently more narrowly distributed ff_d values compared to ff_a in Figure 4.20 *vs.* Figure 4.19, i.e. the histogram metrics for ff_d being overall more narrowly distribured than those for ff_a. This may support the validity of ff_a as a reliable outcome measure in such studies.

Both the ff_a (Figure 4.19) and ff_d (Figure 4.20) results appear to be somehow stratified in two subgroups, of low and high *ff*, above and below 25% and 10% respectively. This distinction together with the corresponding T_{2m} is illustrated in Figure 4.21.

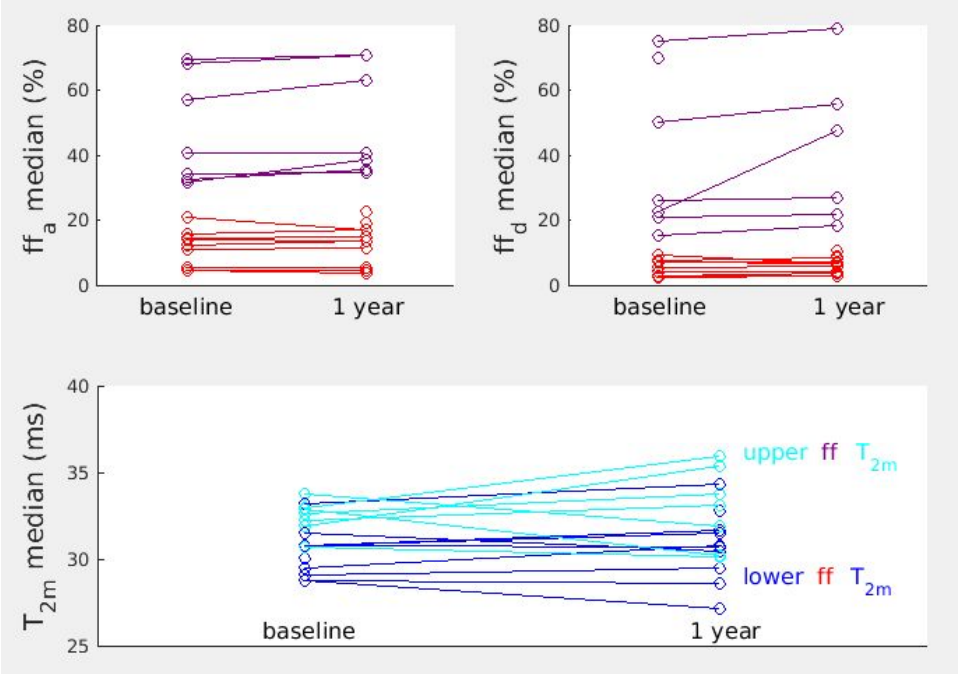


Figure 4.21: Division of ff_a and ff_d in two subgroups of low (in red) and high (in purple) *ff* respectively, and the corresponding T_{2m} (in blue and cyan respectively).

One way ANOVA revealed that T_{2m} of the high *ff* group is also significantly higher than the T_{2m} of the low *ff* group (p=0.0013) suggesting that elevation of T_{2m} correlated with elevated *ff* is a probable manifestation of disease severity.

A comparison of ff_a vs ff_d is shown in Figure 4.22. Each data point (circle) represents an individual subject, by means of the median from the pixels from collectively all the thigh cross-section muscles (after pixel-level quality control) participating in the corresponding ff_d map *vs.* the median from the respective ff_a map.

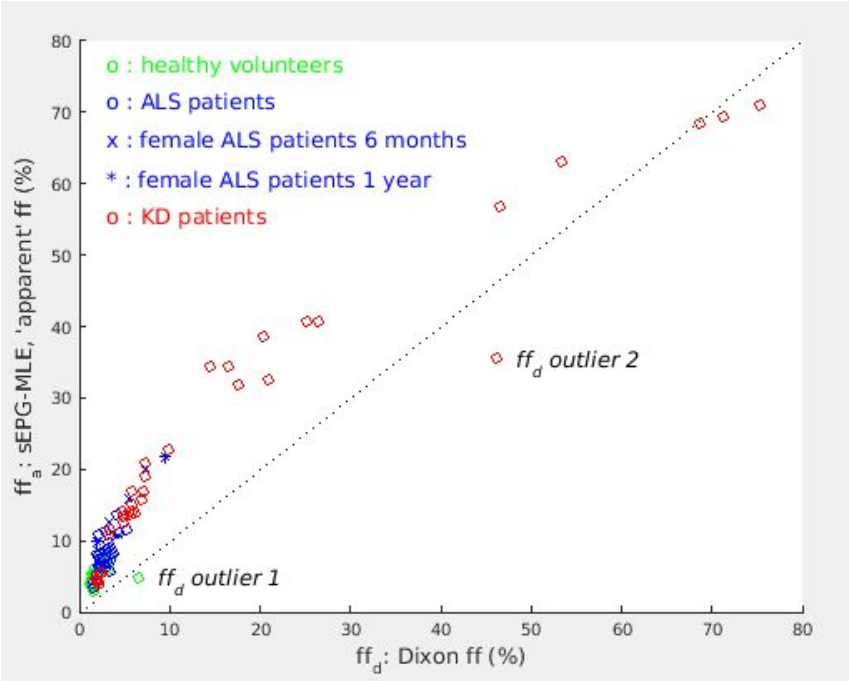


Figure 4.22: Plot of ff_a (%) *vs.* ff_d (%). Each data point represents the thigh cross-section median ff_a value *vs.* the corresponding ff_d value, for all subjects and time points. There appear to be two ff_d outliers, likely due to segmentation and phase unwrapping artefacts.

Similarly, the sEPG-MLE T_{2m} median relationship with the ff_d median is shown in Figure 4.23.

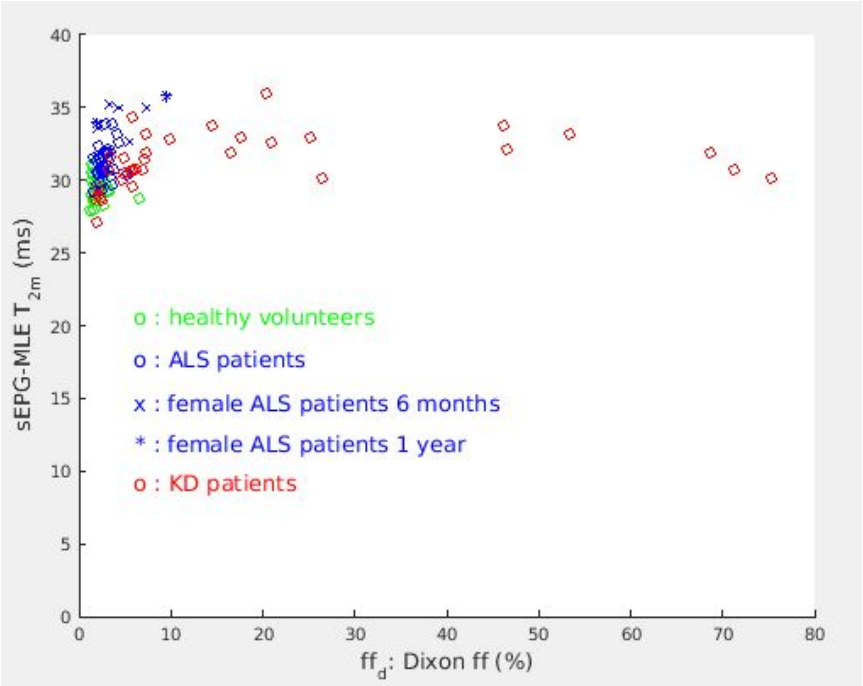


Figure 4.23: Plot of sEPG-MLE T_{2m} (ms) *vs.* ff_d (%). Each data point represents the thigh cross-section median T_{2m} value *vs.* the corresponding ff_d value, for all subjects and time points.

T_{2m} appears largely independent from ff_d, however ff_a markedly deviates from ff_d, exhibiting marked overestimation at the lower end of the range, an effect which gets gradually mitigated with increasing *ff* and finally reversed at the upper end of values. A small number of outlier data points, which appeared on visual assessment to be in an inconsistent position with respect to the line of unity compared with other nearby points are visible in Figure 4.22. Closer inspection of the source data suggested that they are likely to be due to segmentation imperfections (outlier 1) and phase unwrapping inadequacy within the Dixon method algorithm (outlier 2) contaminating the resulting ff_d distributions. This is demonstrated in Figure 4.24, where these potential artefacts linked to them are shown.

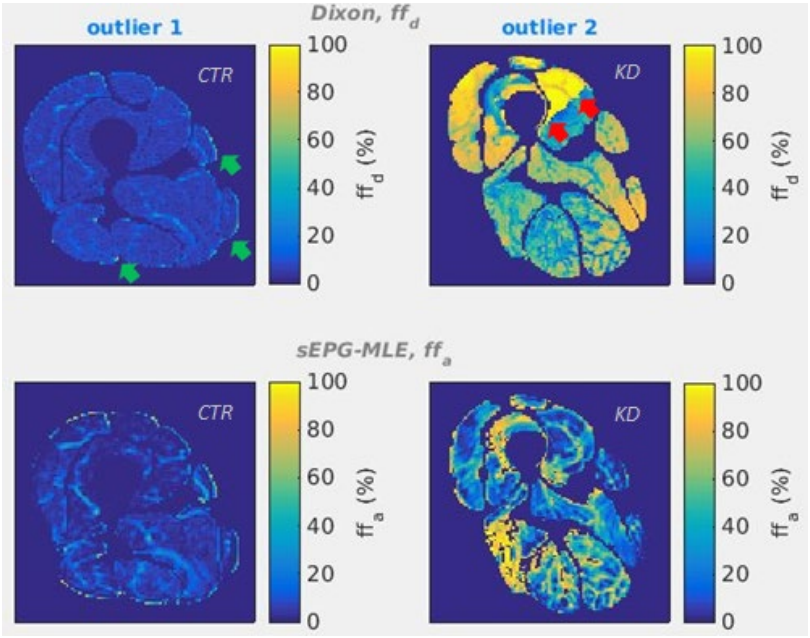


Figure 4.24: The ff_d maps that yielded the outlying results observed in the graph of Figure 4.21, exhibiting segmentation (top left, green arrows) and phase (top right, red arrows) unwrapping imperfections that potentially caused them. The corresponding ff_a maps are also presented (bottom).

Segmentation-related high values (most likely subcutaneous fat contamination) also exist in the corresponding ff_a maps (bottom images), these however have less impact on the overall ff_a distributions and consequently on the cross-section median (outlier 1), because these generally yield increased values compared to ff_d. Besides the apparent localised phase unwrapping imperfection in the ff_d map, this map appears to be inverted throughout when compared to the ff_a map, suggesting a phase-related mislabelling of the water and fat components leading to ff_d elevation and outlier 2.

In Figure 4.25 one-to-one comparisons of sEPG-MLE and Dixon *ff* maps are presented for good quality example datasets from each group (CTR, ALS and KD).

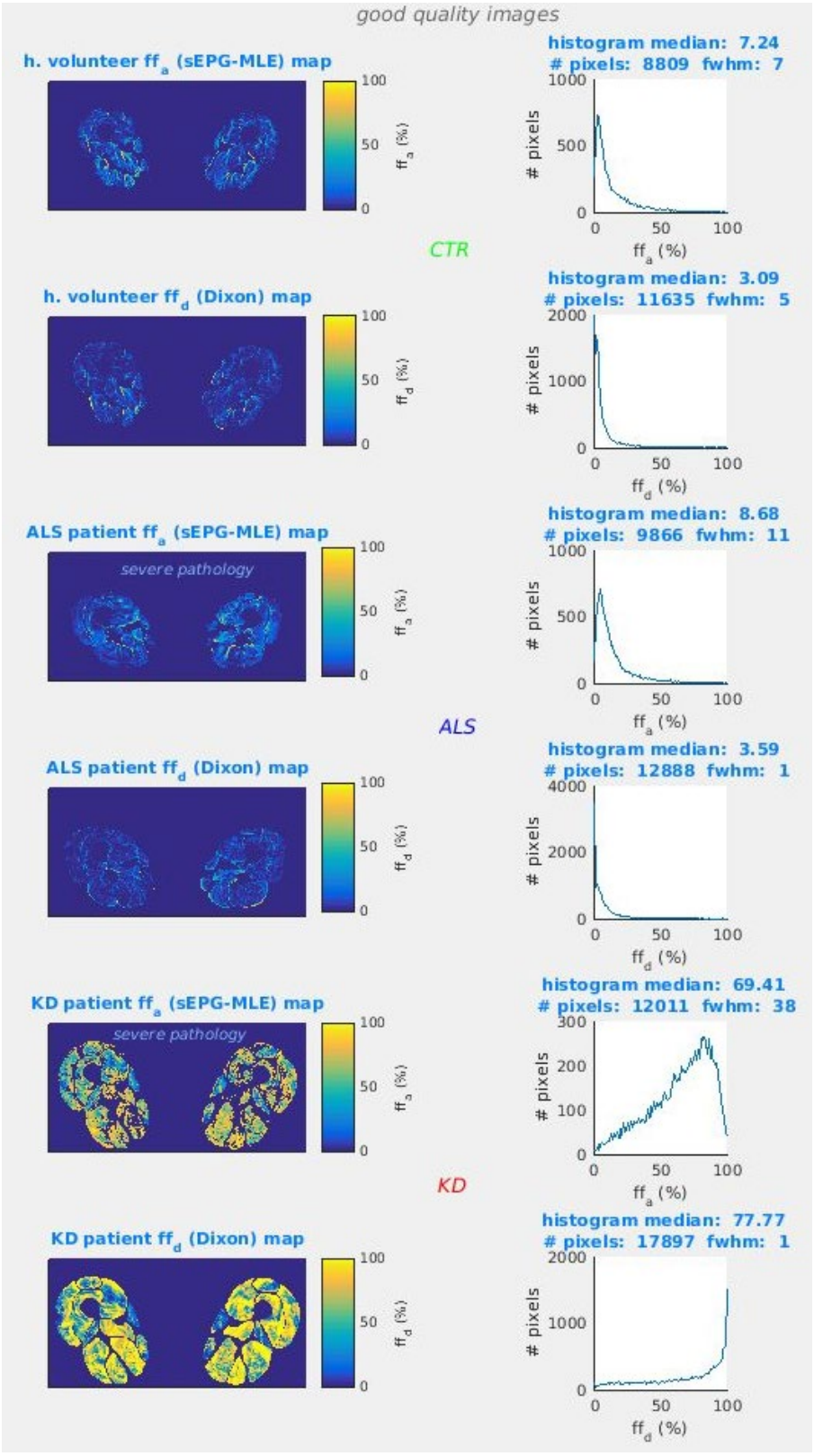


Figure 4.25: One to one comparison of ff_a and ff_d maps and histograms for good quality datasets from healthy volunteers (CTR, top), ALS patients (middle) and KD patients (bottom). The trends seen in Figure 4.21 for overestimated ff_a compared to ff_d at the low end of the ff range and vice versa are seen also in these ff maps.

Higher ff_a estimates in comparison to ff_d are seen in the CTR and ALS maps which predominantly contain values at the lower end of the *ff* spectrum, and conversely the KD ff_d map appears brighter than the ff_a map at the higher end of the spectrum – in agreement with Figure 4.21. The Dixon method *ff* is systematically more narrowly distributed than its sEPG-MLE counterpart.

There are several candidate mechanisms that might lead to the above differences. Dixon underestimation of *ff* is known and attributed to the fact that it relies on the main fat peak frequency (of the methylene group) and the respective correction introduced (Azzabou et al., 2015) – and also used in this work – may not fully address the issue. For the ff_a estimation on the other hand, separated non-exchangeable fat and water compartments in tissue are assumed, which might be an oversimplified model (sections 2.1.3 and 2.1.4). Parameter interaction and uncertainty in the fitting process might be an additional practical issue as seen previously (section 3.3.2). For example the combined ff_a underestimation and T_{2m} overestimation at the upper *ff* range seen in Figure 3.75 might be related to the ff_as being lower than ff_ds at the upper *ff* range. Inaccuracies might also be introduced due to the approximations used in the correction steps used in the ff_a estimation to account for the incomplete T_1 recovery and MT effects. At the pixel level, ff_d map ROIs are in general smaller and potentially different than their ff_a map counterparts, mainly due to the post processing steps used with the sEPG-MLE fitting tool which acts in the main to exclude pure fat or artefactual signals (pixels). Finally, apart from the segmentation and phase unwrapping imperfections seen in ff_d maps, occurrence of calculated ff_ds outside of the expected physically meaningful bounds (between 0 and 100%), in

many cases reaching extreme values (up to thousands), suggests that there might be errors from the same sources in the results also within the useful range. This could possibly explain the uneven density of the ff_d results along the *ff* range, where values appear more clustered towards the two ends of the spectrum and rather sparse in the middle, whereas ff_a shows a more linear distribution (Figure 4.21). These differences cause the curved ('banana'-shaped) appearance above the diagonal equivalence line in their correlation graph.

Exponential models and LSQ minimization

The exponential model used for comparison at this stage was the analogous equivalent of the final muscle-fat component sEPG model, incorporating a single component term for muscle linked through the *ff* parameter to a fixed parameter 2-component model for fat (section 4.2). The fat component fixed parameters were again estimated from subcutaneous fat ROIs, as was the case for a similar approach previously reported (Azzabou et al.). The estimated values used for the fixed parameters were similar to the ones found from the sEPG-MLE method (45ms, 214ms and 70% respectively for T_{2fs}, T_{2fl} and the amplitudes ratio). In order to limit any confounding effects due to artefacts, signal dropouts or obvious segmentation imperfections, good quality datasets only were compared.

In Figure 4.26 one-to-one comparisons of the sEPG-MLE and EXP-LSQ estimated T_{2m} maps and histograms for representative example datasets are presented, and in Figure 4.27 the corresponding comparisons for the estimated *ff* results.

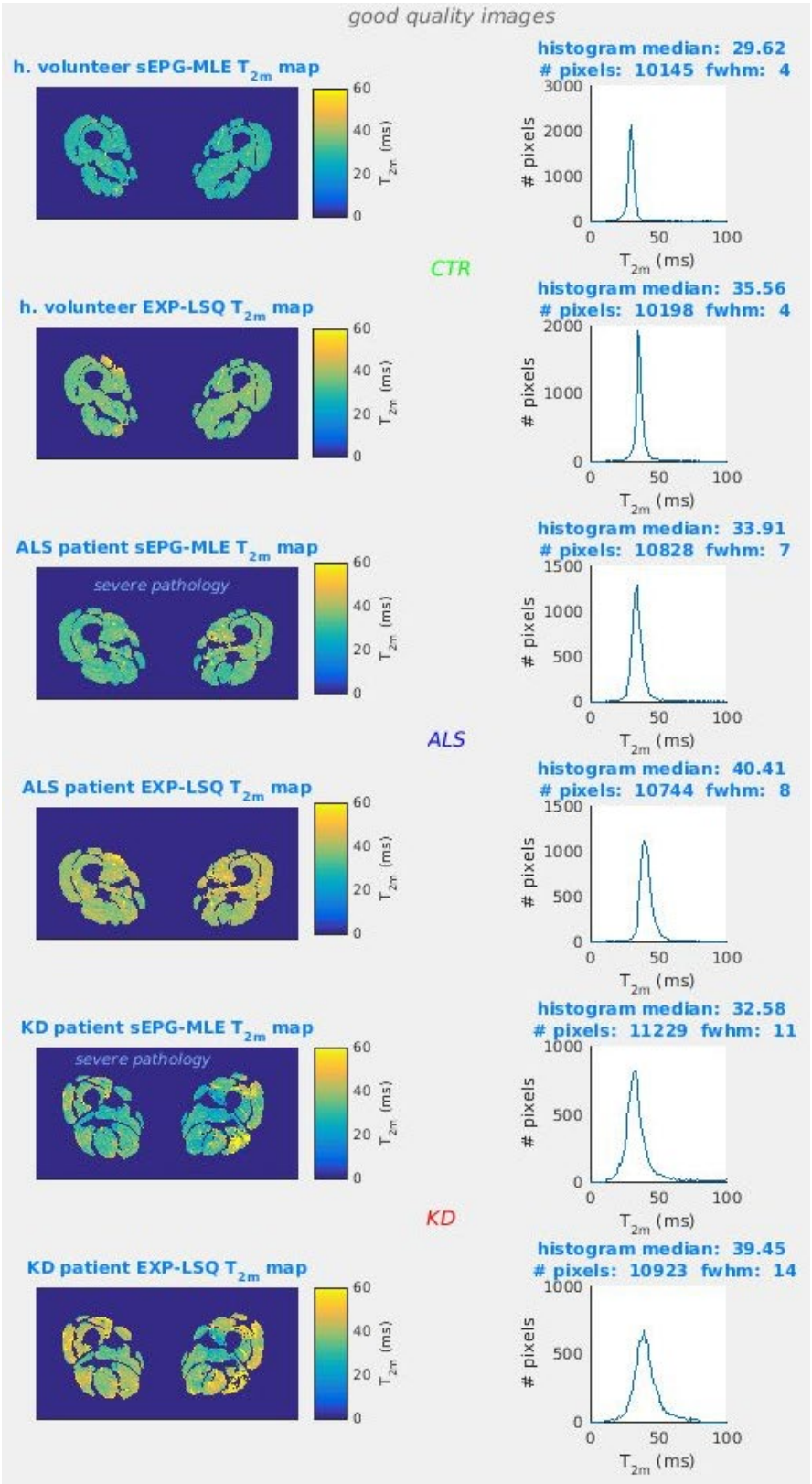


Figure 4.26: One-to-one comparison of sEPG-MLE and EXP-LSQ T_{2m} maps and histograms for representative good quality datasets, for healthy volunteers (CTR, top), ALS patients (middle) and KD patients (bottom). T_{2m} estimates from the EXP-LSQ method are systematically higher and more widely distributed than from sEPG-MLE.

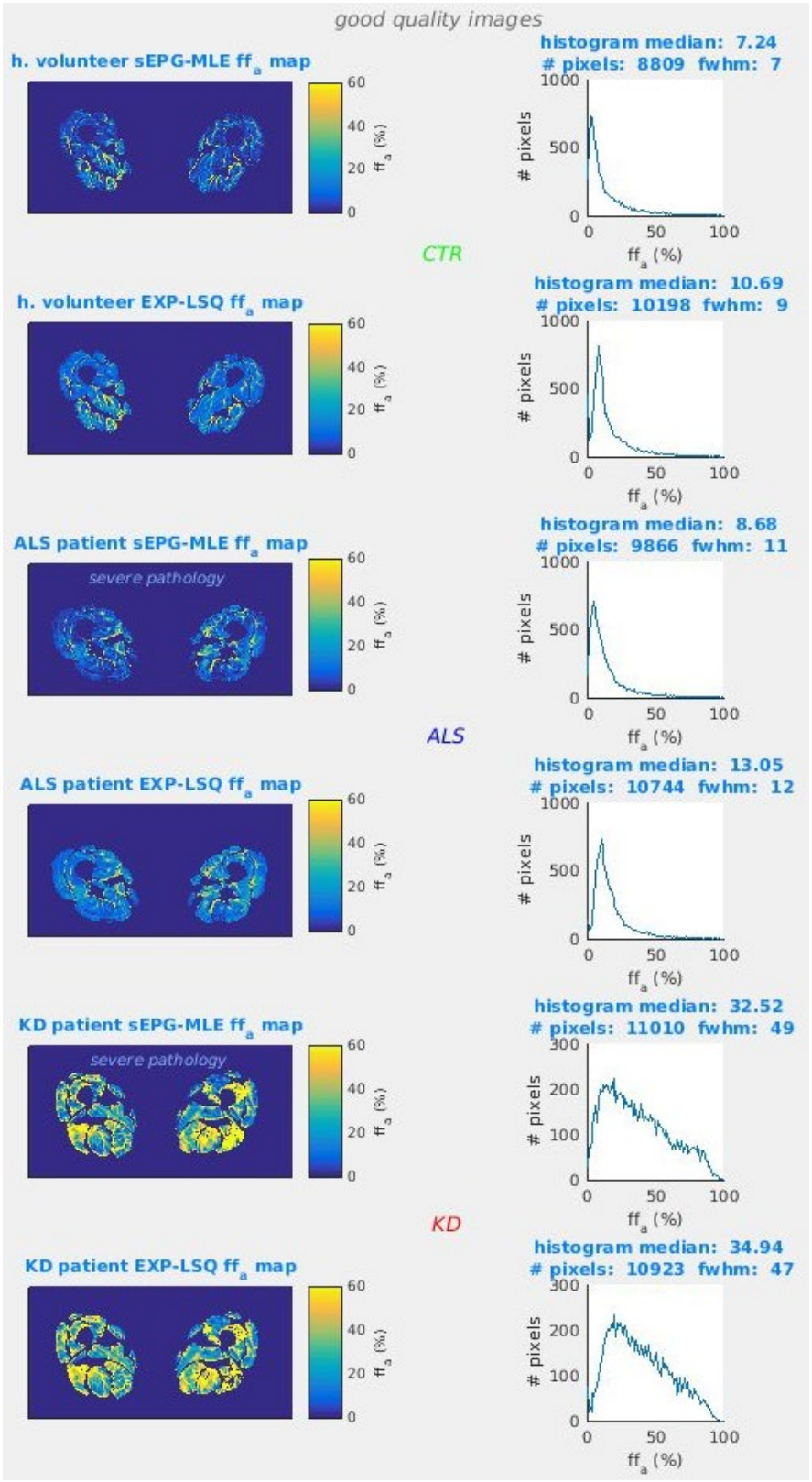


Figure 4.27: One-to-one comparison of sEPG-MLE and EXP-LSQ *ff* maps and histograms for representative good quality datasets, for healthy volunteers (CTR, top), ALS patients (middle) and KD patients (bottom). *ff* estimates from the EXP-LSQ method are overall higher than from sEPG-MLE.

The EXP-LSQ yielded T_{2m} is systematically higher (about 6.5ms) and more widely distributed than its sEPG-MLE counterpart. EXP-LSQ *ff* was also higher than sEPG-MLE *ff*. This important outcome was seen consistently throughout the study data.

The finding of positively biased T_{2m} estimates with the EXP-LSQ approach were consistent with similar results obtained comparing single-component sEPG and exponential models or fitted to phantom data. Re-fitting the 5 disc-shaped phantom T_2 ROIs shown in Figure 3.57, section 3.2.10, using an exponential plus constant baseline model and LSQ minimization, while omitting the 1st echo (Milford et al., 2015), produced the results shown in Table 4.4 (middle column), where for reference the ground truth (calibration) values are also given (left column).

Table 4.4: Comparison of the EXP-LSQ method and sEPG-MLE results, together with the manufacturer's calibration values for the phantom ROIs analysed.

Ground truth T_2 and σ_N (ms)	EXP-LSQ method parameter estimates T_2 and interquartile ranges (ms)	sEPG-MLE method parameter estimates T_2 and interquartile ranges (ms)
31.97 (0.083)	38.47 (37.00 39.65)	32.13 (30.78 33.45)
46.42 (0.014)	54.86 (52.70 57.60)	46.29 (45.12 47.39)
64.07 (0.034)	80.87 (76.51 84.68)	66.24 (64.98 67.64)
96.89 (0.049)	120.21 (115.29 124.58)	96.80 (93.98 99.040
133.27 (0.073)	177.51 (171.32 183.12)	142.00 (138.20 145.69)

The EXP-LSQ T_2 estimates substantially deviate from the reference calibration values – and the respective sEPG-MLE values (shaded column at the right, copied from Table 3.4) – and departures become larger for longer T_2. Comparing the estimates for the two shortest T_2 ROIs (discs), which are comparable to the relaxation times commonly encountered in *in vivo* muscle tissue (Saab et al., 1999) and the real *in vivo* example datasets presented above (Figure 4.25), shows that they are increased by 6.34 and 8.57ms respectively with use of EXP-LSQ fitting instead of the sEPG-MLE tool. This is in good agreement with the 6.5ms approximate increase that was seen in the aforementioned experiment in patients and healthy controls. This is particularly relevant for the shortest T_2 ROI of 32ms (T_1=458ms), because it is reasonably close to the case of lean healthy muscle (T_2≈30ms, T_1≈370ms) where *ff* is very low and the signal decay approaches single component behaviour, as the model used in the phantom experiment. Since the results for the two methods (EXP-LSQ and sEPG-MLE) are very similar for the two experiments (phantom and healthy *in vivo* muscle data fitting) this is a point worth mentioning in terms of support for the validity of the sEPG-MLE fitting tool.

4.5 T_{2m} and STIR contrast mechanisms

Nominally T_2-weighted short tau inversion recovery (STIR) imaging is useful in radiological assessment of neuromuscular pathologies, with hyper-intensity in muscle interpreted as reflecting increased water content due to inflammation or increased blood flow (Kumar et al., 2016, Lovitt et al., 2006). Systematic grading of changes on STIR MRI has been used as a semi-quantitative disease severity index

(Morrow et al., 2013), complementing Mercuri grading of T_1-w images (Mercuri et al., 2002). Important aspects of the technique's contrast dependence and the consequent physical interpretation of contrast changes came to light at this stage of the study, prompting the following separately presented discussion. It is shown that, in fat-infiltrated tissue, STIR signal intensity (SI) may be strongly affected by factors independent of muscle water T_2, specifically the relative proton density (PD) of fat and water in each voxel, as well as imperfect fat signal nulling.

The STIR imaging results at baseline from this patient study were presented in a semi-quantitative classification (Figure 4.28) in a paper published separately to the work of this thesis (Klickovic et al., 2019) by colleagues within the wider collaborative MRI project at our centre mentioned in section 4.1. Muscle tissue STIR hyperintensities were more pronounced in KD than in ALS patients and almost absent in controls; they also increased from the medial to the anterior compartment and further to the posterior compartment.

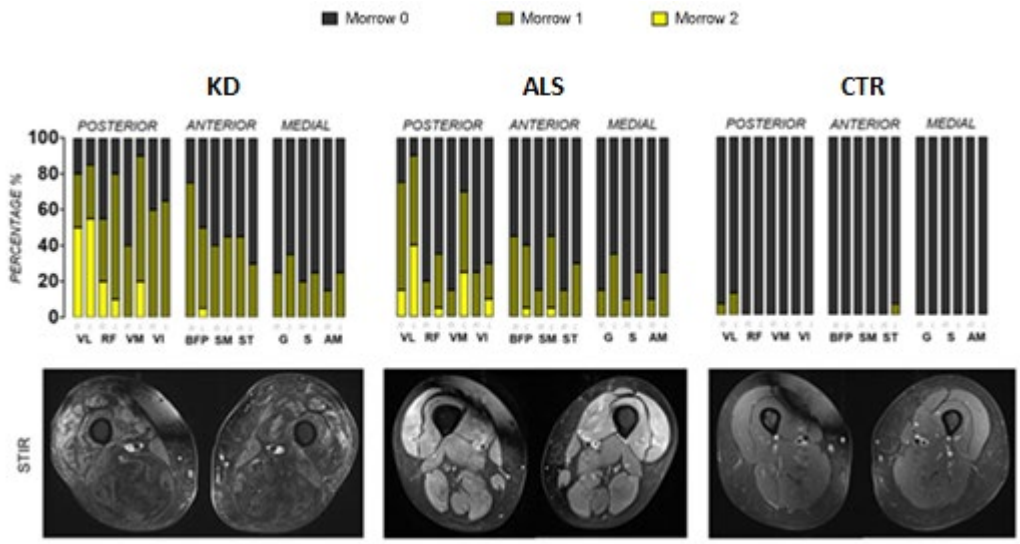


Figure 4.28: At the top row the proportion of participants with each of the Morrow rating scale scores (Morrow et al., 2013) for thigh muscle compartments are shown in all study groups, CTR, ALS and KD. STIR hyperintensities were observed in both patient groups. No marked signal abnormalities were observed in healthy controls. In the bottom row corresponding example STIR axial images are shown for the right and left thigh in KD (left), ALS (middle) and CTR (right) (Klickovic et al., 2019).

The inter-muscle STIR contrast differences observed above are in broad agreement with the trends seen in the T_{2m} results presented in section 4.4 (Figure 4.15, results at baseline, 3 first columns). However a notable difference is that within the patient T_{2m} distributions, as well as a clear tendency for T_{2m} increase, there is also evidence of decrease to a lower value in some individuals, in particular in the KD group, resulting in a higher overall elevation (T_{2m} median per muscle ROI) for the ALS compared to the KD group as well as significantly higher fwqm and skewness (Table 4.1). This suggests that decreases in T_{2m} may also be associated with disease, a tendency which is not shown in the STIR results analysis – indeed the STIR rating scale used in that analysis by design captures only signal hyper-intensities presumed to reflect T_{2m} increases.

In fact there are distinctive regional intensity features in STIR images, different to those apparent in the corresponding sEPG-MLE T_{2m} maps. For instance, in the circled image regions in Figure 4.29, elevated T_{2m} (in the sEPG-MLE T_{2m} map) is not matched by a corresponding hyper-intensity in its respective T_2-w STIR image counterpart, in which this region is actually hypointense.

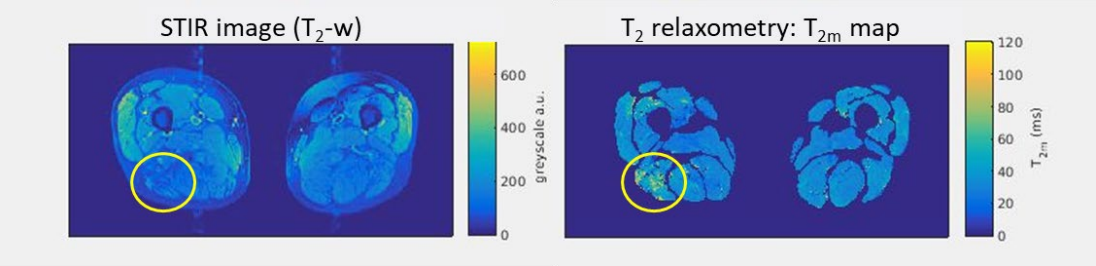


Figure 4.29: Comparison in the case of a KD patient, of the nominally T_2-weighted STIR image, with the corresponding sEPG-MLE method T_{2m} map; An image region where the two methods disagree exists (circled).

The basic equation describing the single component signal behaviour in terms of relaxation effects for an inversion recovery sequence is:

$$s_{IR} = PD\left(1 - 2e^{-\frac{TI}{T_1}} + e^{-\frac{TR}{T_1}}\right)e^{-\frac{TE}{T_2}} \qquad \text{Equation 17}$$

where s_{IR}, PD are respectively the signal and proton density in the image, TI the inversion time, TR the sequence repetition time, TE the echo time, and T_1 and T_2 the decay constants specific to the imaged anatomy (McRobbie et al., 2006). The TI, TR and TE intervals are sequence specific (specified by the operator) and therefore known. For a STIR sequence TI is typically set at a value where the recovering, previously inverted magnetization for fat protons will be approximately zero, i.e. so

that $TI \approx T_{1f} \cdot \log(2)$, where T_{1f} is the longitudinal relaxation decay rate for fat. For the particular STIR sequence used for the above images *TI* was 220ms which signifies that T_{1f} was considered to be around 320ms (at 3T). However – as seen in other parts of this thesis and in published work – this estimation can vary substantially. It follows that if *PD* and T_2 were known the STIR signal intensity could be calculated.

Tissue-specific estimates of relevant parameters obtained by T_2-relaxometry measurements, as already described, and assumed reasonable fixed values for the remaining parameters, allows such calculations. This provides a means to clarify the relationships between nominally T_2-w STIR contrast and corresponding regional T_{2m} variations. To apply this to real world data as in the context of this work, a 2-component signal model is required, accounting for the co-existence of fat in the same pixel region. In that case, the *PD* parameter may be replaced by the global amplitude (intercept) α returned by the sEPG-MLE (or any alternative algorithms for that matter) and an estimate for muscle T_2 is the T_{2m} estimate; T_{2f} can be fixed to a commonly accepted value for fat tissue (e.g. 137ms), similarly to T_{1f}, and the same applies for T_{1m} (e.g. 1400ms). Finally, an estimate for the fat fraction is the *ff* estimate (ff_a). With these modifications Equation 17 becomes

$$s_{STIR} = \alpha \cdot ff_a \cdot \left(1 - 2e^{-\frac{TI}{T_{1f}}} + e^{-\frac{TR}{T_{1f}}}\right) e^{-\frac{TE}{T_{2f}}} + \alpha \cdot (1 - ff_a) \cdot \left(1 - 2e^{-\frac{TI}{T_{1m}}} + e^{-\frac{TR}{T_{1m}}}\right) e^{-\frac{TE}{T_{2m}}} \quad \text{Equation 18}$$

To support a detailed comparison, T_{2m} and T_2-weighted STIR images and the associated T_2-weighted raw CPMG images and *ff* maps – both the apparent (sEPG-

MLE) *ff* map (ff_a) and the Dixon *ff* map (ff_d) – are shown in Figure 4.30 for the previous (Figure 4.29) KD patient. The raw CPMG image corresponds to the 4th echo of the CPMG sequence (i.e. at *TE*=40ms) to match with the 39ms *TE* in the STIR sequence. Included in these figures is also a '$STIR_{CPMG}$' synthetic image (2nd image in the middle row of each figure) calculated by substituting the sEPG-MLE parameter estimates as described above in Equation 18. This was done on a pixel-by-pixel basis from the respective sEPG-MLE parameter estimates maps, i.e. the T_{2m}, ff_a and α maps produced previously.

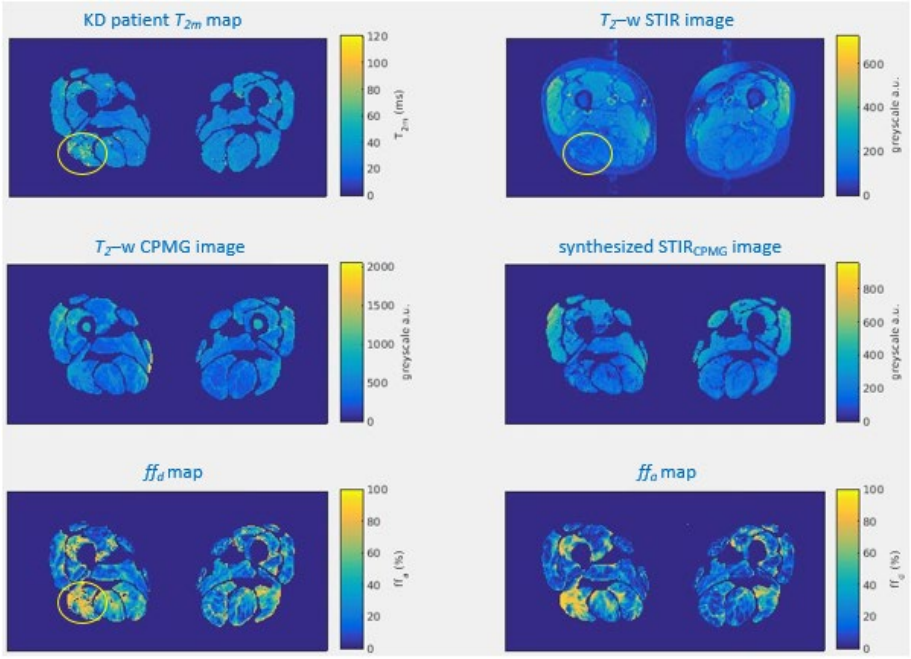


Figure 4.30: Comparison in the case of a KD patient of the sEPG-MLE method T_{2m} map with the corresponding T_2-weighted STIR image (top row); image regions where the two methods disagree exist (circled). Also shown are the corresponding raw CPMG data (4th echo) and a synthesized '$STIR_{CPMG}$' image computed using the sEPG-MLE parameter estimates as described in the main text (middle row), and the apparent (sEPG-MLE) *ff* map (ff_a) and Dixon *ff* map (ff_d)(bottom row).

As can be seen the '$STIR_{CPMG}$' synthetic image corresponds well to the original T_2-weighted STIR image. At the same time they both show distinctive regional intensity features, different to those apparent in the sEPG-MLE T_{2m} map and the corresponding T_2-weighted CPMG image. For instance, in the circled image regions in the top row of Figure 4.30, elevated T_{2m} (in the sEPG-MLE T_{2m} map) is not represented by a corresponding hyper-intensity in its respective T_2-w STIR image counterpart. This suggests that, in addition to the expected T_2-weighting, STIR contrast may be heavily dependent on factors other than local T_2. The signal modelling experiment incorporated in this figure corroborates that total *PD* and *ff* significantly impact STIR contrast: in this case the increased *ff* (bottom row) would result in reduced STIR signal intensity. This strongly suggests that there is danger in STIR image interpretation that increased *ff* may mask oedema effects and vice versa.

The theoretical dependence of STIR signals intensity on fat content (*ff*) for a fixed T_{2m} of 30ms, and for values of T_1 (T_{1f}) that deviate from the inversion recovery nulling condition in muscle tissue is demonstrated in Figure 4.31.

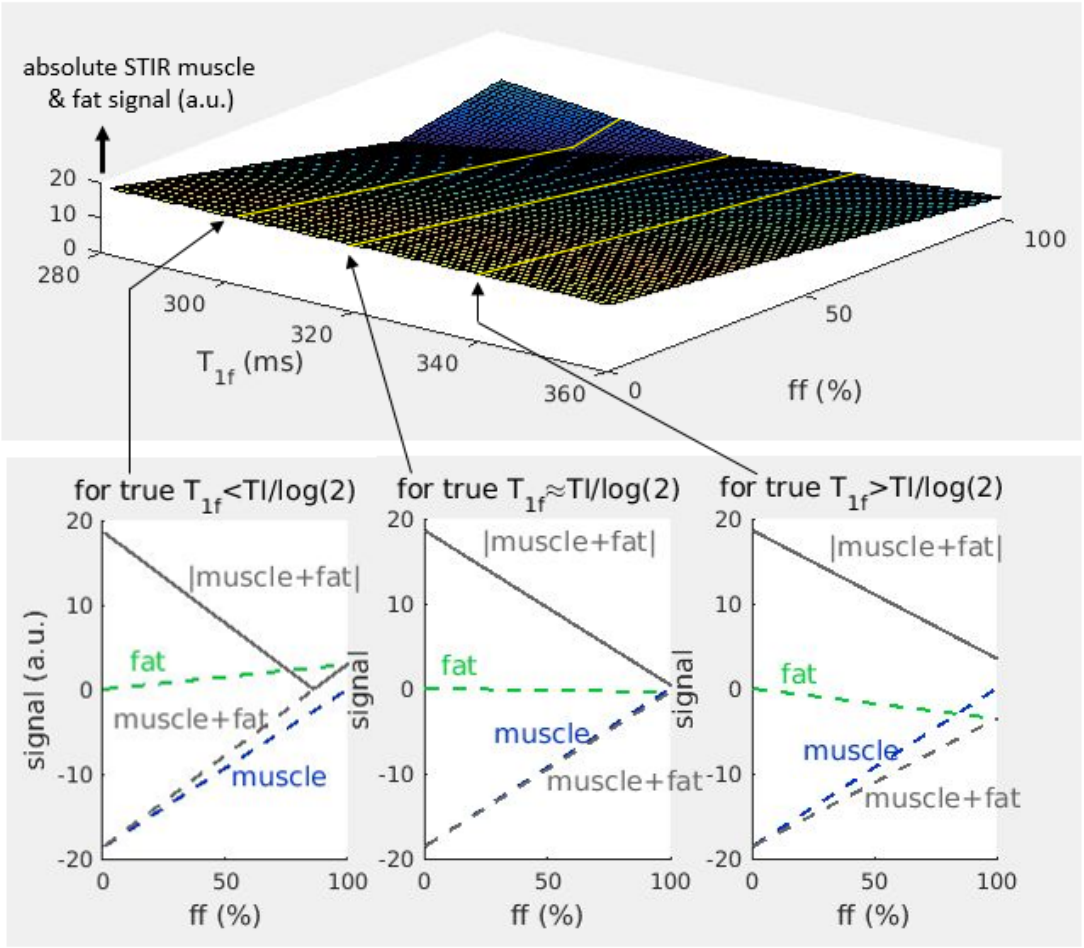


Figure 4.31: Theoretical dependence of STIR imaging signals on fat content (*ff*) and fat T_1 decay rates (T_{1f}) in musculoskeletal tissue. As Equation 18 predicts STIR signals are approximately inversely proportional to *ff*, i.e. have significant *PD* weighting in addition to the expected T_2 weighting. Imperfect fat signal suppression, i.e. when $T_{1f} \neq TI/\log(2)$ also influences the STIR signal intensity and its dependence on *ff*.

The predicted STIR signal intensity decreases approximately linearly with *ff* (fat tissue *PD*), while imperfect fat suppression (when $T_{1f} \neq TI/\log(2)$) is predicted to cause elevated signal strength in regions of high *ff*. These STIR contrast mechanisms will be explicitly explained through the following examples.

The $STIR_{CPMG}$ image in Figure 4.30 was synthesized assuming perfect fat suppression ($T_{1f}=TI/\log(2)$). The good correspondence between the synthetic image and the real

STIR image suggests that this condition held true in this case, and therefore *ff* affected image intensity only via the decreased water *PD* it implies. However according to Equation 18 and as shown in Figure 4.31, mismatch between the *TI* and T_{1f}, should this occur, can also have an additional mild effect, especially for high *ff* ranges, where part of the signal can now emanate from fat protons. The predicted effect is demonstrated in the middle row of synthetic images in Figure 4.32, based on a different KD patient dataset.

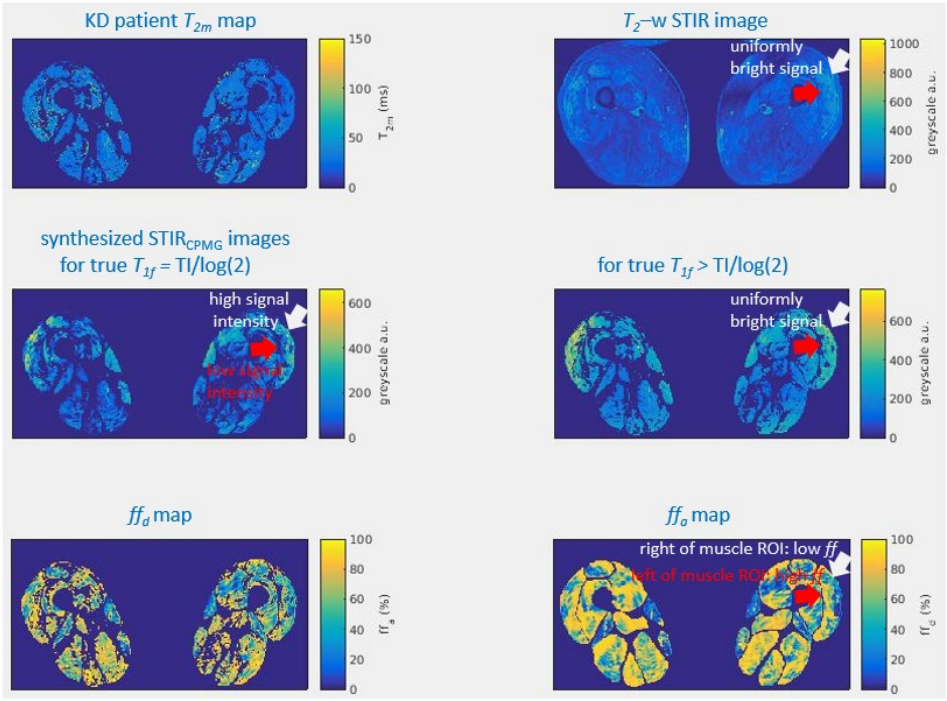


Figure 4.32: Comparison of the sEPG-MLE method T_{2m} map for a KD patient with the corresponding T_2-weighted STIR image (top row). Inaccurate choice of *TI* with respect to T_{1f} would mostly amplify the image intensity in regions of high *ff* as shown by the synthesized '$STIR_{CPMG}$' images comparison for perfect and imperfect fat suppression (middle row, red arrows, see main text); the apparent (sEPG-MLE) *ff* map (ff_a) and Dixon *ff* map (ff_d) are also shown (bottom row).

In this dataset *ff* was overall very high which allows better examination of the signal increases due to incomplete fat suppression which are more intense towards the upper end of the *ff* range. The right hand example in the middle row of Figure 4.32

illustrates the case where true T_{1f} is larger than the value assumed for IR-nulling, where muscle water signals and signals from non-suppressed fat are co-adding mitigating the decreasing water *PD* effect due to increasing *ff*. This behaviour corresponds to the right side of the top surface in Figure 4.31, or equivalently the bottom right graph in that figure.

As it can be seen in the ff_a map at the bottom right of Figure 4.32 the right part of the outer thigh example muscle ROI chosen has low *ff* (light grey arrow) while its left part has high *ff* (red arrow). If fat suppression was optimal, i.e. $T_{1f}=TI/\log(2)$, this would translate to a brighter right side STIR intensity in this ROI, and a darker left side, as shown at the left panel of the middle row of the figure. However as both the synthesized STIR image for $T_{1f}>TI/\log(2)$ (middle row, right) and the true STIR image (top right) show, the entire ROI has a relatively uniform bright appearance. This could be explained by a combination of high water *PD* / low *ff* at the right of the ROI (light grey arrow), and the $T_{1f}>TI/\log(2)$ / high *ff* condition at the left of the ROI (red arrow). In the case that true T_{1f} was smaller than assumed ($T_{1f}<TI/\log(2)$), signal from fat would further reduce the overall observed intensity with increasing *ff* (in addition to the decreasing water *PD* effect), until the point that both *ff* and the T_{1f} discrepancy relative to the choice of *TI* were so large that all apparent signal emanated from fat (folded part of the surface at the top graph of Figure 4.31 and bottom left graph). Therefore for very high *ff*s and large departures of T_{1f} from $TI/\log(2)$, the 2 conditions ($T_{1f}>TI/\log(2)$ and $T_{1f}<TI/\log(2)$) would be equivalent; it is only more probable and thus an assumption that indeed $T_{1f}>TI/\log(2)$, since this condition signifies elevated STIR signal for the entire *ff* range. It should be noted that the main feature in the

Figure 4.32 paradigm remains the increased signal intensity at relatively low *ff* regions, however the overall uniformity of the STIR image – e.g. at adjacent regions of high and low *ff* – may be explained by signal amplification due to imperfect fat nulling where *ff* is high, as described above.

Conversely to the previous paradigm of Figure 4.30, where T_{2m} increase was not matched by STIR hyper-intensity, the opposite condition may prevail, i.e. hyper-intense STIR image regions do not always correlate with elevated T_{2m}, as demonstrated by the circled areas in the top row of Figure 4.33 below.

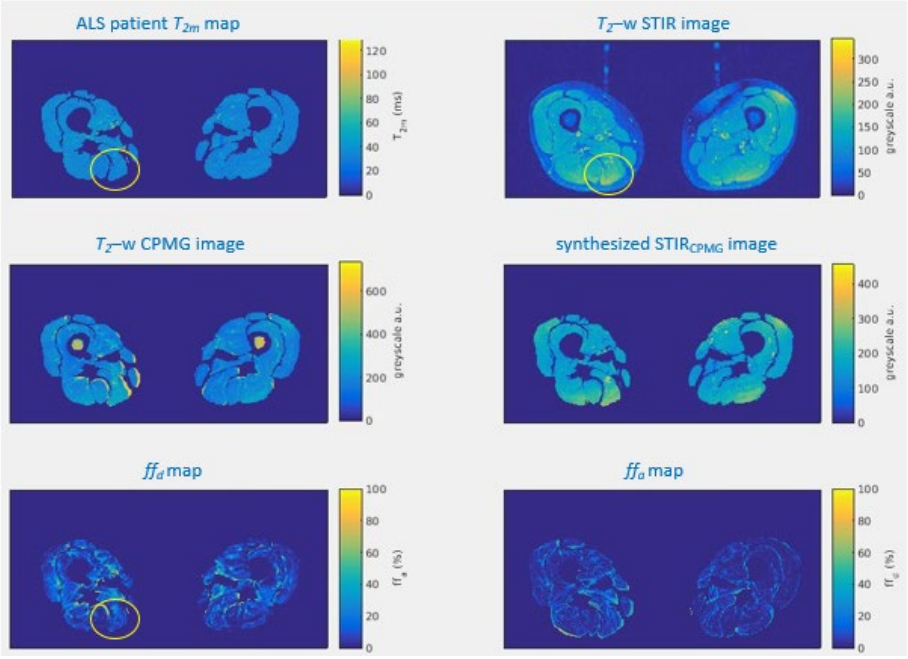


Figure 4.33: Comparison in the case of an ALS patient of the sEPG-MLE method T_{2m} map with the corresponding T_2-weighted STIR image (top row); image regions where the two methods disagree exist (circled). Also shown are the corresponding raw CPMG data (4th echo) and a synthesized '$STIR_{CPMG}$' image computed using the sEPG-MLE parameter estimates as described in the main text (middle row), and the apparent (sEPG-MLE) *ff* map (ff_a) and Dixon *ff* map (ff_d)(bottom row).

The similar appearance of the synthesized STIR image with the true STIR image suggests that the STIR hyper-intensity in that region was due to either locally

increased total *PD* or decreased *ff,* or a receive coil sensitivity effect, since STIR signal hyper-intensity may clearly arise in the absence of increased T_{2m}, i.e. STIR hyper-intensity does not always reflect T_{2m} increases.

These observations suggest that, in addition to the expected T_2-contrast weighting, image intensity on STIR images is inversely proportional to fat content, or in other words, STIR images should also be considered water *PD* weighted. This is corroborated by the $STIR_{CPMG}$ images (which depict the cumulative fat and muscle water signal) which broadly match with the original STIR images and Equation 18. Loss of tissue water proton density due to water replacement by fat in certain regions may in this way mask expected STIR signal intensity increases due to T_{2m} increases in the same area. On the other hand B_1 transmit or receive coil non-uniformity may cause apparent STIR hyperintensity in regions where in fact T_{2m} is normal. Imperfect fat suppression can also have a mild effect, causing STIR hyper-intensity independent of T_{2m}, in regions of high *ff*. STIR contrast in diseased muscle is therefore complex, depending on both the effective muscle water proton density, which decreases as fat content increases, and oedema related changes in T_{2m}.

In conclusion, despite STIR imaging being useful both as a research tool and in radiological assessment, the above findings suggest that STIR contrast interpretation requires caution, and may be improved by comparison with quantitative T_2 maps.

4.6 Discussion

The initial aim of the work presented here was to explore the usefulness of a multi-component sEPG-MLE model method to estimate T_{2m} in this patient study and then

use the data to characterise changes in ALS and KD, assessing its relevance and potential use as a biomarker. As demonstrated in Chapter 3 a fixed parameter secondary 2-component approximation for the fat within a multi-component model would serve this purpose. Application of a set of robust quality control criteria was also shown to be helpful to obtain interpretable results. Estimates obtained for the median of thigh level T_{2m} in both healthy volunteers and from patients with this method were overall in accordance with similar previous measurements, supporting the validity of our approach. In comparison with healthy control findings, substantially wider distributions about these median values were seen in the patients, both in T_{2m} and *ff*. Statistically significant differences (one way ANOVA) were revealed in histogram metrics, between both patients and controls and between ALS and KD patients. Median values for T_{2m} and *ff* were significantly different between healthy controls and patients, predominantly higher in both conditions (ALS and KD), and these differences appear to increase longitudinally.

T_{2m} results were also compared with Dixon *ff* values, available for the same patient dataset from a separate acquisition; Mean T_{2m} appeared broadly independent of fat content, indicating that the aim of obtaining T_{2m} estimates that were independent of fat fraction has been achieved. This is in contrast with global T_2 estimated via a mono-exponential fit which is highly dependent on fat content. Estimation of *ff* from the multi-component (muscle-water and fat) model was also shown to be plausible. Direct correlation with the mean muscle *ff* values yielded by the Dixon data (ff_d) revealed marked departures from the apparent *ff* (ff_a) estimates – even after adjusting corrections – but similar trends overall. There are various explanations for

this disparity, from theoretical model limitations to more practical aspects related to fitting (sEPG-MLE tool) and fat-water separation calculation approaches (Dixon method), and post processing of results, as described in section 4.4 above. This will also be further explored in Chapter 5 (section 5.6). An important feature seen here is the extended range of the Dixon ff_d values below 0% and above 100%, which were discarded. This might be a reason for the uneven distribution of ff_d also within the useful range and could partly explain the non-linear relationship with the sEPG-MLE ff_a estimation, especially near the two ends of the respective distributions. Nevertheless Dixon method obtained ff_d distributions were qualitatively similar to the sEPG-MLE method based ff_a maps.

Our method for improved T_{2m}-mapping, and concomitantly also ff_a-mapping, involved two main aspects. First, with regards to the Physics model, use of the of the specific sEPG formalism used here better accounts explicitly for the effects of B_1 field variations in the slice selection process (Chapter 3, sections 3.2.6 and 3.3.2) compared to other versions presented so far (McPhee and Wilman, 2017). Second, with regards to the statistical modelling, use of the MLE approach explicitly accounts for noise floor effects apparent in this type of acquisitions. These innovations (i.e. the sEPG-MLE fitting tool) led to increased accuracy and precision in parameter estimation when compared to more conventional methods (e.g. the EXP-LSQ approach). The changes in parameter estimation when transitioning from EXP-LSQ to sEPG-MLE using *in vivo* data from healthy muscle closely matched the changes due to the same transition using comparable phantom data and acquisition conditions, validating the abovementioned improvements. It was shown that the widely used

exponential model fitting overestimates T_2 by a factor around 20% or more, which the sEPG-MLE approach can correct for. The healthy muscle sEPG-MLE T_{2m} estimates were also consistent with the widely accepted values, for instance from spectroscopy studies (Forbes et al., 2014, Saab et al., 1999) and T_2-spectrum CPMG techniques (Araujo et al., 2014). Consistency of the MLE estimated noise levels with independent noise measurement methods further supported the validation of our method.

Another novel aspect of this work was to introduce the 2-component model with fixed parameters mentioned at the beginning of this section to describe pure fat signals, and subsequently combine it with a water component to achieve the final CPMG fat-water quantification. While this is a novelty in the context of an EPG application (Marty et al., 2016), the concept in fact is similar to its equivalent exponential version, used in published work and referenced earlier (Azzabou et al., 2015). Unlike the latter exponential approach where there was some correlation found between the dominant fat component and the main methylene group spectroscopic peak – namely about 80ms T_2 and 67% relative amplitude – in both EPG-based approaches mentioned here this was not the case. This might simply reflect an intrinsic unsuitability of MESE data fitting to assign specific T_2 decay constants as per corresponding spectroscopic analyses, for instance as discussed for water T_2 in section 2.1.4. Nevertheless as a fixed model for fat – used in a 'black box' approach – the 2-component model describing fat was shown here to be superior to a single component model and useful in the interpretation of the resulting overall fat-water model fitting behaviour. The role of parameter bound constraints and quality of fit metrics used to optimise the analysis (section 4.3) helped as well in this respect,

predominantly by identifying the regions of pure fat and artefactual signals. There appear to be limited pixel cases where this interpretation becomes challenging, mainly for very high fat content (but not artefactual) pathological tissue ROIs, where more sophisticated experiments could elucidate these observations, for instance comparison with biochemical analyses.

Interesting findings emerged by comparing the sEPG-MLE T_{2m} maps with the corresponding T_2-weighted STIR images. Both the comparisons of these images and a closer examination of the basic assumptions for STIR imaging revealed that this modality's contrast depends in addition to T_2 also on water proton density. This introduces ambiguity and potential difficulty in the interpretation of the features seen in such images. Increased fat content in general leads to reduced image brightness and vice versa, and this can mask the expected T_2-weighted contrast. If additionally fat suppression is imperfect (incomplete) this can reverse or further enhance this effect in variable combinations. Therefore direct comparison of STIR signal intensity with quantitative T_2 imaging results can be challenging – STIR signal hyperintensity is not a direct unambiguous surrogate for increased T_2. An additional important observation is that the way that STIR images are interpreted clinically is that commonly only signal hyper-intensities are considered as a manifestation of abnormalities. This is a clear difference with T_{2m}-mapping where also lower range values are evident and play a substantial role in the overall evaluation of results as discussed in this section and also in Chapter 5.

Many processes can contribute to the aforementioned changes in the T_{2m} distributions reported in this chapter. Muscle denervation due to motor neuron

degeneration is a key characteristic of ALS and KD and at the early stages of disease is associated with oedematous changes which are thought to lead to increased T_2 before they eventually give way to fat replacement (Lovitt et al., 2006, Bryan et al., 1998). This fat deposition also causes T_2 elevation in conventional T_2-mapping (e.g. using single-component or mono-exponential models), and it can be alleviated – in principle – with fat corrected T_2-mapping techniques, in order to consistently depict muscle water T_2 (T_{2m}) (Yao and Gai, 2012). The main purpose of obtaining pure T_{2m} assessment is to efficiently monitor changes and potentially intervene at the early stages of disease before wide spread fat infiltration and deterioration of symptoms (Wary et al., 2015). The acute denervation and inflammatory stages expected in the fast course of ALS seem to correlate with its higher T_{2m} relative to both controls and KD patients, showing promise for the above purpose. Similarly the wider KD T_{2m} distributions (both upwards and downwards) seem to correlate with the intense fat deposition and inter-muscle variations (diffuse patterns and 'mottled' appearance) typical in this condition. This widening may occur both because inflammatory stages and therefore increases in T_{2m} are rather transient, and because of the challenges that high fat content introduces in the water component characterisation in terms of uncertainty and tissue compartmentalisation. Additionally muscle fibre destruction may play a role (Klickovic et al., 2019). These differences were evident also from other histogram metrics; for instance increased fwqm values were seen for KD patients, significantly higher than for the ALS group. Therefore the observed relatively higher T_{2m} distributions in ALS as opposed to wider ones in KD, along also with time, are consistent with dominance of the inflammatory as opposed to fat infiltration effects respectively for each condition, as described at the beginning of this chapter.

The above findings are also consistent with the ff_a estimation results (and to a lesser extend with ff_d). All three groups (ALS, KD and healthy volunteers) showed significantly different ff_a distributions and metrics between each other. ALS patients showed a marginal (not significant) tendency for increase in ff_a longitudinally, which may imply that KD patients reach rather promptly a stage of saturation in the occurring fat deposition and replacement. This may be related to the rather transient aforementioned oedematous stages for KD patients, and does not preclude the varied patterns and spared ROIs seen in their ff_a-maps.

4.7 Summary and Conclusions

As a final conclusion, the findings of this study suggest that inflammation/oedema and fat infiltration both progress with disease evolution but in different proportions and rate between the two diseases studied (ALS and KD). Additional tissue compartmentalisation processes may as well play a role. These findings were demonstrated by the trends observed in the T_{2m} and ff_a estimates' distributions and histogram metrics yielded using the sEPG-MLE algorithm.

In summary:

- While arising from different pathogenic processes, ALS and KD may show similar symptoms at their initial stages and distinguishing the two diseases on imaging is of interest.
- The multi-component sEPG-MLE fitting tool including a fixed parameter secondary 2-component approximation for the fat signal as described in

Chapter 3 was used to estimate T_{2m} and ff_a for a cohort consisting of ALS and KD patients and healthy volunteers, and within time frames of one year.

- Additional post-processing steps were included in order to avoid contamination from pure fat and artefactual signals in the overall results.
- ALS patients manifested significant elevation in both thigh-level muscle T_{2m} and ff_a suggesting that inflammation and fat infiltration both progress with time.
- KD patients manifested predominantly severe fat deposition and replacement in diffuse patterns in the corresponding ff_a maps and significant T_{2m} changes suggesting inflammatory and potentially other effects.
- Despite the MT, T_1-weighting and spectroscopy-based corrections used, ff_a results did not quantitatively match their Dixon (ff_d) counterparts; however ff_a- and ff_d-maps were qualitatively similar. Between group comparison p-values for ff_a were lower than for ff_d, potentially supporting the validity of ff_a as a reliable outcome measure.
- T_{2m} changes appear to be in general independent from fat fraction (ff_a and ff_d) elevation, suggesting that the two processes are not affecting each other due to the estimation (measurement) method.
- The changes seen when shifting from the conventional EXP-LSQ to the sEPG-MLE method in a phantom experiment were consistent with equivalent experiment results using *in vivo* data (in comparable conditions), providing validation for the sEPG-MLE fitting tool.
- Caution should be taken interpreting STIR image contrast in fat infiltrated muscle and comparing STIR acquisition results with sEPG-MLE T_{2m}-mapping.

The relative proton density of fat and water tissue in each voxel, imperfect fat signal nulling and receive coil sensitivity effects are confounding factors, possibly challenging the inherent assumptions in STIR imaging.

- The proposed method could potentially be applied to individual examinations and clinical trials to add to the diagnostic sensitivity of clinical MRI. In the longer term it can aid towards facilitating drug development and validating MRI clinical outcome measures.

Chapter 5. The non-ambulant DMD MRI natural history study

This chapter presents work based on data acquired in an MRI natural history study in non-ambulant DMD boys performed at our institution. While lower limb MRI has been more extensively investigated in DMD as a potential measure of disease progression, upper limb measurements have received little attention to date. Forearm muscle MRI measurements may be important because upper limbs are affected by the disease at a later age than lower limbs thus offering the opportunity to meaningfully monitor disease progression for older patients.

5.1 DMD background

Duchenne muscular dystrophy (DMD) is one of the most common inherited skeletal muscle degenerative diseases, with a frequency of occurrence of one in 3,500 boys. It is caused by mutations in the X-linked dystrophin gene (Xp21), leading to muscle membrane deficiencies which render them susceptible to injury from eccentric muscle contractions, progressing to oedema and finally to irreversible fatty infiltration and fibrosis. Patient symptoms start with weakness by the age of 5 years, will typically progress to wheelchair use in their early teens, and die of cardiorespiratory complications by the second or third decade of their life span (Arpan et al., 2013). The disease is in general considered incurable, however it has been shown that it can be decelerated with early steroid treatment (Kim et al.). Oligonucleotide treatment inducing increased muscle dystrophin expression and benzoquinone administration improving functional late stage cardiopulmonary parameters are recent advances (Fischmann et al.).

The aim of this longitudinal study was to examine the ability of the sEPG-MLE fitting tool presented in the previous chapters of this work to quantify disease progression related changes in the upper limbs of non-ambulant DMD boys, comparing measurements between baseline and 6 and 12 month intervals, and also with those from healthy volunteers. Previous analyses of data from this cohort study performed by colleagues in our centre, the UCL Queen Square Institute of Neurology, have demonstrated the efficiency of dynamic MRI in extracting lung and diaphragm measures to monitor DMD (Bishop et al., 2018), and – more pertinent to this work – the sensitivity of muscle Dixon method *ff* as a biomarker to monitor disease progression in the upper limb in non-ambulant DMD, for a cohort size and time intervals typically used in clinical trials (Ricotti et al., 2016). While fatty infiltration in muscle typically appears at later stages of the condition, muscle water T_2 may be hypothesised to show abnormalities early in the disease course, likely to reflect changes in the water content due or oedema (possibly but not necessarily related to inflammation), denervation, fibrosis and other pathological manifestations (Mercuri et al., 2007, Lovitt et al., 2006, Arpan et al., 2013).

The CPMG MRI data acquisition protocol generating the data for the following analysis (as detailed in section 5.2 below) was in retrospect considerably sub-optimal in comparison to other datasets seen so far. This is for historical reasons – at the time of acquisition *ff* was the primary outcome measure of interest, and the team had not yet considered CPMG sequence optimisation – and also for practical reasons specific to forearm anatomy. The limited echo train length of 12 echoes, and the unavoidable positioning of the scanned anatomy (forearm) at the periphery of the magnet bore

are the main issues. In addition issues of reported patient discomfort causing motion (e.g. tremor), stiffness, physical tension and similar effects may also have had negative impact on the data quality. These challenges offer an opportunity for further evaluation and improvement of the analysis approach developed so far, building on the implementation and optimisation of the proposed method (Chapter 3), and in comparison with its application to better (than the DMD) quality datasets (Chapter 4).

To address the aforementioned specific issues with this study, the credibility of the proposed sEPG-MLE method for these datasets is first explored in terms of model suitability and post-fitting quality control (section 5.3), followed by initial results (section 5.4). Then atypical CPMG signal behaviours encountered in the study are explored (section 5.5), followed by *ff* estimation considerations (section 5.6) and further results (section 5.7) and discussion (sections 5.8 and 5.9).

5.2 Tools and Methods

The multi-echo MRI data analyses described in this chapter were obtained as part of a collaborative study performed by colleagues at the UCL Queen Square Institute of Neurology. A first report describing the study and cross-sectional Dixon *ff* findings has previously been published (Ricotti et al., 2016). 14 non-ambulant DMD boys were recruited with a mean age of 13.3 y, (range: 10.8-17.3 years) mean duration of non-ambulation 20.2 months (range: 4.7-41.6) and mean BMI 26.5, range: 20.8-41.7). Assessments were performed at baseline, 3, 6 and 12 months. All but one subject were receiving glucocorticoids. Nine age and gender-matched healthy control

subjects were scanned once (mean age: 14.6y, range: 13-17; mean BMI 21.5, range: 16.5-25.4). Approval from the local research ethics committee was obtained for this study, which was performed in compliance with the Declaration of Helsinki.

For the T_2 weighted imaging the forearms of all boys were examined at 3T (Siemens Skyra) using a flexible matrix-coil (4-Channel Flex Coil) wrapped around one forearm. For that reason the imaged anatomy was in general away from the optimal position at the centre of the magnet bore and closer to its periphery. A multi-echo spin-echo (MESE) sequence (TR=3000ms, 12 *TE*s from 10-120ms with 10ms interval, 9 x 6 mm slices, matrix 320x320, or 320x190, in-plane resolution 0.5625x0.5625 mm) was used. Ten forearm muscles were manually segmented on the central slice of a reference image derived from the respective unprocessed shortest *TE* Dixon acquisition (*TE* = 3.45ms, see next paragraph) by observers blinded to subject groups, using ITK-snap software (Yushkevich et al., 2006). This central slice was defined as the first axial slice distal to the supinator muscle; for consecutive longitudinal data from the same subject, the slices selected for analysis were matched to the baseline slices by visual inspection and reference to a coronal scout image. Due to the different acquisition spaces of the Dixon and the MESE, it was necessary for the segmented Dixon images to be spatially adjusted by the current author using a custom-written software tool in order to match to the T_2-weighted raw images for this work. An additional tool was created to apply these ROI segmentations to the raw T_2-weighted images. This was the core data set that was subsequently used throughout the work presented in this chapter, the main aim being to determine the most effective strategy to extract useful information by fitting this signal with the methods developed in Chapter 3.

Comparative data were available in terms of images collected using three-point-Dixon technique (Glover, 1991) were acquired (2D gradient-echo *TE1*/*TE2*/*TE3*/TR= 3.45/4.60/5.75/102ms, flip angle 10°, nine 6mm axial slices, slice gap 12mm, FOV 18x18cm, matrix 320x320, pixel size 0.56x0.56mm, NEX=4). Images were post-processed offline with a Python programming language pipeline according to Glover and Schneider's algorithm (Glover and Schneider, 1991) and separated fat (f) and water (w) images were used to calculate pixel-wise fat fraction (*ff*) maps according to *ff*(%)= 100*f/(w+f). Data were extracted from the same segmented muscle ROIs as above.

Finally, to allow comparison of the new sEPG-MLE method to the conventional exponential model approach to MSME data analysis, a multi-exponential function:

$$s(TE) = \alpha \cdot (1 - ff) \cdot e^{-\frac{TE}{T_{2m}}} + \alpha \cdot ff \cdot \left(0.75 \cdot e^{-\frac{TE}{76}} + 0.25\right)$$

where *s* is the signal at each *TE* and muscle water T_2 (T_{2m}), fat fraction (*ff*) and overall amplitude (α) were the parameters to be estimated, was fitted pixel-wise to the data using nonlinear least squares minimization (section 2.2.2) in a custom-written Matlab tool. The fat component parameters were determined separately as the mean values from 4 subcutaneous fat ROIs in 8 representative subjects. In terms of quality control, values were excluded from the maps for pixel data which failed to meet empirically determined fit-quality criteria: R^2 (goodness of fit)>0.8, amplitude α lower confidence interval (CI)>0, amplitude α<10 times the 1st *TE* image amplitude, T_{2m}-CI-width/ T_{2m} <100%, *ff*<50% and T_{2m} <100ms.

5.3 Analysis implementation

To implement the proposed sEPG-MLE method three preliminary steps were taken, following the fitting strategy deployment and core optimisation strategy presented in Chapter 3, before moving onto a full evaluation of the patient study data. First, to validate the calibration of the fat signal model developed in Chapter 3 and in Chapter 4 when applied to this DMD dataset. Second, incorporating the previous findings, to re-examine the behaviour of the overall (fat and water) multi-component model specifically with data from this study. Third, to test the post-fitting quality control steps so that physical meaningfulness of the yielded results is ensured.

5.3.1 Forearm fat compartment calibration

The proposed sEPG-MLE method in its optimised form (equation 14, section 3.3.3) relies on using a fat compartment two component model with fixed parameters, which were empirically pre-determined. The fat signal model consists of two sEPG components the most important parameters of which were a short and long T_2 (T_{2fs} and T_{2fl}), and the ratio of the amplitude factors of their respective sEPG signals. It was considered essential to corroborate that the assumptions developed in earlier chapters supporting the use of such fat-signal model hold also for the data available in this study. For that reason the behaviour of the fat compartment model

$$s_{fat}(TE) = 0.33 \cdot sEPG(B_1 f, T_2 = 40ms, \alpha, , \sigma_N, TE) + 0.67 \cdot sEPG(B_1 f, T_2 = 198ms, \alpha, \sigma_N, TE)$$

established in section 3.3.3 was examined for selected subcutaneous fat ROIs from images from the DMD dataset, using only images which were of good quality, i.e. having relatively high SNR with no obvious B_1 inhomogeneity, signal drop-outs or other artefacts. The workflow routine is summarised in Figure 5.1 including specific settings (bound constraints, start values, number of iterations etc.).

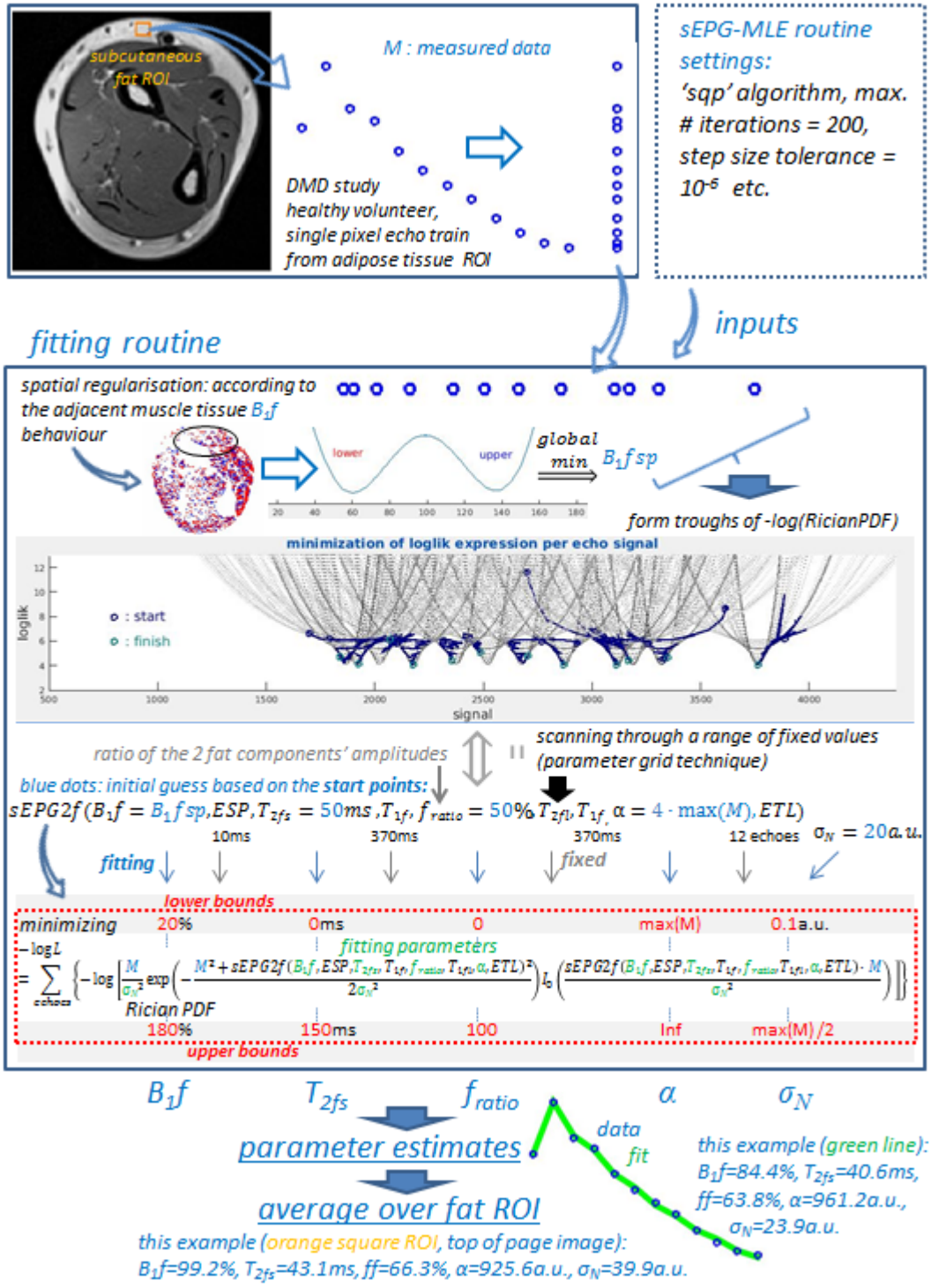


Figure 5.1: A scheme illustrating the implementation of the 2-component fat model sEPG-MLE fitting algorithm, including specific start points, bound constraints and algorithm settings as discussed in Chapter 3 (section 3.3.2). Signals from selected forearm subcutaneous fat ROIs (here shown for an example dataset from a healthy volunteer) are fed into the fitting routine (lower box) where their joint negative log-likelihood is minimized in order to yield the (MLE) parameter estimates, which are subsequently averaged over all pixels of the ROI.

The above model for s_{fat} was used, with this generic mathematical expression (it cannot be written in closed form) describing the aggregated EPG (function) calculations across the slice profile for the fat signals, which is subsequently fitted pixel-wise to the study data using MLE minimization in a custom-written Matlab tool (section 2.2.3). The parameters to be determined for each pixel were the fat short T_2 component T_{2fs}, the ratio of the 2 fat components' amplitudes, B_1 field factor ($B_1 f$), overall amplitude (α) and noise standard deviation (σ_N). T_1 for fat was fixed in accordance to existing literature values, and the remaining sEPG model parameters (*TE*, ETL, RF excitation and refocusing profiles and duration vectors) were set according to sequence and scanner settings, as listed in section 5.2. Finally T_{2fl} was fixed at 198ms, its final estimated value from the discrete values-based technique used previously (Table 3.3). The model was then fitted to forearm subcutaneous fat ROIs from the current study, to test if the remaining free parameter estimates show any significant deviations which would challenge the validity of using the same fat component model also for this study. Examples of the yielded parameter estimates and quality of fit indices are shown in Table 5.1.

Table 5.1: Fitting of the 2-component fat model introduced previously (section 3.3.2), and determined under different acquisition conditions, to subcutaneous fat ROIs from the DMD study, yields similar results (medians) and shows good quality of fitting. Fitting was performed with a fixed T_{2fl} (fat compartment long T_2 component) at 198ms, whereas fitted were $B_1 f$ (B_1 field factor), T_{2fs} (fat compartment short T_2 component), ratio (the 2 fat components' amplitude proportion), amplitude (not shown) and noise (σ_N). Pseudo-R^2 and loglik are as previously defined in sections 3.2.6 and 3.3.2.

	$B_1 f$ (%)	T_{2fs} (ms)	ratio (%)	T_{2fl} (ms)	σ_N	pseudo-R^2	loglik
patients	99.8	**40.3**	**69.3**		23.7	99.7	55
	99.4	**36.6**	**67**	**198**	20.0	99.7	52.9
Healthy	99.2	**43.1**	**66.3**	*fixed*	39.9	99.1	61.3
volunteers	85.6	**38.6**	**66**		18.4	99.8	51.8

Two fat ROIs were analysed from 2 patients and 2 healthy volunteers.

T_{2fs} emerged around 40ms and the two fat components' amplitude ratio was close to 67%, which were closely matched to fat component parameters found previously in Chapter 3, with the minimization routine rapidly converging to these values. An example of performing the fitting with T_{2fl} fixed to a value from 100 to 300ms in steps of 1ms (i.e. the discrete parameter testing technique) for an example ROI – the same as shown in Figure 5.1, corresponding to the 3rd example from Table 5.1 – is shown in Figure 5.2.

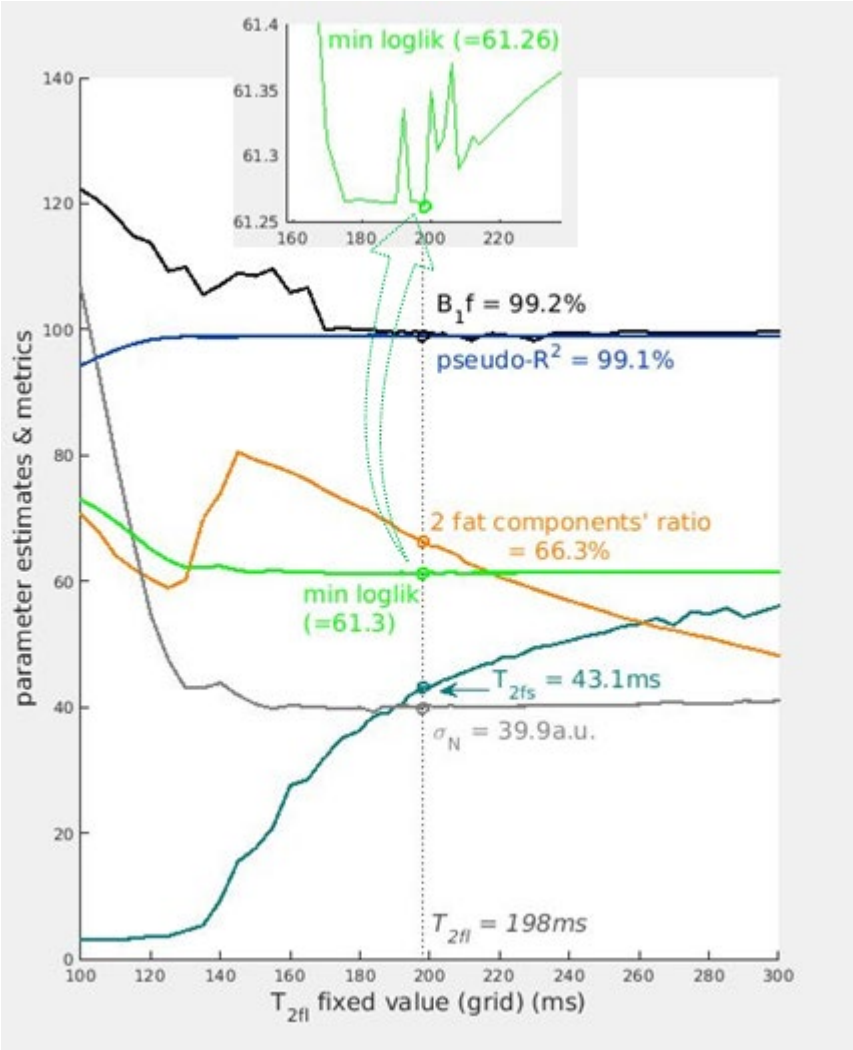


Figure 5.2: A numerical evaluation of the discrete parameter testing method for the ROI of Figure 5.1, showing that the lowest loglik corresponds to 198ms T_{2fl} thus determining the choice of the rest of the parameter estimates used as fixed values in the fat 2-component model.

These results corroborate the choice of 198ms as the T_{2fl} value corresponding to minimum loglik, leading to the optimal remaining parameter values for the fat 2-component model shown above. As can be seen in the inset of Figure 5.2, while the loglik vs. T_{2fl} dependence shows a clear minimum at 198ms it is still a smoothly changing function: this is exactly the reason of taking negative logarithms of the pdf to be minimised (section 2.2.3). Pseudo-R^2 and σ_N also appear to be smoothly varying – also due to the plots of the other parameters in the same graph (y-axis scaling) –

suggesting a sensible minimisation, despite the use of discrete T_{2fl} values. Good quality 2-component fits to the subcutaneous ROI fat signals were thus obtained, despite the age-range and anatomical differences, and lower data quality in general, compared to datasets examined previously, e.g. from the ALS and KD patient study presented in Chapter 4 and the phantom experiments, as mentioned in section 3.1. Interestingly, loglik was here in the same range (between 50 and 70 units) as in the initial experiment in section 3.3.2, which given the non-normalised nature of loglik – since it expresses the pdf utilised for MLE minimization, in general an "unnormalised probability" (Myung, 2003) – supports the suggestion of a good quality of fit, correlated with high pseudo-R^2.

5.3.2 Multi-component model fitting on the DMD data

The next stage was to perform the equivalent overall combined muscle-fat component model testing for subcutaneous fat ROIs as implemented before in section 3.3.3, in order to examine the role of the T_{2m} lower bound in the interpretation of the parameter estimates. The fitting process as described for fat tissue in the previous section was repeated for the same ROIs, this time using the multi-component model expressed by Equation 14, Chapter 3, for 3 different values of the T_{2m} lower bound, namely 5, 10 and 15ms. In this case the parameters to be determined for each pixel are effective muscle water T_2 (T_{2m}), fat fraction (ff), B_1 field factor (B_1f), overall amplitude (α) and noise standard deviation (σ_N). T_1 for muscle and fat were fixed, the fat compartment T_{2fs}, T_{2fl} and the 2 fat components' amplitude ratio as above determined, and the remaining sEPG model parameters (*TE*, ETL, RF

excitation and refocusing profiles and duration vectors) were set to the values given in Figure 5.1.

The purpose of this exercise was, by applying the full signal model including the water compartment to ROIs containing essentially only fat, to characterise and thus distinguish on the basis of the fitting algorithm outputs, pure fat from fat infiltrated muscle. This was necessary since the established ('s_{fat}') fat model has been seen to be interchangeable (in terms of quality of fit) with the multi-component model when *ff* approaches 100% and T_{2m} is equal or below 15ms. This experiment produced the results presented in Table 5.2.

Table 5.2: Dependence of the overall multi-component sEPG-MLE model parameter estimates on the T_{2m} lower bound, set successively at 10, 15 and 20ms, when fitting subcutaneous fat ROI data. T_{2m} estimates coincide with the bound value in most cases. Setting the T_{2m} lower bound at 10ms (results in bold) appears to be a good compromise for fat signals identification and improved quality of fit, as shown by the corresponding loglik, σ_N and pseudo-R^2.

	B_1f (%)	T_{2m} (ms)	ratio (%)	T_{2fl}, T_{2fl} (ms) *fixed*	σ_N	pseudo-R^2	loglik
Patients	108.6	5	100		35	99.4	60.2
	109.7	**10**	**100**	**40 198**	**35.2**	**99.4**	**59.7**
	108.3	15.1	100		35.6	99.4	59.8
	98.4	5	94.3		22	99.7	54.1
	99.9	**10**	**98**	**40 198**	**21.6**	**99.7**	**53.9**
	99.5	15	98.8		20.6	99.7	53.3
Healthy volunteers	110.3	5	93.4		45.7	98.8	62.9
	109.3	**10**	**100**	**40 198**	**43.2**	**98.8**	**62.2**
	107.9	15.1	99.7		42.7	98.9	62
	79.7	5	76.2		27.5	99.5	56.8
	83.5	**10**	**96.9**	**40 198**	**20.9**	**99.7**	**53.7**
	84.7	15	98.5		20.3	99.7	53.1

The apparent T_{2m} estimate returned for pure fat regions mostly coincides with (or is restricted by) its lower bound, acting again as a criterion to identify them. The

experiment once again showed that raising the T_{2m} lower bound upwards (from 5ms to 10 or 15ms) improves results, in terms of both physical meaningfulness of the parameter estimates for B_1f (smaller departures from 100%), *ff* (dominance of the fat component as expected) and σ_N (lower levels), and in terms of goodness of fit via pseudo-R^2 and loglik. On the other hand there is no clear superiority of using 15ms as a lower bound compared to 10ms, plus the former (15ms) would potentially restrict T_{2m} estimation to suboptimally or undesirably high levels. Thus setting it finally at 10ms (as also concluded in Chapter 4) appears to be a good compromise between effectively identifying fat regions, optimising quality of fit and more importantly ensuring the efficiency of the overall model for its main objective, the estimation of muscle water T_2 and muscle fat content.

5.3.3 Post-fitting quality control

Next, two example data sets from this study were analysed in detail to better understand aspects of the fitting process under these specific experimental conditions. Scatter plots representing all pixels in the segmented muscle areas are shown for each case in Figure 5.3. Both high fat or pure fat content pixels, and pixels from regions of sub-optimal data quality (due to e.g. flow or motion artefacts, B_1 inhomogeneities, signal drop-outs, segmentation imperfections) tended to return T_{2m} estimates falling in the 10 to 11ms T_{2m} range (top panels). However of these only the latter poor data-quality subgroup yielded poor quality of fit as indicated by decreased pseudo-R^2, irrespectively of their associated *ff* (middle and lower panels).

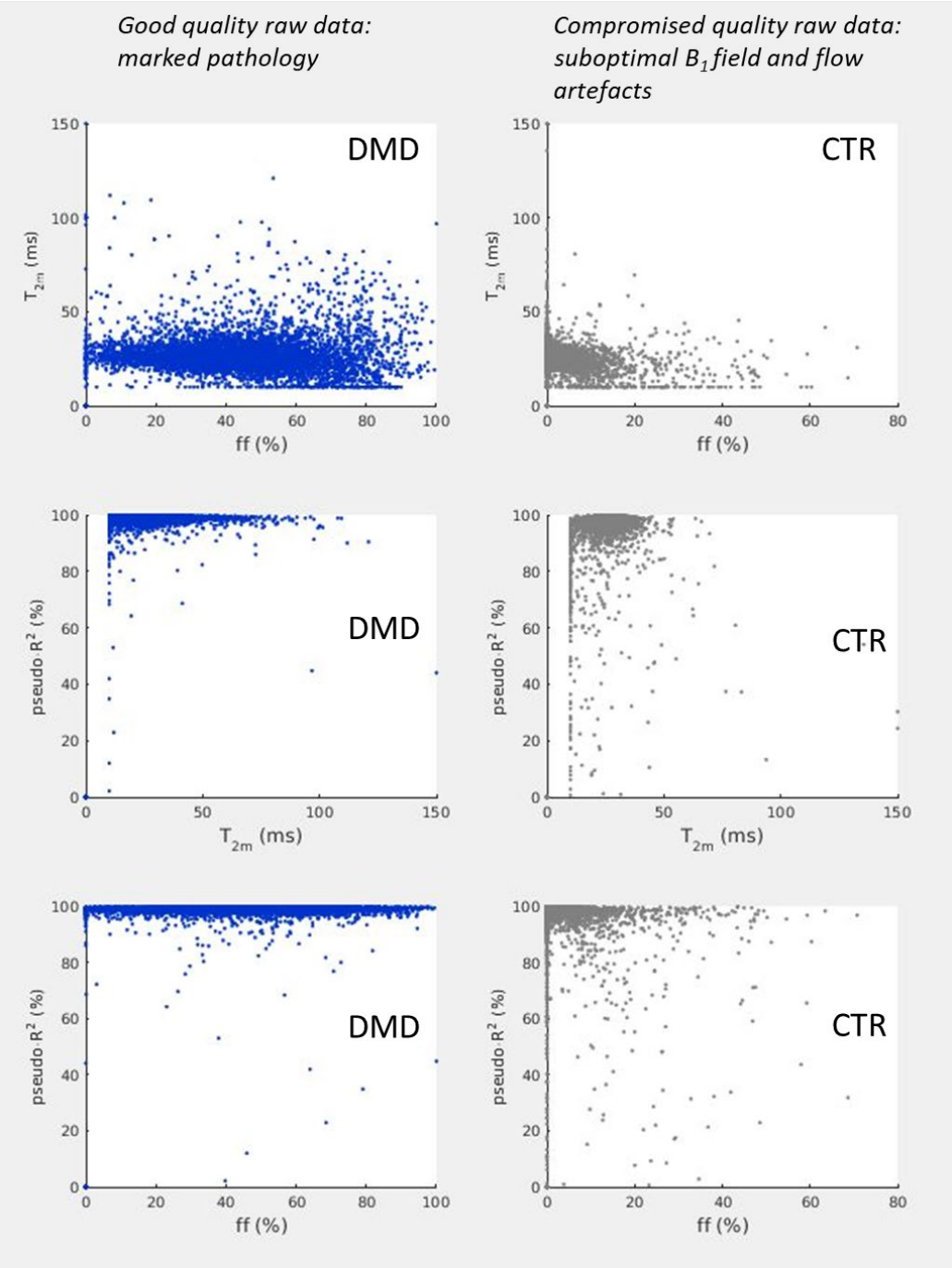


Figure 5.3: The correlation plots for T_{2m} vs *ff* (1st row), pseudo-R^2 vs T_{2m} (2nd row) and pseudo-R^2 vs *ff* (3rd row) estimates for a good quality patient dataset (DMD) and for a compromised quality healthy volunteer (CTR) dataset. T_{2m} is broadly stable with increasing *ff* however on many occasions the minimisation has reached bound constraints, mostly the 10ms lower T_{2m} bound. It can be seen that the pseudo-R^2 distribution shifts towards lower values for the artefacted data set (middle image 2nd column for the healthy volunteer) but does not show a strong dependence upon T_{2m} or *ff* changes caused by pathology (left column, DMD patient).

In contrast, for the good quality patient dataset estimates of T_{2m} at or close to the 10ms lower T_{2m} bound were returned mainly for high *ff* pixels, and were not

associated with decreased pseudo-R^2. These results are consistent with the findings reported in section 4.3 with regards to the role of the T_{2m} lower bound in the parameter estimation and interpretation. This distinction is demonstrated more clearly in Figure 5.4, where for the patient (good quality) dataset in Figure 5.3, the group-T_{2m} estimates at 10 to 11ms corresponds predominantly to high fat content pixels, whereas the respective group for the sub-optimal quality healthy volunteer data corresponds to regions of data of poor quality.

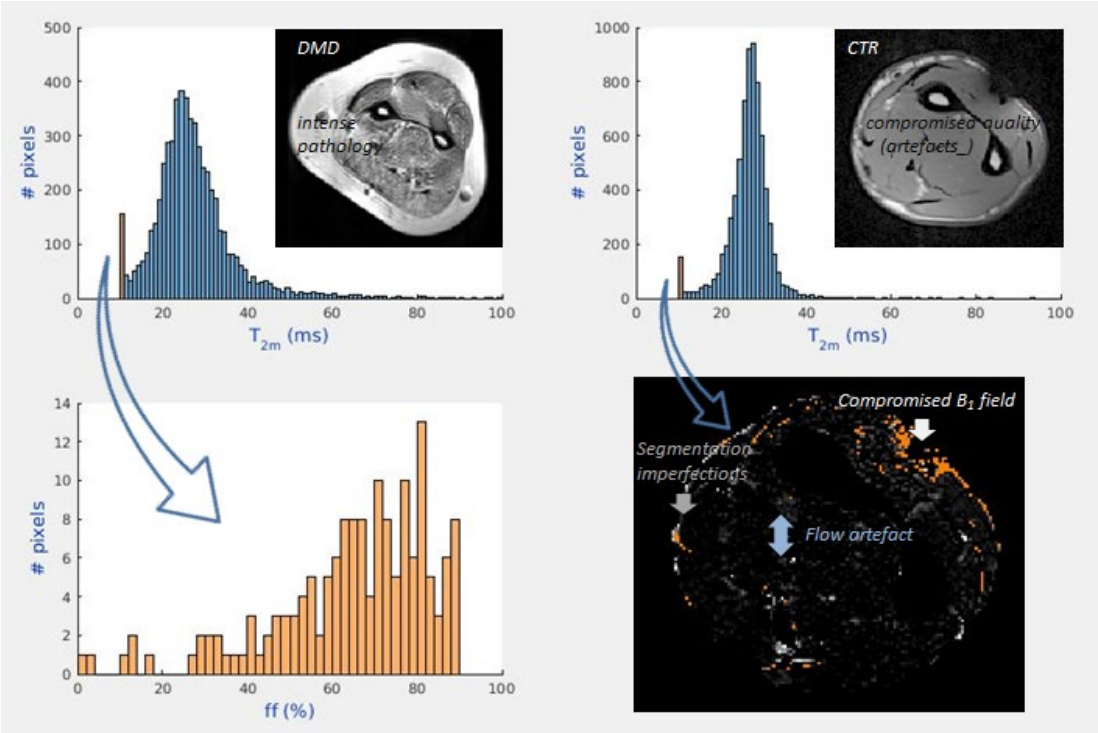


Figure 5.4: The effect of excluding pixels near the 10ms lower T_{2m} bound (10<T_{2m}<11ms) in the 2 example datasets of Figure 5.1, for two subgroups of interest: left (DMD patient), from fat pixels where such values represent in reality an additional fat component with short T_2 as explained in the text in section 5.3.2, and right (healthy volunteer, CTR), from compromised quality data due to artefacts (*ff* map, bottom left). In both cases the previous thresholds for T_{2m}<15ms and *ff*>90%, and pseudo-R^2<80% have already been used (note: the T_{2m} histograms only include pixels from the segmented muscle ROIs).

Taking together these observations, it follows that an additional post-processing step eliminating pixels of pseudo-R^2 below 80% and of T_{2m} below 11ms – as also proposed

in section 4.3 for the ALS and KD data – can be useful in eliminating such unwanted 'pure fat' or heavily artefacted regions from the analysis.

Finally *ff* estimates should be also corrected for the sequence specific TR-dependent T_1 recovery factor difference between muscle and fat – in addition to the necessary MT correction (section 3.3.5). For the CPMG protocol used here TR was 3000ms (section 5.2), therefore *ff* should be corrected by the expression $ff/(1.1326-0.1326 \cdot ff)$ (Table 3.5). The workflow of the final fitting tool pipeline used to analyse the CPMG data acquired in this study is shown in Figure 5.5, including the specified settings discussed in the preceding sections, as well as the $B_1 f$ polarity spatial regularization strategy developed in Chapter 3.

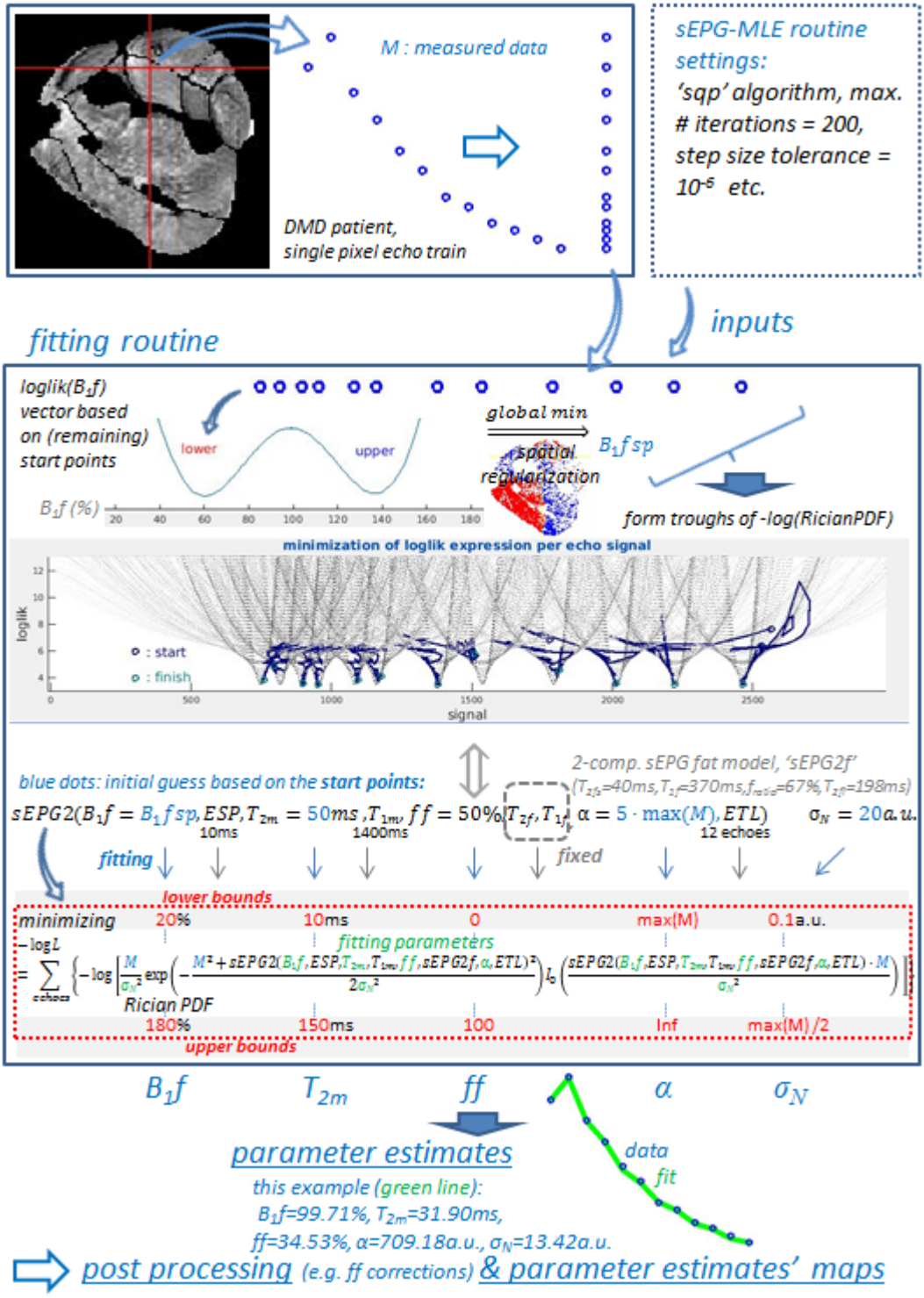


Figure 5.5: A scheme illustrating the implementation of the sEPG-MLE fitting algorithm used for the DMD study data, including start point and bound constraints specifications and algorithm settings.

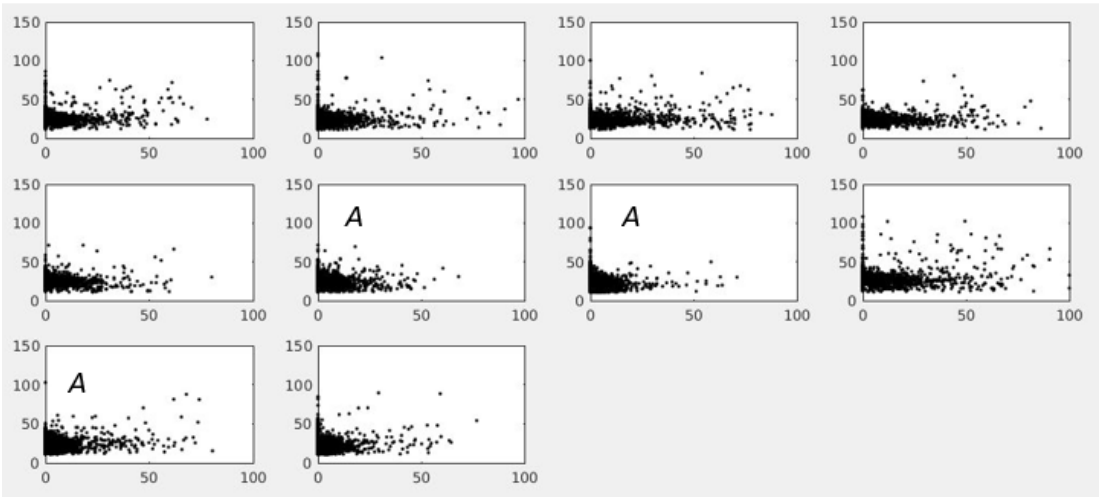
The basic steps of identification and exclusion of fat pixels (i.e. T_{2m}<15ms & *ff*>90% threshold values) are also included as a quality control step at the end of the fitting

routine. The algorithm is otherwise implemented in exactly the same way as in Chapter 4 (Figure 4.2) except for accounting for the different acquisition protocol parameters, notably the echo train length (ETL) and TR.

5.4 Study analysis results I

The parameter estimate distributions for each participant in terms of T_{2m} vs *ff* plots obtained using the fitting procedure and final post-processing steps described in the previous section are shown in Figure 5.6.

Note: Similarly to Chapter 4, the term *ff* is used in general for the sEPG-MLE estimated or 'apparent' fat fraction. This is termed ff_a for comparisons with Dixon ff_d results (section 5.5.2).



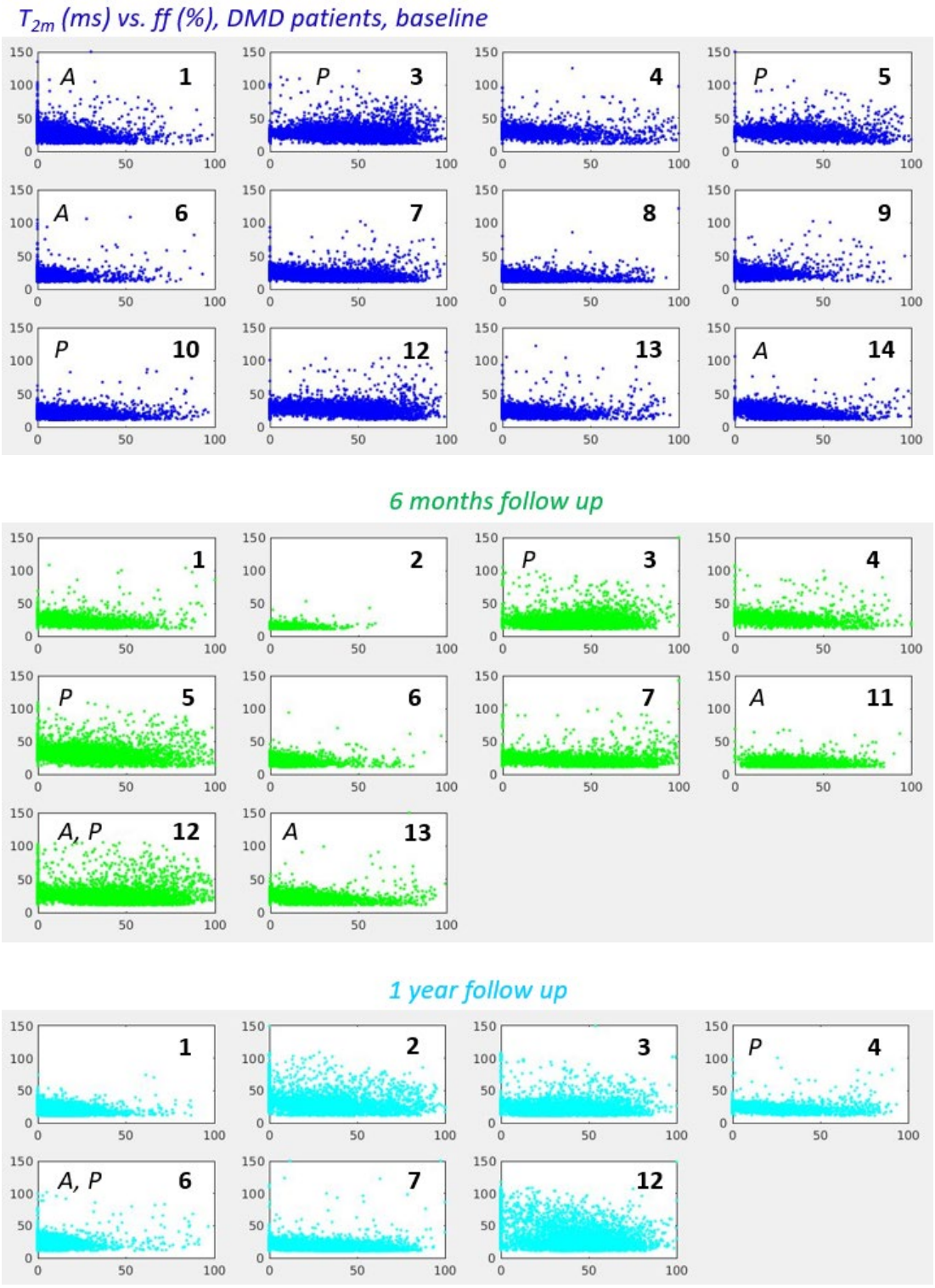


Figure 5.6: Overview of the estimated muscle T_{2m} (0 – 150ms) vs *ff* (0 – 100%) for the segmented maps for DMD patients at baseline (top panel), 6 months (2nd panel from top) and 1 year (3rd panel from top) follow-up scans, and for healthy controls at baseline (top panel). Letter A denotes datasets with clearly visible artefacts (mainly inhomogeneous B_1 field / low SNR, and motion) and P denotes datasets showing the most intense pathology. Integers serve to identify each patient at successive time points.

Shown are scatter plots of T_{2m} against *ff* for all pixels in muscle compartment segmented regions in a single forearm image slice. Considering first the healthy control data sets, T_{2m} largely lay in the range 10-40ms, while *ff* was restricted for the most part to below 20%. In the patient data, marked differences in the *ff* distributions were visually apparent between individuals apparently mildly or more severely affected, based on their imaging appearances. T_{2m} appears to be distributed largely independently from *ff*, exhibiting larger dispersion for higher *ff* ranges associated with more advanced disease, while some mild T_{2m} dispersion can also be seen for low to middle range *ff* likely associated to compromised image quality. There is no clear distinction between these subgroups since T_{2m} changes along the *ff* range appear rather smooth and continuous. There is consistently a similar number of artefacted datasets in the control and patient baseline and follow up groups (3 or 2 for each group, 20-25% of the total number).

Examples of T_{2m} and *ff* maps and their respective histograms from representative datasets are shown respectively in Figure 5.7 and Figure 5.8. Comparing T_{2m} maps from good quality source data, the patient examples showed substantial regions of elevated T_{2m} compared with the control examples. Regions of elevated T_{2m} in patients mostly overlap with regions of elevated *ff*. Again, some increase in the number of pixels at the lower end of values (in the histograms) for both parameters is seen in cases of sub-optimal source data quality.

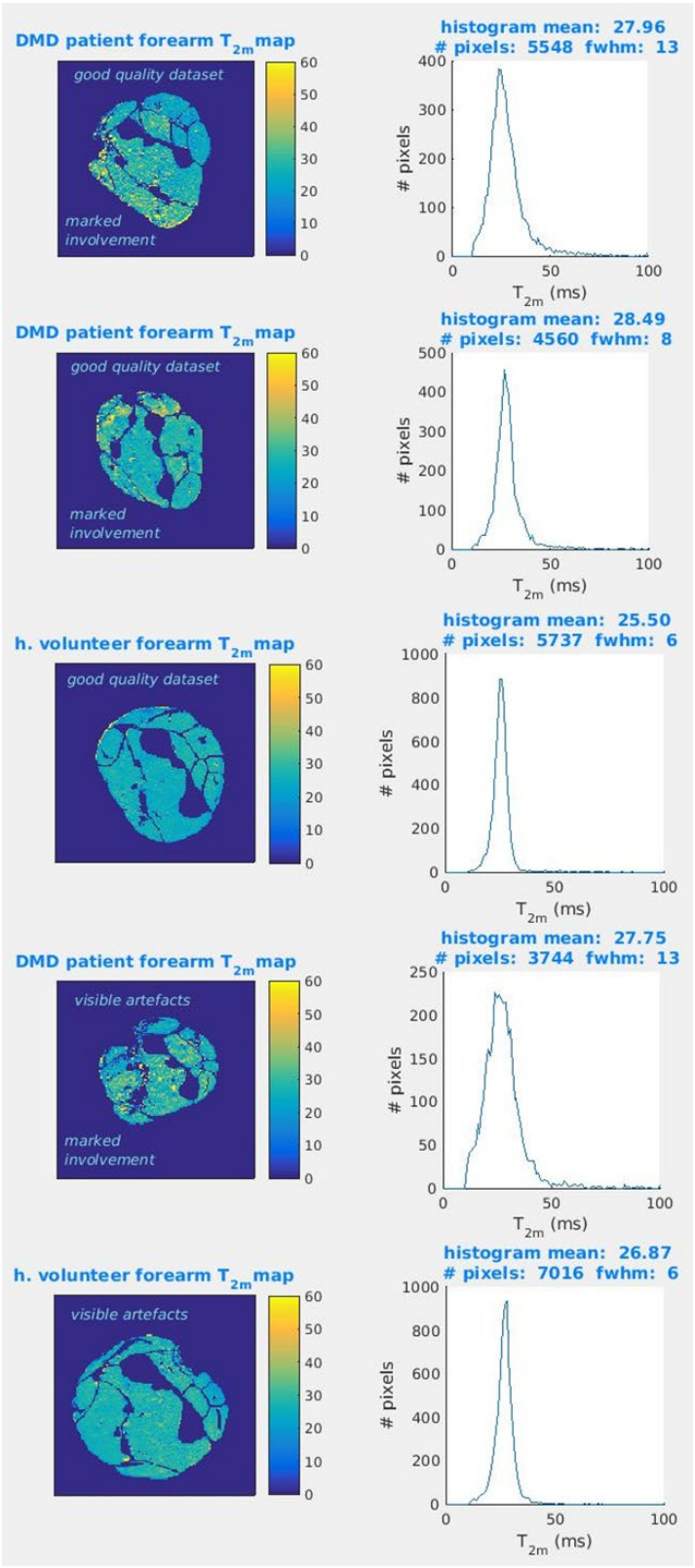


Figure 5.7: Examples of T_{2m} maps and histograms for representative example data, from patients with good image quality and marked pathological involvement (1st and 2nd row from top), a healthy control with good image quality (middle row), a patient with suboptimal image quality and marked pathological involvement (2nd from last row) and from a healthy control with suboptimal image quality (last row).

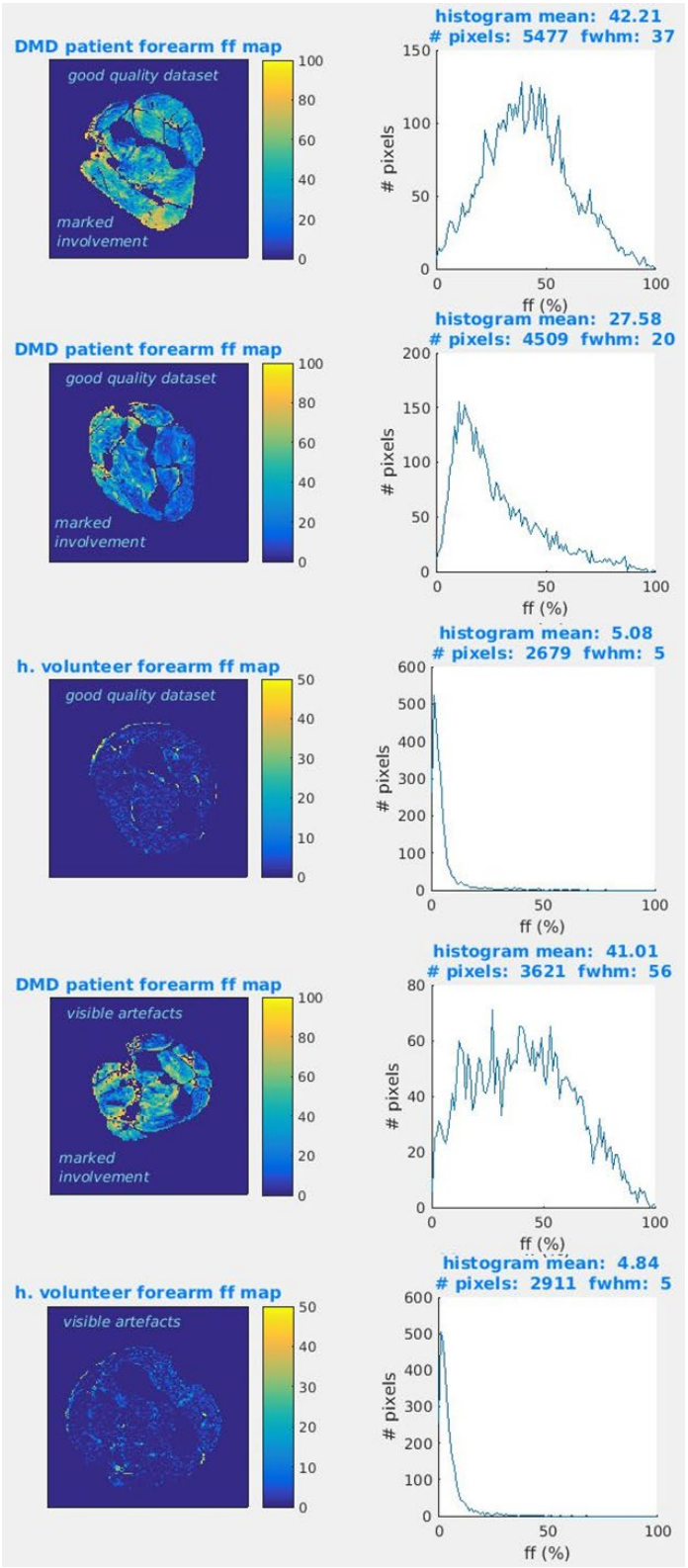
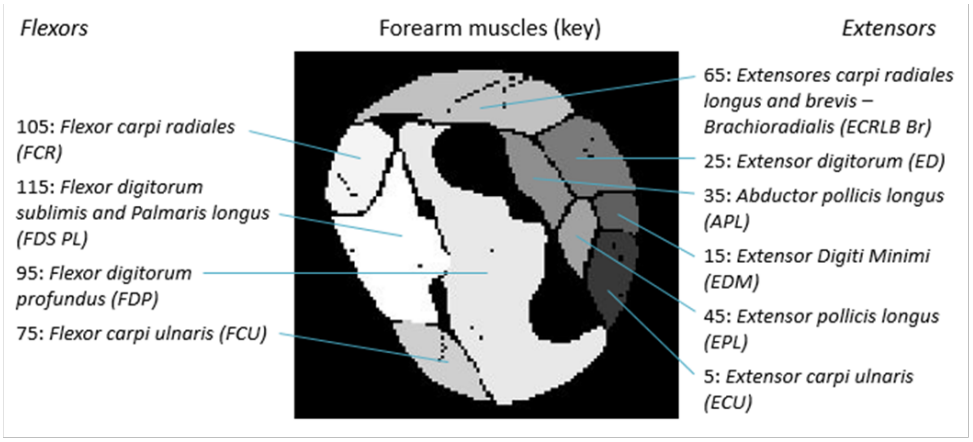


Figure 5.8: Examples of *ff* maps and histograms for representative example data, from patients with good image quality and marked pathological involvement (1st and 2nd row from top), a healthy control with good image quality (middle row), a patient with suboptimal image quality and marked pathological involvement (2nd from last row) and from a healthy control with suboptimal image quality (last row).

T_{2m} and *ff* boxplots per individual muscle ROI are shown for all participants in Figure 5.9 (T_{2m} for healthy controls, including the muscle ROI key), Figure 5.10 (T_{2m} for patients) and Figure 5.11 (*ff* for all groups).



T_{2m} (ms) per muscle ROI, healthy controls, baseline

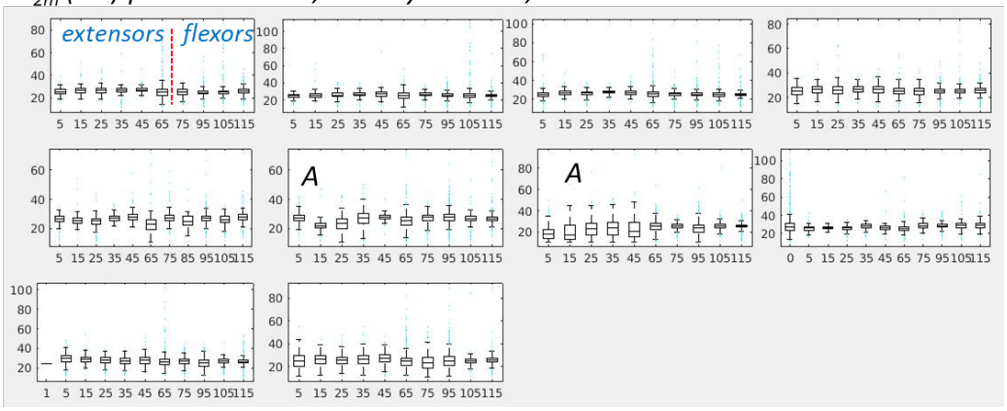


Figure 5.9: Overview of the estimated T_{2m} boxplots for the individual muscle ROIs for healthy controls at baseline; 10 or 11 muscles ROIs per subject are shown, from extensors (left) to flexors (right). The numbers along the x-axes identifying the individual muscles according to the key at the top were as specified by the segmentation experts (section 5.2). Outliers are identified by the cyan dots. While T_{2m} lay mostly between 25-30ms (normative range) differences between individual muscles also exist.

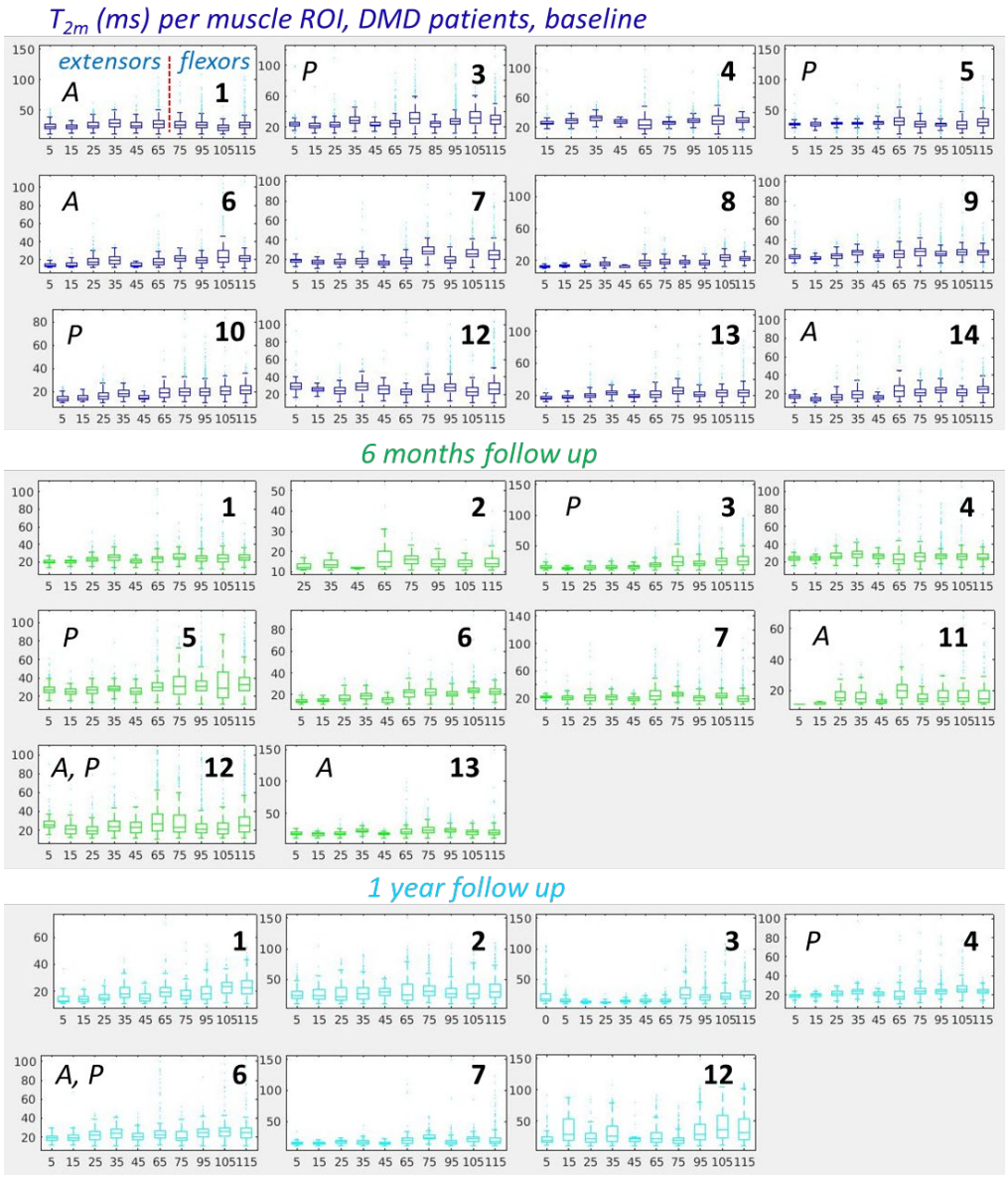


Figure 5.10: Overview of the estimated T_{2m} boxplots for the individual muscle ROIs for DMD patients at baseline and at 6 months and 1 year follow-up; 10 or 11 muscles ROIs per subject are shown, from extensors (left) to flexors (right). Outliers are identified by the cyan dots. Individual muscles do not seem to be affected by disease all in the same manner in terms of both T_{2m} median values and distributions.

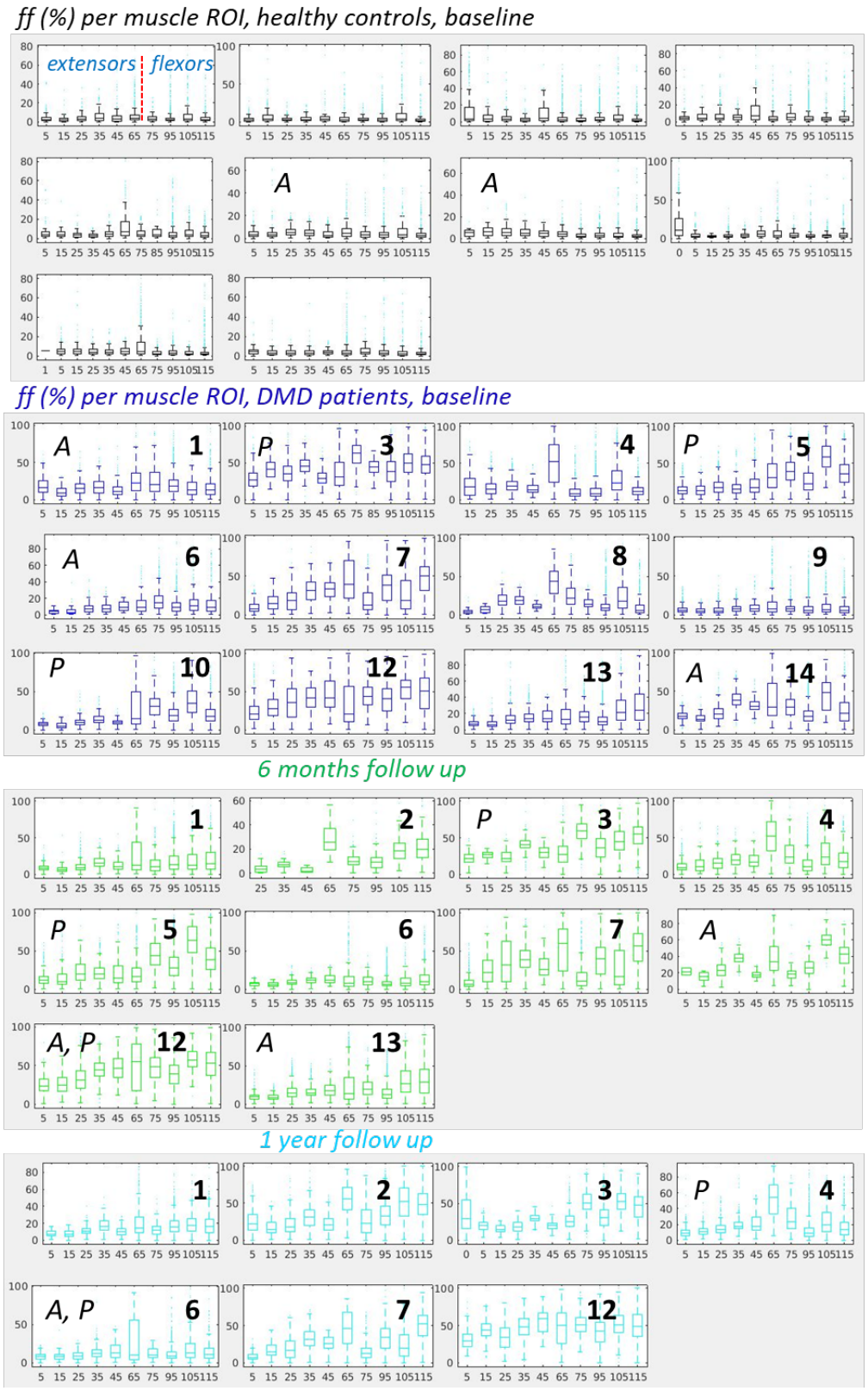


Figure 5.11: Overview of the estimated *ff* boxplots for the individual muscle ROIs for healthy controls (1st panel) at baseline and for DMD patients at baseline and at 6 months and 1 year follow-up scans (2nd, 3rd and last panel); 10 or 11 muscles ROIs per subject are shown, from extensors (left) to flexors (right). Outliers are identified by the cyan dots. Individual muscles do not seem to be affected by disease in the same manner in terms of both *ff* median values and distributions

Clear variation is observed between different regions of the forearm cross section in patients, predominantly between flexors and extensors as compartments. There is evidence for more marked T_{2m} and *ff* elevation in flexors than in extensors in general.

This variation becomes masked to a certain extent when the entire forearm cross-sectional T_{2m} and *ff* boxplot comparisons are displayed (Figure 5.12).

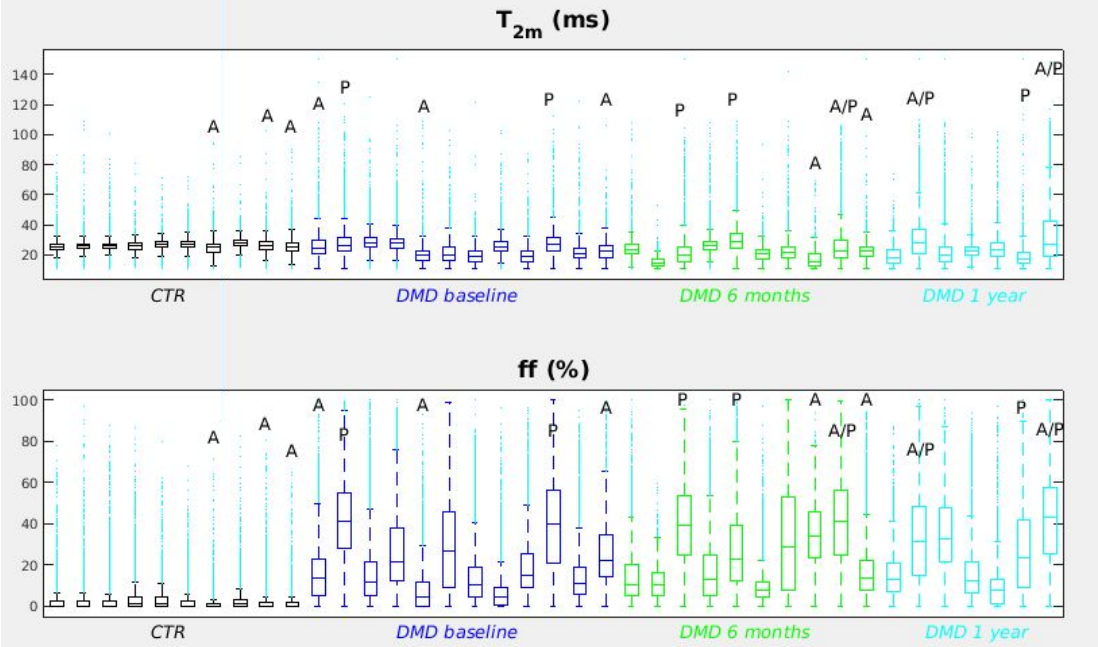


Figure 5.12: Boxplots of the estimated *ff* (0 – 100%) and T_{2m} (0 – 150ms) from all muscle groups combined across the forearm cross-section for DMD patients longitudinally at baseline (blue), 6 months (green) and 1 year (cyan) follow-up scans and for healthy controls (CTR, black); outliers are represented by the cyan dots. Letter A denotes datasets with visible artefacts, P denotes marked pathological involvement and A/P datasets with both in the same DMD subject. Forearm cross-sectional results follow patterns in parameter elevation and dispersion due to pathology broadly consistent with those seen in individual muscles in Figure 5.9, Figure 5.10 and Figure 5.11.

Each boxplot here represents the total distribution of T_{2m} and *ff* estimates from all forearm muscle pixels for each individual, and at the different time points for the patients. For healthy controls T_{2m} and *ff* are tightly clustered between 25-30ms and 0-5% respectively. For patients T_{2m} and *ff* appear to change in opposite directions

with T_{2m} mainly decreasing and *ff* markedly decreasing. Distributions also appear wider for patients, with T_{2m} showing a tendency to continue dispersing longitudinally. More detailed comparisons will be presented with the revised analysis results in section 5.6.

5.5 Data quality and atypical CPMG signal behaviours

5.5.1 Signal dropouts and motion artefacts

In Figure 5.13 median T_{2m} values (without whisker and box plots) are shown per muscle for all individual subjects – including follow up scans; and in Figure 5.14 the respective equivalent estimated median *ff* values are shown. One healthy control dataset, indicated in Figure 5.13, was excluded from the subsequent analysis due to regions of severe signal drop-out causing complications in the interpretation of those results.

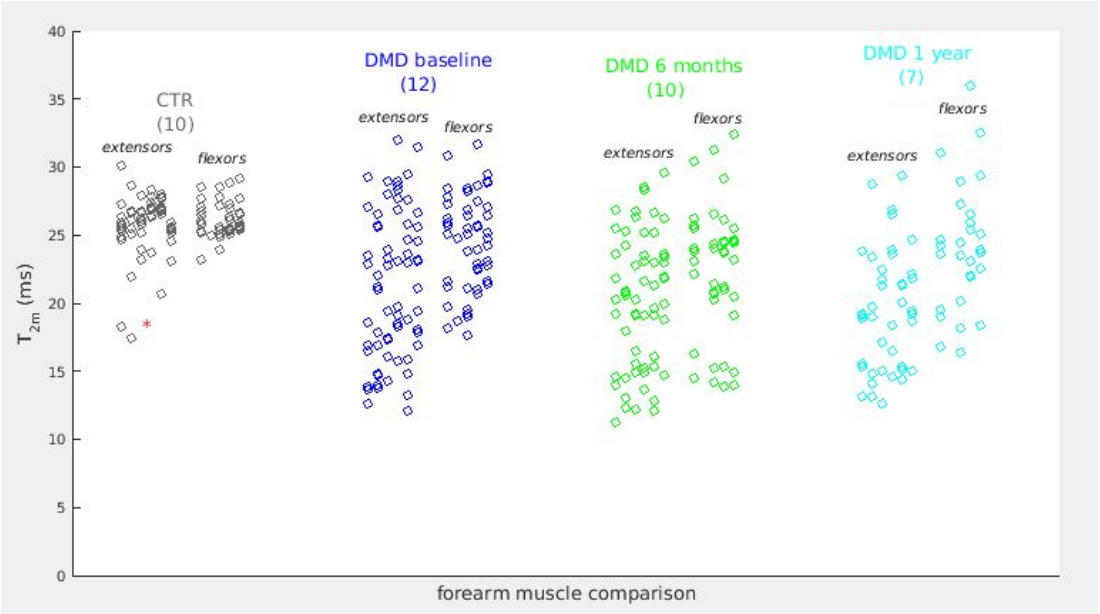


Figure 5.13: Individual forearm muscle median T_{2m} for the controls (CTR) and patients (DMD) at baseline, and the DMD group at the follow-up time points. In brackets are the number of subjects for each group and time point. Note: The CTR data points indicated with an asterisk (*) are from an individual dataset manifesting severe signal drop-out (SNR < 10) and low SNR, where although the MLE parameter estimation performed reasonably well, there was a clear data quality influence on the distribution and interpretation of these results, which were therefore excluded from the subsequent analysis.

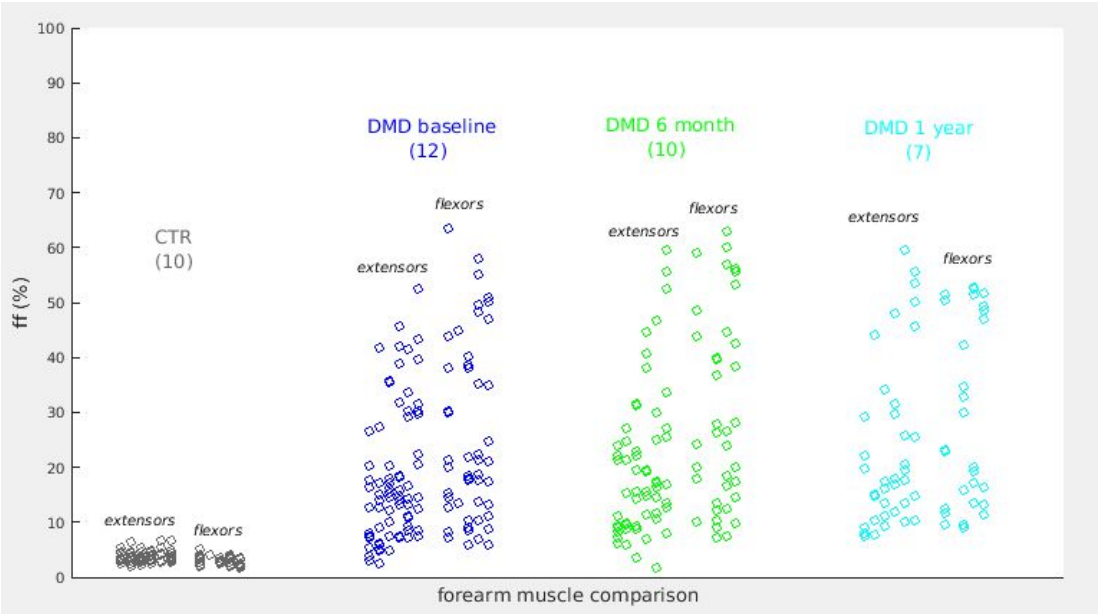


Figure 5.14: Individual forearm muscle median *ff* for the controls (CTR) and patients (DMD) at baseline, and the DMD group at the follow-up time points. In brackets are the number of subjects for each group and time point.

Similarly there were occasional data quality issues in the DMD group, which were however much milder and did not appear to affect the estimates markedly. This is indicated in Figure 5.15 where data points identified as potentially influenced by motion or $B_1 f$ artefacts are seen to be consistent with the general T_{2m} distributions, suggesting it is unlikely that in these cases the data quality introduced parameter estimation bias.

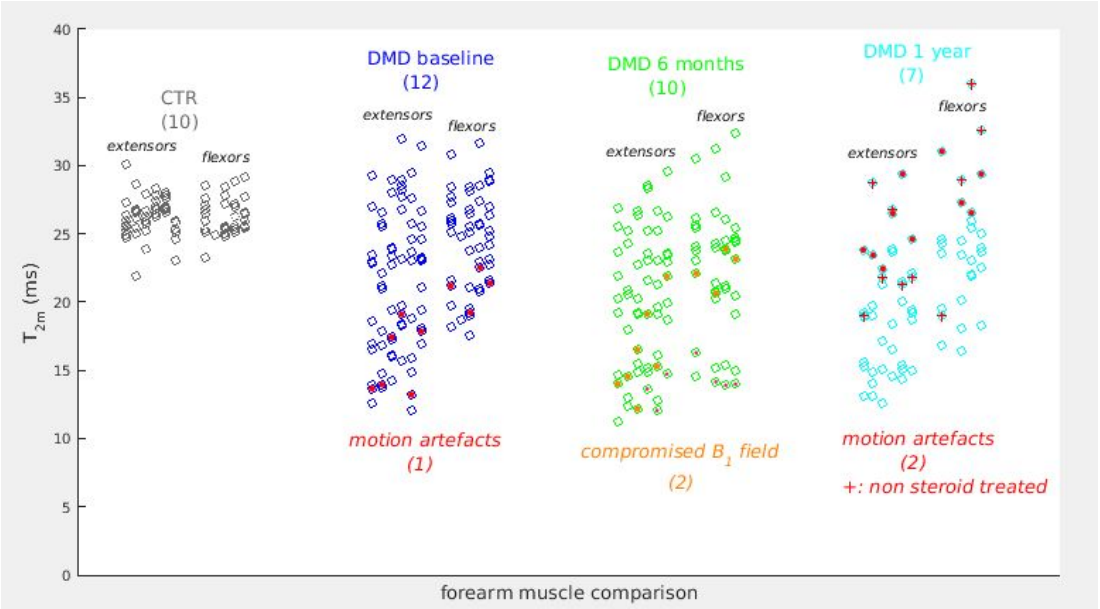


Figure 5.15: Examining the effect of suboptimal raw data quality in the DMD group: points with a red dot were estimated from datasets with apparent motion artefacts and those with an orange dot from data with suboptimal B_1 distributions. These appear distributed across the entire range of T_{2m} and thus are unlikely to introduce bias. Datapoints indicated with a red cross ('+') are from the only patient that was not steroid treated. In brackets are indicated the number of subjects for each group/timepoint.

A possible exception is one of the two mildly artefacted datasets at the 1 year follow up, where in this case high T_{2m} estimates appeared to result. In fact however this data was from the only subject in the group that was not steroid treated, and thus it is also possible that this was a genuine T_{2m} elevation due to alternative factors (uncontrolled oedema).

5.5.2 Atypical CPMG signal behaviours seen in this study data

An additional complication in this study is the effect of acquiring data at the periphery of the magnet bore. This was unavoidable since, for reasons of patient comfort it was necessary that the participants' forearms were positioned to the side of the abdomen for imaging, parallel to the body, with a flexible surface coil wrapped around them. This may be a contributing cause of the unexpected signal behaviour seen in wide areas of several datasets (Figure 5.4). In these regions spin-echo trains with signal intensity rapidly diminishing from the first to the second echo were obtained, often returning T_{2m} estimates well below the expected range for muscle water (i.e. < 20ms). Recalling the basic concept of stimulated and alternate echoes characterizing the shape of CPMG sequence signals, such that the second echo amplitude commonly exceeds the first (Chapter 2, Figure 2.30), and that common experimental errors, e.g. erroneous B_1f intensity reinforce this effect, suggests that this entirely different signal behaviour in these regions might be caused by an additional experimental confound associated with off-centre scanning.

In Figure 5.16 examples of such fast decaying signals from the indicated image areas are shown together with sEPG model parameter estimates – manifesting short estimated T_{2m} and low *ff*. For these pixels although the returned B_1f estimate was 100%, the model clearly cannot adequately describe the initial signal collapse (i.e. the situation where the second echo amplitude height is markedly less than the first). This is because even in the case of overall ideal B_1 amplitudes, unavoidable variation in flip angles across the slice profile will generate additive signal from the slice edges and side lobes after the second echo, causing the total signal to be modulated with

respect to the ideally exponential (for $B_1 f$≈100% as in this case) decay from the middle of the slice (section 3.2.5, Figure 3.22).

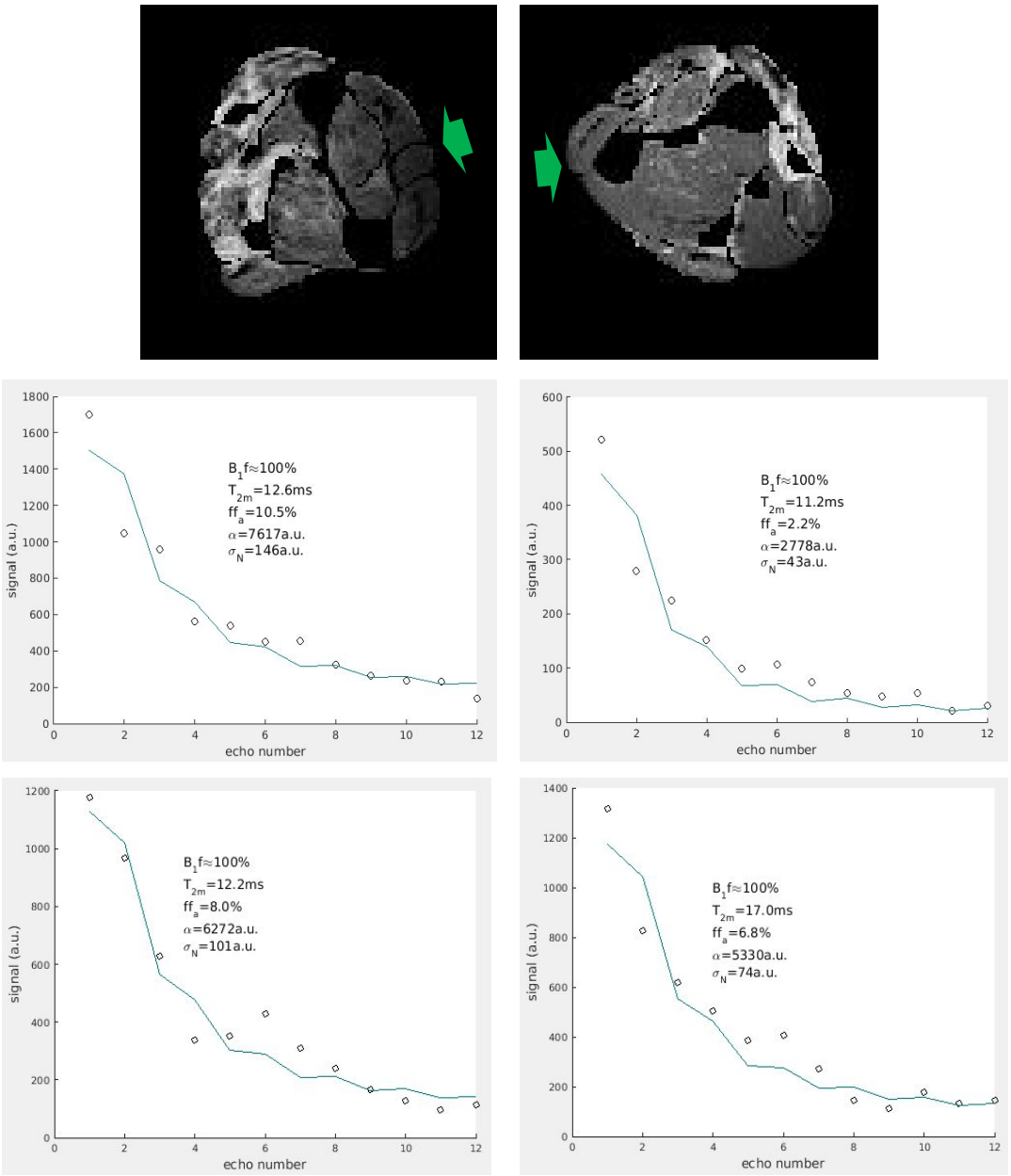


Figure 5.16: Top: example raw images where considerable areas (green arrows) of rapidly decaying signal intensity was observed. Middle: CPMG echo amplitudes from representative pixels showing marked modulation of the echo train with lack or reversal of the expected stimulated echo effect (1st < 2nd echo magnitude). Bottom: lack or severe alteration of both the stimulated echo and the alternate echoes effect. The parameters listed in each case are estimates returned by the sEPG-MLE algorithm assuming correct gradient waveforms. This behaviour was observed for many pixels returning short T_{2m} estimates.

Since the position of the target anatomy was at the edge of the scanner magnet bore, we may speculate that non ideal magnetic field properties in this region may influence the signal evolution. This could cause the gradient amplitudes, timing and linearity experienced by the spins to differ from those specified in the pulse sequence design. A full simulation of these effects is beyond the scope of this work; however, to test the plausibility of this as explanation of the signal behaviour, the effects of an incorrect ratio between the magnitudes of the gradients associated with the excitation and refocusing RF pulses, will now be briefly explored. Figure 5.17 shows sEPG simulations of the echo trains for 3 different ground truth parameter sets where the magnitude of the refocusing RF pulse slice select gradient was changed to values below and above the sequence design value (algorithm, settings and post-processing, as in section 5.3). In Figure 5.17 panel (a), the examples with T_{2m}=25ms clearly show an elevated amplitude of the second echo relative to the first, and alternating echo amplitudes along the echo train, whereas in the case of the lowest T_{2m} and *ff* combinations (11ms and 0% respectively, bottom echo train) observed in practice, the latter effect was maintained, while the *TE*1 and *TE*2 echo amplitudes where reversed.

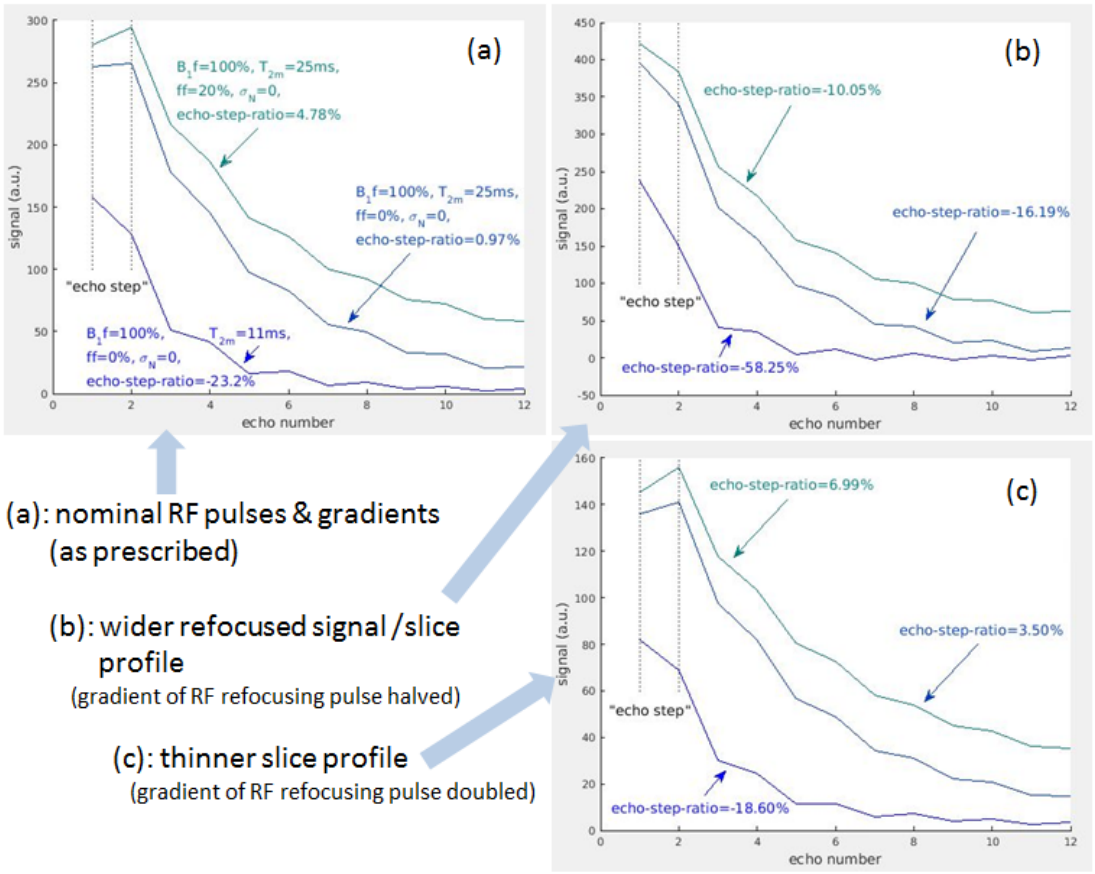


Figure 5.17: The theoretically expected effect of altered gradient amplitude ratio between excitation and refocusing pulses for typical sEPG model parameters encountered in this study data. The "echo-step-ratio" is the fractional difference (%) between the second and first echo amplitudes. A wider than expected slice profile due to departures in the prescribed gradients may be a plausible reason for the exacerbated decay apparent in parts of several images: a high degree of similarity is seen between the signals of Figure 5.15 and panel (b) of Figure 5.17

However, it might be that the slice width corresponding to the refocusing RF pulse is effectively much wider potentially driven by gradient non-linearity towards the periphery of the magnet bore the effects of which are demonstrated in Figure 5.17 panel (b), where for the T_{2m}=25ms signals the "echo-step-ratio" for the first two echoes is reversed, creating a signal very similar to that for the T_{2m}=11ms case in panel (a). The effect is reversed in the case of an effectively increased refocusing pulse gradient magnitude (panel(c)), with intensified stimulated and potentially also alternate echoes.

Therefore we conclude that an instrumental effect causing the situation in Figure 5.17(b), i.e. an effective decrease in the refocusing slice gradient magnitude with respect to the excitation pulse gradient, is indeed a physically plausible explanation for the signal behaviours seen in some regions of this study data. The ratio of these gradient amplitudes differs from unity by design, accounting for the excitation and refocusing RF pulses being different, both in amplitude waveform and duration. Any deviations from the prescribed gradient amplitudes and their ratio, caused e.g. by gradient field non-linearity, eddy current gradients, or B_0 non uniformity will result in slice selection distortions and echo train alterations. A number of other authors have highlighted the importance of the role of refocusing widths in T_2 relaxometry parameter estimation, in historical and recent publications (Lebel and Wilman, 2010, McPhee and Wilman, 2017, Keene et al., 2020). This was the conceptual basis also for the work presented here.

Alternative physical effects, in particular macroscopic and microscopic motion – including tremor and localised motor unit activation – may also be candidate mechanisms behind such signal alterations. The deviations in the echo heights seen in last 2 examples of Figure 5.16 – possibly due to consecutive dephasing effects – may be due to such mechanisms. On the other hand, noise does not seem to be a major contributor as SNR levels for such signals are reasonably high in general.

While the precise physical origin of these atypical signal behaviours remains unknown, because of their prevalence in this study data we aimed to better characterise such processes and assess their impact on the parameter estimation; distinguishing them from sub-optimal $B_1 f$ performance is also of interest.

The echo train shape dependence on factors which are not included in the standard sEPG model has direct consequences in the sEPG parameter estimation. While departures in $B_1 f$ are accounted for in the sEPG-MLE algorithm, signal deviations either caused by, or similar to those caused by slice profile deviations due to gradient non-linearity are not, and this poses challenges in interpreting the resulting parameter estimates. The rapidly diminishing 2nd *vs.* 1st echo, i.e. the reversal of the expected stimulated echo effect appears to be a 1st order effect when compared to the alternate echoes' and/or echo periodicity alterations which rather may be considered 2nd order effects. The signal height change in the first case can exceed 50%: commonly a 2nd echo amplitude would be 10-20% higher than the 1st, while in the observed collapsing signals it often rather becomes 50% smaller: a 60-70% net decrease; whereas further along the echo train alternate echoes typically show more limited amplitude changes (oscillations) up to an order of magnitude smaller. Importantly, these two regimes probably represent independent processes: collapsing signals don't necessarily lose their alternate echo behaviour, at least not entirely, as in the 2 middle examples in Figure 5.16.

Since it was the largest effect, it is therefore instructive to examine the T_{2m} and *ff* estimation dependence on the signal behaviour characterised in terms of the relative amplitude difference ("step") between the 2nd and 1st echo, defined for this purpose as the 'echo-step ratio', $\frac{echo\ 2 - echo\ 1}{echo\ 2} \cdot 100\%$.

This was initially investigated in simulation, with signals exhibiting this behaviour generated with forward sEPG models with varying excitation and refocusing gradient

amplitudes, accepting that the true origin of this signal behaviour may be different. Results shown in Figure 5.18 for a ground truth signal of 30ms T_{2m} and 10% *ff* – to simulate values typical of real datasets.

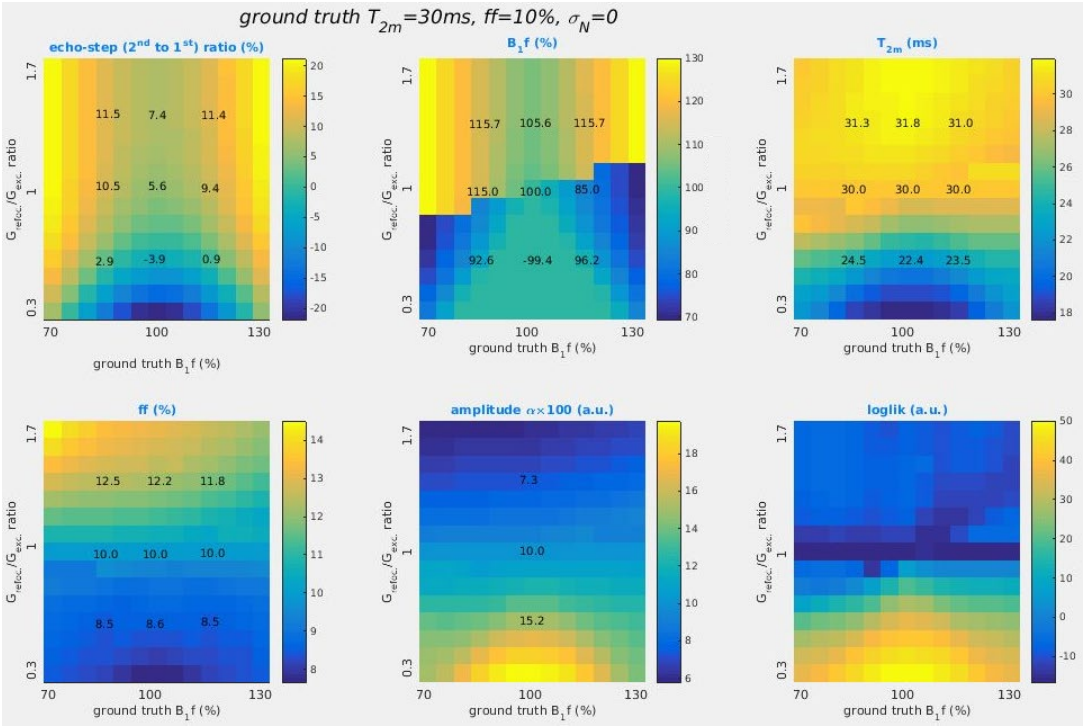


Figure 5.18: Dependence of the echo-step ratio (1st panel), the parameter estimates (2nd through 5th panel, excluding noise σ_N) and loglik on B_1f (x-axis) upon the refocusing *vs.* excitation pulse gradients' ratio (y-axis), for a noiseless ground truth signal with T_{2m}=30ms and *ff*=10%. Note: the G_{refoc}/G_{exc} ratio value of 1 corresponds to a normalised representation of the correct design value.

The sEPG-MLE algorithm was implemented as described in sections 5.2 and 5.3 and outlined in Figure 5.5, initially for noiseless signals. The RF excitation and refocusing profiles and duration vectors and the associated gradient pulses' amplitudes were set according to sequence and scanner settings, as provided by the vendor (section 2.2.3). The effect of fitting to signals with echo-step ratios different to those that would arise from the assumed ground truth signal model fixed parameters was tested by systematically varying the refocusing and excitation slice select gradient

magnitudes in the model generating the simulated data, while keeping the fixed parameters in the fitted model at their design values.

More specifically the correct (design specified) refocusing gradient amplitude was 0.7829 times the excitation gradient amplitude; this corresponds to the unit value (i.e. the prescribed value) for the refocusing and excitation gradient ratio ("G_{refoc}/G_{exc} ratio") in Figure 5.18. The echo-step ratio for the forward model, the parameter estimates (B_1f, T_{2m}, ff and overall amplitude) and loglik are plotted as functions of varying ground truth B_1f and the ratio of the forward model refocusing and excitation gradients amplitudes. Consistent with Figure 5.17, variation due, in this case, to incorrect slice selective gradient echo amplitude has a substantial impact on the echo train shape, exemplified by the 2nd to 1st echo step ratio (top left panel) and the subsequent fitting parameter estimates (middle 4 panels), and the goodness of fit as expressed by the loglik (last panel). This is evident from the features observed in the lower regions of the panels compared to the upper relatively more homogeneous parts. A typical feature is that the increased signal within the acquired slice for normalised $G_{refoc}/G_{exc} < 1$, leads to saturation of the B_1f estimate, consistently tending to 100% as shown in the lower part of the 2nd panel.

This is the probable reason that, as shown in Figure 5.19, image noise does not affect much the lower parts of the panels (normalised $G_{refoc}/G_{exc} < 1$) while it clearly has an impact in the upper parts. In this figure the same simulation experiment was repeated this time adding Rician noise to achieve an average SNR of 50 for all signals in the raw image (adjusted to their relative amplitudes). Importantly, for both experiments the lower parts of the B_1f, T_{2m} and ff panels – corresponding to

normalised $G_{refoc}/G_{exc} < 1$ – show markedly decreased values (appear darker) corresponding to significant underestimates *vs.* the ground truth T_{2m} and *ff* values. The echo-step-ratio also becomes negative (i.e. 2nd echo lower than the 1st) only across the lower part of the respective panel.

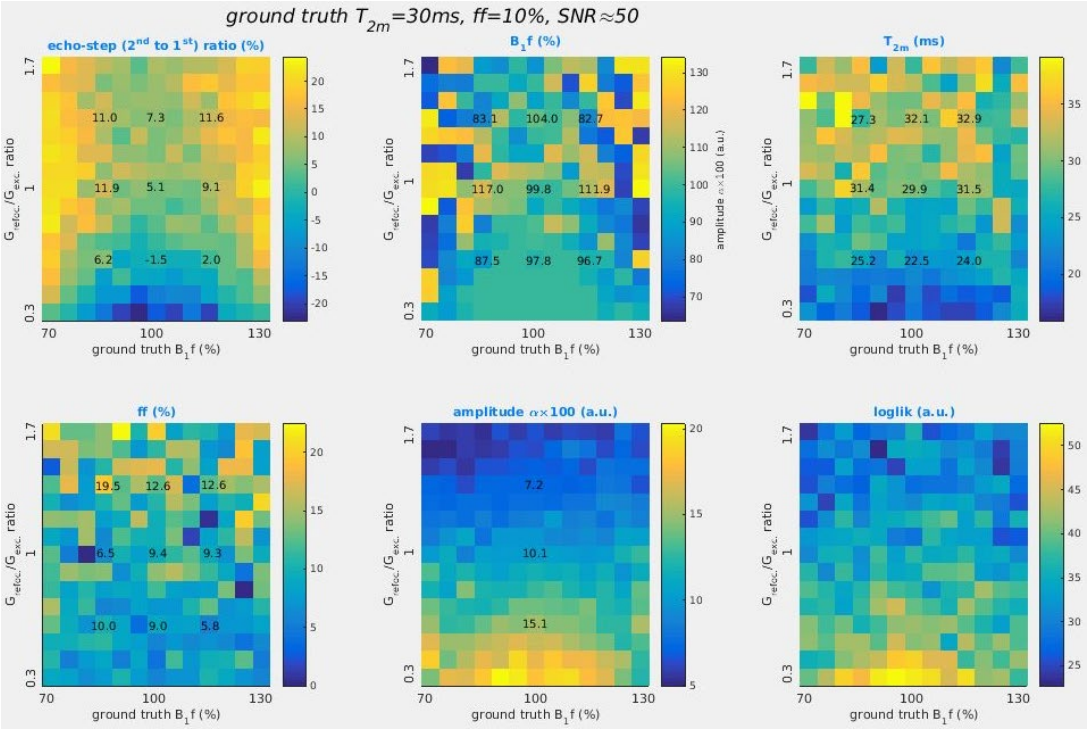


Figure 5.19: Dependence of the echo-step ratio (1st panel), the parameter estimates (2nd through 5th panel, excluding noise σ_N) and loglik on B_1f (x-axis) upon the refocusing *vs.* excitation pulse gradients' scaling (y-axis), for a ground truth signal of SNR≈50, with T_{2m}=30ms and *ff*=10%.

Taking together the above suggests that the rapidly diminishing echo trains and decreased parameter estimates seen in certain regions of the *in vivo* datasets could be due to a decreased rather than increased ratio of the refocusing and excitation gradient amplitudes.

This tendency was corroborated in real data (sections 5.2 to 5.4) as shown in Figure 5.20, where the echo-step ratio and B_1f distributions are plotted for two good quality

healthy volunteer datasets from both the DMD study and the ALS-KD (Chapter 4) study.

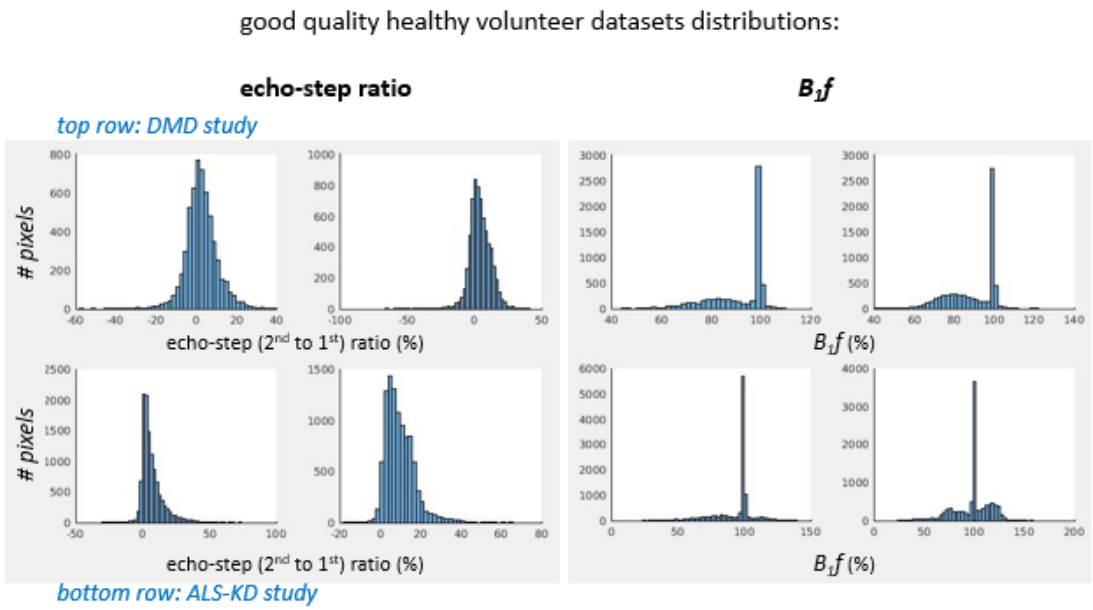


Figure 5.20: A comparison of the echo-step ratio and B_1f distributions for good quality healthy control datasets between the DMD forearm study (top row) – where bore periphery effects are expected to play a role – and the ALS-KD thigh study (bottom row) where the acquisition was of higher quality overall. Both quantities' distributions extend to lower range values for the DMD data compared to ALS-KD.

The DMD datasets (acquired at the periphery of a 70cm bore magnet) show27 an extended left tail in both distributions compared to the ALS-KD study datasets (acquired close to the centre of a 60cm bore magnet) suggesting that there was a systematic effect in the DMD data consistent with the above mechanism. The dependence of the T_{2m}, ff, B_1f and amplitude α estimates on the echo-step ratio across the entire forearm segmented muscle area is shown respectively in Figure 5.21 and Figure 5.22 for healthy volunteers and (DMD) patients at baseline.

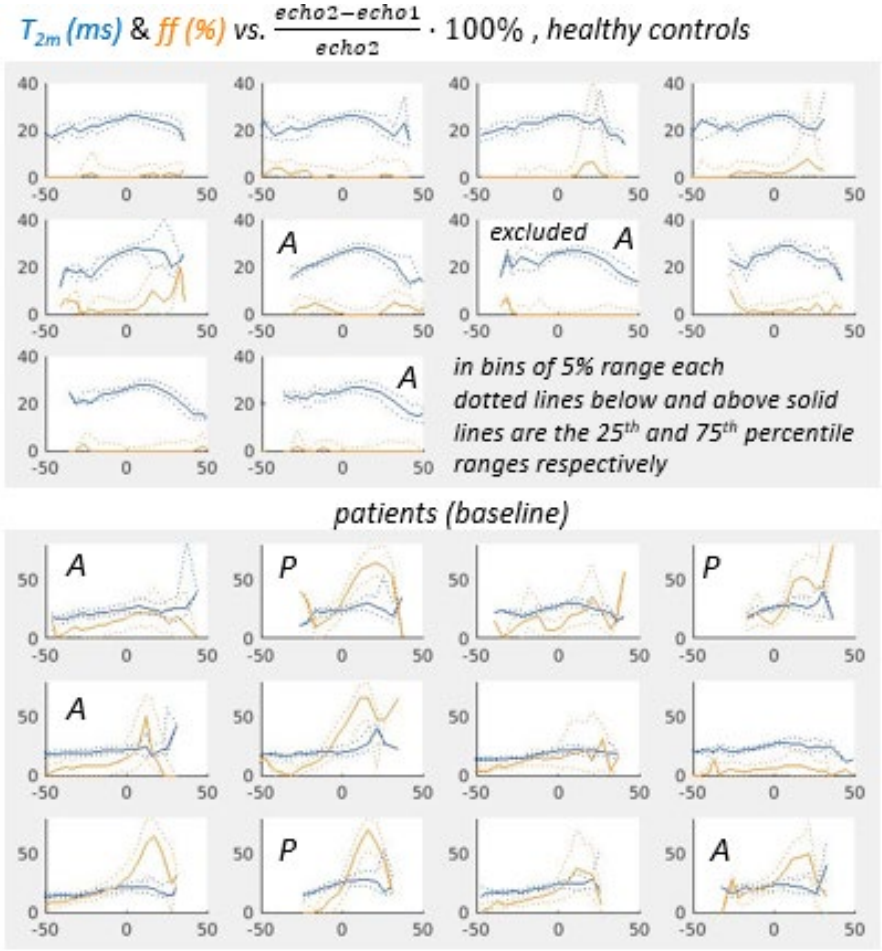


Figure 5.21: Overview of the dependence of T_{2m} (blue colour) and *ff* (orange colour) on the signal step between the 2nd and the 1st echo ($\frac{echo\ 2-echo\ 1}{echo\ 2} \cdot 100\%$), plotted in bins of 5% range of this ratio for the range -50 to +50% and reporting also the 25th and 75th percentiles (dotted curves). Above: healthy volunteers, below: patients at baseline. While various processes may affect this relationship (as discussed in the text) parameter estimates are in general mildly affected from the signals' echo-step ratio.

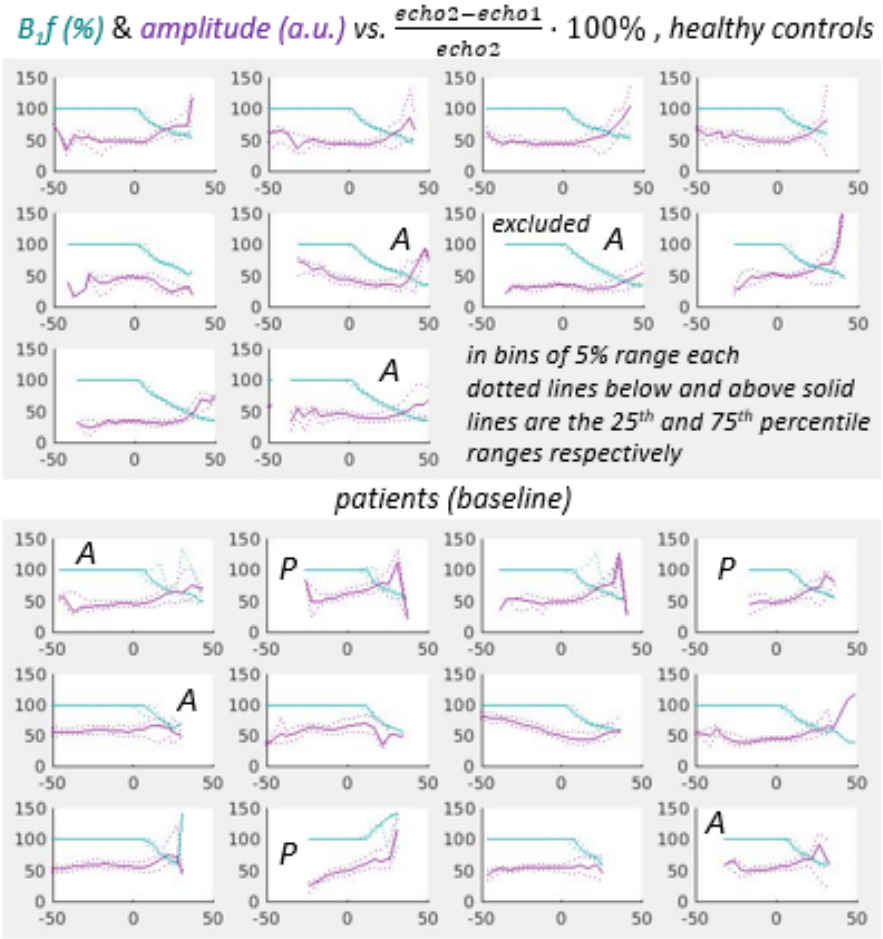


Figure 5.22: Overview of the dependence of B_1f (green colour) and amplitude α (purple colour) on the signal step between the 2nd and the 1st echo ($\frac{echo\ 2-echo\ 1}{echo\ 2} \cdot 100\%$), plotted in bins of 5% range of this ratio for the range -50 to +50% and reporting also the 25th and 75th percentiles (dotted curves). Above: healthy volunteers, below: patients at baseline. At lower, negative range of the echo-step ratio B_1f takes maximum values near 100% in agreement with the simulation results while the amplitude α estimates are largely independent from the echo-step ratio.

Comparing these parameter estimates *vs.* echo-step ratio curves with the simulation experiment results suggests milder overall associations in real data, particularly for T_{2m} and α – while of course additional, more complex and possibly competing processes that affect the acquired signal behaviour may exist. As can be seen short T_{2m} is not associated with negative echo-step ratio (2nd echo < 1st echo) but is encountered also at the upper echo-step ratio ranges (2nd echo > 1st echo) corresponding to typical CPMG echo trains. In contrast the association between B_1f

estimates returned at 100% and negative echo-step ratio, is entirely consistent with the simulation results.

Based on the observations so far the following points can be made:

- For negative echo step ratio (i.e. 1st echo higher than 2nd, consistent with a hypothetical slice select gradient amplitude ratio error), T_{2m} and *ff* estimation appears markedly more self-consistent in the real results (Figure 5.21) than might be predicted from the simulations (lower regions of the 3rd and 4th panel in Figure 5.18 and Figure 5.19 respectively). Small (5-10%) downward changes in B_1f or the gradients' ratio predict several ms changes in T_{2m} and 1-2% in *ff* estimates in simulation, whereas in real data there is a much milder decrease in estimates with increasingly negative echo step ratios; this behaviour on the other hand is expected since it naturally corresponds to faster decays (shorter T_{2m} and smaller *ff,* Figure 5.17).
- The amplitude intercept α in particular clearly increases in simulation for increasingly negative echo step ratios, which is not the case in real datasets (Figure 5.22); B_1f as predicted by the simulations is consistently close to 100% in that region, supporting the possibility of a slice-select gradient error effect.
- Importantly, the lower range T_{2m} estimates towards 11ms (as also the higher *ff*) are only seen in patients and not healthy controls, suggesting that, unless the patients found it more challenging to keep their arms away from the edge of the bore than controls, they are not driven by acquisition imperfections (Figure 5.21).

- Finally this hypothesised scanner settings errors at the magnet periphery might act in more complex ways than the simple changes in the slice profile and width due to gradients considered above.

The previously established post-fitting quality control of returned parameter estimates also impacts upon the above observations: excluding non-meaningful or bad fits as described earlier (section 5.3) has already removed the most severely compromised fits. Fast T_{2m} (and possibly relatively low *ff*) estimation is not only seen to emanate from markedly negative echo-step ratio signals, as it has also been demonstrated through the post-fitting quality control steps and the hypothesised role of *ff*, noise and artefacts. While echo-step ratio curves for healthy volunteers in Figure 5.21 are quite similar to the patient group ones, the latter manifest a slightly higher gradient in estimate magnitudes from lower to higher step-ratio values, as well as lower overall T_{2m} values, higher *ff* and a lower absolute step-ratio values range, suggesting that these observations are associated with disease rather than artefactually collapsing signals. Values towards the upper end of the step-ratio range are expected to predominantly correlate with compromised B_1 field effects, where the stimulated and Hahn echo phenomena are intense and typically the 2nd echo significantly exceeds the 1st. There is no clear trend towards lower or higher parameter estimates there, suggesting again that there is no marked bias introduced due to $B_1 f$ deviation effects – unlike the clear increase that was seen in simulation for both T_{2m} and *ff* (Figure 5.18, upper parts of the 3rd and 4th panels), even in the presence of noise (Figure 5.19). The only healthy volunteer dataset for which substantial numbers of fast T_{2m} estimates were yielded is the one that was excluded

(Figure 5.21, upper panel), and these values were correlated to high echo step-ratio values corroborating their association with compromised B_1f effects – as was seen also from the marked signal dropouts in the raw images.

In summary, two subgroups of steep decaying ("collapsing") signals were found in this study data. The first, associated with suboptimal acquisition conditions, was largely excluded from the analysis after the post-fitting quality control. The second, mainly found in the patient datasets was likely to be predominantly due to disease involvement, with acquisition imperfections having a rather minor effect in the overall results.

5.6 Factors affecting the absolute accuracy of ff_d

Another issue important to discuss is the interpretation of the Dixon *ff* (ff_d) results and their comparison with ff_a – the 'apparent' sEPG-MLE-derived *ff*. There are a number of reasons that ff_d might deviate from the notional "ground-truth" *ff*, for instance, the multiple spectral peaks of lipid resonances and the T_2^* weighting in the series of gradient echo images used for the Dixon imaging in this work, are potential sources of errors, while differences in T_1 weighting between muscle and fat signal sources, and noise bias due to using magnitude reconstructed images are also potential factors (Burakiewicz et al., 2017). The last two may be determined by the acquisition settings (e.g. flip angle and readout/reconstruction technique) while the first two can be addressed via more sophisticated signal modelling. The closed form ff_d correction used in Chapter 4 is an attempt to correct for the hypothesized underestimation of ff_d due to its single peak representation within the Dixon

technique (Azzabou et al., 2015). However it is unclear if that correction takes into account T_2^* effects and, even if that is so, only a common decay rate for all tissue compartments is assumed. Such 'global' T_2^* corrections have been shown to be beneficial in chemical shift based fat quantification methods (Loughran et al., 2015, Hernando et al., 2010), however the additional step of using different T_2^* constants for the water and fat compartments (equal for all the fat resonance peaks) offers further scope for improvement (Siracusano et al., 2017, Chebrolu et al., 2010).

The following paragraphs address theoretically the likely impact of such an approach in the context of the present study. The DMD study results presented so far showed a clear tendency for lower (than normative) range T_2 values in the forearm muscles, both in healthy volunteers and particularly in patients, consistent with a previous publication (Wary et al., 2015). As mentioned by these authors, and also here, forearm T_2 values are lower than commonly reported. This generally lower T_2 suggests that T_2^* may also be lower, given that T_2 and T_2^* contrasts are considered to be very similar in general (McRobbie et al., 2006), and therefore T_2^* may play a more important role in ff_d bias compared with previously reported situations. Therefore we will now consider in more detail a difference between the T_2^* of fat and water that could affect ff_d. The effect of T_2^* in this context can be expressed via the equation describing the signal s_d from a voxel containing water and fat with T_2^* differences between the water and fat compartments:

$$s_d(TE) = W_0 e^{-\frac{TE}{T^*_{2m}}} + F_0 e^{-\frac{TE}{T^*_{2f}}} \sum_{k=1}^{5} \alpha_k e^{2\pi i \Delta f_k TE} \quad \text{Equation 19}$$

where W_0 and F_0 are the water and fat signal amplitudes, T_{2m}^* and T_{2f}^* the T_2^* of water and fat respectively, Δf_k the (central) frequency offset of the k^{th} fat peak, α_k its relative proportion ($\sum_{k=1}^{5} \alpha_k = 1$) and TE the echo time – taking values corresponding to the in- and out-of-phase conditions required within the Dixon method. Magnetic field inhomogeneities are ignored and Δf_k and α_k are presumed known (Azzabou et al., 2015) – the values proposed in that work were used here. Taking advantage of the results presented so far in this work as an approximation we can set T_{2m}^* and T_{2f}^* at T_2 values previously seen to describe muscle water and fat compartments, e.g. at 30 and 140ms respectively (assuming a mono-exponential decay for fat), based on the above assumption of similar $e^{-\frac{TE}{T^*_{2m}}}$ / $e^{-\frac{TE}{T^*_{2f}}}$ and $e^{-\frac{TE}{T_{2m}}}$ / $e^{-\frac{TE}{T_{2f}}}$ ratios. The TEs corresponding to the in and out of phase images $s_d(TE_{IN})$ and $s_d(TE_{OUT})$, can be calculated as follows: with a 435 Hz difference at 3.0 T, the signals are in phase 435 times per second, so after the start, they are in phase again after 1/435s=2.3ms; thus, they will be out of phase for the first time at 1.15ms and every 2.3ms after that. In this calculation 4.6ms and 3.45ms were the TEs used for $s_d(TE_{IN})$ and $s_d(TE_{OUT})$ respectively – however any other combinations (multiples) produce identical results. Then by substituting in Equation 19 and solving the equations $F_d = 0.5(s_d(TE_{IN}) - s_d(TE_{OUT}))$ and $W_d = 0.5(s_d(TE_{IN}) + s_d(TE_{OUT}))$ (where F_d, W_d the fat and water signals and TE_{IN} and TE_{OUT} the echo times for in and out of phase water and fat signals respectively according to the Dixon method, section 2.1.2) we obtain the relationships $F_0 = \frac{F_d + 0.017}{0.425}$ and $W_0 = \frac{W_d - 0.014}{0.861}$, and finally $ff_{d-corrected} =$

$\frac{ff_d+0.017}{0.506ff_d+0.504}$ as a correction expression for ff_d accounting for both the multi-peak issue and differing T_2^*s between the compartments. This correction resembles the expression ($ff_{d-corrected} = \frac{ff_d}{0.45ff_d+0.55}$) used previously (section 4.3.2) which accounts only for the multiple lipid peaks, and has a similar yet stronger effect on the ff_d estimates.

The impact of these successive corrective steps on ff_d is demonstrated in Figure 5.23 where a clear overall negative bias in ff_d compared to ff_a in panel (a) becomes substantially mitigated in panels (c) and (d) on applying the respective corrections. An additional simple correction was to discard any *ff* values below 0% or above 100% (the physically meaningful range), applied after panel (b).

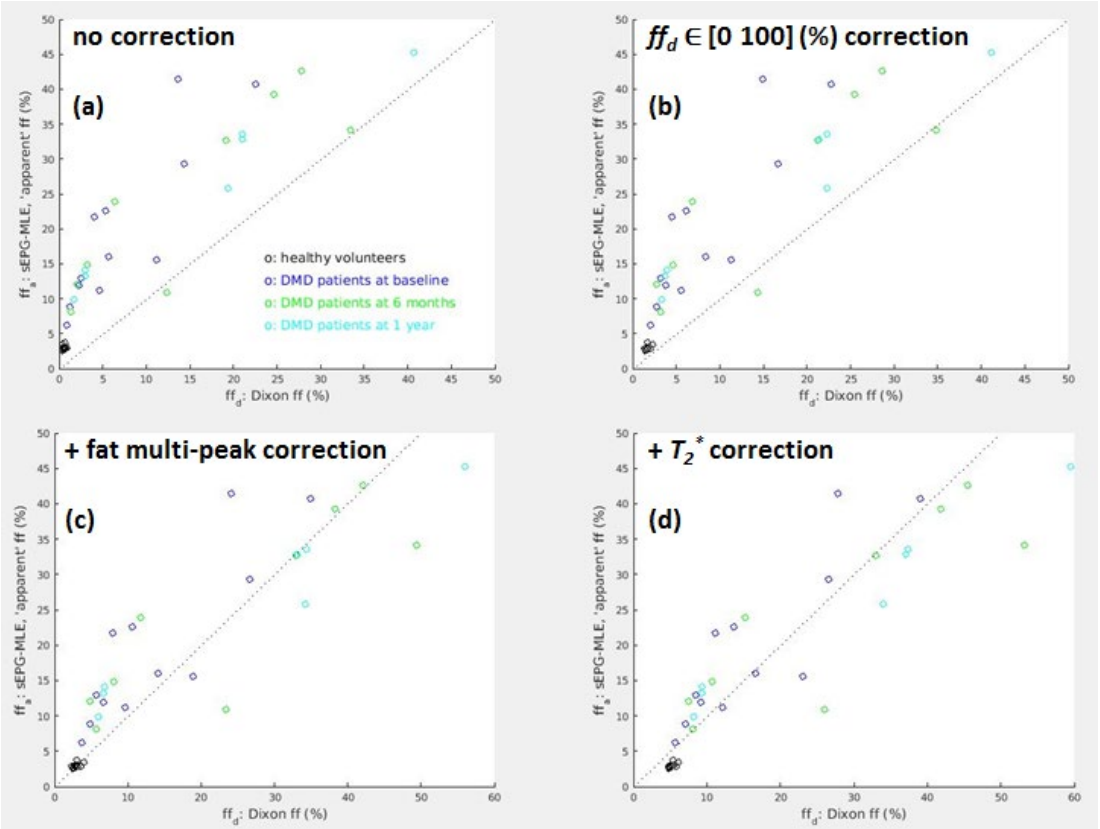


Figure 5.23: The effect of applying corrections (described in the main text) to ff_d: (a) ff_a vs ff_d without any ff_d correction, (b) with any ff_d values beyond the meaningful 0 to 100% range being discarded, (c) with additional correction for dephasing due to the fat multi-peak spectrum and (d) with additional correction for differential T_2^* weighting between fat and water; ff_a has been corrected for T_1 weighting and MT effects (section 3.3.2)

Notably the bias near the two extremes of the *ff* range continues to exist after all corrections (as in the last two panels at the bottom of the figure) as may be expected because it is associated with the noise properties of the magnitude images reconstruction (Burakiewicz et al., 2017) which has not been accounted for. By modelling the Dixon method acquired signal according to the above expressions (in- and out-of-phase signal, T_2^* decay and Equation 19) while considering a 2-component model for fat and for water and a corresponding ground truth *ff*, the theoretical plot shown in Figure 5.24 can be constructed. In this simulation ff_d is thus calculated taking into account the effects of the fat multi-peak spectrum and differential T_2^* weighting between fat and water on the ground truth *ff*.

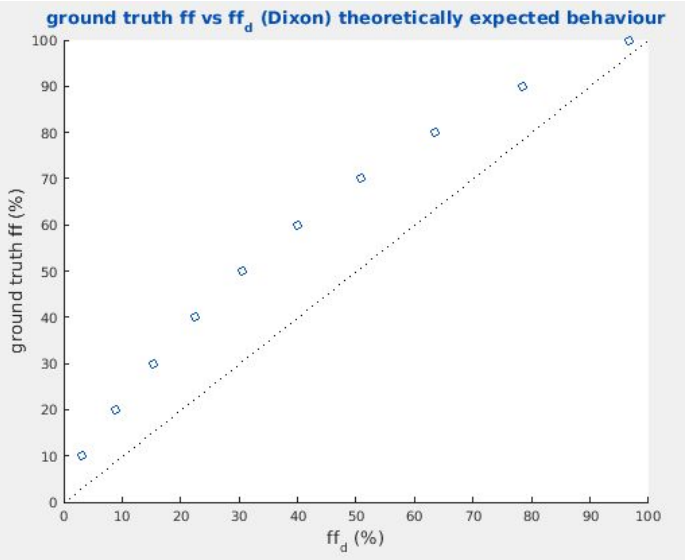


Figure 5.24: Comparison of ground truth and theoretically predicted measured Dixon *ff*, with the latter modelled to include the multipeak fat spectrum and differential T_2^* decay during acquisition.

This is broadly consistent with the divergence seen experimentally in real data (i.e. Figure 4.22 and Figure 5.23(a) between ff_a and ff_d.

The assumptions underlying the hypothetical ff_d correction approach explored above regarding T_2^* effects are approximate, and further work is needed to demonstrate its validity, including direct determination of the relevant T_2^* values. For these reasons only the previous corrections (section 4.3) are effectively applied in the results presented here, in consistency with the previous chapters.

5.7 Study analysis results II

It should be noted that the post-processing quality control steps described in section 5.3, by which parameter estimates from pure fat, and problematic or physically non-meaningful signals are already eliminated from the analysis, can significantly affect the different weighting of individual muscle ROIs' contributions to the summary

measures in the quantitative comparisons. Typically in artefactual or problematic regions relatively less and in some cases very few pixels are left in the final parameter estimate maps. This affects the interpretation of the individual muscle ROI results presented in section 5.4 above. For example several ROIs from the datasets that yielded the lower range T_{2m} estimates near 11ms, discussed in the previous section and evident in Figure 5.10, were – mainly due to the post-processing strategy – an order of magnitude smaller in size (0 to 100 pixels) than the average size (several hundreds of pixels) of individual muscle ROIs across all subjects. The segmentation process, deliberately excluding non-muscle tissue types as fat, nerves and fascia, has also had an impact on this by further reducing the size of muscle ROIs in many cases.

Histogram metrics

There is substantial variability observed between whole cross-sectional muscle median values (Figure 5.13) and distributions in individual ROIs (Figure 5.9 and Figure 5.10) created in part by variations in pixel numbers of individual muscle ROIs. These data also imply that disease progression can cause both increase and decreased T_{2m} as well as affecting *ff*, therefore impacting their distributions' shape mainly by increasing their width. These observations are quantified in more detail in the histogram metrics analysis presented in this section. In addition to the excluded healthy volunteer dataset mentioned in the previous section, ff_d for one patient (at the 6 months follow up scan) showed marked inconsistencies both quantitatively and qualitatively with the remaining data, possibly due to an error in the fat-water separation process (water and fat signal components' mislabelled (swapping)), and this subject's ff_d data was therefore also excluded from the following analysis. In

Figure 5.25 summary histogram metrics are shown from all participants and study scanning time points, at this stage ignoring differences in the patients' disease severity at baseline.

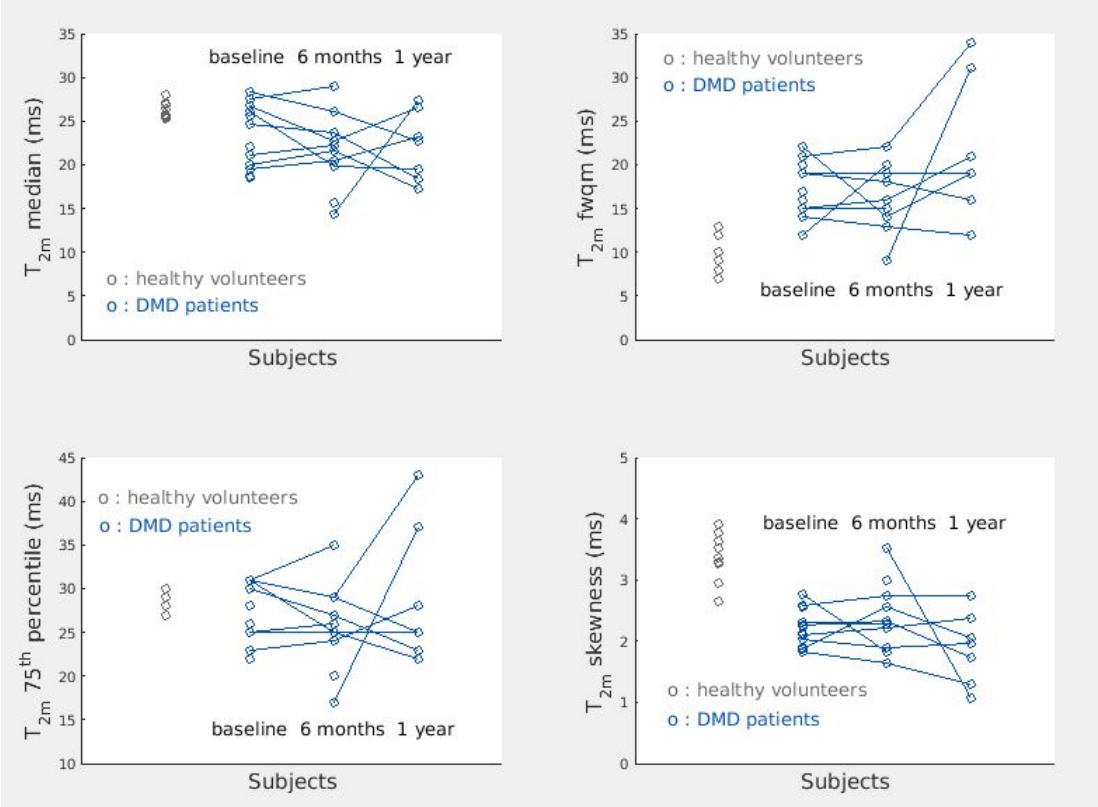


Figure 5.25: Forearm single-slice all muscles T_{2m} histogram metrics comparison for healthy controls (in grey) and DMD patients (blue) at baseline, 6 months and 1 year follow-up showing median (top left), full width at quarter maximum (top right), 75th percentile (bottom left) and skewness (bottom right) of the estimated T_{2m} distributions.

The metrics plotted are the median, full width at quarter maximum (fwqm), 75th percentile and skewness from the distributions of estimated T_{2m} from the entire forearm cross-sectional area (using the parameter estimate maps of the segmented slice produced as described in section 5.2 and after the quality control steps described in section 5.3). Figure 5.26 and Figure 5.27 show the equivalent ff_a and ff_d histogram metrics. To compute and extract these histogram metrics and the corresponding text files containing them, custom-written software tools were

created by the current author using Matlab code, using default functions wherever possible (i.e. the main histogram chart, skewness and kurtosis). A unit binwidth was used for both T_{2m} and *ff* (ms and % respectively).

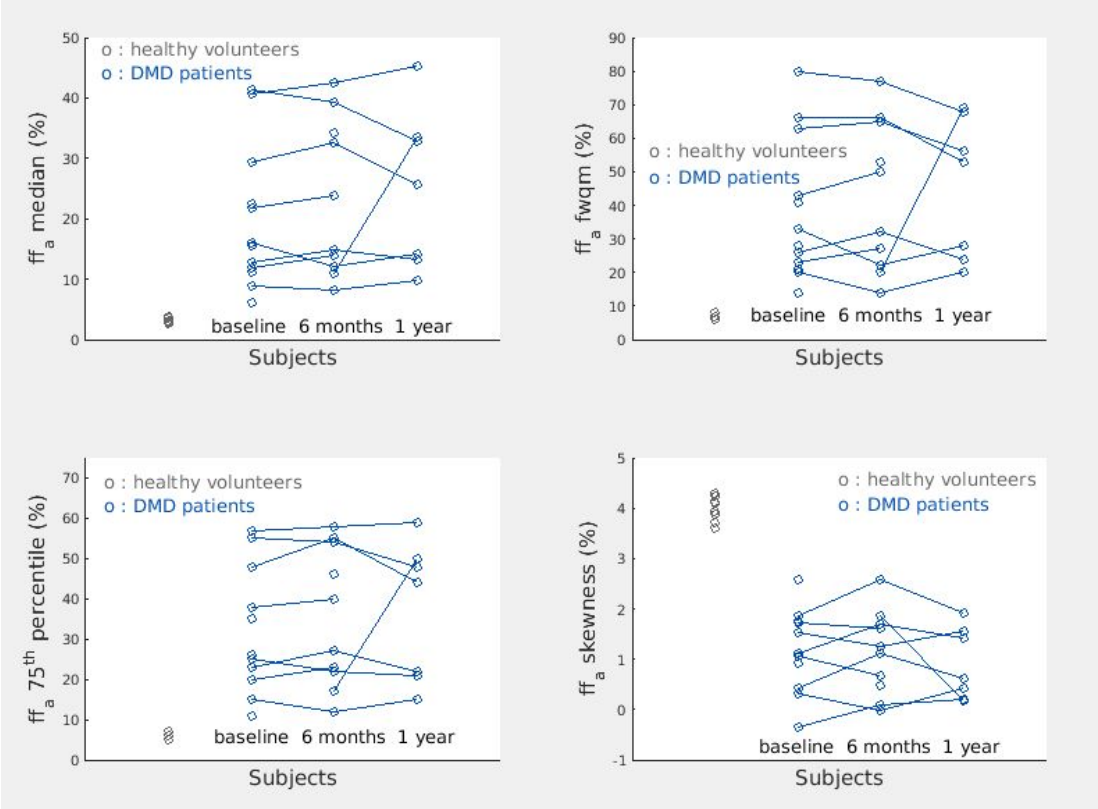


Figure 5.26: Forearm single-slice all muscles ff_a histogram metrics for healthy controls (in grey) and DMD patients (blue) at baseline, 6 months and 1 year follow-up. Shown are median (top left), full width at quarter maximum (top right), 75th percentile (bottom left) and skewness (bottom right) of the ff_a (apparent, sEPG-MLE *ff*) estimates' distributions.

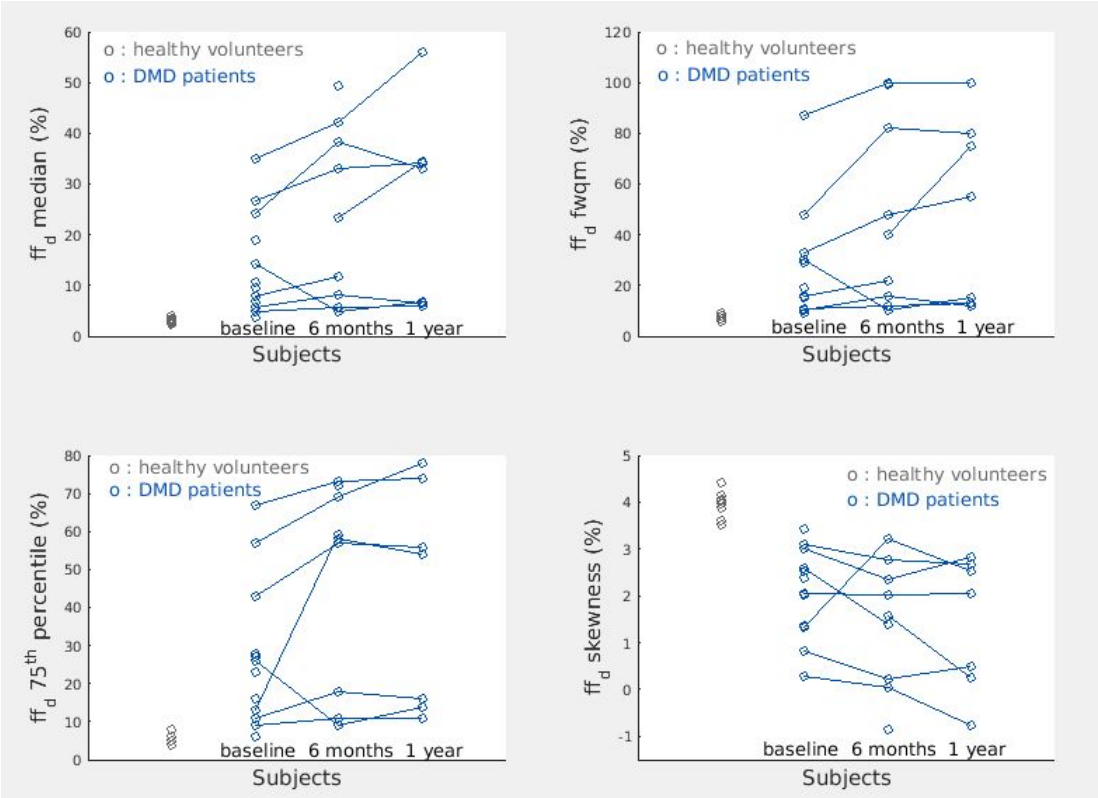


Figure 5.27: Forearm single-slice all muscles histogram metrics for healthy controls (in grey) and DMD patients (blue) at baseline, 6 months and 1 year follow-up. Shown are median (top left), full width at quarter maximum (top right), 75th percentile (bottom left) and skewness (bottom right) of the ff_d (Dixon *ff*) estimates' distributions.

The main observations were decreased T_{2m} and increased ff_a and ff_d for patients compared to healthy controls, with wider (increased fwqm) and more skewed (reduced skewness) distributions, and more dispersed ranges of values (broader 75th percentiles' range). Results showed also a tendency to decrease longitudinally – except for skewness where numerical results are somehow reversed since higher skewness is denoted by smaller numbers. However, for the individuals for which longitudinal data were available, 4 subjects showed increases, and from the 4 that decreased 2 subjects had their minimum and maximum values respectively at the intermediate time point at 6 months. Increases were more marked for ff_d, clearly manifested for 5 subjects while 2 appeared rather stable and 1 showed decrease.

ANOVA revealed significant differences between healthy volunteers and patients at baseline for all 3 parameters, with greater significance for skewness and fwqm than for 75th percentile and medians (Table 5.3). No significant differences were found between the patients' longitudinal measurements.

Table 5.3: One way ANOVA comparisons between healthy volunteers and patients and for patients longitudinally (baseline – 6 months – 1 year examination times) for the T_{2m}, ff_a and ff_d histogram metrics; p-values for statistically significant differences (p<0.05) are shown in blue colour, otherwise in red.

Histogram metrics ANOVA (p-values)		***CTR vs. DMD baseline***	***DMD baseline – 6m – 1y***
T_{2m}	median	0.0263	0.6121
	fwqm	0 ($<10^{-5}$)	0.0604
	75th percentile	0.4315	0.4696
	skewness	0 ($<10^{-5}$)	0.1168
ff_a	median	0.0004	0.6633
	fwqm	0.0003	0.7628
	75th percentile	0.0001	0.7156
	skewness	0 ($<10^{-5}$)	0.7237
ff_d	median	0.004	0.1383
	fwqm	0.0165	0.2106
	75th percentile	0.003	0.1964
	skewness	0 ($<10^{-5}$)	0.4009

The longitudinal forearm cross-sectional median T_{2m} plotted against the time for which each patient was recorded as being non-ambulant, a potential surrogate for disease severity, are shown in Figure 5.28 and the corresponding apparent (sEPG-MLE) *ff* values in Figure 5.29.

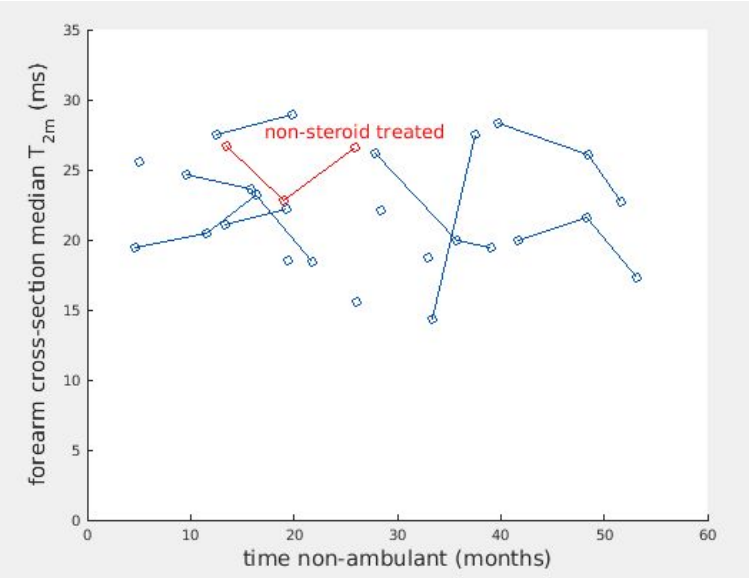


Figure 5.28: Dependence of the forearm cross-sectional median T_{2m} on time non-ambulant for DMD patients. Lines connect longitudinal results (between baseline, 6 months and 1 year scans) for the same individual where they exist. The red points are data for the non-steroid treated patient.

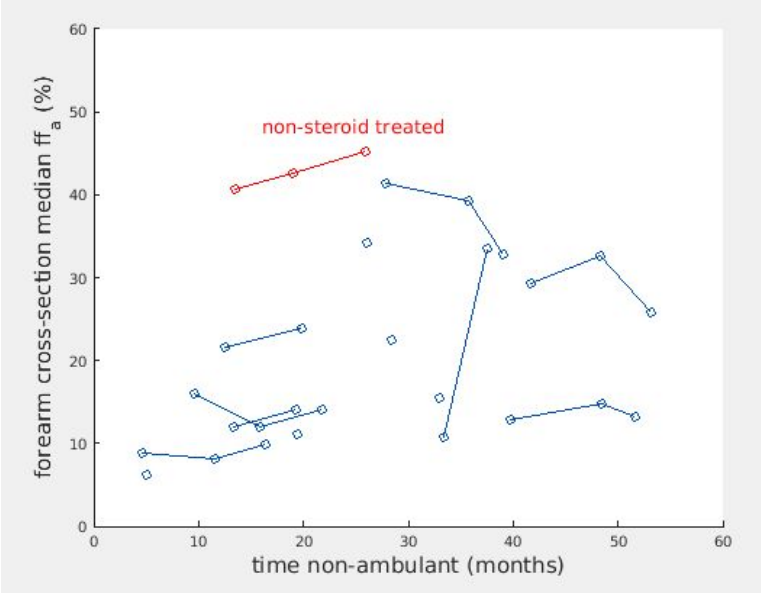


Figure 5.29: Dependence of the forearm cross-sectional median ff_a on time non-ambulant for DMD patients. Lines connect longitudinal results (between baseline, 6 months and 1 year scans) for the same individual where they exist. The red points are data for the non-steroid treated patient.

The changes with time described previously (Figure 5.25) are more clearly shown here: a tendency for decreasing T_{2m} with months non ambulant in 4 of the 8

individuals for which longitudinal data were available, and an apparent increase for the remaining 4. Additionally, in 3 of the 4 subjects for which T_{2m} increased this occurred in the first 24 months non ambulant, while in 3 of the 4 subjects for which T_{2m} decreased this occurred after the first 24 months non ambulant. Similar tendencies are observed in the median ff_a results, suggesting transient changes in the dependence of disease involvement on the duration of time non ambulant.

Representative T_{2m} and ff_a maps from the different time points of the study are shown below. The examples shown in Figure 5.30 are from relatively good quality datasets of the study. While small numbers of pixels were excluded in both due to the post-fitting quality control, results for both subjects followed the main tendencies described above. For instance T_{2m} for subject 3 (top row maps) show a decrease longitudinally (medians: 26.1, 19.9 and 19.5ms, all scans after 24 months non ambulant).

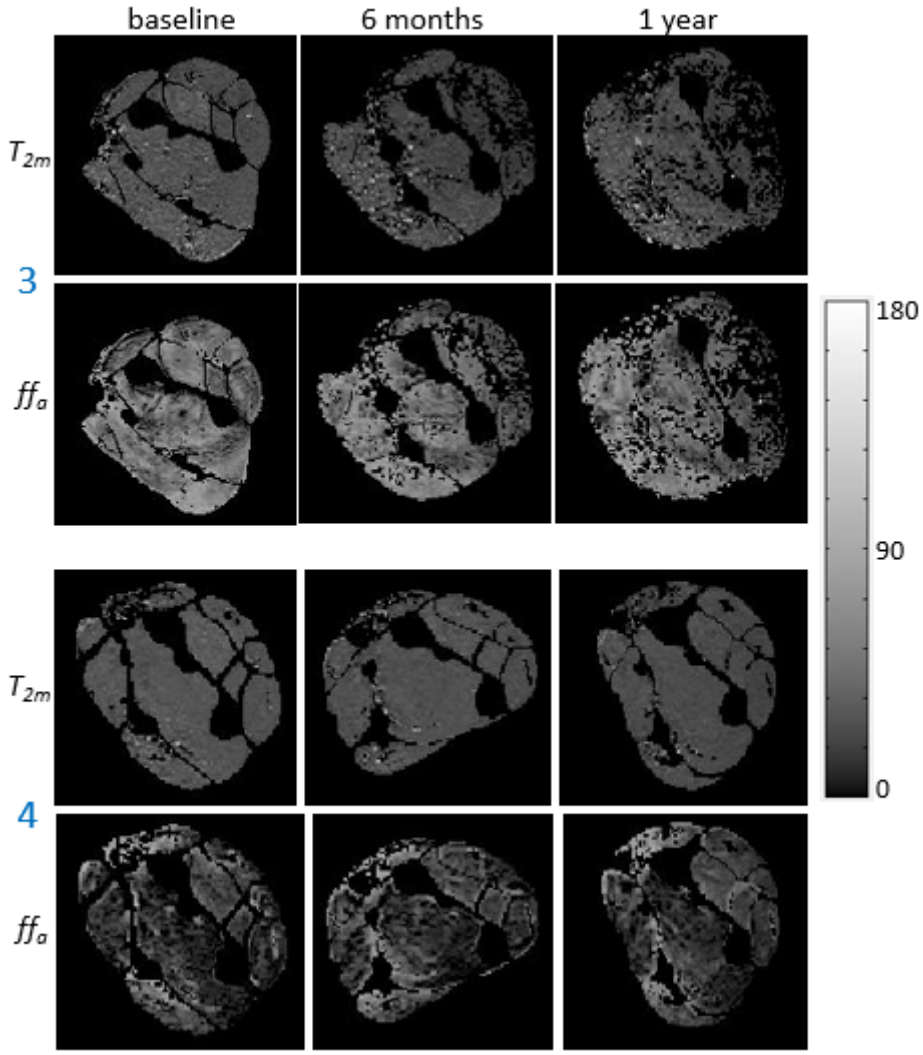


Figure 5.30: Examples T_{2m} and ff_a maps from relatively good quality DMD study datasets (subject 3 above and subject 4 below, as referenced in Figure 5.6, Figure 5.10 and Figure 5.11) along the different time points (columns). These may be considered as typical parameter maps in terms of quality of fitting and physiological results.

In contrast the cases presented in Figure 5.31 however may be considered atypical in terms of quality of fit, and pathologically. Visible effects of motion artefacts and of signal dropouts are observed in the parameter maps of subject 6, especially at baseline, and this may be related to its somewhat reduced T_{2m} values at the disease onset stages (only a few months non ambulant). Large numbers of pixels were excluded for subject 2 at the 6 months measurement due to the collapsing signals effect (section 5.5.1) while the reduced median T_{2m} atvbaseline compared to 1 year

for this participant was visually inconsistent in magnitude with the T_{2m} trajectories for all the other subjects.

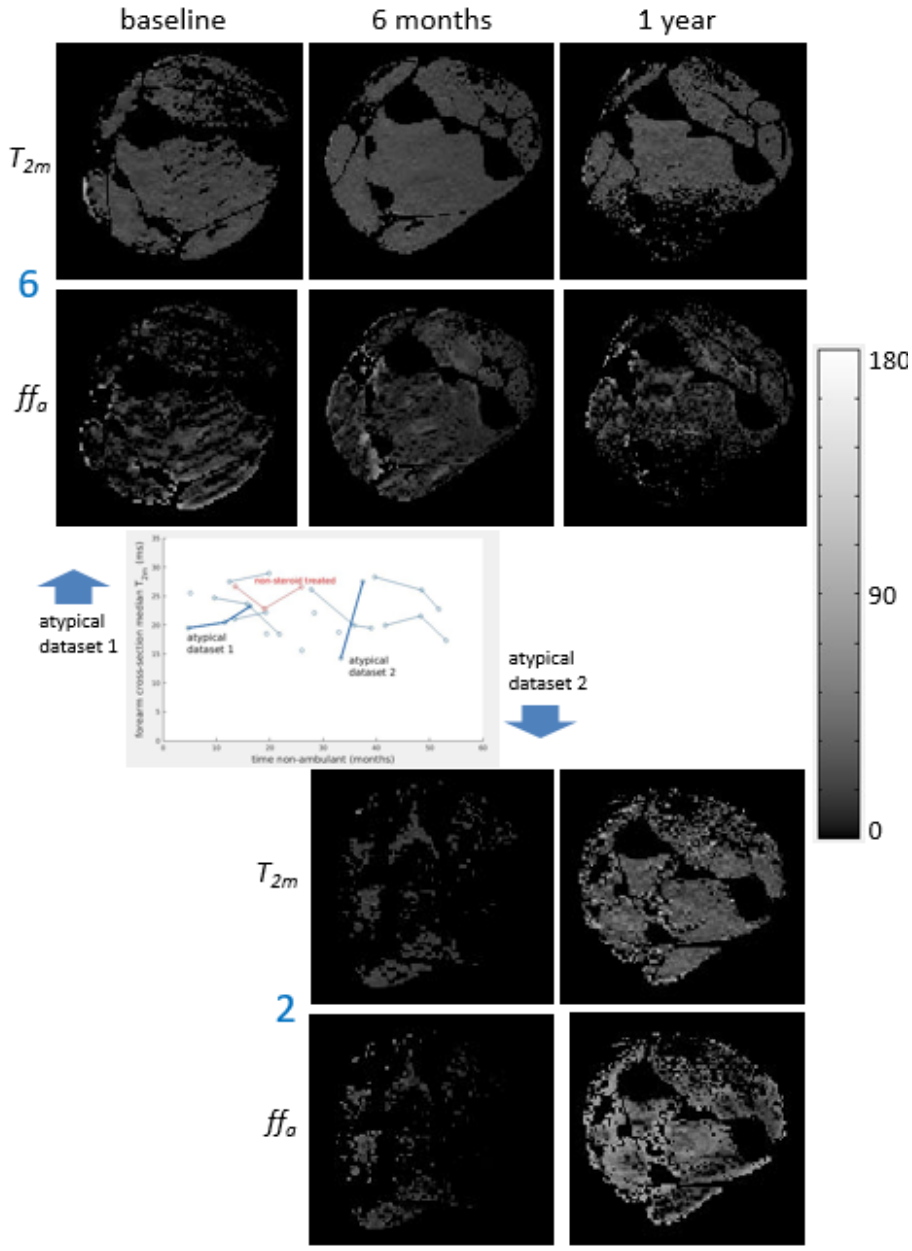


Figure 5.31: Examples T_{2m} and ff_a maps from atypical DMD study datasets (subject 6 above and subject 2 below, as referenced in Figure 5.6, Figure 5.10 and Figure 5.11) along the different time points (columns). Sub-optimal quality of fitting was observed for subject 6 with visible motion artefacts and signal drop-outs, while large numbers of pixels were excluded for subject 2 due to the collapsing signals effects described in section 5.5.1. They were also rather different in terms of physiological interpretation (inset, based on Figure 5.29) compared to the rest of the subjects.

However the quality of both the fitting and the raw data for the remaining pixels was at acceptable levels suggesting that this observation may be due to physiological processes (e.g. muscle compression, tremor, steroid treatment effects, unknown exercise changes etc.).

Comparisons with alternative established methods

Dixon method

As previously mentioned, a first report describing Dixon *ff* findings in this study has previously been published (section 5.1). However in that work no corrections for ff_d were considered and additionally different data inclusion criteria (e.g. due to data quality) were used. This may be taken into consideration when comparing results and conclusions between that and the present work.

The results presented in Figure 5.23 and in Figure 5.27 (as compared to Figure 5.26) give a detailed comparison between ff_a and ff_d. For completeness the dependence of T_{2m} vs ff_d is additionally shown in Figure 5.32 below.

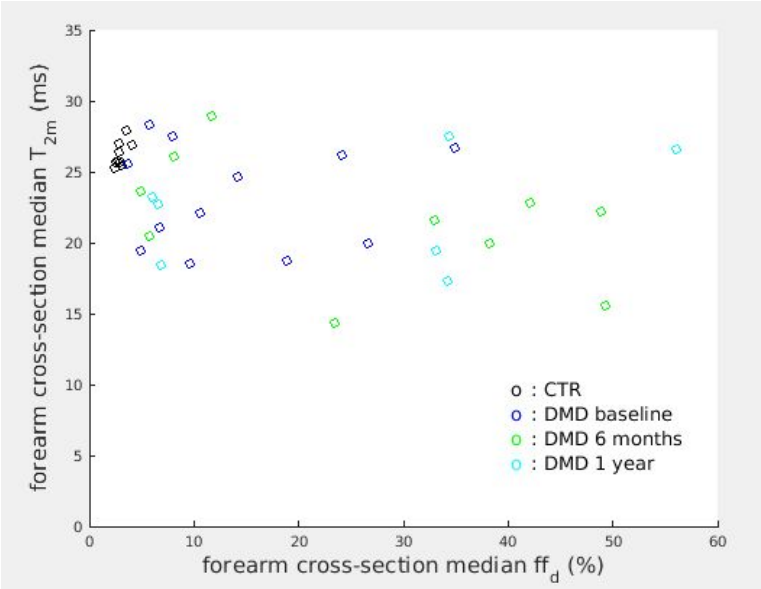


Figure 5.32: Plot of the sEPG-MLE T_{2m} (ms) vs ff_d (%). Values shown are media values across the entire segmented muscle area

The above graph reveals for the patients a largely uniform broad distribution of the estimated T_{2m} *vs.* the independently determined ff_d with a tendency to decrease as ff_d increases. This is in agreement with the results shown previously for both T_{2m} *vs.* ff_a (Figure 5.6) and the dependence of T_{2m} and ff_a on time non ambulant (Figure 5.28 and Figure 5.29). Taken together these results suggest that this marginal tendency for decreasing T_{2m} correlates with both increasing *ff* and with time non ambulant, since increasing *ff* is also a manifestation of disease progression.

Exponential models and LSQ minimization

It is of interest, in particular to allow comparison of our results with those of older studies, to compare for this dataset the performance of previously used exponential T_2 decay models (section 2.3) with the sEPG MLE approach developed here. The exponential model used here for comparison was the equivalent of the multi-component sEPG model used to produce the above results, incorporating a single

component term for muscle linked through the *ff* parameter to a fixed parameter mono-exponential plus baseline (constant) model for fat (section 5.2). This model was fitted to the CPMG data from this study using a LSQ algorithm used in previously presented work by the author (Zafeiropoulos, 2016). Representative results comparing estimates obtained using the two methods, namely EXP-LSQ and sEPG-MLE as they previously have been labelled (Chapter 4, section 4.4), are presented in Figure 5.33.

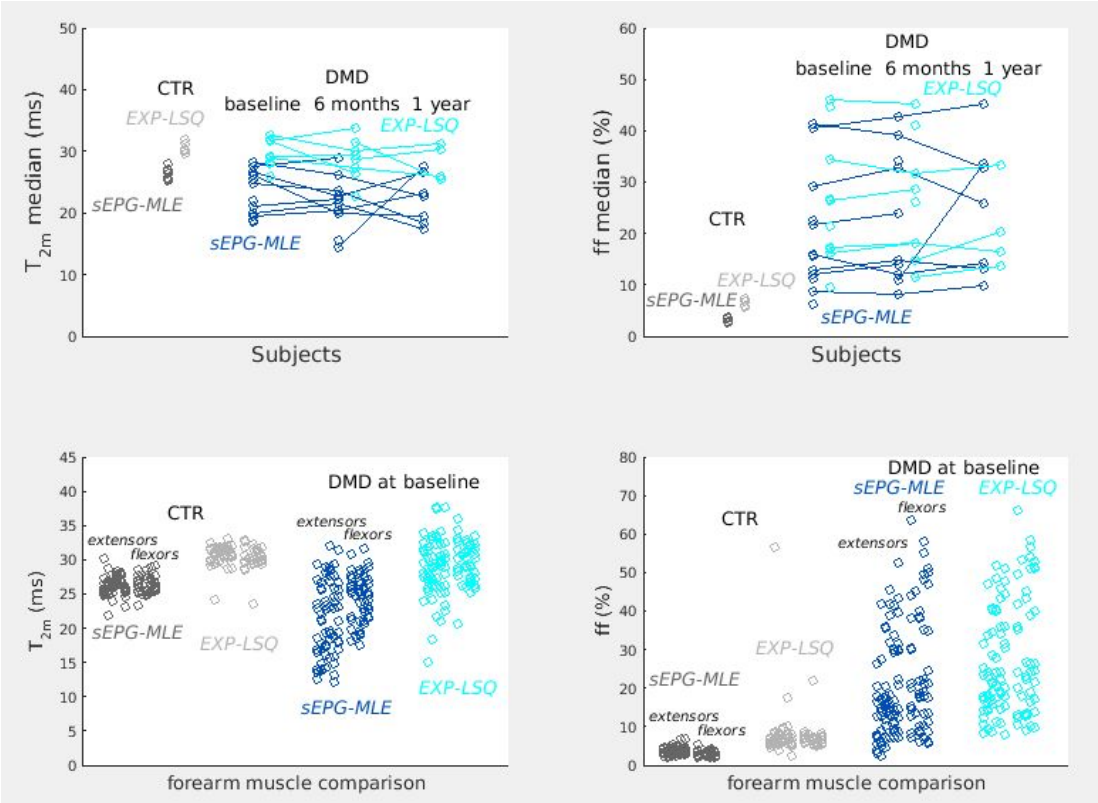


Figure 5.33: Comparison between the EXP-LSQ and sEPG-MLE methods' parameter estimate results, plotted in dark and light grey for healthy volunteers, and in blue and cyan for DMD patients respectively. Top: forearm cross-sectional median values for all time points available (for patients); bottom: individual muscle ROIs for controls and patients at baseline. Left column: T_{2m} results; right column: (apparent) *ff* results. EXP-LSQ estimation produces higher overall values than the respective sEPG-MLE results.

The clear elevation of the EXP-LSQ results compared to the sEPG-MLE estimates is consistent with the results from the respective experiment with phantom data and

with ALS-KD study data as described in Chapter 4, section 4.4. It can be seen that the LSQ approach leads to overestimated T_{2m} compared with the sEPG MLE method, although the degree of dispersion and change with disease were generally similar.

5.8 Discussion

The sEPG-MLE fitting tool developed and presented in this work relies on a fixed parameter subsidiary 2-component model approximating the fat signal, combined with a single-component sEPG model for the water fraction signal to estimate T_{2m} and *ff* as the main parameters of interest. However the remaining model parameters – global amplitude α, the $B_1 f$ field factor and the noise σ_N – are also of key importance in the optimisation of the method. In this work improvement in estimation accuracy has been achieved by the approaches used accounting for the behaviour of the latter two, $B_1 f$ and σ_N. As shown in Chapter 3 and also in published work (McPhee and Wilman, 2017), the slice-selective RF pulse and gradient combination used to produce CPMG signals cause non-symmetric behaviour for symmetric $B_1 f$ deviations above or below its ideal 100% value, and the sEPG model implemented here accounts for that. The use of the MLE method to fit to that model then properly accounts for the Rician noise floor in magnitude-reconstructed image data. Additionally, spatial regularisation of $B_1 f$ polarity, careful post-fitting quality control and parameter space gridding strategies helped to improve the reliability of the finally presented parameter estimate maps.

A main finding in this work is the overall reduction seen in T_{2m} due to disease progression. This is primarily demonstrated by the significant reduction of T_{2m}

between patients and controls (Table 5.3) and an additional tendency for mild decrease seen in both the follow up scans' results (Figure 5.25) and the dependence on time non-ambulant results (Figure 5.28). These observations are consistent with the findings published in a quantitative MRI study on forearm data from DMD patients (Wary et al., 2015). One of the main conclusions from that work is that T_{2m} in the more severely affected non ambulant patients, similar to the cohort examined here, decreases both with time and compared to controls; since fat content in contrast increases, this demonstrates that "inflammation/oedema and fat infiltration progress in opposite directions with disease evolution" as those authors concluded. An additional feature is that T_{2m} in flexors appears slightly elevated compared to extensors along with time non ambulant (Figure 5.10, Figure 5.13, Figure 5.33), but this pattern becomes diffuse for the later disease stages. Other than fat content, T_{2m} has been reported to be restricted also by treatment with steroids (Hogrel et al., 2016). All these observations show clear similarities with the results presented in this chapter. The results from both previous findings and this work are also corroborated from spectroscopic studies showing that water T_2 decreases with increasing ff in general (Schlaeger et al., 2019). While the overall trends shown in this work are similar with those presented in the above forearm DMD study (Wary et al., 2015), the range of T_{2m} values with the sEPG-MLE algorithm used here are about 10ms lower overall compared to the results from the aforementioned paper which may be explained – at least partly – from the apparent overestimation of T_2 when using exponential models as has been previously demonstrated (sections 5.6 and 4.4). An additional interesting feature revealed here is the dependence of the estimated parameters on time non ambulant, showing a tendency to also increase initially (i.e.

within the first 24 months) and then decrease (T_{2m}, Figure 5.28) or potentially stabilise (*ff*, Figure 5.29) at later stages, suggesting transient changes during disease progression. Importantly, such transient changes may be convolved in the measurements with factors relating to muscle behaviours such as motion, both macroscopic (e.g. tremor) and microscopic, muscle compression or stiffness, effects of exercise, which were not controlled for here, and treatment effects. These may explain the large differences between the measurements at different time point seen in some cases.

Similar findings have been reported in DMD MRI studies of the lower limbs, however a main difference in those cases is that despite disease progression causing progressive T_{2m} reduction, it remains higher for patients than for controls in contrast with the results presented here. Increasing fibrosis – in addition to steroid use and fat infiltration – has been considered as an additional potential mechanism for the observed T_{2m} reduction with time (Forbes et al., 2014, Arpan et al., 2014). Fibrosis therefore may be one of the physiological processes causing the fast decaying T_{2m} signals in some muscle regions in patients, examined in section 5.5. An additional process might be disease related muscle motor unit fasciculation effects (Whittaker et al., 2019) causing loss of signal in a similar way to motion effects (artefacts). Even when attempting to correct for fat content, different modalities produce different absolute results; this can be more clearly seen in healthy controls where yielded values span between 25ms e.g. spectroscopic T_{2m} in the thigh (Forbes et al., 2014) – consistent as well with the results presented here – and 35ms or above e.g. relaxometry T_{2m} in the forearm (Wary et al., 2015). On the other hand, changes due

to disease often do not exceed a range of 2 to 3ms, thus obscuring the apparent trends. Overlapping ranges and opposite directions in the T_{2m} longitudinal evolution are also seen in the results both here and in published work. For those reasons histogram metrics other than the median or the mean could more usefully describe T_{2m} distribution changes, which might not be fully represented by measures of central tendency (e.g. Figure 5.25).

Both the validation in this data set of the secondary 2-component fat model and the comparisons with established methods were seen to be largely successful. The fixed parameters for the fat model emerged virtually identical to the ones estimated when initially developing the method (Chapter 3) – a process accomplished with use of a higher quality dataset from a completely different anatomical location and participant demographic. The corrections used for ff_d in order to better match it with ff_a on the one hand confirmed the findings supporting their similarities already published (Azzabou et al., 2015), and on the other hand may indicate useful directions for further improvement. Finally, comparisons of the sEPG-MLE with the more conventional exponential LSQ signal model and parameter estimation methods were closely similar to the behaviour seen previously in overall higher quality *in vivo* and phantom datasets. Despite the encountered issues regarding the quality of data analysed here, i.e. shorter echo train, signal dropouts and artefacts etc., results were interpretable and consistent with other studies' outcomes; careful attention to post-fitting quality control seemed to play a key role in that achievement. Effects of the acquisition conditions and settings on the parameter estimation are further explored in a more general sense in the next chapter.

5.9 Summary and Conclusions

Overview of the results (as in section 5.7 mainly) suggests that competing processes contribute to the T_{2m} reduction in DMD patients compared to healthy controls, and to further decreases with time non ambulant: increasing fat infiltration, steroid treatment combined with receding inflammation as disease progresses and possibly also increasing fibrosis, appear to be the main physiological mechanisms behind these observations. Additional factors such as patient discomfort, stiffness and other motion-including conditions may also play a role. On the other hand fat infiltration appears to increase steadily towards stable levels at later stages of the disease.

In summary:

- Calibration of the secondary 2-component fat model used within the sEPG-MLE method was consistent with the behaviour seen during development of the main algorithm (Chapter 3).
- Post processing of the results, i.e. discarding extreme or physically not meaningful values, is an important step in the interpretation and validation of results – for instance in comparisons with more conventional or established methods. It also appears to largely account for the acquisition imperfections (shorter echo train, motion artefacts, $B_1 f$ and gradient errors etc.) existing in this dataset.
- T_{2m} emerges significantly lower for DMD patients than for healthy controls, largely stable across the *ff* spectrum, showing however a tendency to further decrease with time non ambulant and the underlying physiological processes.

- Comparison with Dixon method *ff* revealed a closer correlation when using corrections for both methods: MT and T_1-weighting corrections for ff_a and fat multi-peak spectrum orientated and T_2*-weighting corrections for ff_d.
- sEPG-MLE T_{2m} estimation is shown consistent with existing methods: both qualitatively with the trends seen in multi-component exponential model-based approaches – which nevertheless are seen to generally overestimate results (in absolute values) – and quantitatively with spectroscopic methods.
- Examination of the results longitudinally with time non-ambulant confirms that T_{2m} generally decreases whereas ff_a increases with disease progression.

Chapter 6. Acquisition design and optimisation

In this chapter the impact of acquisition parameter choice upon CPMG parameter estimate bias and variance, as well as the impact of coil signal combination and imaging acceleration strategies will be discussed in detail. The intention is to better characterize these data collection aspects and to provide guidance helpful in the optimal design of practical imaging protocols for future clinical studies.

Echo train length (ETL, number of echoes) and echo spacing (ΔTE, in ms) can be readily specified by the operator (at the scanner console) and their impact is explored in sections 6.2 and 6.3 respectively. Bandwidth is also operator-controlled and directly affects SNR as presented in section 6.4. Parallel imaging techniques and acquisition acceleration choices, which depend on the scanner manufacturer's software, settings for this, and coils and coil combination methods used are discussed in section 6.5. Overall discussion and conclusions are given in section 6.6.

6.1 Introduction

Recalling the discussion regarding SNR in section 2.1.9, the generic expression describing the factors affecting this was given as

$$SNR \propto \Delta x \cdot \Delta y \cdot \Delta z \cdot F_{SE} \sqrt{\frac{NSA \cdot N_{PE} \cdot N_{FE}}{BW}}$$

where $F_{SE} \propto \left(1 - e^{-\frac{TR}{T_1}}\right) \cdot e^{-\frac{TE}{T_2}}$ is the sequence contrast-specific signal decay factor, in this case for a MESE type sequence. In practical terms the parameters that can

directly and easily be adjusted for a given MESE acquisition are the echo spacing (ΔTE) and echo train length (ETL), and the chosen bandwidth (*BW*). *TR* plays also a role via T_1 contrast effects (section 3.3.2) and the total scan duration. Choice of the pixel dimensions affects SNR and influences partial volume and chemical shift effects, thus affecting also the parameter estimation. Multi-element coil use and parallel imaging reconstruction affect the acquired data quality predominantly by influencing the noise statistical and spatial distribution (compared to simple single coil 2-channel signal detection) and potential errors due to the k-space sampling trajectory.

This chapter investigates the impact upon sEPG model parameter estimation of these acquisition parameter choices using simulation experiments, test object acquisitions and examples of *in vivo* data. For the phantom experiments described in this chapter the ISMRM/NIST System Phantom as described in sections 3.1 and 3.2.10 was examined again at 3T (Siemens Magnetom Prisma) using a transmit/receive ('TxRx') head coil, or alternatively for comparison a 20-channel head & neck receive coil (section 6.5), with a conventional MESE imaging sequence (TR= 3500ms, varying ΔTEs and ETLs as described in the following sections, with a single 2 mm slice at the T_2 spheres level of the phantom, matrix 210x210, in-plane resolution 0.8x0.8mm). For the *in vivo* experiments data from the motor neuron diseases study patient acquisitions (Chapter 4) were used with varying ETLs as presented at the end of section 6.2.

6.2 Echo train length

It is important to establish the effect, assuming for the time being a constant in*ter*-echo spacing, of ETL upon parameter estimation. It might be assumed that increasing ETL will improve parameter estimation accuracy, but a compromise may be required since extended ETLs may reduce the number of slices that fit into a given *TR* in multi-slice imaging, thereby reducing anatomical coverage.

As a first step to explore the impact of varying ETL upon practical CPMG data acquisition parameter estimation accuracy, results from a single component sEPG model simulation experiment are presented in Figure 6.1 below. T_2, B_1f, α and noise σ_N were estimated by MLE minimization for a single component sEPG signal model using 1000 replicates at each SNR (25, 50 and 75, after adding Rician noise) for two representative ETLs (12 and 64) and with 10ms *TE* to match practical clinical acquisitions. The ETL of 12 was chosen to match the shortest test-object experiment value, and 64 as the longest likely to be practical for *in vivo* imaging, e.g. the longest that can fit into a reasonable TR to achieve the desired (multi-slice) anatomical coverage. An important additional reason for examining such long echo trains where practically all signal will have decayed after the 20th echo (for T_2=30ms) is investigate the effect of more fully sampling the Rician noise distribution which plays an important role under the MLE approach proposed in this work. As can be seen in both cases there was a negative bias in median estimated T_2 relative to the ground truth value, which decreased as SNR increased. In this case it appeared that this bias was slightly worse for the longer ETL. The comparative dispersion of the T_2 estimates (inter-percentile ranges) was similar for both ETLs.

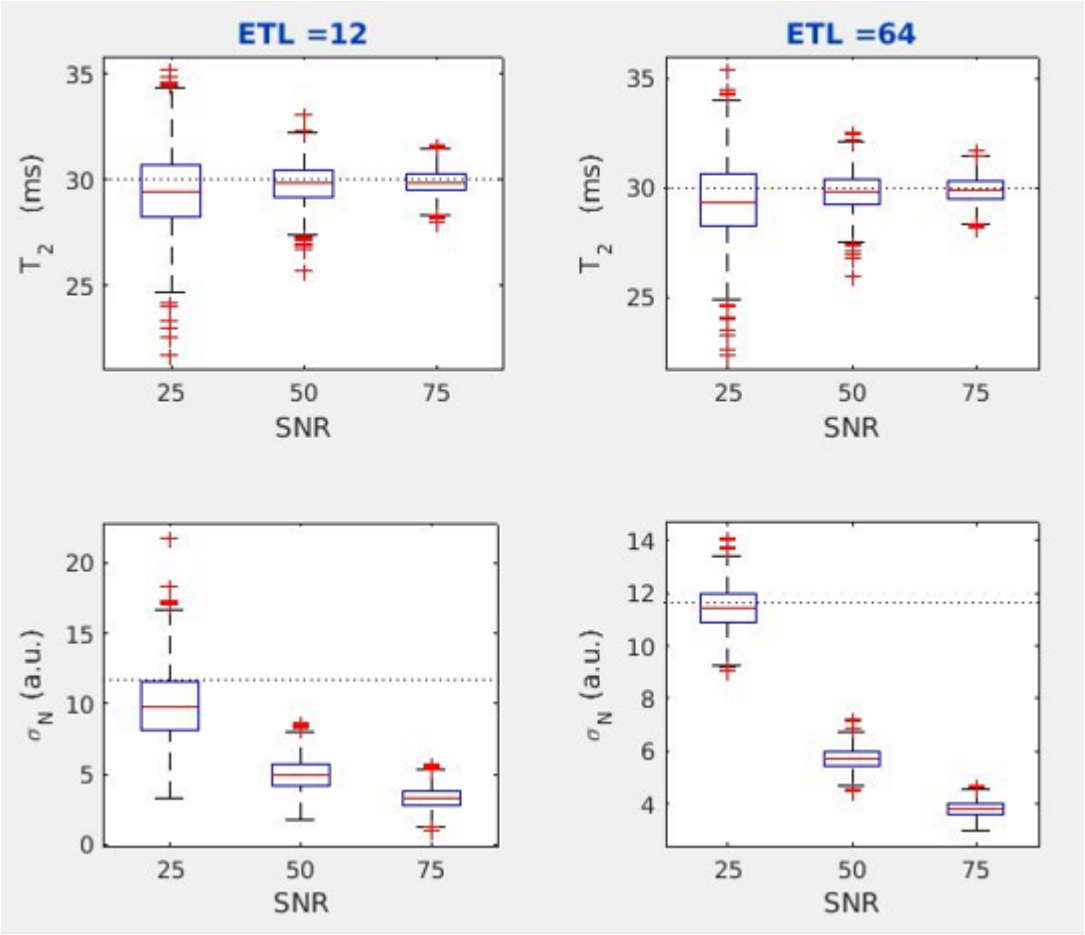


Figure 6.1: Comparison of the effects of the largest available differences in ETL (between 64 and 12) for 3 different SNRs in a single component (T_2) model simulations experiment. Lower SNR appears to compromise estimation in terms of bias and precision more than reduced ETL and longer ETL improves noise σ_N estimation but not T_2 estimation. Dotted lines represent the ground truth T_2 (30ms) (top panels) and ground truth highest noise σ_N (for SNR 25) (bottom panels).

On the other hand the model Rician noise σ_N is more accurately estimated with longer ETL. A plausible explanation for these opposite trends in T_2 and noise σ_N estimation bias may be that for shorter ETL, overall signal along the echo train is higher, since for the ground truth T_2 in these experiments the signal reaches the noise floor after the 20th echo approximately. This would be expected to directly affect the cumulative loglik as the sum of the logarithms of the Rician PDFs of the entire echo train (section 2.2.3), since noise σ_N impacts their widths.

To provide validation of the single T_2 component simulations experiment results, the effect of ETL on parameter estimation was investigated experimentally in test objects acquisitions, fitting the data pixel-wise again with a single component model appropriate to the available phantom. A single component sEPG model and the MLE approach was used as described at the phantom data validation stage (section 3.2.10) of the proposed sEPG-MLE method introduced in Chapter 3. ΔTE was 9.9ms and ETL values of 17 and 12, chosen to match the two *in vivo* protocols used in Chapter 4 and in Chapter 5, were compared yielding the results presented in Figure 6.2. Shortening the ETL from 17 to 12 caused a marginal increase (≤ 1ms) in the median estimated T_2, for nominal T_2s < 97ms, while overall estimates remain largely consistent with the manufacturer-provided calibration values. This tendency is consistent with the results of the simulations experiment (Figure 6.1) and additionally suggests that the phantom ground truth T_2 values may in reality be higher than those estimated (and not lower). The larger deviation for the highest T_2 133ms sphere may be due to a phantom calibration error or with slower decays (longer T_2) being intrinsically associated with greater T_2 uncertainty (section 3.2.10). The inter-percentile ranges appear to become larger with higher T_2 and wider for the shorter ETL. The average SNR for this experiment (with respect to the highest echo, section 3.1) was estimated empirically to be 50, comparable level with the simulation experiments described above.

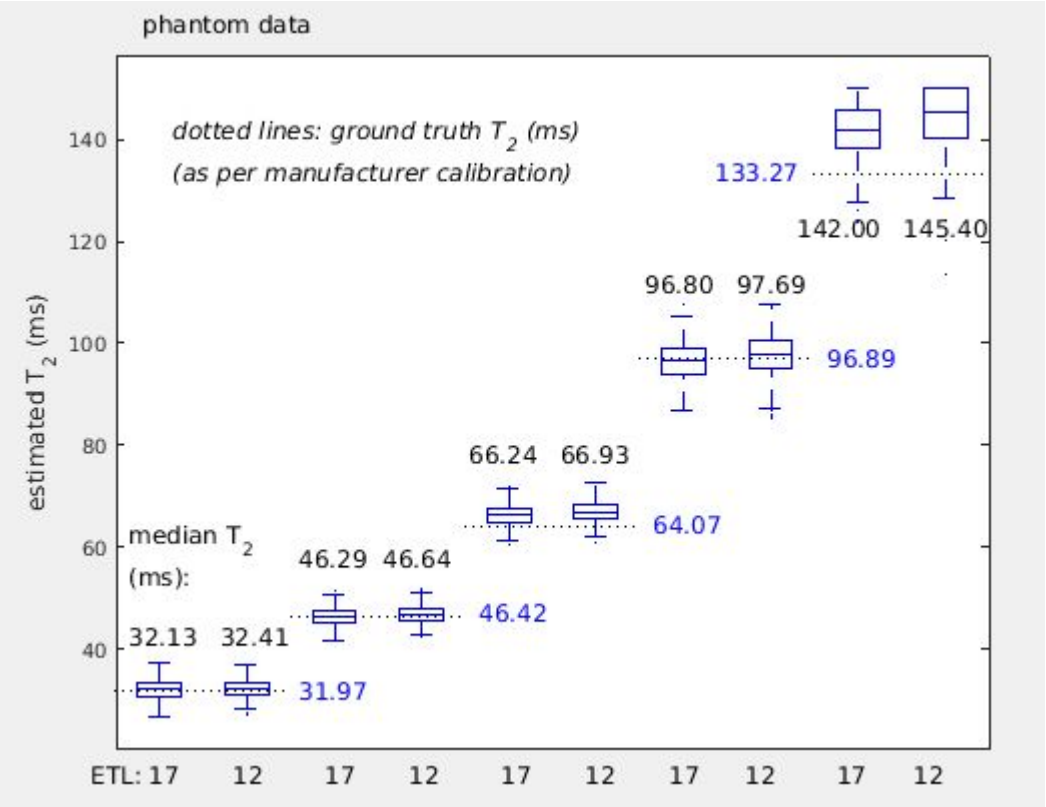


Figure 6.2: The effect of reducing the echo train length from 17 to 12 echoes in phantom CPMG data fitting for five different T_2 m spheres. T_2 estimates (reporting the median from each phantom T_2 sphere/ROI) remain close to the ground truth values (in blue, dotted lines) however marginal increase is observed for the shorter echo train.

An important point that should be taken into account here is that the accuracy and precision of the ground truth T_2 values (as per manufacturer calibration) is not known and they can only be considered as the best available approximation.

To proceed to cases comparable to real *in vivo* studies, further simulation experiments were conducted where, for a multi-component model including both water and fat signals as developed in Chapter 3, 1000 sEPG signal replicas were produced per ETL investigated and Rician noise added to achieve three different levels of SNR (25, 50 and 75), for a range of *ff* between 0 and 100%. ETL was initially varied across successive values between 12 and 64 and *TE* was 10ms to match typical protocol settings in clinical studies presented in this thesis and in the literature.

Simulations were performed in MATLAB using MLE minimization and the multi-component sEPG model described in section 3.3.2, Equation 14, to estimate muscle water T_2 (T_{2m}), apparent fat fraction (ff_a), B_1 field factor (B_1f), overall amplitude (α) and noise standard deviation (σ_N). This was also the forward model, with an incorporated secondary 2-component model describing fat signals.

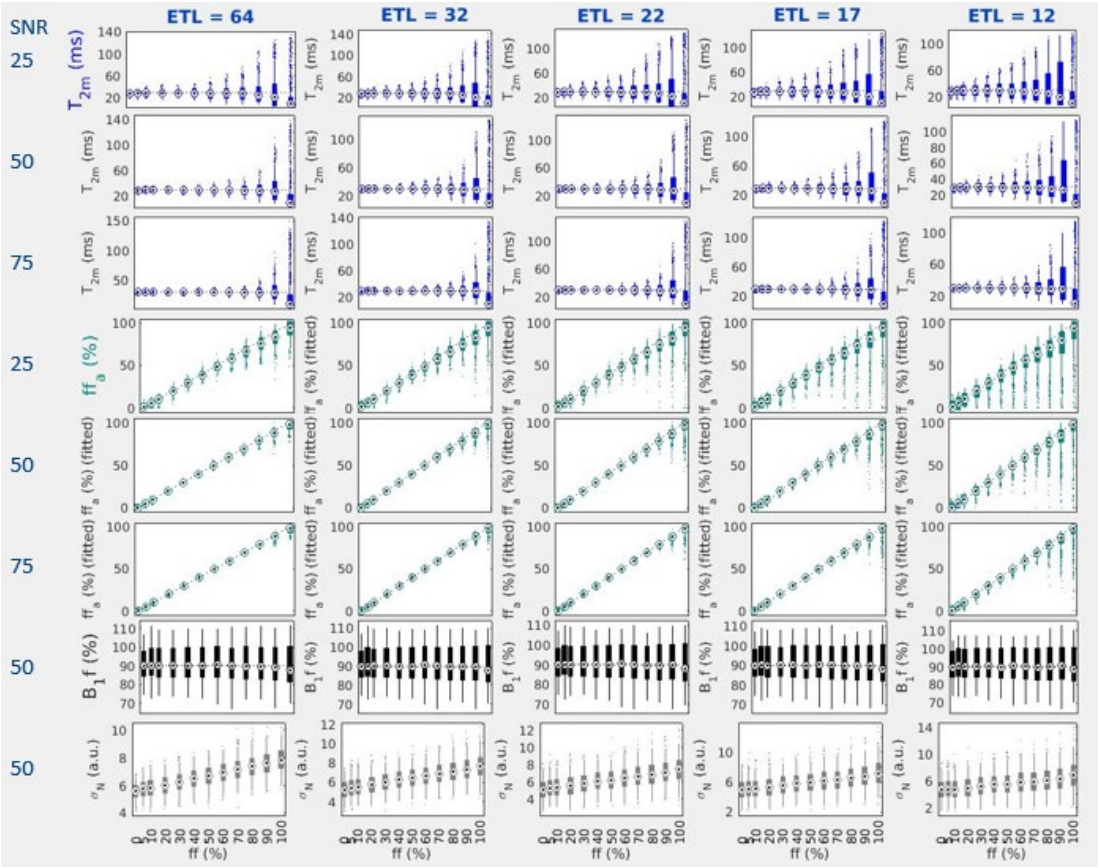


Figure 6.3: Overview of the dependence of the T_{2m} and ff_a estimates on varying ETL (in a range between 64 and 12 echoes) and 3 different SNRs (25, 50 and 75), and the corresponding B_1f and σ_N estimates (here shown only for SNR 50), based on simulations (1000 repetitions for each set). Ground truth values were 30ms T_{2m}, ff varying between 0 and 100% as shown on the x-axes of the last (σ_N) row, σ_N varied as per highest echo (between 5.8 and 8.1a.u.) to produce the desired SNRs mentioned above and B_1f was 90%.

Ground truth values were 30ms T_{2m}, ff varying as above, σ_N varied as per highest echo (between 5.8 and 8.1a.u.) to produce the aforementioned desired SNRs and B_1f was 90%, as a representative value typical in real acquisitions and in agreement with the

experiments of Chapter 3. An overview of the resulting parameter estimation behaviour is shown in Figure 6.3.

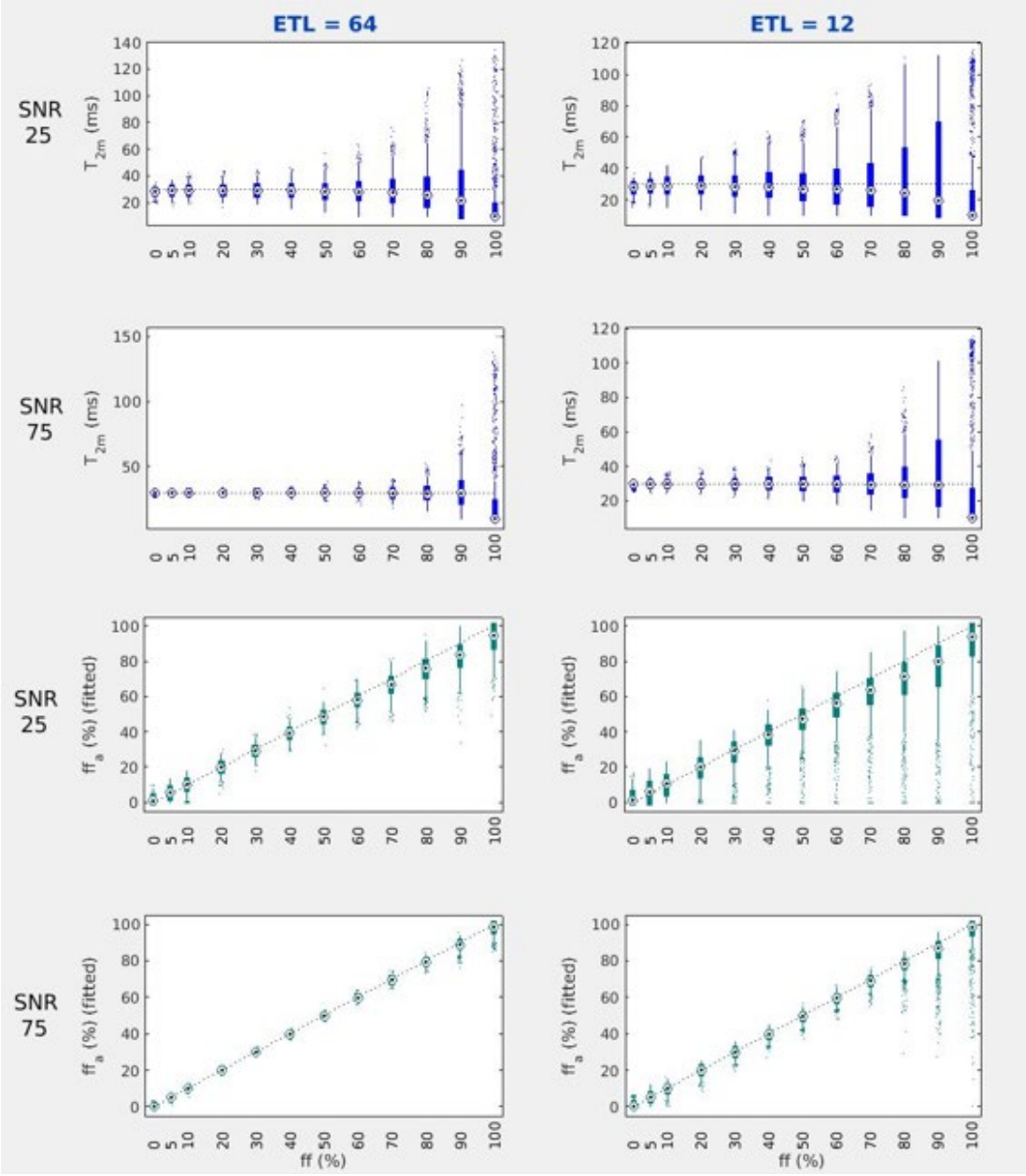


Figure 6.4: Comparison of the effects upon multi-component sEPG model parameter estimates of the most extreme in ETL and SNR conditions in the simulations experiment described in Figure 6.3. In general lowering SNR appears to impact on estimate bias and precision more than reducing ETL. The whisker and box plots are displayed in compact formatting where solid (middle) bars indicate the 25th and 75th percentiles, the whiskers are displayed as lines (extending to the most extreme data points not considered outliers), and the outliers are plotted as dots.

In Figure 6.4 results are shown in closer detail to compare in more detail the simulations most different in terms of the ETL and SNR, and in Figure 6.5 the corresponding differences in bias and SD are shown.

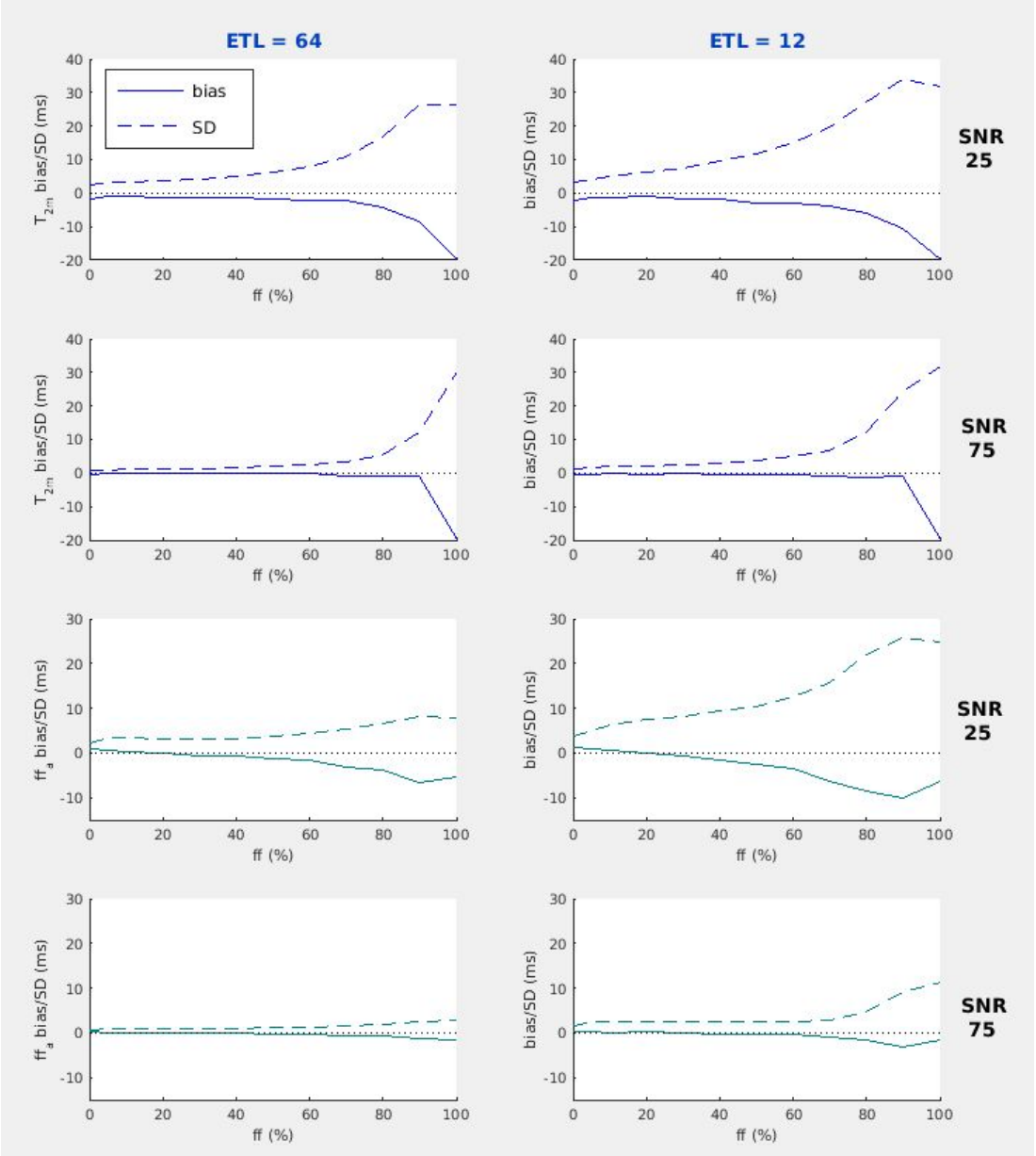


Figure 6.5: The bias and SD in the parameter estimates corresponding to the results presented in Figure 6.4.

SNR clearly appears to have a bigger impact upon the parameter estimation than ETL, and more in terms of accuracy (bias) than in precision (SD).

Bias with respect to T_{2m} estimation is more marked for high *ff* values where it appears negative (T_{2m} underestimation), and similar trends are seen for ff_a where additionally some positive bias appears at low *ff* ranges.

Analogous results for *in vivo* data acquisitions are demonstrated in Figure 6.7., using MLE minimization and the same multi-component sEPG model as described in the initial simulation experiment above (Figure 6.3) to estimate thigh cross-section median T_{2m}, ff_a, B_1f, α and noise σ_N (Figure 6.6) for a healthy volunteer and an ALS patient (Chapter 4). This is to represent both the situation where variance is predominantly due to fitting variations (narrow parameter distributions, healthy volunteer) and the case where disease also impacts results (wider parameter distributions, patient).

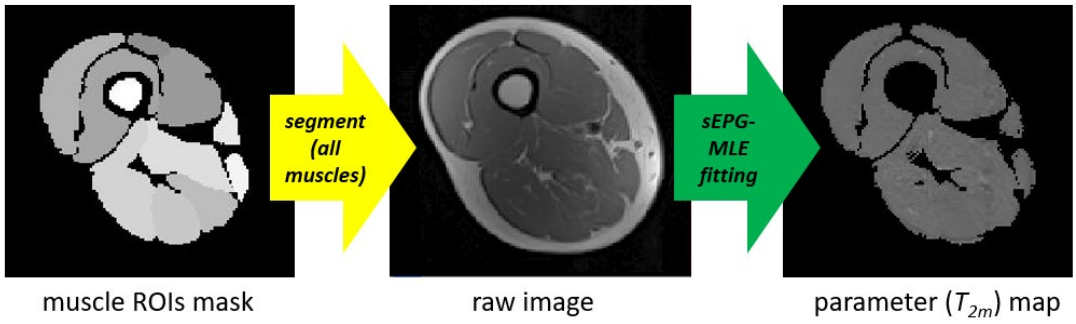


Figure 6.6: In this experiment all muscle ROIs were used to produce parameter estimate maps and provide the respective thigh cross-sectional ('global segmentation') median.

SNR in these datasets was mostly above 50 (sections 3.1 and 4.4). Three different ETLs (22, 17 and 12) were investigated, selected as subsets of the echoes from a single ETL=22 acquisition.

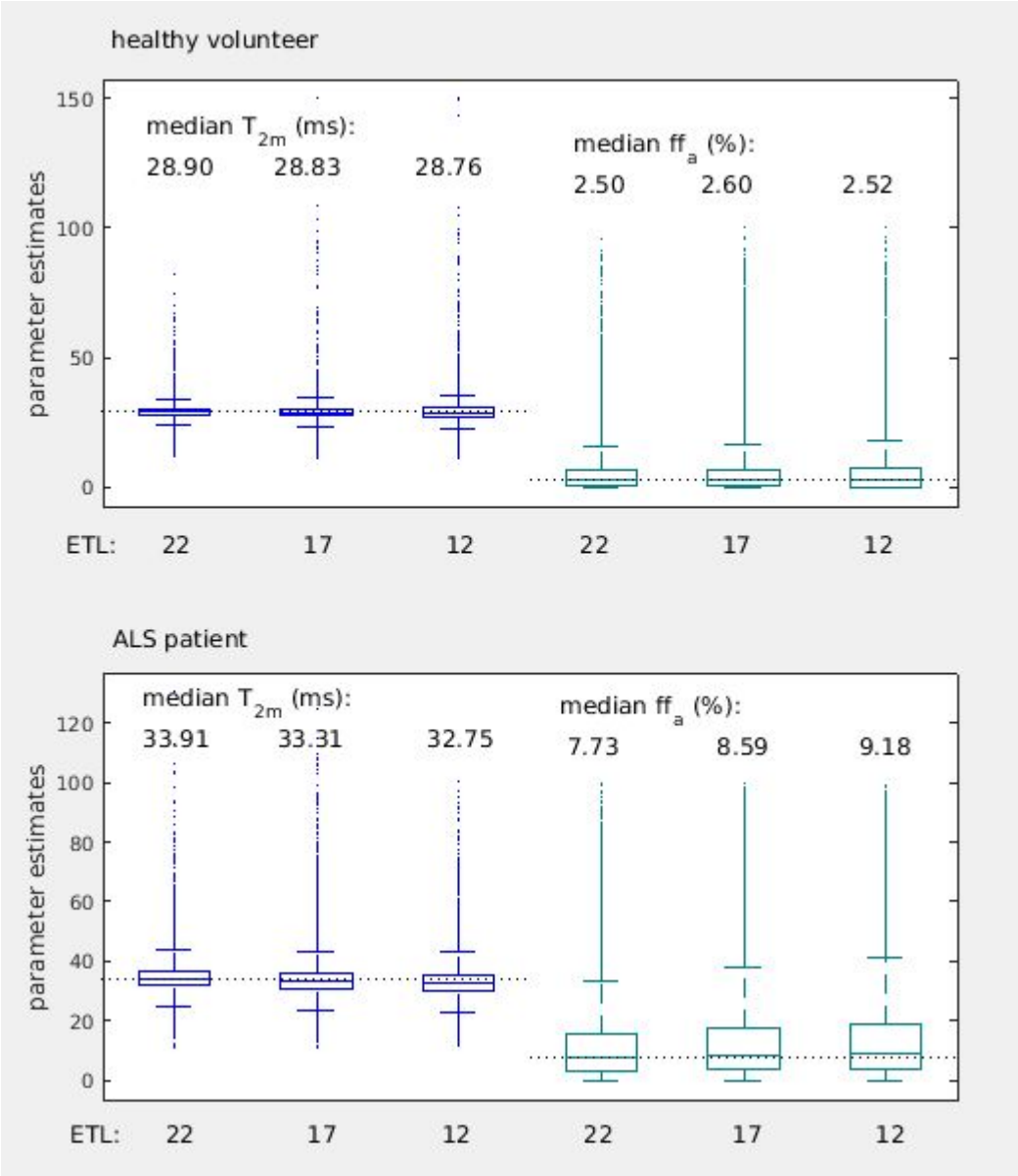


Figure 6.7: The effect of reducing ETL from 22 to 17 and then to 12 echoes for CPMG *in vivo* data on muscle T_{2m} and ff_a estimates for a healthy volunteer and an ALS patient. Parameter estimates remain within the expected ranges, but with estimated T_{2m} decreasing and ff_a conversely increasing when reducing the ETL used in the fitting.

Median estimated T_{2m} remains very close to the expected values with and without presence of disease, showing a slight decrease for shorter echo trains – mostly less than 1ms – in healthy muscle, and more for the patient data. Conversely estimated ff_a shows a clearer tendency to increase at lower ETL for the patient data while differences in the healthy muscle (very low ff_a range) are rather imperceptible. The opposite trends seen in the single component T_2 (phantom and simulations) vs the

multi component T_{2m} (in vivo) behaviour (i.e. increase vs decrease) here are likely due to parameter correlation (interaction), predominantly between T_2, ff_a and σ_N, as this is demonstrated particularly for low *ff*s in Figure 6.5, and as was described in greater detail in previous chapters.

The main outcome from the above experiments is that SNR appears to impact parameter estimation more drastically than ETL mainly in terms of bias. As was seen for many sub-optimal quality datasets in both Chapter 4 and Chapter 5, artefacts (due to B_1 inhomogeneity, motion etc.) are in general associated with loss of SNR and negative bias in the parameter estimation. For reduced ETL the multi component MLE fitting to the sEPG model shows mostly negative bias in T_{2m} estimation and positive bias in ff_a at low *ff* ranges (up to 20% roughly) and changing to negative bias at higher *ff* as ETL is decreased. Decreasing SNR caused increasing T_2 underestimation, consistent with observations for the single component sEPG model fitted to single component signals, however in contrast decreasing ETL in this case caused reduced bias in estimated T_2, likely due to the effectively higher overall SNR of a shorter echo train, bringing the estimate closer to the ground truth values. Parameter estimate correlation and trade-off also occurs and depends in turn on the T_2 and *ff* ranges of interest and on SNR; this may explain the differences between the single- and multi-component model fitting results. Overall, intermediate ETLs (i.e. above 15) may best serve the purpose of achieving adequate precision while avoiding excessively long echo trains which may be impractical in terms of sequence design, power deposition and anatomical coverage, and typically provide low signal at the later echoes.

6.3 Echo spacing

This section investigates the effect upon T_{2m} and ff_a estimation of changing the inter-echo spacing ΔTE for a fixed ETL. The approach adopted in this section is conceptually based on the work related to echo spacing optimisation published by Dula et al. (Dula et al., 2009). In that work the precision of the parameter estimation is assessed via each parameter estimate's 'SNR', defined as its mean over its SD. This was investigated by those authors using both Cramer-Rao lower bound related calculations and simulations fitting to simulated signals with Gaussian noise added (similar to the work presented in previous chapters of this thesis). The model used in that work was bi-exponential with T_{2a}, T_{2b} and M_a, M_b the T_2s and amplitudes of the two respective exponential components. Examples of their results are shown in Figure 6.8.

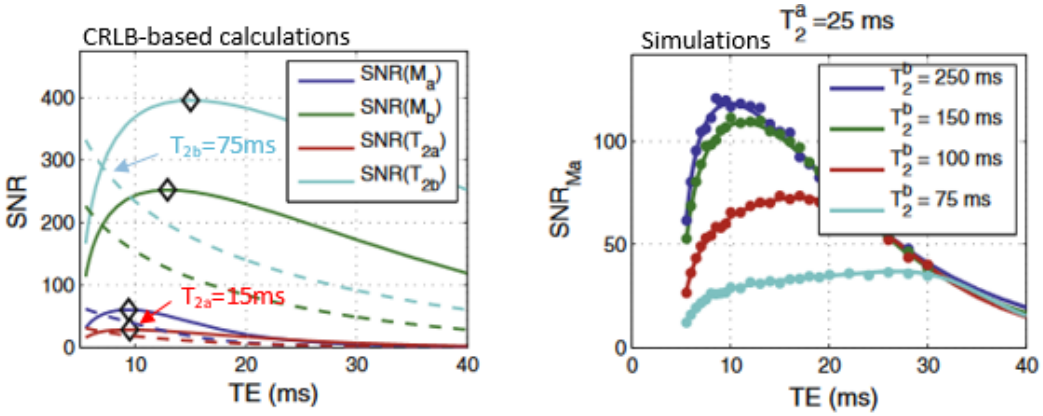


Figure 6.8: Left: plots of SNR of four estimated parameters (indicated by line colour) for a bi-exponential model of transverse relaxation, as a function of ΔTE and with fixed ETL = 200. Also shown as dashed lines are the results from the same calculations made without considering the minimum sequence 'play-out time' determined from the duration of the RF and gradients pulses required for a real world sequence used. Right: example plots of SNR of the short T_2 component amplitude estimate as a function of TE, taken from a range of different T_{2a} and T_{2b} values (Dula et al., 2009)

It was demonstrated that the dependence of estimated parameter SNR upon ΔTE showed quite broad maxima, and the optimal *TE*s, i.e. those providing the highest parameter estimate SNRs, were independent of the two components' relative amplitudes, the baseline SNR, and the number of echoes. Therefore a similar dependence of estimated T_2 SNR for the single component case, where the amplitude of the second bi-exponential component is effectively fixed at zero, on ΔTE may be expected. Key to this approach however were the assumptions related to the available ΔTE choice. This is considered to be limited by the time required for fixed events like the RF refocusing pulse, spoiler gradients and ramp-time delays, such that the only way to minimise ΔTE is by reducing the acquisition time which will in turn increase bandwidth (BW). The inherent trade-off in this acquisition time reduction is that because the image SNR is inversely proportional to the square root of BW, so reducing ΔTE through increased BW also decreases the SNR. Reducing ΔTE in principle improves the parameter estimates' precision, however the concomitantly increased BW leads to decreased SNR which deteriorates this precision. This appears to be the mechanism driving the behaviour of the parameter estimate SNRs which under this condition show a maximum (denoted by the diamond symbols) with respect to ΔTE, including in particular for the solid red and light blue curves at the left of Figure 6.8 corresponding to the behaviour of the two components' respective T_2s. On the other hand the dashed red and light blue lines show the ΔTE dependence of the two T_2s SNR in the theoretical case that the finite RF pulse and gradient durations may be ignored, where the estimate's SNRs simply rise to maximum as *TE* tends to zero.

When attempting to use the above findings to inform the effect of ΔTE upon parameter estimation for real CPMG imaging data, the effects of finite RF and gradient duration are clearly significant, defining both the range of practically accessible minimum *TE*s for a particular sequence, and the effect of that choice upon estimate precision. In the following experiments acquiring data from a test object it was seen that in practical terms, due to the minimum RF and gradient pulse durations, the minimum *TE* available was 8ms for imaging parameters chosen (454 Hz/pixel bandwidth, 2 mm slice thickness, matrix 210x210, in-plane resolution 0.8x0.8 mm) (Siemens Magnetom Prisma scanner) since the sampling time must play out between the RF and non-read gradient pulses, the durations of which are fixed for a given bandwidth and slice thickness. Four data sets were acquired with ΔTE varied between 8, 10, 12 and 15ms, TR=3500ms, ETL=22, a single 2 mm slice at the T_2 spheres level from which five T_2 ROIs (discs) of 113 pixels each were extracted, matrix 210x210, in-plane resolution 0.8x0.8 mm using a simple 2-channel quadrature transmit/receive ('TxRx') head coil. T_2s were estimated by fitting a single component sEPG model to the data using the MLE method. The dependence of the estimated T_2 SNR on varying *TE* is shown in Figure 6.9 (estimated T_2 SNR=$\frac{median\ T_2}{deviation\ T_2}$, where '*deviation* T_2' is defined as half the width between the 15.9 and 84.1 percentile ≈ 68.2% of the data, equivalent to 1 SD of a Gaussian distribution, to account for non-normally distributed data).

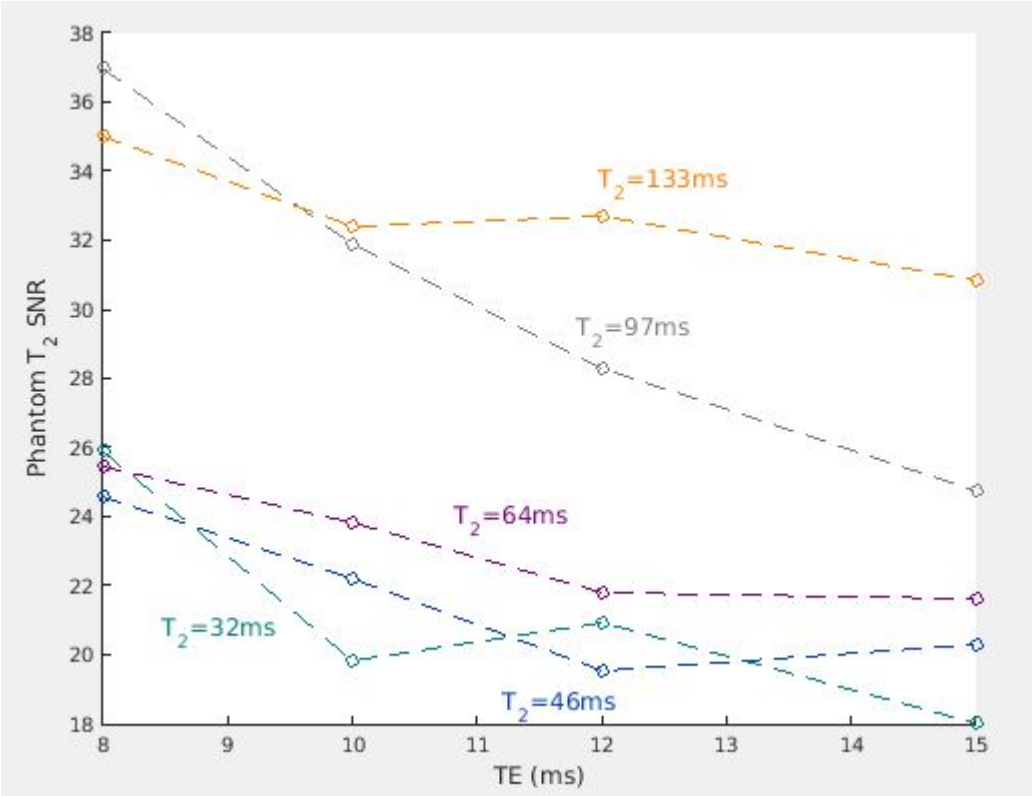


Figure 6.9: The effect of varying ΔTE between 8, 10, 12 and 15ms (circles) on T_2 estimation, in phantom CPMG data fitting for five different T_2 spheres. Estimated T_2 SNR ($\frac{median\ T_2}{deviation\ T_2}$) appears to decrease in general when increasing ΔTE.

As can be seen the effect of varying *TE* in these acquisitions broadly mimics the behaviour shown by the dashed curves in the left panel of Figure 6.8, corroborating the validity of the results of the previous calculations in real conditions. Estimated T_2 SNR ($\frac{median\ T_2}{deviation\ T_2}$) decreases with increasing ΔTE, even though not strictly monotonically, which is likely to be related to variabilities involved in real experimental settings and the fact that T_2 SNR is a ratio of large numbers (T_2 estimates) over small (their SDs) with relatively small ROIs of 112 pixels each (section 6.1). Importantly, Gibbs ringing effects were visible in most raw images, which combined with motion/object stability and RF standing wave effects may impact the

above ratio (T_2 SNR). Nevertheless the overall behaviour seen in these experimental results is consistent with the theoretical predictions.

In general shorter *TE*s are estimated more precisely, as expected due to the higher SNR at those times. The ability to study the effects of further decreasing *TE* obviously depends on the settings accessible within the software of individual scanners due to the time frames of the sequence dependent fixed events (RF pulse shapes and duration, spoiler gradients, slew rates etc.). Given the limitations of the available systems, investigating this further is beyond the scope of this work.

6.4 Bandwidth and SNR

Continuing now to a discussion of the effect of BW on parameter estimate precision and bias, in this section an experiment investigating the effects of varying the receive bandwidth is presented. A phantom acquisition as described in the previous section, with *TE*=10ms and all other parameters identical was conducted. This was followed by two additional acquisitions, one with a halved and one with a roughly doubled receive bandwidth, keeping all other parameters the same. The BWs for the 3 sequences were 454, 227 and 930 Hz per pixel respectively (nominal, half and double bandwidth). Image noise levels were estimated, firstly directly from the images using the signal background method (section 3.1) and secondly as an estimated parameter on fitting the data using the sEPG-MLE method proposed in this thesis. The results are presented in Table 6.1 below.

Table 6.1: Noise levels dependence on sequence receive bandwidth

Bandwidth	**Nominal**	**Half**	**Double**
Noise σ_N (a.u.) – direct measurement	35.3	25.3	50.6
'$\sqrt{2}$ rule' based expectation	SD_0=35.3	$SD_0/\sqrt{2}$=24.9	$SD_0 \cdot \sqrt{2}$=49.9
Noise σ_N (a.u.) – sEPG-MLE estimated	17.5	12.5	23.8

The expected '$\sqrt{2}$ rule' relating noise σ_N and BW based on the factors affecting SNR previously mentioned (sections 2.1.9 and 6.1) appears to accurately hold between the three acquisitions, both for the empirical/experimental measurements and for the sEPG-MLE estimation. While in general the sEPG-MLE method was already seen in previous chapters to mildly underestimate noise levels, in this case its noise σ_N estimates were consistently a factor of 2 smaller than the directly measured values. This may be related to the potential bias both in the sEPG-MLE estimation – where σ_N is also one of the estimated parameters – due to imperfections as motion artefacts (e.g. test object vibration effects) and magnetic field inhomogeneity, and in the direct measurement where the assumptions of invariant noise σ_N across the image background may not hold due to the implied coil conditioning and image reconstruction technique.

The effect of increasing the acquisition bandwidth while maintaining all other acquisition parameters constant – on T_2 estimation is demonstrated for the five phantom sphere-ROIs in Figure 6.10 below.

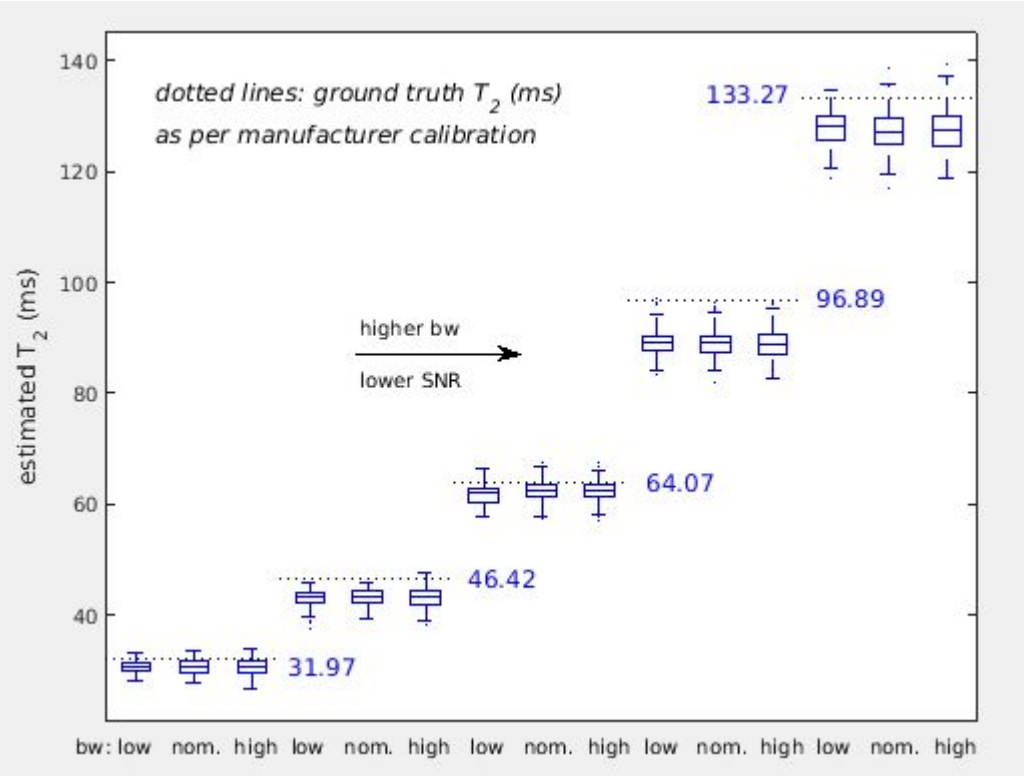


Figure 6.10: The effect of increasing bandwidth by a factor of 2 below and above the default settings in phantom CPMG data fitting for five different T_2 spheres.

The behaviour seen appears to be very similar with the respective (SNR related) simulations results produced in sections 3.2.5 and 3.3.1 where small changes in the T_2/T_{2m} medians were seen, predominantly an increasing negative bias as SNR decreases, accompanied by noticeable broadening of the parameter estimate distributions.

Note: the T_2 estimates yielded in this experiment were 3 to 8% lower in general compared to the phantom validation results of section 3.3.4. Temperature differences are likely to play a role in this. The temperature of the object during the initial validation acquisition was recorded as 19.9°C, very close to the 20°C (293°K) temperature where the ground truth values calibration was conducted according to the manufacturer, whereas during the latter acquisition the recorded object

temperature was 19°C. T_2 values for this phantom have been shown to decrease with decreasing temperature (Keenan et al., 2016). Additional processes related to diffusion effects (temperature and bulk motion dependent) and phantom positioning and geometry, are also likely to contribute to the aforementioned T_2 estimation variations between different scans.

6.5 Coil combination and acquisition acceleration effects

In routine *in vivo* acquisitions it is common to improve SNR and reduce scan duration by using different coil combinations methods (presupposing the use of multi-element coils) and image acceleration techniques, which may impact on parameter estimation, in addition to the effects of sequence parameter choices previously described. These effects are explored in this section.

In acquisitions using multi-element coil combinations the signal in each image pixel after reconstruction is expected to follow a non central χ^2 distribution with degrees of freedom depending on the number of coil elements, assuming equal and non-correlated noise σ_N between them (Bevan, 2013). Within the sEPG-MLE algorithm developed in this thesis the signal is assumed to follow a Rician distribution , which is the simplest case of a non central χ^2 distribution with 2 degrees of freedom, corresponding to the 2 quadrature modes of a 2-channel quadrature transmit/receive ('TxRx') coil used in the simplest type of magnitude reconstructed image acquisitions. In Figure 6.11 the theoretical dependence of the reconstructed signal distribution on the number of coil elements and SNR (based on the ground truth noiseless signal) is shown.

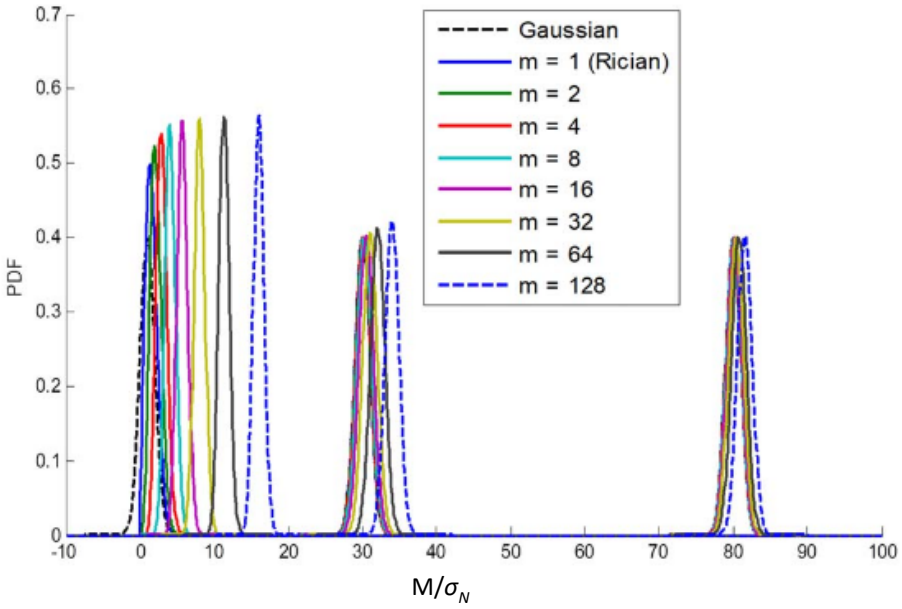


Figure 6.11: The Gaussian and non central χ^2 signal distributions as a function of SNR (M/σ_N, where M is the measured signal and σ_N the noise SD) for different numbers of coil elements m. Results are shown for ground truth noiseless signal of 1, 30, and 80 units and unit σ_N, corresponding to low, moderate and high SNR respectively (Bouhrara and Spencer, 2018)

As can be seen for SNR above 30 and number of elements less than 32 the signal distribution approximates the Rician reasonably closely. Since this corresponds roughly to the type of acquisitions (coils used and image quality) most commonly encountered in the datasets analysed in this thesis, the Rician pdf assumption within the algorithm used will also be considered to be an adequate approximation in the experimental work presented in this section.

In the test-object acquisition experiments presented here the following steps were taken:

1. Use of the simple 2-channel quadrature transmit/receive ('TxRx') head coil, and 'performance' gradient option offered at the scanner interface. This system gradient option allowed the minimum TE of 8ms used in section 6.3.
2. The acquisition was then repeated with the scanner 'fast' gradient option – the more commonly used option – selected, and everything else identical.
3. The acquisition was then repeated using a 20-channel head & neck receive coil using the scanner sum of squares (SoS) option for combination of the signals from the multi-element coil (MC) and everything else identical.
4. Repetition of the MC acquisition combined with GRAPPA and R=2 (phase encoding reduction factor due to k-space under sampling) parallel imaging mode.
5. Alternative reconstruction of the above MC images using the on-scanner 'adaptive coil combine' (ACC) reconstruction (instead of SoS).
6. Reconstruction also of the MC-GRAPPA images using ACC.

All other acquisition conditions were as described in section 6.3 above. Parameters were estimated using the single-component sEPG-MLE method and model described in section 3.2.10. A comparison of the result for each of the above described procedures is summarised in Table 6.2. Parameter estimate SDs for all acquisition methods were comparable and stable overall, and thus are not shown for the sake of clarity.

Table 6.2: Dependence of the sEPG-MLE parameter estimates (reporting the median from each phantom T_2 sphere/ROI) upon RF coil-element signal combination method, gradient mode, and parallel imaging acceleration method

Nominal T_2 sphere (ms)	**31.97**	**46.42**	**64.07**	**96.89**	**133.27**
	Estimated T_2 (ms)				
'performance' gradient	30.59	43.08	61.53	88.85	126.80
'fast' gradient (TxRx)	30.90	43.26	61.78	89.37	126.00
+ multi-coil (sum of sq.)	30.78	43.48	62.99	90.06	129.08
+ GRAPPA (R=2)	30.61	43.48	61.89	89.36	126.61
multi-coil (ACC)	30.67	43.79	61.93	88.90	126.89
+ GRAPPA (R=2)	30.87	44.02	61.36	87.48	125.94
	Estimated B_1f (%)				
'performance' gradient	106.52	110.45	107.54	99.20	108.09
'fast' gradient (TxRx)	104.94	109.69	107.79	98.95	107.33
+ multi-coil (sum of sq.)	111.58	112.05	106.53	99.60	95.16
+ GRAPPA (R=2)	110.54	111.09	108.68	93.98	108.98
multi-coil (ACC)	87.17	85.85	107.54	101.01	108.28
+ GRAPPA (R=2)	88.06	86.86	109.28	108.40	109.10
	Estimated amplitude α (intercept) (a.u.)				
'performance' gradient	5024.73	5104.51	4897.87	4628.56	4642.37
'fast' gradient (TxRx)	4990.74	5059.61	4903.55	4631.22	4615.74
+ multi-coil (sum of sq.)	9138.60	9073.79	7388.88	6464.88	7081.43
+ GRAPPA (R=2)	9123.29	9031.92	7514.61	6518.64	7178.67
multi-coil (ACC)	9067.80	8964.97	7477.99	6497.07	7161.19
+ GRAPPA (R=2)	8993.74	8907.05	7543.43	6623.40	7185.91
	Estimated noise σ_N (a.u.)				
'performance' gradient	18.23	17.16	16.84	16.70	16.50
'fast' gradient (TxRx)	17.88	17.47	16.64	17.28	16.25
+ multi-coil (sum of sq.)	30.49	17.84	14.88	14.16	15.02
+ GRAPPA (R=2)	29.21	24.18	21.38	21.92	22.87
multi-coil (ACC)	20.22	17.91	15.21	14.36	15.15
+ GRAPPA (R=2)	28.12	26.08	22.06	22.88	22.95

As can be seen T_2 estimates were the most stable compared to the remaining estimated parameters – despite the overall temperature-dependent reduction (section 6.4) relative to the calibration values. B_1f estimation was seen to be most affected when switching (from TxRx or SoS-MC) to ACC-MC acquisitions and to a lesser extend due to use of GRAPPA and sum of squares MC techniques. Due to the different coils used, amplitude (intercept) α estimates were markedly increased when using MC acquisitions, and more so for the shorter T_2s, whereas these appeared

broadly stable when using the simple 2-channel coil. The latter is also observed for the noise σ_N estimates which additionally appear to increase mainly due to GRAPPA and again more for the shorter T_2 spheres. Choice of gradient type ('fast' or 'performance') is not seen to significantly affect estimation, for all parameters.

Repeating this experiment for *in vivo* data (thigh images from a healthy volunteer using the sequence described in section 4.2), this time using the multi-component model described in section 3.3.4, produced the results presented in Table 6.3.

Table 6.3: Dependence of the sEPG-MLE parameters (reporting the median from each anatomical ROI) on parallel imaging and acceleration techniques from *in vivo* data

Anatomical ROI	**anterior**	**medial**	**posterior**
	Estimated T_2 (ms)		
multi-coil – sum of sq.	29.98	30.23	28.82
+ GRAPPA (R=2)	29.45	29.50	28.12
multi-coil – ACC	30.92	31.64	30.93
+ GRAPPA (R=2)	30.50	30.83	30.49
	Estimated ff_a (%)		
multi-coil – sum of sq.	4.81	8.93	8.36
+ GRAPPA (R=2)	4.91	8.56	9.65
multi-coil – ACC	2.82	6.39	5.06
+ GRAPPA (R=2)	2.67	5.99	5.14
	Estimated B_1f (%)		
multi-coil – sum of sq.	99.67	99.95	102.33
+ GRAPPA (R=2)	99.73	100.11	104.08
multi-coil – ACC	99.52	99.40	99.39
+ GRAPPA (R=2)	99.52	99.50	99.43
	Est. amplitude α (intercept) (a.u.)		
multi-coil – sum of sq.	1814.18	1349.17	1178.53
+ GRAPPA (R=2)	1806.63	1384.85	1193.64
multi-coil – ACC	1798.64	1330.15	1139.32
+ GRAPPA (R=2)	1786.78	1356.11	1146.54
	Estimated noise σ_N (a.u.)		
multi-coil – sum of sq.	3.26	2.91	2.62
+ GRAPPA (R=2)	4.14	5.22	3.73
multi-coil – ACC	3.32	2.95	2.55
+ GRAPPA (R=2)	4.19	5.44	3.61

In this experiment 3 representative uniform ROIs were analysed, enclosing the anterior, medial and posterior thigh muscle compartments, following the classification adopted in the collaborative work in our centre (Klickovic et al., 2019) mentioned in sections 4.1 and 4.5 (Chapter 4). Only MC data using the 20-channel head & neck receive coil were collected (the previous two first steps using a simple 2-channel coil were omitted as such a coil suitable for limb imaging was not available) and the multi-component sEPG model used in all previous *in vivo* data analyses in this work was used again. Similarly to the phantom results, using ACC-MC instead of sum of squares image reconstruction appears to affect parameter estimation more than the choice of using GRAPPA or not. However in this case it is T_{2m}, ff_a and to a lesser extend B_1f that are affected more, while noise σ_N and α appear largely stable. In this more complex model than that used for the phantom data, parameter interdependence between T_{2m}, ff_a and B_1f may also potentially amplify the acquisition related disparities seen in their estimates. For instance relatively higher T_{2m} values in general correlate with lower range ff_a estimates. B_1f appears relatively more stable through all of the acquisition technique combinations, and also consistent with the expected behaviour seen for example in Figure 3.45, section 3.2.6, where typically B_1 field deviations are seen at particular regions such as the posterior or anterior medial areas. This may introduce additional parameter estimation complexity in such regions, as might be the case for the last column ('posterior' ROI) of Table 6.3, where for instance, differences in the aforementioned parameters appear greater compared to the remaining results. The behaviour of B_1f for all types of acquisitions is shown in Figure 6.12, highlighting a posterior anatomical region where estimation results behaved oppositely for sum of squares and ACC-MC reconstructions. Here

there were marked increases in B_1f and ff_a, and conversely a clear decrease in T_{2m} correlates with sum of squares, and the converse for ACC-MC, while grappa introduced only marginal differences.

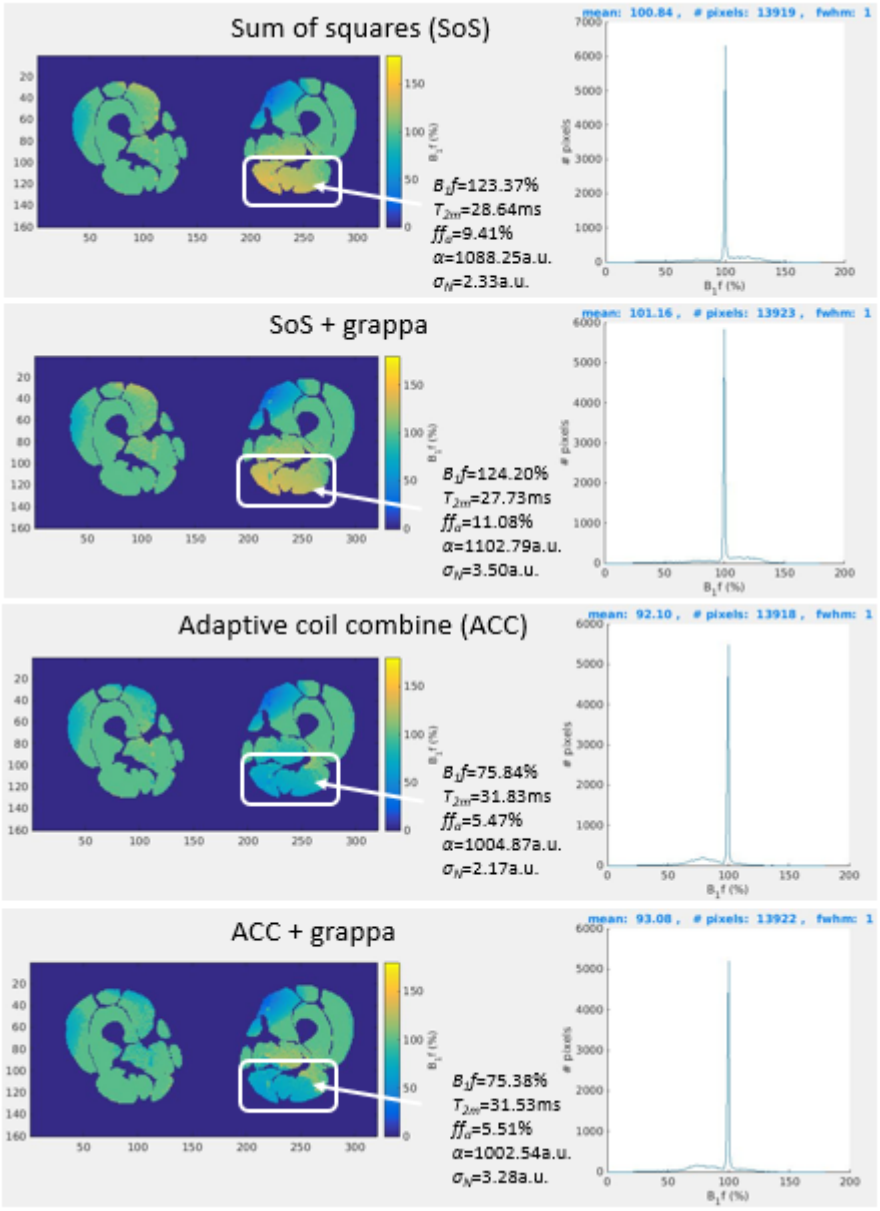


Figure 6.12: B_1f maps and histograms for all four types of *in vivo* data acquisition and reconstruction conditions. In the delineated posterior region B_1f is clearly overestimated for sum of squares reconstruction while it is underestimated for ACC-MC, and this is accompanied by clear disparities between the associated T_{2m} and ff_a parameter estimates.

Interestingly, in the case that B_1f is deliberately switched from the lower to the upper, overestimated region for the lower two ACC-MC reconstruction cases, median parameter estimates for that ROI remain relatively close to the initial ACC-MCC results, at 30.90 and 30.61ms for T_{2m} and at 4.38 and 4.60% for ff_a respectively. Equivalently in the case that B_1f is deliberately switched to the lower, underestimated region for the upper two sum of squares-MC reconstruction cases, median estimates for that ROI remain close to the previous sum of squares-MC values at 29.64 and 28.70ms for T_{2m} and at 10.87 and 12.71% for ff_a respectively. These observations suggest substantial differences between the two reconstruction methods which affect parameter estimation beyond any parameter interdependence effects. Since the *in vivo* studies presented in this thesis (Chapter 4 & Chapter 5) were based on acquisitions using ACC and GRAPPA it follows that this introduces an additional plausible source of bias in those results. Even if ACC is indeed in general superior to SoS – which itself is a topic for further research – this may not necessarily be true in the case of marked B_1 inhomogneities. The choice of method may introduce bias in the order of 1ms or one percentage unit in the T_{2m} and ff_a estimates respectively. Finally, in both phantom and *in vivo* data, noise σ_N appears to increase consistently with the expected $\sqrt{2}$-factor increase (roughly) when using GRAPPA due to the reduction (approximately halving) of the scan time.

6.6 Overall considerations and conclusions

Acquisition conditions and sequence design choices seem to be influenced by the estimated parameter interdependence discussed in other parts of this thesis (mainly Chapter 3) as well as with the available SNR and data quality. In general they are seen

to affect the estimated parameters' medians and distributions to a smaller degree than the changes caused by disease which are the primary target of the quantification. A clear exception is the disparity introduced in the parameter estimation by choosing adaptive coil combine (ACC) *vs.* sum of squares image reconstruction, where differences for T_{2m} and ff_a were seen to reach 2ms and 4.5% respectively on average, with even higher differences in particular cases of smaller ROIs. While a detailed analysis of the differences of the two methods (SoS vs ACC) in theory, design and application is beyond the scope of this work, it may be beneficial to suggest candidate reasons for the observed differences:

- The ACC technique – unlike SoS – aims to address noise variations (e.g. at dark areas of the image) and propagation of motion or flow phenomena while taking also into account spatial phase variations in the underlying signal (Walsh et al., 2000)
- This is particularly relevant for CPMG acquisitions where loss of signal may occur not only in regions of artefact or low coil sensitivity, but also at later echoes.
- It follows that in regions of compromised B_1, fast T_{2m} and low ff it may be expected to obtain relatively higher SNR under ACC than SoS image reconstruction.
- This may in turn explain the systematically longer ACC T_{2m} estimates compared to SoS T_{2m}. Similarly the concomitant lower ACC ff_a (compared to SoS ff_a) may be a result of the above signal modulations and additionally of parameter interdependence (since it signifies less signal in principle).

- The underlying assumption of simple quadrature Rician noise in the fitting algorithm instead of (either SoS or ACC) MC noise following a non-central χ^2 distribution (Bouhrara and Spencer, 2018) may play a role in the above processes.

For the remaining available options (such as ETL and ΔTE) even drastic changes in the operator-controlled sequence parameters will hardly cause differences exceeding 1ms in T_2/T_{2m} or 1% in ff_a medians or clear/perceptible alterations in their distributions. Considering also that sequence settings are typically confined to relatively limited ranges and options for reproducibility purposes, leaves rather few and clear choices for establishing optimised and at the same time pragmatically useful acquisition settings and sequence versions.

The above considerations seem to broadly follow the recommendations available from published work. A minimum *TE* is in general preferred for the reasons presented in section 6.3 and in the literature (Dula et al., 2009), and an additional advantage of minimising *TE* is to minimise diffusion weighting effects. However the scanner hardware and software determine the *TE* lower limit in practice, and therefore in order to allow reproducibility between different centres and vendors, *TE* might need to be compromised towards relatively higher values than the available minimum for a particular system. Longer *ETL* is typically preferred for improved accuracy and precision (section 6.2), depending also on the anatomical features of interest (e.g. volume to be scanned) (Hollingsworth et al., 2012, Uddin et al., 2013) however scan time and patient comfort restrictions can limit this (Burakiewicz et al., 2017). Multi-coil (MC) acquisition techniques can both improve SNR and reduce scan duration

however they might pose challenges in the interpretation of the $B_1 f$ behaviour and noise effects observed in such images (section 6.5). Additionally it has been shown that in MC acquisitions departures from normally distributed noise (additive noise effects following a non-central χ^2 distribution) are more pronounced (Bouhrara and Spencer, 2018) which is a probable reason for the departures in the noise σ_N estimation observed with the proposed sEPG-MLE method.

In more practical terms it seems that in most contemporary quantitative MRI work (section 2.3) sequence settings are usually confined within the ranges suggested above. Commonly recommended *TE* is between 8 – 10ms and ETL above 15 echoes (Hollingsworth et al., 2012), dependent also on acceptable scan durations, power deposition limits and anatomy volume choices. Receive bandwidth is typically kept at intermediate ranges as a compromise between reduced SNR at higher bandwidths and avoidance of artefacts due to chemical shift, motion, B_0 field inhomogeneities etc. at lower bandwidths. Finally, combined coil and acceleration reconstruction techniques via MC use seem to be the norm in most acquisitions. The work presented in section 6.5 have shown that the choice of the coil combination method can impact on the parameter estimation.

Chapter 7. Discussion and Conclusions

7.1 General considerations

The three main aims of this thesis (Chapter 1) were to deploy and optimise the proposed methods, to test their performance on real test object and *in vivo* data and to develop useful recommendations and guidelines for future acquisition optimisation. The purpose of this chapter is to summarise and assess the corresponding outcomes that this work produced. Additionally, with regards to application to *in vivo* data, it is of interest to evaluate the methods' clinical validity and overall findings in the patient studies.

In evaluating and optimising quantitative MRI methods, various, possibly conflicting factors arise, e.g. measurement precision *vs.* acquisition duration, or for parameter estimation, the use of prior information *vs.* the bias that this may introduce. The main goal of developing such methods is to obtain accurate, precise and reproducible parameter estimates suitable as outcome measures. The work presented in this thesis has shown that using the proposed sEPG-MLE method it is possible from a single acquisition, to simultaneously estimate several important quantitative MRI parameters. Of these, $T_{2\mathrm{m}}$ and ff_a are the main parameters of interest, i.e. the candidate outcome measures or potential biomarkers, while $B_1 f$ and the noise σ_N may be considered nuisance factors which typically must be accounted for, for the purposes of optimal fitting and quality control (McPhee and Wilman, 2017, Azzabou et al., 2015). The amplitude α as a surrogate for proton density is both a parameter of interest and useful in terms of interpreting fitting behaviour.

Comparison of a new approach with existing methods is integral to such analyses and a common way to assess the accuracy and precision of the method. This process can reveal shortcomings and peculiarities of both the proposed and established methods, and additionally suggest ways of making them compatible (e.g. with use of corrections). It can also be used to improve data interpretation in light of the patho-physiological system being investigated, for instance compartmentalisation of tissue and aqueous *vs.* fatty tissue components.

Optimisation pertains also to acquisition design and tailoring (Chapter 6). Scanner engineering and software design defined the capabilities and limitations of the MR system in question, and operator parameter choices with these then impact the final measurement performance. Therefore knowledge of both system specification and technical characteristics, such as RF pulse design, on the one hand, and on the other hand their interaction with operator sequence and parameter choices is required to optimise performance. With regards to parameter estimation, computational demands and processing requirements impact upon the time-efficiency and reliability of processing data, and therefore good practice in programming and data handling is needed.

7.2 Method optimisation

Detailed consideration of the RF transmit field scaling factor $B_1 f$ estimation is one of the novel aspects of this work, both in terms of construction of the signal model used, and in the way that parameters were estimated and interpreted. Slice profile corrections incorporated in EPG model applications presented in previous published

work have used either the Fourier transform (FT) of RF pulse shapes (Lebel and Wilman, 2010) or the Shinnar-Le Roux (SLR) algorithm (Pauly et al., 1991). In the sEPG model used here slice profile correction was done using Bloch equations-based calculations (section 2.1.7) since these results were seen to closely match the full Bloch equation-based MESE signal simulations (Hargreaves, 2016). Bloch equation-based models are considered theoretically (McRobbie et al., 2006) to fully account for the RF pulse responses and this has been corroborated in practice (McPhee and Wilman, 2017). It should be noted that the authors of the original work on EPG fitting with FT-based slice profile correction (Lebel and Wilman, 2010) subsequently also adopted the Bloch equations-based approach in the Matlab code for their algorithm which is freely available (Lebel, 2012). Under this approach slice profile asymmetry around perfect (close to 180˚) refocusing angles is observed, and this can lead to degenerate solutions for the B_1 field factor estimates (McPhee and Wilman, 2017, Lebel and Wilman, 2010), as also described in this work (sections 3.2.6 and 3.3.2). Unaddressed, this translated to significant potential bias in both T_{2m} and ff_a estimates (up to several ms or percentage units respectively) and led to the introduction of the B_1f spatial regularisation strategy into the fitting algorithm in the present work in order to address the problem. This was seen to solve the discrepancy for the vast majority of redundant (dual) B_1f solutions – however, in future work, more sophisticated spatial regularisation techniques could further optimise the estimation method and improve results (Basiri et al., 2019). Our internal B_1f determination approach has the advantage of bypassing the need to use separately acquired B_1f maps and the uncertainties or discrepancies they might introduce (Lankford and Does, 2018).

The internal Rician noise parameter estimation is another novel aspect of the proposed method, since in applications of MLE in the MRI field published so far, the Rician noise standard deviation has been either pre-calculated (Walker-Samuel et al., 2009) or the data have simply been considered to be normally distributed (Hardy and Andersen, 2009, Bonny et al., 1996). Our algorithm provided a broadly reliable Rician noise σ_N estimation, with a slight tendency for underestimation in simulation (section 3.3.1, Figure 3.58) and a tendency mostly for overestimation in real datasets compared to the empirical methods presented in section 3.1. The impact of this σ_N estimate bias was however limited: importantly, a simulation experiment fixing σ_N at its nominal value in the fitted model, was not seen to improve results relative to a model with freely estimated σ_N (section 3.2.4, Figure 3.20). The MLE method used here, unlike the more widely used LSQ minimisation approaches, appears to better account for the (rectified) noise floor in magnitude reconstructed images rather than attempting to approximate it by assigning it a constant value (Milford et al., 2015), or discarding later echoes altogether (McPhee and Wilman, 2017). The proposed sEPG-MLE method is therefore a less cumbersome but nevertheless robust alternative to methods where additional steps to account for the noise behaviour are needed.

The algorithm predicted the phantom manufacturer-provided calibration values with high accuracy and precision overall. This was in marked contrast with previous attempts using an exponential or a non-slice profile corrected EPG model, where large differences were seen in the T_2 estimates, in the range of 20-25% from the nominal values. This suggests that taking into account the slice profile information – requiring accurate knowledge of the shape and characteristics of the RF pulses and

their combination with the gradients – is essential for robust estimation of the desired parameters. However the assumed RF pulse and gradient waveforms (specified by the MR system manufacturer) might deviate from those achieved in real conditions, leading to mismatch between the assumed model and actual data compromising fitting. Such cases were encountered in some of the forearm DMD datasets (section 5.3) where challenging MR conditions at the periphery of the magnet bore may have led to erroneous slice profile definition or other errors. The suggested post processing steps helped to largely exclude pixel regions where this occurred, to keep the analysis meaningful and interpretable.

7.3 Clinical application

The final multi component approach used as the final fat-water signal model emerged after the inadequacy of a single component sEPG model to robustly describe fat signals was demonstrated (section 3.3.2). Two-component exponential models have been used in published work over the past decade (Kan et al., 2009, Yao and Gai, 2012) and recently also a three-component exponential model with two fixed parameter components representing the fat compartment was proposed (Azzabou et al., 2015), in what could be regarded as an exponential version of the model introduced here. In that paper, only pixels appearing to experience B_1 field factors above a certain threshold were kept in the analysis, which was seen to improve results, but resulted in relatively large image regions being omitted; this approach also required availability of independently obtained B_1f maps. Addressing the B_1f limitation issue was attempted in subsequent work from the same group (Marty et al., 2016) which used a 2-component sEPG model in a dictionary-based fitting

method, with a single sEPG component corresponding to each assumed compartment, muscle water and fat. The work presented here attempted to address both the impact of acquisition factors, in particular B_1f inhomogeneities, using a sEPG model, and physical meaningfulness using an optimised multi-component fat-signal model.

While fit quality significantly improves when using a 2-component sEPG model for fatty tissue, the question of physical interpretation of the results remains. Careful calibration of this fat model led to a dominant component (67%) of long T_2 (198ms) and a smaller component (33%) of shorter T_2 (40ms). Including this model to the final multi-component sEPG model and fitting it to the available *in vivo* datasets revealed that for adipose tissue estimated $T_{2m} < 15$ms and $ff > 90\%$, which led to one of the main quality control criteria to distinguish muscle tissue from fat. These results are consistent with histological results where the proportion of pure fat in subcutaneous 'fat tissue' is in fact broadly below 90% (Thomas et al., 1962), with a substantial fraction of the remaining 10% presumably comprising water. Both marked $T_{2\mathrm{m}}$ elevation and marked decrease was observed in some regions of very high fat content, even for good quality images and optimal acquisition strategies, such as the datasets analysed in Chapter 4, as indicated by the corresponding ff_a and ff_d estimates (e.g. above 80%). This diverse $T_{2\mathrm{m}}$ and high ff are consistent with the respective oedematous/inflammatory and fat replacement effects commonly reported in the literature. On the other hand with the proposed method ff estimation can be extended to its entire range, resolving the limitation for ff only below approximately 50% in the published work mentioned above.

Comparison of the sEPG-MLE algorithm results with the exponential model and LSQ fitting revealed a consistent systematic 15 to 20% decrease in T_{2m} in all ALS, KD, DMD and healthy volunteer cohorts, in good agreement with the results seen both in simulation (Chapter 3) and in phantom validation (section 3.2.10 and Table 4.3). Slice profile effects can largely explain this difference due to the echo train modulation they cause (section 3.2.5) and the sEPG model as expected more accurately describes the cumulative slice profile signal decay. It should be noted that in similar work mentioned earlier (Marty et al., 2016), slice profile effects were not reported to be a major confounding factor. This may therefore also depend to specific scanner settings, i.e. RF pulse characteristics or magnetic field imperfections. T_{2m} histogram distributions were markedly narrower with the sEPG-MLE approach (e.g. for healthy control data), suggesting overall improved parameter estimate precision, and may thus further support the validity of the method. This is also consistent with the theoretical assumption that ML estimators asymptotically approach the lowest possible variance – or equivalently increase the possibility for an estimator to be unbiased – in a given experiment. It should be noted that although this work has shown that exponential model and LSQ fitting approaches generally produce heavily biased parameter estimates, they still may be pragmatically useful not least because they are much less demanding in computational power and readily available – e.g. via built-in scanner software. However the bias in these methods makes it difficult to compare results from different studies, and the sEPG-MLE method developed here is therefore to be preferred.

In an exploration of more qualitative conventional radiological imaging in section 4.5, T_{2m} maps were also compared with STIR images, which are of interest in this context since STIR images are commonly interpreted as providing fat suppressed T_2-weighted contrast. Modelling STIR contrast based on the signal behaviour in terms of Physics, tissue compartmentalisation and the sEPG-MLE parameter estimates, revealed that STIR image contrast appears to be substantially affected additionally by fat content: increasing *ff* leads to water signal reduction by reducing its relative PD. In addition, if fat signal nulling is inaccurate, part of the remaining signal might be emanating from fat protons. These factors add to other complications (Brown and Semelka, 2005), such as the potentially complex contrast dependence of STIR images on both T_1 and T_2 (as also seen here, section 4.5). B_1 field inhomogeneities can also markedly affect STIR imaging (Bley et al., 2010), something that was observed in several of the images examined here, whereas with the sEPG-MLE method they are explicitly accounted for. Therefore sEPG-MLE T_{2m} is more likely to express true muscle water T_2 behaviour, and these findings may have important implications for the radiological interpretation of STIR images in this context.

Simple acquisition and fitting methods may cause artefactual correlation between the estimated parameters. Lack of such correlation provides evidence that yielded results are independent of the estimation method. In this work, this independence of the sEPG-MLE yielded T_{2m} from fat content was demonstrated also using the Dixon method, with T_{2m} variations appearing stable along the entire ff_d spectrum. While in earlier publications this independence was not explicitly examined but rather taken for granted (Kan et al., 2009, Yao and Gai, 2012) more recent work related to multi-

component models and quantitative MRI has shown that it is not always the case. For example independence was seen in application of a tri-exponential model on thigh data (Azzabou et al., 2015) but not when applying the same method on forearm data (Wary et al., 2015). In the present work overall independence of T_{2m} from Dixon *ff* was observed in both thigh and forearm datasets. It should be noted that caution is required regarding this correlation since it also depends on the characteristics of the disease in question. For instance DMD progression appears to correlate with increase in fat content and decrease in water T_2, therefore it would not be surprising for T_{2m} to also show a tendency to decrease with increasing fat fraction. Additionally there is evidence from spectroscopic studies that decreasing water T_2 with increasing *ff* is a general attribute of many neuromuscular disorders (Schlaeger et al., 2019), especially at the very high fat content ranges. These findings suggest that apparent associations between the estimated parameters, predominantly T_{2m} and *ff*, do not necessarily reflect bias introduced by the estimation method.

While sEPG-MLE method based *ff* maps were qualitatively similar to the Dixon method obtained *ff* distributions, numerical differences were quite marked ranging from multiple times higher (ff_a than ff_d) for the lower *ff* range, with this ratio decreasing towards the middle range *ff* and becoming a few percent lower for very high *ff*. This was manifested as a 'banana' shaped curve of the ff_a values over the line of direct equivalence with the ff_d spectrum of values. The sEPG-MLE *ff* was in turn also markedly lower than the EXP-LSQ estimated *ff* (about 10 to 25%) and closer to the Dixon *ff* values. While some basic strategies exist to correct *ff* estimates for experimental differences between the two approaches, namely for MT effects and

differential T_{1f} *vs.* T_{1m} weighting with regards to ff_a, and for single *vs.* multi-peak fat spectrum with regards to ff_d, in reality experimental conditions might be more complex. As mentioned before the assumptions behind the two-component transverse magnetisation decay model used for fat might not always hold and this would directly affect fat content quantification; and T_1 effects might not be as negligible in sEPG models as in general considered. The sEPG model for fat used here doesn't take j-coupling into account, and this may also contribute to differences and errors.

7.4 Recommendations and future directions

A challenge to MESE data fitting may arise from the difficulty of accurately modelling the slice-selective profile response, an issue often mentioned in the literature and central also in this work. Inclusion of additional parameters in the model could account for imperfections in the assumed CPMG signal behaviour, e.g. the excitation and refocusing widths relationship and phase errors propagation. However adding additional parameters to the specific model in this way would represent additional degrees of freedom, affecting the statistical stability of the method, and in practice this may not commonly be a major confounder (e.g. as discussed in section 5.5).

Nevertheless introducing more sophisticated physical models in general is of great interest – at least theoretically – and this may suggest promising future directions. Macro- and microscopic level motion of muscle tissue may have influenced the results presented here, particularly in the DMD study. Recent work for example on muscle motor unit fasciculation effects suggests that physiological mechanisms can

directly impact the signal evolution (Whittaker et al., 2019) in diffusion- and likely also T_2-weighted imaging. Modelling such diffusion effects or phase errors could help to both further optimise the method proposed here and broaden its range of applications.

With regards to the operator controlled, sequence parameter choices, there appear to be practical limitations which do not in practice leave much room for improvement. This is mainly due to the limited options available on commercial MRI systems, scan duration concerns and the issue of reproducibility between different scanners and centres, as discussed in Chapter 6.

The practical effectiveness of the method developed in this thesis may be further enhanced via improving the available computational power and software engineering. Useful advances – used already in this work – are parallel computing and use of compiled instead of interpreted programming code. In several parts of the algorithms used here replacing Matlab functions with compiled C code was seen to greatly accelerate fitting. Promising, more advanced computing approaches which may be helpful in future work are machine learning applications – with the $B_1 f$ spatial regularisation concept being a concrete example – and trial of library-based fitting by pattern recognition.

7.5 Addressing the overall objectives

Despite the above limitations and the need for further work, the aim of obtaining reliable quantitative MRI outcome measures (Chapter 1) has been accomplished to a satisfactory level. Combination of the sEPG formalism with the MLE statistical

approach adequately addressed the core issues of B_1 field inhomogeneity and ambiguity and the Rician noise effects. These were the fundamental steps that led to marked increase in accuracy, leading to correction of the 15-20% T_2 overestimation manifested when using the established EXP-LSQ based methods, and good agreement with the best available 'gold standards', namely calibrated phantom measurements and literature values from MR spectroscopy. Accuracy was accompanied with increased precision which led to the detection of statistically significant T_{2m} changes between patients and healthy volunteers as well as both increases or decreases longitudinally relating to disease progression. Interestingly, comparing diseases such as ALS and DMD revealed that the time course and direction of T_{2m} changes differed in these patient groups, suggesting that the observed T_{2m} changes may reflect more than one underlying pathological process. These findings demonstrate that T_{2m} should be interpreted with caution, while at the same time reinforcing its potential usefulness as a marker of such processes.

The multi-component (fat-water) representation used in the final model led to a concomitant estimation of *ff* as an additional outcome measure. The latter was compared with Dixon method results producing encouraging findings at least qualitatively. Similarly to T_{2m}, yielding and comparing study group results was directly related to the post-processing and spatial regularisation strategies used. The outcomes corroborated the potential of the suggested method to better characterise subtle or subclinical changes (e.g. relying to oedematous or fibrotic effects as opposed to mere fat infiltration), and monitor their progression, including response to therapies.

7.6 Conclusions

The potential of the proposed sEPG-MLE method as a novel and optimised technique for quantitative neuromuscular (MR) imaging has been demonstrated in this thesis. It was shown to offer adequate precision and accuracy in the estimated parameters to reproducibly characterise and distinguish disease from healthy presentation, thus suggesting these are useful outcome measures.

Further optimisation may be achieved with the inclusion of additional factors in the physical modelling, more sophisticated regularisation methods for parameter estimation and the potential use of machine learning approaches. The proposed methods are also useful practically in testing specific acquisition settings and sequence design optimisation. As such, although these methods were developed here for the specific application of neuromuscular imaging, they may in future be adapted and extended to serve a broader field of research than the scope of this thesis.

References

ADAMS, G. R., DUVOISIN, M. R. & DUDLEY, G. A. 1992. Magnetic resonance imaging and electromyography as indexes of muscle function. *J Appl Physiol (1985),* 73**,** 1578-83.

ARAUJO, E. C., FROMES, Y. & CARLIER, P. G. 2014. New insights on human skeletal muscle tissue compartments revealed by in vivo t2 NMR relaxometry. *Biophys J,* 106**,** 2267-74.

ARPAN, I., FORBES, S. C., LOTT, D. J., SENESAC, C. R., DANIELS, M. J., TRIPLETT, W. T., DEOL, J. K., SWEENEY, H. L., WALTER, G. A. & VANDENBORNE, K. 2013. T2 mapping provides multiple approaches for the characterization of muscle involvement in neuromuscular diseases: a cross-sectional study of lower leg muscles in 5–15-year-old boys with Duchenne muscular dystrophy. *NMR Biomed,* 26**,** 320-328.

ARPAN, I., WILLCOCKS, R. J., FORBES, S. C., FINKEL, R. S., LOTT, D. J., ROONEY, W. D., TRIPLETT, W. T., SENESAC, C. R., DANIELS, M. J., BYRNE, B. J., FINANGER, E. L., RUSSMAN, B. S., WANG, D. J., TENNEKOON, G. I., WALTER, G. A., SWEENEY, H. L. & VANDENBORNE, K. 2014. Examination of effects of corticosteroids on skeletal muscles of boys with DMD using MRI and MRS. *Neurology,* 83**,** 974-80.

AZZABOU, N., LOUREIRO DE SOUSA, P., CALDAS, E. & CARLIER, P. G. 2015. Validation of a generic approach to muscle water T2 determination at 3T in fat-infiltrated skeletal muscle. *J Magn Reson Imaging,* 41**,** 645-53.

BASIRI, R., FEDERICO, P., LEBEL, R. M. J. M. R. M. I. P., BIOLOGY & MEDICINE 2019. Transverse relaxometry with transmit field-constrained stimulated echo compensation. 32**,** 669-677.

BATES, D. M. 1988. *Nonlinear regression analysis and its applications,* New York, Wiley.

BEDE, P. & HARDIMAN, O. 2014. Lessons of ALS imaging: Pitfalls and future directions — A critical review. *NeuroImage: Clinical,* 4**,** 436-443.

BEN-ELIEZER, N., SODICKSON, D. K. & BLOCK, K. T. 2015. Rapid and accurate T2 mapping from multi-spin-echo data using Bloch-simulation-based reconstruction. *Magn Reson Med,* 73**,** 809-17.

BEVAN, A. 2013. *Statistical data analysis for the physical sciences,* Cambridge, Cambridge University Press.

BISHOP, C. A., RICOTTI, V., SINCLAIR, C. D. J., EVANS, M. R. B., BUTLER, J. W., MORROW, J. M., HANNA, M. G., MATTHEWS, P. M., YOUSRY, T. A., MUNTONI, F., THORNTON, J. S., NEWBOULD, R. D. & JANICZEK, R. L. 2018. Semi-automated analysis of diaphragmatic motion with dynamic Magnetic Resonance Imaging in healthy controls and non-ambulant subjects with Duchenne Muscular Dystrophy. *Frontiers in Neurology,* 9.

BLEY, T. A., WIEBEN, O., FRANÇOIS, C. J., BRITTAIN, J. H. & REEDER, S. B. 2010. Fat and water magnetic resonance imaging. Hoboken.

BONNY, J. M., ZANCA, M., BOIRE, J. Y. & VEYRE, A. 1996. T2 maximum likelihood estimation from multiple spin-echo magnitude images. *Magnetic Resonance in Medicine,* 36**,** 287-293.

BOUHRARA, M. & SPENCER, R. G. 2018. Fisher information and Cramér-Rao lower bound for experimental design in parallel imaging. *Magnetic Resonance in Medicine,* 79**,** 3249-3255.

BOX, G. E., JENKINS, G. M., REINSEL, G. C. & LJUNG, G. M. 2015. *Time series analysis: forecasting and control*, John Wiley & Sons.

BROWN, M. A. & SEMELKA, R. C. 2005. *MRI: Basic Principles and Applications,* Hoboken, NJ, USA, Hoboken, NJ, USA: John Wiley & Sons, Inc.

BRYAN, W. W., REISCH, S. J., MCDONALD, L. G., HERBELIN, J. L., BAROHN, L. R. & FLECKENSTEIN, L. J. 1998. Magnetic resonance imaging of muscle in amyotrophic lateral sclerosis. *Neurology,* 51**,** 110-113.

BURAKIEWICZ, J., SINCLAIR, C. D. J., FISCHER, D., WALTER, G. A., KAN, H. E. & HOLLINGSWORTH, K. G. 2017. Quantifying fat replacement of muscle by quantitative MRI in muscular dystrophy. *Journal of Neurology,* 264**,** 2053-2067.

CARLIER, P. G. J. N. D. 2014. Global T2 versus water T2 in NMR imaging of fatty infiltrated muscles: different methodology, different information and different implications. 24**,** 390-392.

CARR, H. Y. & PURCELL, E. M. 1954. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review,* 94**,** 630-638.

CHEBROLU, V. V., HINES, C. D. G., YU, H., PINEDA, A. R., SHIMAKAWA, A., MCKENZIE, C. A., SAMSONOV, A., BRITTAIN, J. H. & REEDER, S. B. 2010. Independent estimation of T*2 for water and fat for improved accuracy of fat quantification. *Magnetic Resonance in Medicine,* 63**,** 849-857.

COLE, W. C., LEBLANC, A. D. & JHINGRAN, S. G. 1993. The origin of biexponential T2 relaxation in muscle water. *Magn Reson Med,* 29**,** 19-24.

CRUZ, P. M. R., PALACE, J. & BEESON, D. 2014. Inherited disorders of the neuromuscular junction: an update. *Journal of neurology,* 261**,** 2234-2243.

DAHLQVIST, R. J., OESTERGAARD, T. S., POULSEN, S. N., THOMSEN, S. C. & VISSING, S. J. 2019. Refining the spinobulbar muscular atrophy phenotype by quantitative MRI and clinical assessments. *Neurology,* 92**,** e548-e559.

DARDZINSKI, B. J., LAOR, T., SCHMITHORST, V. J., KLOSTERMAN, L. & GRAHAM, T. B. 2002. Mapping T2 relaxation time in the pediatric knee: feasibility with a clinical 1.5-T MR imaging system. *Radiology,* 225**,** 233-9.

DIXON, W. T. 1984. Simple proton spectroscopic imaging. *Radiology,* 153**,** 189-94.

DULA, A. N., GOCHBERG, D. F. & DOES, M. D. 2009. Optimal echo spacing for multi-echo imaging measurements of Bi-exponential T 2 relaxation. *Journal of Magnetic Resonance,* 196**,** 149-156.

EMERY, A. E. J. N. D. 1991. Population frequencies of inherited neuromuscular diseases—a world survey. 1**,** 19-29.

EVANS, M. C., SERRES, S., KHRAPITCHEV, A. A., STOLP, H. B., ANTHONY, D. C., TALBOT, K., TURNER, M. R. & SIBSON, N. R. 2014. T2-Weighted MRI Detects Presymptomatic Pathology in the SOD1 Mouse Model of ALS. *Journal of Cerebral Blood Flow & Metabolism,* 34**,** 785-793.

FERRARO, P., AGOSTA, F., QUERIN, G., RIVA, N., BERTOLIN, C., DA RE, E., COPETTI, M., COMI, G., FALINI, A., SORARU, G. & FILIPPI, M. 2016. Structural Brain MRI Abnormalities in Kennedy's Disease. *Neurology,* 86.

FISCHMANN, A., HAFNER, P., FASLER, S., GLOOR, M., BIERI, O., STUDLER, U. & FISCHER, D. 2012. Quantitative MRI can detect subclinical disease progression in muscular dystrophy. *J Neurol,* 259**,** 1648-1654.

FORBES, S. C., WILLCOCKS, R. J., TRIPLETT, W. T., ROONEY, W. D., LOTT, D. J., WANG, D. J., POLLARO, J., SENESAC, C. R., DANIELS, M. J., FINKEL, R. S., RUSSMAN, B. S., BYRNE, B. J., FINANGER, E. L., TENNEKOON, G. I., WALTER, G. A., SWEENEY, H. L. & VANDENBORNE, K. 2014. Magnetic resonance imaging and spectroscopy assessment of lower extremity skeletal muscles in boys with Duchenne muscular dystrophy: a multicenter cross sectional study. *PLoS One,* 9**,** e106435.

FOSTER, M. A. & HUTCHISON, J. M. S. 1987. *Practical NMR imaging / edited by M.A. Foster & J.M.S. Hutchison,* Oxford, England ; Washington, DC, IRL Press.

FRIEDMAN, S. D., POLIACHIK, S. L., CARTER, G. T., BUDECH, C. B., BIRD, T. D. & SHAW, D. W. 2012. The magnetic resonance imaging spectrum of facioscapulohumeral muscular dystrophy. *Muscle Nerve,* 45**,** 500-6.

FSL. 2015. *prelude* [Online]. Available: http://fsl.fmrib.ox.ac.uk [Accessed 16/10 2015].
GAMBAROTA, G., CAIRNS, B. E., BERDE, C. B. & MULKERN, R. V. 2001. Osmotic effects on the T2 relaxation decay of in vivo muscle. *Magn Reson Med,* 46**,** 592-9.
GLOOR, M., FASLER, S., FISCHMANN, A., HAAS, T., BIERI, O., HEINIMANN, K., WETZEL, S. G., SCHEFFLER, K. & FISCHER, D. 2011. Quantification of fat infiltration in oculopharyngeal muscular dystrophy: comparison of three MR imaging methods. *J Magn Reson Imaging,* 33**,** 203-10.
GLOVER, G. H. & SCHNEIDER, E. 1991. Three-point dixon technique for true water/fat decomposition with B0 inhomogeneity correction. *Magnetic Resonance in Medicine,* 18 371-383.
GOLD, G. E., HAN, E., STAINSBY, J., WRIGHT, G., BRITTAIN, J. & BEAULIEU, C. 2004. Musculoskeletal MRI at 3.0 T: relaxation times and image contrast. *AJR Am J Roentgenol,* 183**,** 343.
GRUNSEICH, C., RINALDI, C. & FISCHBECK, K. 2014. Spinal and bulbar muscular atrophy: pathogenesis and clinical management. *Oral Diseases,* 20**,** 6-9.
GUDBJARTSSON, H. & PATZ, S. 1995. The rician distribution of noisy MRI data. *Magnetic Resonance in Medicine,* 34**,** 910-914.
GUÉRIN, B., GEBHARDT, M., CAULEY, S., ADALSTEINSSON, E. & WALD, L. L. J. M. R. I. M. 2014. Local specific absorption rate (SAR), global SAR, transmitter power, and excitation accuracy trade-offs in low flip-angle parallel transmit pulse design. 71**,** 1446-1457.
HAHN, E. L. 1950. Spin Echoes. *American Physical Society,* 80**,** 14.
HAMANO, T., MUTOH, T., HIRAYAMA, M., KAWAMURA, Y., NAGATA, M., FUJIYAMA, J. & KURIYAMA, M. 2004. Muscle MRI findings of X-linked spinal and bulbar muscular atrophy. *Journal of the Neurological Sciences,* 222**,** 93-97.
HAN, E., G., G., STAINSBY, J., WRIGHT, G., BEAULIEU, C. & BRITTAIN, J. In-vivo T1 and T2 measurements of musculoskeletal tissue at 3T and 1.5T. Proceedings of the 11th Annual Meeting ISMRM, 2003 Toronto, ON, Canadea.
HARDY, P. A. & ANDERSEN, A. H. 2009. Calculating T2 in images from a phased array receiver. *Magn Reson Med,* 61**,** 962-9.
HARGREAVES, B. 2012. *Extended Phase Graph functions and examples* [Online]. Available: https://web.stanford.edu/~bah/software/epg/EPG-Lecture.pdf [Accessed March 10 2017].
HARGREAVES, B. 2016. *Bloch Equation Simulation* [Online]. Available: http://mrsrl.stanford.edu/~brian/bloch/ [Accessed November 7 2016].
HATAKENAKA, M., UEDA, M., ISHIGAMI, K., OTSUKA, M. & MASUDA, K. 2001. Effects of aging on muscle T2 relaxation time: difference between fast- and slow-twitch muscles. *Invest Radiol,* 36**,** 692-8.
HENKELMAN, R. M. J. M. P. 1985. Measurement of signal intensities in the presence of noise in MR images. 12**,** 232-233.
HENNIG, J. 1991. 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**,** 125-143.
HERNANDO, D., LIANG, Z. P. & KELLMAN, P. 2010. Chemical shift–based water/fat separation: A comparison of signal models. *Magnetic Resonance in Medicine,* 64**,** 811-822.
HOGREL, J.-Y., WARY, C., MORAUX, A., AZZABOU, N., DECOSTRE, V., OLLIVIER, G., CANAL, A., LILIEN, C., LEDOUX, I., ANNOUSSAMY, M., REGUIBA, N., GIDARO, T., LE MOING, A. G., CARDAS, R., VOIT, T., CARLIER, P. G. & SERVAIS, L. 2016. Longitudinal functional and NMR assessment of upper limbs in Duchenne muscular dystrophy. *Neurology,* 86**,** 1022-1030.
HOLLINGSWORTH, K. G., DE SOUSA, P. L., STRAUB, V. & CARLIER, P. G. 2012. Towards harmonization of protocols for MRI outcome measures in skeletal muscle studies: consensus recommendations from two TREAT-NMD NMR workshops, 2 May 2010,

Stockholm, Sweden, 1-2 October 2009, Paris, France. *Neuromuscular disorders : NMD,* 22 Suppl 2**,** S54.

HUANG, Y., MAJUMDAR, S., GENANT, H. K., CHAN, W. P., SHARMA, K. R., YU, P., MYNHIER, M. & MILLER, R. G. 1994. Quantitative MR relaxometry study of muscle composition and function in Duchenne muscular dystrophy. *J Magn Reson Imaging,* 4**,** 59-64.

JANICZEK, R. L., GAMBAROTA, G., SINCLAIR, C. D. J., YOUSRY, T. A., THORNTON, J. S., GOLAY, X. & NEWBOULD, R. D. 2011. Simultaneous T2 and lipid quantitation using IDEAL-CPMG. *Magnetic Resonance in Medicine,* 66**,** 1293-1302.

JEHENSON, P., LEROY-WILLIG, A., DE KERVILER, E., DUBOC, D. & SYROTA, A. 1993. MR imaging as a potential diagnostic test for metabolic myopathies: importance of variations in the T2 of muscle with exercise. *AJR Am J Roentgenol,* 161**,** 347-51.

JENKINS, T. M., ALIX, J. J. P., DAVID, C., PEARSON, E., RAO, D. G., HOGGARD, N., O'BRIEN, E., BASTER, K., BRADBURN, M., BIGLEY, J., MCDERMOTT, C. J., WILKINSON, I. D. & SHAW, P. J. 2018. Imaging muscle as a potential biomarker of denervation in motor neuron disease. *J Neurol Neurosurg Psychiatry,* 89**,** 248.

JORDAN, C. D., SARANATHAN, M., BANGERTER, N. K., HARGREAVES, B. A. & GOLD, G. E. 2013. Musculoskeletal MRI at 3.0 T and 7.0 T: a comparison of relaxation times and image contrast. *Eur J Radiol,* 82**,** 734-9.

KAISER, R., BARTHOLDI, E. & ERNST, R. R. 1974. Diffusion and field-gradient effects in NMR Fourier spectroscopy. *The Journal of Chemical Physics,* 60**,** 13.

KAMMAN, R. L., BAKKER, C. J., VAN DIJK, P., STOMP, G. P., HEINER, A. P. & BERENDSEN, H. J. 1987. Multi-exponential relaxation analysis with MR imaging and NMR spectroscopy using fat-water systems. *Magn Reson Imaging,* 5**,** 381-92.

KAN, H. E., SCHEENEN, T. W., WOHLGEMUTH, M., KLOMP, D. W., VAN LOOSBROEK-WAGENMANS, I., PADBERG, G. W. & HEERSCHAP, A. 2009. Quantitative MR imaging of individual muscle involvement in facioscapulohumeral muscular dystrophy. *Neuromuscul Disord,* 19**,** 357-62.

KEENAN, K., STUPIC, K., BOSS, M., RUSSEK, S., CHENEVERT, T., PRASAD, P., REDDICK, W., CECIL, K., ZHENG, J., HU, P., JACKSON, E. & MRI, T. A. H. C. O. S. F. Q. Multi-site, multi-vendor comparison of T1 measurement using ISMRM/NIST system phantom. ISMRM, 2016 Singapore.

KEENE, K. R., BEENAKKER, J. W. M., HOOIJMANS, M. T., NAARDING, K. J., NIKS, E. H., OTTO, L. A., VAN DER POL, W. L., TANNEMAAT, M. R., KAN, H. E. & FROELING, M. J. M. R. I. M. 2020. T2 relaxation-time mapping in healthy and diseased skeletal muscle using extended phase graph algorithms.

KEREN, N., SCOTT, K. M., TSUDA, M., BARNWELL, J., KNIBB, J. A., ELLIS, C. M., LEIGH, P. N., SHAW, C. E. & AL-CHALABI, A. 2014. Evidence of an environmental effect on survival in ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration,* 15**,** 528-533.

KIM, H. K., LAOR, T., GRAHAM, T. B., ANTON, C. G., SALISBURY, S. R., RACADIO, J. M. & DARDZINSKI, B. J. 2010a. T2 relaxation time changes in distal femoral articular cartilage in children with juvenile idiopathic arthritis: a 3-year longitudinal study. *AJR Am J Roentgenol,* 195**,** 1021-5.

KIM, H. K., LAOR, T., HORN, P. S., RACADIO, J. M., WONG, B. & DARDZINSKI, B. J. 2010b. T2 mapping in Duchenne muscular dystrophy: distribution of disease activity and correlation with clinical assessments. *Radiology,* 255**,** 899-908.

KIM, H. K., SERAI, S., LINDQUIST, D., MERROW, A. C., HORN, P. S., KIM, D. H. & WONG, B. L. 2015. Quantitative Skeletal Muscle MRI: Part 2, MR Spectroscopy and T2 Relaxation Time Mapping-Comparison Between Boys With Duchenne Muscular Dystrophy and Healthy Boys. *AJR Am J Roentgenol,* 205**,** W216-23.

KIM, H. K., SERAI, S., MERROW, A. C., WANG, L., HORN, P. S. & LAOR, T. 2014. Objective measurement of minimal fat in normal skeletal muscles of healthy children using T2 relaxation time mapping (T2 maps) and MR spectroscopy. *Pediatr Radiol,* 44**,** 149-57.

KIRSHNER, B. & GUYATT, G. 1985. A methodological framework for assessing health indices. *Journal of Chronic Diseases,* 38**,** 27-36.

KLICKOVIC, U., ZAMPEDRI, L., SINCLAIR, C. D., WASTLING, S. J., TRIMMEL, K., HOWARD, R. S., MALASPINA, A., SHARMA, N., SIDLE, K. & EMIRA, A. J. N. 2019. Skeletal muscle MRI differentiates SBMA and ALS and correlates with disease severity. 93**,** e895-e907.

KRSSÁK, M., MLYNÁRIK, V., MEYERSPEER, M., MOSER, E. & RODEN, M. 2004. 1H NMR relaxation times of skeletal muscle metabolites at 3 T. *Magma (New York, N.Y.),* 16**,** 155.

KUMAR, Y., WADHWA, V., PHILLIPS, L., PEZESHK, P. & CHHABRA, A. J. E. J. O. R. 2016. MR imaging of skeletal muscle signal alterations: systematic approach to evaluation. 85**,** 922-935.

LA SPADA, A. & WILSON, E. 1991. Androgen Receptor Gene Mutations in X-Linked Spinal and Bulbar Muscular Atrophy. *Nature,* 352**,** 77-9.

LANKFORD, C. L. & DOES, M. D. J. M. R. I. M. 2018. Propagation of error from parameter constraints in quantitative MRI: Example application of multiple spin echo T2 mapping. 79**,** 673-682.

LEBEL, R. StimFit: a toolbox for robust T2 mapping with stimulated echo compensation Proceedings of the 20th annual meeting of ISMRM, 2012 Melbourne, Australia.

LEBEL, R. M. & WILMAN, A. H. 2010. Transverse relaxometry with stimulated echo compensation. *Magn Reson Med,* 64**,** 1005-14.

LIVINGSTON, B. P., SEGAL, R. L., SONG, A., HOPKINS, K., ENGLISH, A. W. & MANNING, C. C. 2001. Functional activation of the extensor carpi radialis muscles in humans. *Arch Phys Med Rehabil,* 82**,** 1164-70.

LOUGHRAN, T., HIGGINS, D. M., MCCALLUM, M., COOMBS, A., STRAUB, V. & HOLLINGSWORTH, K. G. 2015. Improving highly accelerated fat fraction measurements for clinical trials in muscular dystrophy: origin and quantitative effect of R2* changes. *Radiology,* 275**,** 570.

LOVITT, S., MOORE, S. L. & MARDEN, F. A. 2006. The use of MRI in the evaluation of myopathy. *Clinical Neurophysiology,* 117**,** 486-495.

MAILLARD, S. M., JONES, R., OWENS, C., PILKINGTON, C., WOO, P., WEDDERBURN, L. R. & MURRAY, K. J. 2004. Quantitative assessment of MRI T2 relaxation time of thigh muscles in juvenile dermatomyositis. *Rheumatology (Oxford),* 43**,** 603-8.

MAJUMDAR, S., ORPHANOUDAKIS, S. C., GMITRO, A., O'DONNELL, M. & GORE, J. C. 1986. Errors in the measurements of T2 using multiple-echo MRI techniques. I. Effects of radiofrequency pulse imperfections. *Magn Reson Med,* 3**,** 397-417.

MANKODI, A., AZZABOU, N., BULEA, T., REYNGOUDT, H., SHIMELLIS, H., REN, Y., KIM, E., FISCHBECK, K. H. & CARLIER, P. G. 2017. Skeletal muscle water T2 as a biomarker of disease status and exercise effects in patients with Duchenne muscular dystrophy. *Neuromuscul Disord,* 27**,** 705-714.

MARTY, B., BAUDIN, P. Y., REYNGOUDT, H., AZZABOU, N., ARAUJO, E. C., CARLIER, P. G. & DE SOUSA, P. L. 2016. Simultaneous muscle water T2 and fat fraction mapping using transverse relaxometry with stimulated echo compensation. *NMR Biomed,* 29**,** 431-43.

MCPHEE, K. C. & WILMAN, A. H. 2017. Transverse relaxation and flip angle mapping: Evaluation of simultaneous and independent methods using multiple spin echoes. *Magn Reson Med,* 77**,** 2057-2065.

MCROBBIE, D. J. T. B. J. O. R. 1996. The absolute signal-to-noise ratio in MRI acceptance testing. 69**,** 1045-1048.

MCROBBIE, D. W., MOORE, E. A., GRAVES, M. J. & PRINCE, M. R. 2006. *MRI From Picture to Proton,* Cambridge, Cambridge : Cambridge University Press.

MEIBOOM, S. & GILL, D. 1958. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* 29**,** 688-691.

MERCURI, E., CINI, C., COUNSELL, S., ALLSOP, J., ZOLKIPLI, Z., JUNGBLUTH, H., SEWRY, C., BROWN, S. C., PEPE, G. & MUNTONI, F. J. E. J. O. P. N. 2002. Muscle MRI findings in a three-generation family affected by Bethlem myopathy. 6**,** 309-314.

MERCURI, E., PICHIECCHIO, A., ALLSOP, J., MESSINA, S., PANE, M. & MUNTONI, F. 2007. Muscle MRI in inherited neuromuscular disorders: Past, present, and future. *Journal of Magnetic Resonance Imaging,* 25**,** 433-440.

MILFORD, D., BENDSZUS, M. & HEILAND, S. 2018. A novel method for T2 quantification in presence of B1 inhomogeneities. *Z Med Phys,* 28**,** 63-72.

MILFORD, D., ROSBACH, N., BENDSZUS, M. & HEILAND, S. 2015. Mono-Exponential Fitting in T2-Relaxometry: Relevance of Offset and First Echo. *PLoS One,* 10**,** e0145255.

MITSUMOTO, H., BROOKS, B. R. & SILANI, V. 2014. Clinical trials in amyotrophic lateral sclerosis: why so many negative trials and how can trials be improved? *The Lancet Neurology,* 13**,** 1127-1138.

MOHAN, J., KRISHNAVENI, V. & GUO, Y. 2014. A survey on the magnetic resonance image denoising methods. *Biomedical Signal Processing and Control,* 9**,** 56-69.

MORRIS, T. P., WHITE, I. R. & CROWTHER, M. J. 2017. Using simulation studies to evaluate statistical methods.

MORROW, J. M., MATTHEWS, E., RAJA RAYAN, D. L., FISCHMANN, A., SINCLAIR, C. D., REILLY, M. M., THORNTON, J. S., HANNA, M. G. & YOUSRY, T. A. 2013. Muscle MRI reveals distinct abnormalities in genetically proven non-dystrophic myotonias. *Neuromuscular Disorders , 23 (8) pp. 637-646. (2013)*.

MORROW, J. M., SINCLAIR, C. D., FISCHMANN, A., MACHADO, P. M., REILLY, M. M., YOUSRY, T. A., THORNTON, J. S. & HANNA, M. G. 2016. MRI biomarker assessment of neuromuscular disease progression: a prospective observational cohort study. *Lancet Neurology , 15 (1) pp. 65-77. (2016)*.

MYUNG, I. J. 2003. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology,* 47**,** 90-100.

NAVAL POSTGRADUATE SCHOOL. 2004. *Determining the number of iterations fo Monte Carlo simulations of weapon effectiveness* [Online]. Available: https://calhoun.nps.edu [Accessed 23 Jan 2018].

NEMA. 2015. *Determination of Signal-to-Noise Ratio and Image Uniformity for Single-Channel, Non-Volume Coils in Diagnostic Magnetic Resonance Imaging (MRI)* [Online]. Available: https://www.nema.org/Standards/Pages/Determination-of-Signal-to-Noise-Ratio-and-Image-Uniformity-for-Single-Channel-Non-Volume-Coils-in-Diagnostic-Magnetic.aspx [Accessed 30 January 2019].

NEUMANN, D., BLAIMER, M., JAKOB, P. & BREUER, F. 2014. Simple recipe for accurate T2 quantification with multi spin-echo acquisitions. *Magnetic Resonance Materials in Physics, Biology and Medicine* 27**,** 567-577.

NYGREN, A. T. 2006. Shortened T2 after exercise in ischemic skeletal muscle. *J Magn Reson Imaging,* 23**,** 177-82.

PAPOULIS, A., S. UNNIKRISHNA PILLAI 2002. *Probability, random variables, and stochastic processes,* Boston; London, McGraw-Hill.

PARBOOSINGH, S. J., FIGLEWICZ, A. D., KRIZUS, A. A., MEININGER, S. V., AZAD, A. N., NEWMAN, A. D. & ROULEAU, A. G. 1997. Spinobulbar muscular atrophy can mimic ALS: The importance of genetic testing in male patients with atypical ALS. *Neurology,* 49**,** 568-572.

PAULY, J., LE ROUX, P., NISHIMURA, D. & MACOVSKI, A. 1991. Parameter relations for the Shinnar-Le Roux selective excitation pulse design algorithm (NMR imaging). *Medical Imaging, IEEE Transactions on,* 10**,** 53-65.

PENNSTATE. 2018. *Maximum Likelihood Estimation* [Online]. Available: https://newonlinecourses.science.psu.edu/stat414/node/191/ [Accessed 2019 23 Jan].

PETROVIC, A., SCHEURER, E. & STOLLBERGER, R. 2015. Closed-form solution for T2 mapping with nonideal refocusing of slice selective CPMG sequences. *Magnetic Resonance in Medicine,* 73**,** 818-827.

PHILLIPS, B. A., CALA, L. A., THICKBROOM, G. W., MELSOM, A., ZILKO, P. J. & MASTAGLIA, F. L. 2001. Patterns of muscle involvement in inclusion body myositis: clinical and magnetic resonance imaging study. *Muscle Nerve,* 24**,** 1526-34.

PHOENIX, J., BETAL, D., ROBERTS, N., HELLIWELL, T. R. & EDWARDS, R. H. 1996. Objective quantification of muscle and fat in human dystrophic muscle by magnetic resonance image analysis. *Muscle Nerve,* 19**,** 302-10.

POON, C. S., BRONSKILL, M. J., HENKELMAN, R. M. & BOYD, N. F. 1992. Quantitative magnetic resonance imaging parameters and their relationship to mammographic pattern. *J Natl Cancer Inst,* 84**,** 777-81.

POOT, D. H., KOTEK, G., NIESSEN, W. J. & KLEIN, S. Bias correction of maximum likelihood estimation in quantitative MRI. Medical Imaging 2013: Image Processing, 2013. International Society for Optics and Photonics, 86691F.

PRASLOSKI, T., MADLER, B., XIANG, Q. S., MACKAY, A. & JONES, C. 2012. Applications of stimulated echo correction to multicomponent T2 analysis. *Magn Reson Med,* 67**,** 1803-14.

RICE, S. O. 1944. Mathematical Analysis of Random Noise. *Bell System Technical Journal,* 23**,** 282-332.

RICOTTI, V., EVANS, M. R. B., SINCLAIR, C. D. J., BUTLER, J. W., RIDOUT, D. A., HOGREL, J. Y., EMIRA, A., MORROW, J. M., REILLY, M. M., HANNA, M. G., JANICZEK, R. L., MATTHEWS, P. M., YOUSRY, T. A., MUNTONI, F. & THORNTON, J. S. 2016. Upper Limb Evaluation in Duchenne Muscular Dystrophy: Fat-Water Quantification by MRI, Muscle Force and Function Define Endpoints for Clinical Trials. *PLOS ONE , 11 (9) (2016)*.

RINALDI, C., MALIK, B. & GREENSMITH, L. 2016. Targeted Molecular Therapies for SBMA. *J Mol Neurosci,* 58**,** 335-342.

SAAB, G., THOMPSON, R. T. & MARSH, G. D. 1999. Multicomponent T2 relaxation of in vivo skeletal muscle. *Magn Reson Med,* 42**,** 150-7.

SALEH ELESSAWY, S., ABD EL-GHAFFAR BORG, M., ABD EL-SALAM MOHAMED, M., ELSAYED ELHAWARY, G. & MOHAMED ABD EL-SALAM, E. 2013. The role of MRI in the evaluation of muscle diseases. *The Egyptian Journal of Radiology and Nuclear Medicine,* 44**,** 607-615.

SCHEFFLER, K. 1999. A pictorial description of steady-states in rapid magnetic resonance imaging. *Concepts in Magnetic Resonance,* 11**,** 291-304.

SCHLAEGER, S., WEIDLICH, D., KLUPP, E., MONTAGNESE, F., DESCHAUER, M., SCHOSER, B., BUBLITZ, S., RUSCHKE, S., ZIMMER, C. & RUMMENY, E. J. 2019. Decreased water T2 in fatty infiltrated skeletal muscles of patients with neuromuscular diseases. *NMR in Biomedicine***,** e4111.

SESTO, M. E., CHOURASIA, A. O., BLOCK, W. F. & RADWIN, R. G. 2008. Mechanical and magnetic resonance imaging changes following eccentric or concentric exertions. *Clin Biomech (Bristol, Avon),* 23**,** 961-8.

SHAFER-CRANE, G. A., MEYER, R. A., SCHLINGER, M. C., BENNETT, D. L., ROBINSON, K. K. & RECHTIEN, J. J. 2005. Effect of occupational keyboard typing on magnetic resonance imaging of the median nerve in subjects with and without symptoms of carpal tunnel syndrome. *Am J Phys Med Rehabil,* 84**,** 258-66.

SHARAFI, A., CHANG, G. & REGATTE, R. R. 2017. Bi-component T1ρ and T2 Relaxation Mapping of Skeletal Muscle In-Vivo. *Scientific Reports,* 7**,** 14115.

SIJBERS, J. & DEN DEKKER, A. J. 2004. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magnetic resonance in medicine,* 51**,** 586-594.

SIJBERS, J., DEN DEKKER, A. J., SCHEUNDERS, P. & VAN DYCK, D. 1998a. Maximum-likelihood estimation of Rician distribution parameters. *Medical Imaging, IEEE Transactions on,* 17**,** 357-361.

SIJBERS, J., DEN DEKKER, A. J., VERHOYE, M., RAMAN, E. R. & VAN DYCK, D. 1998b. Optimal estimation of T2 maps from magnitude MR images. *Medical Imaging 1998: Image Processing.* International Society for Optics and Photonics.

SINCLAIR, C. D., MORROW, J. M., JANICZEK, R. L., EVANS, M. R., RAWAH, E., SHAH, S., HANNA, M. G., REILLY, M. M., YOUSRY, T. A. & THORNTON, J. S. 2016. Stability and sensitivity of water T2 obtained with IDEAL-CPMG in healthy and fat-infiltrated skeletal muscle. *NMR in Biomedicine , 29 (12) pp. 1800-1812. (2016)*.

SINCLAIR, C. D. J., MORROW, J. M., HANNA, M. G., REILLY, M. M., YOUSRY, T. A., GOLAY, X. & THORNTON, J. S. 2012. Correcting radiofrequency inhomogeneity effects in skeletal muscle magnetisation transfer maps. *NMR Biomed,* 25**,** 262-270.

SIRACUSANO, G., LA CORTE, A., MILAZZO, C., ANASTASI, G. P., FINOCCHIO, G. & GAETA, M. 2017. On the R2$*$ relaxometry in complex multi-peak multi-Echo chemical shift-based water-fat quantification: Applications to the neuromuscular diseases. *Magn Reson Imaging,* 35**,** 4-14.

STOLLBERGER, R. & WACH, P. 1996. Imaging of the active B 1 field in vivo. *Magnetic Resonance in Medicine,* 35**,** 246-251.

STRAUB, V., CARLIER, P. G. & MERCURI, E. 2012. TREAT-NMD workshop: pattern recognition in genetic muscle diseases using muscle MRI: 25-26 February 2011, Rome, Italy. *Neuromuscul Disord,* 22 Suppl 2**,** S42.

SULLIVAN, D. C., OBUCHOWSKI, N. A., KESSLER, L. G., RAUNIG, D. L., GATSONIS, C., HUANG, E. P., KONDRATOVICH, M., MCSHANE, L. M., REEVES, A. P. & BARBORIAK, D. P. 2015. Metrology standards for quantitative imaging biomarkers. *Radiology,* 277**,** 813-825.

THOMAS, L. W. J. Q. J. O. E. P., TRANSLATION, C. M. S. & INTEGRATION 1962. The chemical composition of adipose tissue of man and mice. 47**,** 179-188.

TOFTS, P. 2004. *Quantitative MRI of the Brain,* Chichester, UK, John Wiley & Sons.

UCLA. 2011. *FAQ: What are pseudo R-squareds?* [Online]. Available: https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/ [Accessed 19 June 2020].

UDDIN, M. N., MARC LEBEL, R. & WILMAN, A. H. 2013. Transverse relaxometry with reduced echo train lengths via stimulated echo compensation. *Magnetic Resonance in Medicine,* 70.

WALKER-SAMUEL, S., ORTON, M., MCPHAIL, L. D. & ROBINSON, S. P. 2009. Robust estimation of the apparent diffusion coefficient (ADC) in heterogeneous solid tumors. *Magn Reson Med,* 62**,** 420-9.

WALKER-SAMUEL, S., ORTON, M., MCPHAIL, L. D., BOULT, J. K., BOX, G., ECCLES, S. A. & ROBINSON, S. P. J. M. R. I. M. 2010. Bayesian estimation of changes in transverse relaxation rates. 64**,** 914-921.

WALKER-SAMUEL, S., ORTON, M., MCPHAIL, L. D. & ROBINSON, S. P. 2009. Robust estimation of the apparent diffusion coefficient (ADC) in heterogeneous solid tumors. *Magnetic Resonance in Medicine,* 62**,** 420-429.

WALSH, D., GMITRO, A. & MARCELLIN, M. 2000. Adaptive reconstruction of phased array MR imagery. *Magnetic Resonance In Medicine,* 43**,** 682-690.

WARY, C., AZZABOU, N., GIRAUDEAU, C., LE LOUER, J., MONTUS, M., VOIT, T., SERVAIS, L. & CARLIER, P. 2015. Quantitative NMRI and NMRS identify augmented disease progression after loss of ambulation in forearms of boys with Duchenne muscular dystrophy. *NMR Biomed,* 28**,** 1150-62.

WATTJES, M., KLEY, R. & FISCHER, D. 2010. Neuromuscular imaging in inherited muscle diseases. *Eur Radiol,* 20**,** 2447-2460.

WEIGEL, M. 2015. Extended phase graphs: Dephasing, RF pulses, and echoes - pure and simple. *Journal of Magnetic Resonance Imaging,* 41**,** 266-295.

WHITTAKER, R. G., PORCARI, P., BRAZ, L., WILLIAMS, T. L., SCHOFIELD, I. S. & BLAMIRE, A. M. 2019. Functional magnetic resonance imaging of human motor unit fasciculation in amyotrophic lateral sclerosis. *Annals of Neurology,* 85**,** 455-459.

WILLIS, T. A., HOLLINGSWORTH, K. G., COOMBS, A., SVEEN, M. L., ANDERSEN, S., STOJKOVIC, T., EAGLE, M., MAYHEW, A., DE SOUSA, P. L., DEWAR, L., MORROW, J. M., SINCLAIR, C. D., THORNTON, J. S., BUSHBY, K., LOCHMÜLLER, H., HANNA, M. G., HOGREL, J. Y., CARLIER, P. G., VISSING, J. & STRAUB, V. 2013. Quantitative Muscle MRI as an Assessment Tool for Monitoring Disease Progression in LGMD2I: A Multicentre Longitudinal Study. *PLoS One , 8 (8) , Article e70993. (2013)*.

WOKKE, B. H., BOS, C., REIJNIERSE, M., RIJSWIJK, C. S., EGGERS, H., WEBB, A., VERSCHUUREN, J. J. & KAN, H. E. 2013. Comparison of dixon and T1-weighted MR methods to assess the degree of fat infiltration in duchenne muscular dystrophy patients. *Journal of Magnetic Resonance Imaging,* 38**,** 619-624.

YAO, L. & GAI, N. 2012. Fat-corrected T2 measurement as a marker of active muscle disease in inflammatory myopathy. *AJR Am J Roentgenol,* 198**,** W475-81.

YUSHKEVICH, P. A., PIVEN, J., HAZLETT, H. C., SMITH, R. G., HO, S., GEE, J. C. & GERIG, G. 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage,* 31**,** 1116-1128.

ZAFEIROPOULOS, N. A simplified method to determine tissue-water T2 from CPMG image data in fat infiltrated skeletal muscle: application in the forearm in Duchenne muscular dystrophy. ISMRM, 2016 Singapore.

ZAIDMAN, M. C. & VAN ALFEN, M. N. 2016. Ultrasound in the Assessment of Myopathic Disorders. *Journal of Clinical Neurophysiology,* 33**,** 103-111.