LUND UNIVERSITY

Development of machine learning-based analysis and processing tools applied to magnetization transfer MRI

Z-spectral denoising and fitting

Mohammed Ali, Sajad

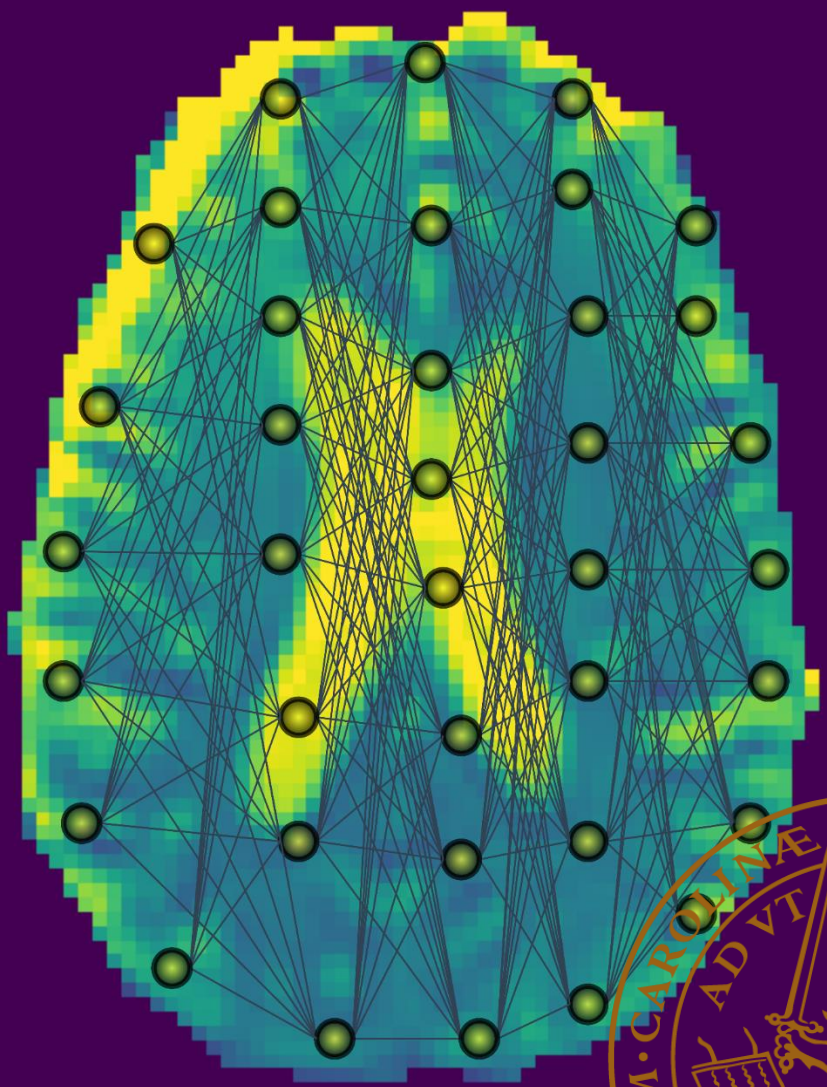
2025

Document Version:
Publisher's PDF, also known as Version of record

Development of machine learning-based analysis and processing tools applied to magnetization transfer MRI

Z-spectral denoising and fitting

SAJAD MOHAMMED ALI
MEDICAL RADIATION PHYSICS, LUND
FACULTY OF SCIENCE | LUND UNIVERSITY



Development of machine learning-based analysis and processing tools applied to magnetization transfer MRI

Development of machine learning-based analysis and processing tools applied to magnetization transfer MRI

Z-spectral denoising and fitting

Sajad Mohammed Ali

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy at the Faculty of Science at Lund University to be publicly defended on 14th of October at 09.00 in Lecture Hall 5, Skåne University Hospital, Entrégatan 7, 222 42 Lund

Faculty opponent
Professor Michael Chappell, School of Medicine, University of Nottingham, Nottingham, United Kingdom

Organization: LUND UNIVERSITY Faculty of Science Department of Medical Radiation Physics	**Document name:** DOCTORAL DISSERTATION	
	Date of issue: 2025-10-14	
Author: Sajad Mohammed Ali	**Sponsoring organization:**	

Title and subtitle:
Development of machine learning-based analysis and processing tools applied to magnetization transfer MRI: Z-spectral denoising and fitting

Abstract:
Chemical exchange saturation transfer MRI is a promising *magnetization transfer* (MT) imaging technique that utilizes, by *radiofrequency* irradiation, selective saturation of exchangeable protons in solutes. This enables the indirect detection of the solutes through the water signal after saturation transfer by chemical exchange, *cross-relaxation,* or the combination of the two. Typically, the water-signal change is studied within a frequency range where different molecular protons have visible resonances in the proton *nuclear magnetic resonance* spectrum due to the molecular mobility in the solution. In saturation transfer experiments, a normalized water signal as a function of saturation frequency, also referred to as a *Z-spectrum,* is obtained. Like any other medical imaging modality, the MT imaging technique may be hampered by several drawbacks and limitations. A small effect size, together with the unavoidable noise in digital imaging, might render Z-spectra useless, and various data processing strategies are thus warranted. In the work described in this thesis the *deep learning* (DL)-based *constrained loss autoencoder residual denoiser* was developed, which combined the strength of latent mapping by autoencoders with subtractive residual denoising to obtain improved denoising performance and enable an increased level of signal recovery compared to other state-of-the-art approaches. Moreover, even for processed or high-quality Z-spectra, valuable biochemical information from various convolved contributions is commonly extracted by fitting the spectral data to a model. Conventional algorithms such as *least squares* (LS) fitting hamper efficient analysis due to inherent limitations such as a high dependence on data quality and sampling density, as well as long fitting times. To increase the feasibility of implementing MT-techniques, promote their clinical applications, and allow for studies of larger cohorts, it is necessary to streamline and standardize the analysis. A DL-based fitting approach for *direct water saturation* (DS) Z-spectra was developed and provided increased robustness and accelerated fitting compared to LS. The developed method found an application in a subsequent study by showing tangible differences in *linewidth* changes of DS spectra (pre- and post-glucose infusion) for healthy brain tissue and tumor. Finally, a multi-pool *machine learning*-based fitting approach using *gradient boosted decision trees* was also developed in the work of this thesis. The reduction in algorithmic complexity resulted in training times of approximately one minute, thus providing more freedom to change acquisition protocols. The fitting time was also reduced to approximately one second per brain compared to several hours with LS. The goodness-of-fit of four components was also empirically compared across the Lorentzian and Voigt spectral models, showing a statistically significant improvement for the latter.

Key words:

Classification system and/or index terms (if any):

Supplementary bibliographical information:	**Language:** English
ISSN and key title:	**ISBN:** 978-91-8104-634-2 (print) 978-91-8104-635-9 (electronic)
Recipient´s notes:	**Number of pages:** 111 **Price:** **Security classification:**

Signature: _____ **Date:** _____

Development of machine learning-based analysis and processing tools applied to magnetization transfer MRI

Z-spectral denoising and fitting

Sajad Mohammed Ali

LUND
UNIVERSITY

Faculty of Science
Department of Medical Radiation Physics

ISBN 978-91-8104-634-2 (print)
ISBN 978-91-8104-635-9 (electronic)

Printed in Sweden by E-huset, Lund University
Lund 2025

بسم الله الرحمن الرحيم

وما توفيقي الا بالله عليه توكلت و اليه انيب

Table of contents

Abstract

Chemical exchange saturation transfer MRI is a promising *magnetization transfer* (MT) imaging technique that utilizes, by *radiofrequency* irradiation, selective saturation of exchangeable protons in solutes. This enables the indirect detection of the solutes through the water signal after saturation transfer by chemical exchange, *cross-relaxation,* or the combination of the two. Typically, the water-signal change is studied within a frequency range where different molecular protons have visible resonances in the proton *nuclear magnetic resonance* spectrum due to the molecular mobility in the solution. In saturation transfer experiments, a normalized water signal as a function of saturation frequency, also referred to as a *Z-spectrum,* is obtained. Like any other medical imaging modality, the MT imaging technique may be hampered by several drawbacks and limitations. A small effect size, together with the unavoidable noise in digital imaging, might render Z-spectra useless, and various data processing strategies are thus warranted. In the work described in this thesis the *deep learning* (DL)-based *constrained loss autoencoder residual denoiser* was developed, which combined the strength of latent mapping by autoencoders with subtractive denoising in the latent dimensions to obtain improved denoising performance and enable an increased level of signal recovery compared to other state-of-the-art approaches. Moreover, even for processed or high-quality Z-spectra, valuable biochemical information from various convolved contributions is commonly extracted by fitting the spectral data to a model. Conventional algorithms such as *least squares* (LS) fitting hamper efficient analysis due to inherent limitations such as a high dependence on data quality and sampling density, as well as long fitting times. To increase the feasibility of implementing MT-techniques, promote their clinical applications, and allow for studies of larger cohorts, it is necessary to streamline and standardize the analysis. A DL-based fitting approach for *direct water saturation* (DS) Z-spectra was developed and provided increased robustness and accelerated fitting compared to LS. The developed method found an application in a subsequent study by showing tangible differences in *linewidth* changes of DS spectra (pre- and post-glucose infusion) for healthy brain tissue and tumor. Finally, a multi-pool *machine learning*-based fitting approach using *gradient boosted decision trees* was also developed in the work of this thesis. The reduction in algorithmic complexity resulted in training times of approximately one minute, thus providing more freedom to change acquisition protocols. The fitting time was also reduced to approximately one second per brain compared to several hours with LS. The goodness-of-fit of four components was also empirically compared across

the Lorentzian and Voigt spectral models, showing a statistically significant improvement for the latter.

Populärvetenskaplig sammanfattning

Magnetiseringsöverföring (eng. *magnetization transfer, MT*) är ett fysikaliskt/kemiskt fenomen som kan utnyttjas inom *magnetresonanstomografi* (MRT) för att icke-invasivt undersöka substanser som förekommer i låga koncentrationer. Dessa substanser kan finnas i kroppen, *endogena*, eller så kan de administreras utifrån före undersökningen, *exogena*. En magnetisk märkning (s.k. saturering) av protonerna i dessa substanser åstadkoms genom att de exponeras för radiofrekventa vågor med en frekvens som är unik för varje typ av substans. Denna saturering följs av ett utbyte av protoner, genom olika fysikaliska eller kemiska processer, mellan substansen och det fria vatten som finns i hög koncentration i kroppen. När omärkta protoner i vattnet ersatts med substansens märkta protoner kan dess koncentration indirekt mätas genom den resulterande förändringen i vattensignalen. Den uppmäta vattensignalen studeras över ett relevant frekvensomfång i det som kallas Z-spektrum. Z-spektra innehåller riklig information av fysiologisk relevans och denna information kan till exempel användas för att skilja normal vävnad från patologisk. För att kunna extrahera den värdefulla informationen krävs ofta en inledande behandling av rådata (insamlade obearbetade Z-spektra), eftersom effektstorleken från MT är liten samtidigt som data oundvikligen innehåller brus. Efter databehandlingen följer analys av stora mängder Z-spektra. *Artificiell intelligens* (AI) har det senaste årtiondet uppvisat en enorm utveckling och smugit sig in nästan överallt i vardagen, även i tillämpningar inom medicinsk avbildning. I denna doktorsavhandling har olika lösningar baserade på *maskininlärning* (ML), som är en gren inom AI, utvecklats, dels för att förbättra och effektivisera analysen av MT-data, dels för att förbättra datakvaliteten med avseende på brus. Mer explicit har ML-lösningar utvecklats för att analysera Z-spektra genom att extrahera relevanta parametrar av fysiologisk betydelse. Jämfört med konventionella metoder för motsvarande analyser har de utvecklade ML-metoderna visat ökad robusthet (mot diverse faktorer som påverkar datakvaliteten) och framför allt kortare analystider. För en mängd Z-spektra från en hel hjärnavbildning tar analysen ungefär en sekund jämfört med flera timmar med konventionella metoder. En ML-baserad brusreduceringsmetod har också utvecklats som resulterat i lägre brusnivå med högre grad av återhämtning av förlorad kontrast jämfört med tidigare motsvarande metoder.

Original papers

This doctoral thesis is based on three publications and one manuscript, which are referred to below by their Roman numerals.

I. **Sajad Mohammed Ali**, Nirbhay N. Yadav, Ronnie Wirestam, Munendra Singh, Hye-Young Heo, Peter C. van Zijl, Linda Knutsson. Deep learning-based Lorentzian fitting of water saturation shift referencing spectra in MRI. *Magn Reson Med*. 2023;90(4):1610-1624. doi: 10.1002/mrm.29718

II. **Sajad Mohammed Ali**, Peter C.M. van Zijl, Jannik Prasuhn, Ronnie Wirestam, Linda Knutsson, Nirbhay N. Yadav. Machine learning-based multi-pool Voigt fitting of CEST, rNOE and MTC in Z-spectra. *Magn Reson Med.* 2025;94(1):346-361. doi: 10.1002/mrm.30460

III. Linda Knutsson, Nirbhay N. Yadav, **Sajad Mohammed Ali**, David Olayinka Kamson, Eleni Demetriou, Anina Seidemo, Lindsay Blair, Doris D. Lin, John Laterra, Peter C.M. van Zijl. Dynamic glucose enhanced imaging using direct water saturation. *Magn Reson Med.* 2025;94(1):15-27. doi: 10.1002/mrm.30447

IV. **Sajad Mohammed Ali**, Patrick Schunke, Peter C. van Zijl, Ronnie Wirestam, Jannik Prasuhn, Hye-Young Heo, Pia C. Sundgren, Linda Knutsson, Nirbhay N. Yadav. A constrained loss autoencoder residual denoiser (CLAERD): Application to saturation transfer spectra. *Manuscript*.

List of Contributions

A summary of my contributions to the papers of this thesis is listed below.

I. I participated in the study design, studied several fitting solutions, designed and optimized the single-Lorentzian fitting neural network (sLoFNet) used for fitting the direct water saturation Z-spectra. I evaluated the performance using simulated and in vivo data. I conducted the statistical analysis and was the first author of the paper.

II. Together with Dr. Nirbhay Yadav, I conceived the study and formulated the study design. I pre-processed the datasets for training and testing. I trained the machine learning models, evaluated their performance and conducted the statistical analysis. I was the first author of the paper.

III. I developed the deep learning (DL)-based Lorentzian fitting approach used in the study and assisted in its implementation for the data analysis. I reviewed and provided edits to the paper. I was the third author of the paper.

IV. I invented the constrained loss autoencoder residual denoiser (CLAERD) and conceived the study. I trained, optimized, and evaluated the DL-based denoiser. I compared it to previous conventional and DL-based approaches. I made a thorough analysis of the model's latent functionality. I was the first author of the paper.

Abbreviations and symbols

A	Amplitude, i.e., the peak area
AE	Autoencoder
AI	Artificial intelligence
APT-w	Amide proton transfer weighted
BMC	Bloch McConnell
CAE	Convolutional autoencoder
CART	Classification and regression trees
CESL	Chemical exchange sensitive spin lock
CEST	Chemical exchange saturation transfer
CNN	Convolutional neural network
CPU	Central processing unit
CSA	Chemical shift anisotropy
CSF	Cerebrospinal fluid
DCAE	Denoising convolutional autoencoder
DGE	Dynamic glucose enhanced
DL	Deep learning
DNN	Deep neural network
DS	Direct water saturation
DT	Decision tree
ELU	Exponential linear unit
FA	Flip angle
FDA	Food and drug administration
FWHM	Full width at half maximum

GBDT	Gradient boosted decision trees
GD	Gradient descent
GM	Gray matter
GNA	Gauss-Newton algorithm
GPU	Graphics processing unit
GT	Ground truth
H	Height, i.e., linewidth-weighted amplitude
HPO	Hyperparameter optimization
HWHM	Half-width at half maximum
LD	Lorentzian difference
LMA	Levenberg-Marquart algorithm
LS	Least squares
LW	Linewidth
MAE	Mean absolute error
MAP	Maximum a posteriori estimation
ML	Machine learning
MLE	Maximum likelihood estimation
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MSE	Mean squared error
MT	Magnetization transfer
MTR_{asym}	Magnetization transfer ratio asymmetry
MTR_{Rex}	MTR exchange dependent relaxation contribution
MTC	Magnetization transfer contrast
MVN	Multivariate Newton's method
NMR	Nuclear magnetic resonance
NN	Neural network
NOE	Nuclear Overhauser enhancement
PC	Principal component

PDF	Probability density function
PTR	Proton transfer ratio
ReLu	Rectified linear unit
RF	Radiofrequency (Chapter 3 and 4)
RF	Random forests (Chapter 5)
rNOE	NOE-relayed exchange
SAR	Specific absorption rate
SNR	Signal-to-noise ratio
Tanh	Hyperbolic tangent
TE	Echo time
TR	Repetition time
TV	Total variation
WASSR	Water saturation shift referencing
WM	White matter

1. Introduction

In the last few decades, non-invasive *magnetization transfer* (MT) *magnetic resonance imaging* (MRI) techniques have emerged enabling indirect probing of naturally occurring (endogenous) or externally administered (exogenous) solutes, even at low concentration (mM), through the effects on the free water signal. This is achievable through the selective saturation of the compound protons by *radiofrequency* (RF) field irradiation at their specific resonance frequency, followed by the transfer of the saturation through an exchange with the bulk water protons.[1–6] The MT effects are commonly observed using a so-called *Z-spectrum,* which is the normalized water signal as a function of saturation frequency.[3,4] Retrieval of this information has shown promising potential in several clinical applications, such as stroke detection,[7–9] brain cancer[9–12] and non-brain cancer diagnoses,[13–15] as well as cancer therapy monitoring.[16] The responsible physician's interpretation and understanding of the MT data are crucial for diagnosis, treatment planning and follow-up. However, several limitations and challenges still need to be overcome to unlock the full potential of these techniques, both in data-quality improvement by post-processing, as well as in standardization and increased efficiency of analyses.

Artificial intelligence (AI), with its many different fields − not least *machine learning* (ML) − has grown vastly over the last few decades as a consequence of improved hardware and increased data availability.[17–22] This has led to various everyday implementations and a plethora of applications within medical imaging. In data science and technology, the growth of the number of publications related to AI and ML has resembled an exponential over the last few years, a trend also observed in natural science- and medicine-based journals.[23] Not surprisingly, ML has thus been implemented in various ways within MT-based imaging. Successful implementations can be found both in terms of analysis, such as in *deep learning* (DL)-based parameter estimation from *chemical exchange saturation transfer* (CEST) images,[24] and processing, such as in the DL-based *super-resolution* approach for reconstructing high-resolution CEST images from fast low-resolution acquisitions.[25] Despite the increasing number of ML-based publications on MT imaging, it is not far-fetched to state that the field is still in its infancy due to the numerous possibilities provided by ML and the continued need for improved or innovative analysis and processing methods for MT imaging. In light of this development, the work of this thesis has focused on finding ML-based solutions to

some of the current limitations and challenges faced in MT-based imaging in the realms of both analysis and processing.

For processing specifically, efforts were directed toward addressing the inherent problem of noise in digital imaging. Due to the generally small effect size in MT imaging, the unavoidable noise may result in a loss of observable contrast, thus rendering the data useless. While several denoising approaches have been applied to MT data, both conventional[26] and DL-based,[27,28] the ability to restore an observable effect on the signal (especially signals that are small to begin with) with high fidelity remains challenging. With this as a motivational factor, an innovative DL-based approach was developed (described in Paper IV). In the attempt to streamline the analysis of MT data, focus was placed on fitting Z-spectra, which is a central approach for extracting the wealth of biochemical information that the spectra hold. Firstly, a study primarily concerned with increasing robustness and reducing time consumption when fitting *direct water saturation* (DS) spectra was presented in Paper I. Secondly, an investigation that focused on reducing time complexity, both in terms of training and inference, as well as comparing different multi-pool spectral profiles for modeling Z-spectral components, was described in Paper II. The need for effective and robust fitting is also essential in research applications, as exemplified in Paper III, where the approach developed in the work of Paper I was applied.

This thesis is structured in the following manner: Chapter 3 provides a general theoretical background regarding MT MRI and noise in MRI, and the Z-spectrum is also described. Chapter 4 focuses on various spectral models used for Z-spectra and provides a comparison based on the empirical findings reported in Paper II. Chapter 5 gives an overview of conventional fitting approaches and delves deep into the ML-based Z-spectral fitting solutions proposed in Papers I and II. Chapter 6 investigates available DL-based retrospective denoising approaches, including the proposed solution described in Paper IV. Finally, Chapter 7 includes a critical review of the completed work along with conclusions and future perspectives.

2. Aims

The projects described in this thesis focused primarily on developing and applying ML-based solutions to streamline and standardize the processing and analysis of MT-based images and to solve their inherent challenges.

The specific aims of the projects described in this thesis were as follows:

- Developing and testing ML-based methods for a robust and streamlined analysis of MT MRI data in the form of Z-spectral fitting (Papers I and II).
- Implementing and applying a developed method in clinical research applications (Paper III).
- Developing, testing and implementing an innovative DL-based approach for denoising in CEST-MRI (Paper IV).

3. Magnetic resonance imaging and magnetization transfer

3.1 Magnetic resonance imaging

In 1946, Bloch and Purcell observed the first *nuclear magnetic resonance* (NMR) phenomena in bulk matter.[29,30] Nuclei with nonzero spin, like ^{1}H and ^{31}P, have a magnetic dipole moment associated with their spins. When an external static magnetic field is applied, the spin or magnetic dipole vectors will orient with a precessional motion either parallel to the magnetic field orientation, corresponding to a lower energy state, or antiparallel, corresponding to a higher energy state.[31,32] In the quantum mechanical description, the spins can also exist in superposition states, where their orientation is not strictly defined until measured.[33] Nevertheless, the difference in the populations of the two energy states results in a so-called *polarization,* which is proportional to the magnetization measured in MRI. The relation between the populations of the parallel (N_+) and anti-parallel (N_-) spins is given by *Boltzmann's distribution*[34]:

$$\frac{N_+}{N_-} = e^{\frac{2\mu_p B_0}{kT}}, \tag{3.1}$$

where $\mu_p = 1.4 \cdot 10^{-26}$ J/T is the intrinsic magnetic dipole moment for a proton, B_0 is the magnetic flux density in T, $k = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann constant, and T is the absolute temperature in K. The highest magnetic field strength for clinical use approved by the U.S. *Food and Drug Administration* (FDA) is 7 T,[35] and the typical room temperature is 293 K. Insertion of these values into Equation 3.1 results in a difference of less than 50 ppm between the two spin populations, indicating that the resulting net magnetization vector, M, has a small magnitude. Since the spins have a precession frequency when residing in an external field, M will precess around the direction of the external magnetic field lines with a resonance frequency referred to as the *Larmor frequency*, ω_0. This frequency is given as the product of the magnetic flux density, B_0, and the *gyromagnetic constant*, γ, which is an element-dependent quantity ($\gamma/2\pi$ = 42.6 MHz/T for ^{1}H).[32]

It is the existence of a net magnetization vector that makes MR imaging possible. By using an RF pulse to apply an orthogonal time-varying magnetic field, B_1, to the sample in a static B_0-field, the net magnetization vector can be "flipped" towards the xy-plane (Figure 3.1). Using a receiver coil, a *free induction decay* signal can be recorded due to the precession of the flipped magnetization vector, in accordance with *Faraday's law of induction*,[36] thus yielding a time-varying signal.

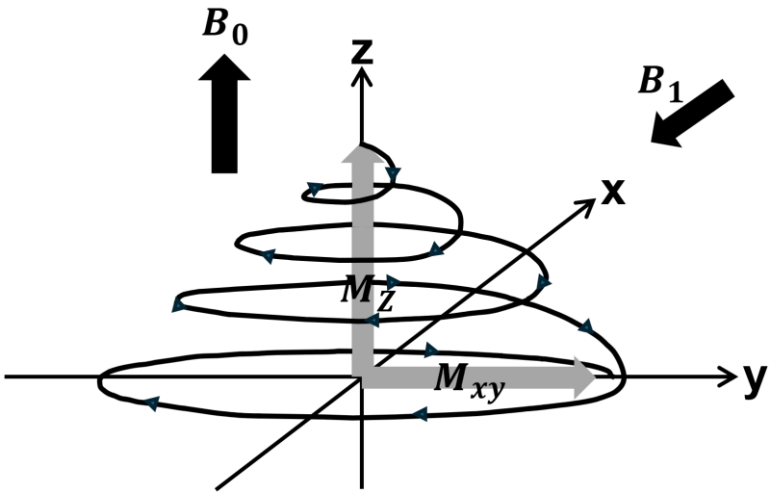


Figure 3.1. Flipping of the net magnetization vector. The nutational motion of the net magnetization vector towards the xy-plane from the equilibrium state in the direction of the applied static magnetic field under the influence of a time-varying magnetic field, B_1.

After the RF pulse is terminated, the flipped magnetization vector will recover to its thermal equilibrium state, a process referred to as relaxation. The *longitudinal* (spin-lattice) relaxation time is referred to as T_1, while the *transverse* (spin-spin) relaxation time is referred to as T_2. The time evolution of the magnetization in a static magnetic field is given by the *Bloch equations*, which are defined as:

$$\frac{dM_x(t)}{dt} = \gamma\left(M_y(t)\cdot B_z(t) - M_z(t)\cdot B_y(t)\right) - R2\cdot M_x(t)\,, \qquad (3.2)$$

$$\frac{dM_y(t)}{dt} = \gamma\left(M_z(t)\cdot B_x(t) - M_x(t)\cdot B_z(t)\right) - R2\cdot M_y(t)\,, \qquad (3.3)$$

$$\frac{dM_z(t)}{dt} = \gamma\left(M_x(t)\cdot B_y(t) - M_y(t)\cdot B_x(t)\right) - R1(M_z(t) - M_0)\,, \qquad (3.4)$$

where $M_x, M_y \; and \; M_z$ indicate the components in the x, y and z directions of the time-dependent magnetization vector, respectively, while M_0 is the magnitude of the magnetization at thermal equilibrium in the static magnetic field. Similarly, $B_x, B_y \; and \; B_z$ are the components of the applied magnetic field experienced by the

nuclei. Note that R_1 and R_2 are the *longitudinal* and *transverse relaxation* rates, respectively, defined as $R_1 = 1/T_1$ and $R_2 = 1/T_2$.

The proton of the hydrogen nucleus, which is highly abundant in water, is the primary target in MRI. Depending on the tissue compartment or molecular environment in which the water resides, different degrees of binding exist (e.g., free in the cytoplasm or restricted in semisolid environments such as membranes), resulting in a variation of molecular motions affecting the relaxation properties. Moreover, various chemical, physical and magnetic interactions (more detailed discussion in the following sections) occur for the water proton spins, ultimately affecting the measured signal. Hence, the properties of the water protons reflect both the macroscopic and microscopic organization of tissue.[1] By varying acquisition settings such as *flip angle* (FA), *repetition time* (TR) and *echo time* (TE), various pulse sequences can be created to produce contrast patterns that reflect the different inherent contrast mechanisms. The acquired MR signal is stored in the so-called *k-space*. To obtain the MR image, the signal data need to be *Fourier-transformed*.[31,32] By direct inverse transformation of the complex k-space signal, a *complex image* is generated. However, *magnitude* and *phase* images are the most common[37,38] since they provide more diagnostic value due to their direct interpretability. It should be noted that these images are obtained by nonlinear operations applied to the complex signal before the subsequent inverse Fourier transform.

3.1.1 Noise in magnetic resonance imaging

The main contributor to noise in MRI images is the thermal motion of the electrons and ions in the studied object. Charged particles in motion will exchange energy upon interactions with surrounding atoms, which may result in (signal-interfering) electromagnetic emissions referred to as *thermal noise*.[39–41] Other significant contributions to the observed interference arise from the electronic components of the system itself, including the pre-amplifier, receiver coil and analogue-to-digital converter[41]. These have a variety of sources, not limited to but including thermal noise, as well as *flicker noise,* which is partly explained by imperfections in the electronic components.[42] The latter, also referred to as $1/f - noise$, where f is the frequency, is thus more notable at lower frequencies. While some sources such as flicker noise yield noise that adheres to non-Gaussian distributions, most of the noise contributing to the MRI signal (including thermal noise) follow a Gaussian *probability density function* (PDF).[43] Furthermore, a linear combination of Gaussians will result in another Gaussian probability function, with the mean and variance equal to the linear combination of the individual means and variances. For the raw MR data collected in k-space, the governing PDF is thus a Gaussian.[37,38,43–45] To obtain complex MR images, the raw k-space data are transformed through the discrete inverse Fourier transform. Due to the linearity and orthogonality of the transform, the PDFs governing the noise in complex (i.e., real and imaginary) MR

images are maintained.[37,38,43,44] However, the more common magnitude and phase images in MRI[37,38] are obtained using nonlinear operations. A magnitude MR image X is generated by taking the pixel-by-pixel root sum of squares of the real and imaginary components of a complex image[38,43–46], expressed as:

$$X = \sqrt{(A_R + n_R)^2 + (A_I + n_I)^2}\,, \tag{3.5}$$

where A_R and A_I are the real and imaginary signal components while n_I and n_R are the real and imaginary uncorrelated Gaussian noise variables. For the resulting image, the PDF governing the pixel values is transformed to a *Rician distribution*,[37,45,46] which is defined as:

$$p(x) = \frac{x}{\sigma^2} exp\left(-\frac{x^2+\nu^2}{2\sigma^2}\right) \cdot I_0\left(\frac{x\nu}{\sigma^2}\right) H(x)\,, \tag{3.6}$$

where x is the magnitude image signal, ν the non-centrality parameter given by $\nu = \sqrt{A_R^2 + A_I^2}$, σ is the standard deviation of the noise, I_0 the zeroth-order Bessel function and H the Heaviside function, thus ensuring the PDF expression for x is only valid for non-negative values.[43] In the image regions where the signal components are zero ($\nu = 0$), the Rician simplifies to a so-called *Rayleigh distribution*. Figure 3.2 provides an overview comparing the main characteristics of three PDFs. The Gaussian probability function is symmetrically centered around a mean value, μ, and its tails stretches infinitely on both sides. However, 68% of the data falls within one standard deviation (std). Since the Rician distribution describes the magnitude data, it cannot take negative values (Equation 3.5) and the same applies to the Rayleigh distribution. The scaling parameter (σ) of the Rician and Rayleigh distributions is analogous to the std of the Gaussian and controls the spread of the distributions, while the noncentrality parameter of the Rician distribution controls the position. It is worth mentioning that subtracting two Rician distributions results in a Gaussian-like distribution.[47] This becomes relevant, for example, in *arterial spin labeling* imaging, where a labeled image is subtracted from a control image. It is also relevant in the novel denoising method described in Paper IV, where a subtraction is applied between the latent representations of noisy signal and pure noise (a more detailed discussion is provided in Chapter 6.2).

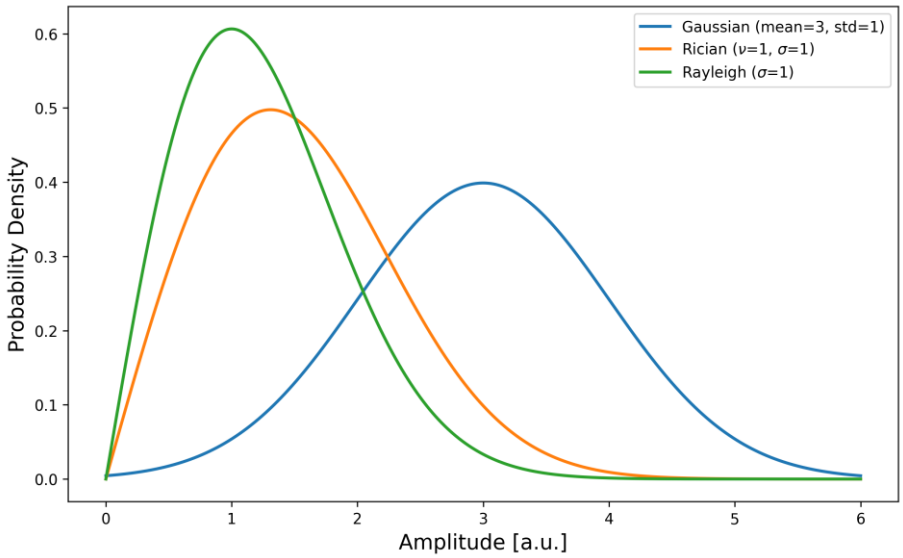


Figure 3.2. Comparison of the characteristics of Gaussian, Rician and Rayleigh distributions. The Gaussian (blue line) is centered about a mean value ($\mu = 3$ in this case) with symmetric tails and can thus have negative values. The Rician and Rayleigh probability functions cannot have negative values. Note that the scaling parameter is the same for the Rician (orange line) and the Rayleigh (green line), and that the difference is caused by the nonzero value of the noncentrality parameter for the Rician.

3.2 Magnetization transfer

Conventional MRI relies on the Bloch equations[29] (Equations 3.2 – 3.4) for describing the time evolution of the collective nuclear magnetization in an external static magnetic field (B_0) after RF excitation by applying a time-varying magnetic field (B_1) under the influence of relaxation processes (longitudinal and transversal). However, physical (e.g., *diffusion* and *flow*), chemical (e.g., *substrate binding* and *chemical exchange*), and magnetic interactions (e.g., *scalar* and *dipolar coupling*) also take place and affect the magnetization.[1,48,49] Chemical exchange and magnetic interactions are especially important to MT, a term that includes processes occurring when the effect of saturation on solute proton magnetization is transferred within or between molecules. Chemical exchange is the process wherein the transfer of magnetization occurs through the physical exchange of protons between molecular situations in which they experience a different proton chemical shift. This can be due to transfer between different molecules, or within molecules between different magnetic environments.[1–4] As for magnetic interactions, scalar coupling occurs due to the interaction of spins through the electrons in the chemical bond of the molecules,[49] which can be exploited for MT between nuclei. For dipolar coupling, the interactions between spins occur through space, and the transfer of

magnetization can occur through a process called *cross-relaxation*.[1,48,50] It should be noted that dipolar coupling, due to its dependence on motion, is not detectable in MRI because of the rapid molecular tumbling rate in liquids, which averages out the effect. However, it does contribute to spin relaxation.[1] Chemical exchange and dipolar coupling underlie several processes (Figure 3.3) that contribute to the MT effects detectable by MRI.

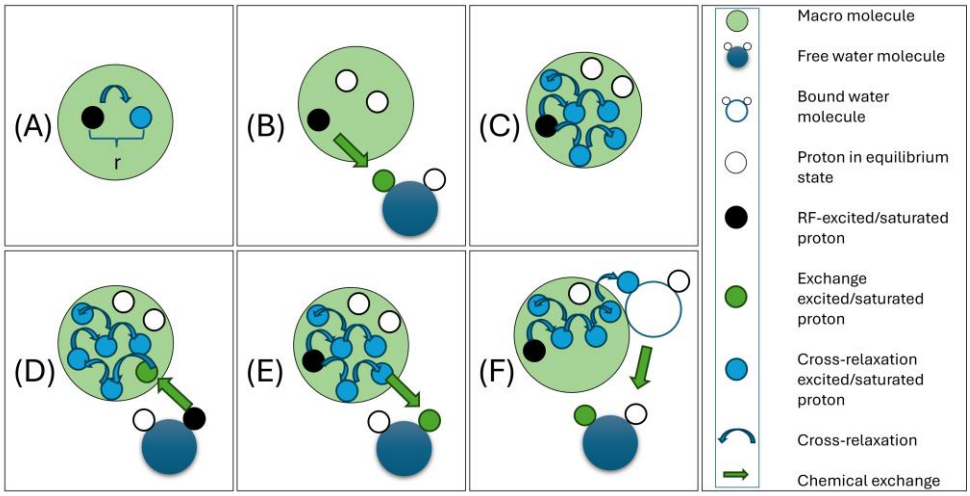


Figure 3.3. Magnetization transfer processes. (A) Magnetization transfer via cross-relaxation from an RF-saturated proton (black) to another molecular proton in proximity (distance r). (B) Chemical exchange between an RF-saturated proton in a macromolecule and a free water proton. (C) Intramolecular (relayed) NOE, also referred to as spin-diffusion or relayed dipolar transfer. (D) Exchange-relayed NOE, where a saturated proton is transferred via chemical exchange to a mobile molecule in which NOE proceeds. (E) NOE relayed exchange (rNOE), wherein saturation is transferred via NOE in a mobile molecule, followed by chemical exchange. (F) MTC, where spin-diffusion occurs in the backbone of the semisolid macromolecule and then continues through intermolecular NOE to bound water, with subsequent chemical exchange of whole molecules or protons occurring with free water. Alternatively, the fast transfer from the semisolid to free water via proton exchange contributes to the MTC effect. Figure adapted from reference [1] with permission from the publisher.

3.2.1 Different types of magnetization transfer

3.2.1.1 Chemical exchange

Chemical exchange for the purpose of CEST MRI refers to the mechanism by which the effect of saturation of a proton pool is transferred by the physical exchange of protons from one molecule to another (Figure 3.3B and more detailed in 3.4A). Conversely, the effect of exchange on relaxation is exploited in *chemical exchange sensitive spin lock* (CESL) MRI.

In CEST imaging, either endogenous or exogenous solutes are probed by selectively saturating their exchangeable protons via irradiation with RF pulses at their specific

resonance frequencies. This is followed by the solute protons exchanging with the bulk water protons (Figure 3.4A), which resonate at a different frequency due to their different chemical shift. Therefore, the solutes are detected indirectly through the reduction of the water signal. In practice, the low concentration (typically in the millimolar range) of the solutes would not be detectable due to the much higher proton concentration in bulk water (110 M). However, this method relies on a successive build-up of the water-signal loss due to the repeated transfer of saturated protons from the solute pool to free water and, inversely, the transfer of unsaturated protons from the water pool to the solute pool.[1,4] In this way, signal enhancements exceeding a factor of 1,000 can be obtained under the right conditions.[1] In particular, the exchange rate k_{sw} should be within the correct order of magnitude in relation to the saturation efficiency α. An analytical expression for the *proton transfer ratio* (PTR) for a two-pool model under conditions of slow exchange on the MR time scale can be expressed as[51]:

$$PTR = x_s \alpha_s k_{sw} T_{1w} \left(1 - e^{-\frac{t_{sat}}{T_{1w}}}\right), \tag{3.7}$$

where x_s is the ratio of the concentration of protons in the exchangeable solute to the concentration of water protons, given as $x_s = \frac{[exchangable\ protons]}{[water\ protons]} = \frac{k_{ws}}{k_{sw}}$. The saturation efficiency can be approximated as:

$$\alpha_s \approx \frac{(\gamma B_1)^2}{(\gamma B_1)^2 + k_{sw}^2}. \tag{3.8}$$

From the above, it is clear that a higher concentration of solute protons and a higher saturation efficiency are favorable for signal enhancement, as well as a longer T_1 for water (T_{1w}) since it allows the saturation to continue longer. It can also be inferred from Equation. 3.8 that for faster exchange rates, a stronger B_1-field strength is required to obtain sufficient saturation efficiency before exchange occurs. Another favorable condition for an enhanced signal is for the exchangeable protons to be in the slow to intermediate MR exchange regime ($\Delta\omega \geq k_{sw}$).

CEST effects are studied using the Z-spectrum (Figure 3.4B), which is defined as the normalized water signal intensity as a function of saturation frequency, i.e., $Z(\Delta\omega) = S_{sat}(\Delta\omega)/S_0$, in which S_{sat} is the water signal intensity at the irradiation offset $\Delta\omega$ from the water proton resonance frequency, and S_0 is the signal intensity without saturation. The CESL method is conceptually similar to CEST in terms of the mechanism of exchangeable protons. The main difference, however, lies in the pulse sequences employed and how the exchange is probed. As previously mentioned, imaging in CEST is preceded by a saturation pulse applied for a time

t_{sat}. During this time, saturation of the target proton pool occurs and is followed by the exchange with water protons. In contrast, the imaging in CESL is preceded by flipping the magnetization with an angle $\theta = \arctan(\omega_{1,SL}/\Delta\omega)$, where $\omega_{1,SL}$ is the so-called *Rabi frequency*, into the yz-plane and, subsequently, locking it with a *spin-lock pulse* ($B_{1,\text{eff}} = \sqrt{(\omega_{1,SL}^2 + \Delta\omega^2\ )}/\gamma$) for a time t_{SL} before flipping it back and proceeding with the imaging. During t_{SL}, a so-called "spin-lattice relaxation in the rotating frame" occurs, which is given as:

$$R_{1p} = R_1 \cdot \cos^2\theta + (R_2 + R_{ex}) \cdot \sin^2\theta\ , \tag{3.9}$$

where, R_{ex} describes the transverse relaxation contribution due to exchange, which is given as:

$$R_{ex}\big(\Delta\omega, \omega_{1,SL}\big) = \frac{p_s \delta^2 k_{ex}}{(\delta-\Delta\omega)^2+\omega_{1,SL}^2+k_{ex}^2}\ , \tag{3.10}$$

where p_s is the population fraction of the solute, δ is the chemical shift difference between the solute and water resonance frequencies and $k_{ex} = k_{ws} + k_{sw}$. The relaxation effects due to exchange can thus be studied either as a function of the Rabi frequency when $\Delta\omega$ is kept constant, for example, on-resonance with $\Delta\omega = 0$ or $\Delta\omega = \delta$, which yields a so-called *relaxation dispersion curve*, or off-resonance as a function of $\Delta\omega$ ($\omega_{1,SL}$ kept constant) in a so-called *SL Z-spectrum* (similar to the CEST Z-spectrum).[52] Simply put, the main difference is thus relaxation versus saturation for CESL and CEST, respectively. As a result, CEST is suitable for slow exchanging protons (ms-to-s range) on the MR timescale, while CESL is more suitable for protons exchanging in the intermediate-to-fast regime (ms-to-μs range).[52–54]

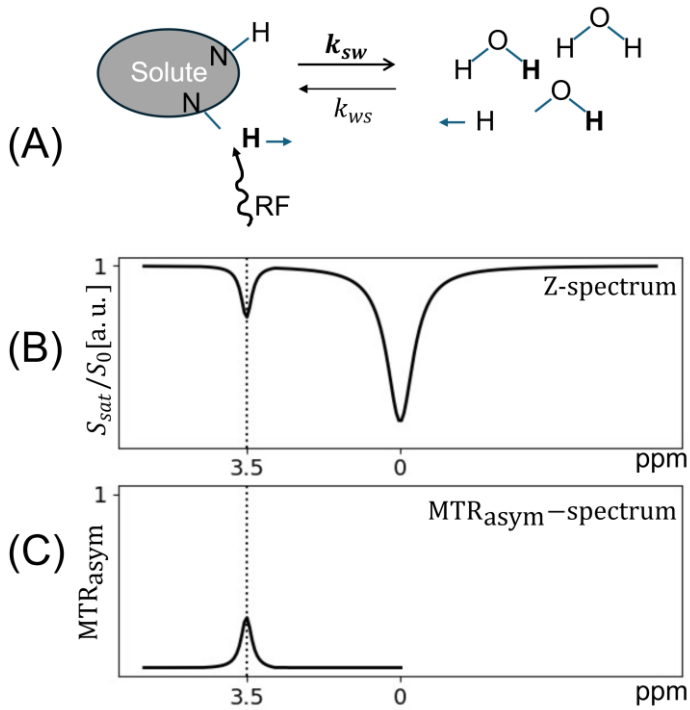


Figure 3.4. The principle of chemical exchange saturation transfer and examples of a Z-spectrum and an MTR_{asym} spectrum. (A) A solute proton (in this case an amide proton) is selectively saturated by irradiation using RF pulses at its MR resonance frequency offset (3.5 ppm from the water resonance frequency for amide protons). The saturated protons exchange with the protons of the water pool. (B) A Z-spectrum showing the normalized water signal as a function of saturation frequency. (C) The corresponding MTR_{asym} spectrum is generated through the subtraction of signals at a positive frequency offset from those at a negative frequency offset (Equation. 3.18). Since there are no signals at the negative offset, the MTR_{asym} spectrum visualizes the effects observed at 3.5 ppm. The figure was adapted from reference [4] with permission from the publisher.

3.2.1.2 Cross-relaxation

Another mechanism that causes MT, although without the physical exchange of protons, is cross-relaxation (Figure 3.3A). Dipolar coupling of neighboring spins may cause a saturated (or excited) spin population to transfer its magnetization to a neighboring population while the original population relaxes to equilibrium state. This so-called cross-relaxation process can be much faster than T_1 relaxation. The magnitude of dipolar coupling is inversely proportional to the cube of the distance between coupled nuclei and is also dependent on the squares of their gyromagnetic ratio as well as their orientation.[1,50] Consequently, the efficiency of cross-relaxation has the same dependencies as dipolar coupling, in addition to the requirement of a matching frequency range for the two spin pools, as described by the so-called $J(\omega)$- or *spectral density functions* (see references [50] and [56] for a theoretical and experimental understanding, respectively). With the reduced rotational motion of

molecules, the cross-relaxation rate increases, thus making it more prevalent in larger molecules and extremely fast in semisolid structures.[1]

3.2.1.3 Nuclear Overhauser enhancement

With the underlying mechanism of cross-relaxation, yet another MT process between nuclei is the *nuclear Overhauser enhancement* (NOE),[57] which leads to the spread of magnetization throughout a macromolecule (Figure 3.3C). This process is thus referred to as *relayed NOE*, *relayed dipolar transfer* or *spin-diffusion*, all of which are reflective of the effective spread of the magnetization within a molecule.[1] Relayed NOE, in combination with the chemical exchange mechanism, can result in two transfer processes. The first process is referred to as *exchange-relayed NOE* (Figure 3.3D), wherein an initially saturated solute proton transfers its magnetization through chemical exchange to another molecule, in which relayed NOE subsequently proceeds. The second process is referred to as *NOE-relayed exchange* (rNOE), wherein magnetization is transferred first within a molecule via relayed-NOE, with a subsequent chemical exchange between the saturated molecule protons and protons of free water (Figure 3.3E).[1,58] Thus, rNOE is a process that contributes to the Z-spectrum. Since the underlying mechanism for NOE is cross-relaxation, the efficiency of the process increases with the reduction of rotational molecular motion. Another important note is that these processes need to be faster than R_1 of the target molecule to be measurable.[1]

3.2.1.4 Magnetization transfer contrast

The relayed NOE process itself is the underlying mechanism for another type of MT, which is *magnetization transfer contrast* (MTC). MTC is not to be confused with MT, which is the umbrella expression for all magnetization transfer processes. As the efficiency of NOE increases with reduced molecular motion, it becomes most efficient in semisolid environments, allowing for rapid intermolecular NOE to occur (intermolecular NOE is insignificant in mobile macromolecules). The MTC processes (Figure 3.3F) describe the transfer of magnetization via either (i) intermolecular NOE from semisolids to bound water and, subsequently, molecular or proton transfer to free water via exchange,[1,59,60] or alternatively, (ii) the fast transfer from the semisolid to free water via proton exchange[61] (for example, in myelin, which has a cerebroside group containing a large number of *hydroxyl*, OH, groups). It should be noted that MTC is limited to saturation transfer since excitation removes the detectable signal in semisolid structures due to the fast dephasing of transverse magnetization (T_2 of the order of a few μs).

3.2.2 The Bloch-McConnell equations

It is important to recognize that in MT data from sources such as CEST experiments, multiple effects can coincide. These include chemical exchange, NOEs, MTC, and DS, where the saturation pulse is applied on or close to the water resonance

frequency. The overlap of these effects is particularly pronounced when irradiation occurs between 5 to -5 ppm relative to the water resonance wherein most of these effects contribute. Nevertheless, the exchange of magnetization between free water protons and the protons bound to solutes or macromolecules can be exploited to provide an additional MRI contrast with the possibility of detecting biomarkers and studying metabolic processes.[3,62–65] To account for the additional mechanisms that affect magnetization, a modification to the Bloch equations (Equations 3.2 – 3.4) including exchange parameters is required. This yields the so-called *Bloch-McConnell* (BMC) equations.[66] In addition to the dependence on B_0, B_1 and relaxation effects (T_1 and T_2), in the context of MT, the magnetization vector also depends on molecular motion, exchange rate, and molecular interactions.[1] Therefore, all these quantities are needed to properly describe the magnetization vector's time evolution. For the simplest MT case, which consists of only a two-site exchange model including a large water pool (*w*) and a small solute pool (*s*), and under the assumption that the RF field is applied in the x direction in the rotating frame, the BMC equations are given as[51,67]:

$$\frac{dM_{xs}}{dt} = -\Delta\omega_s M_{ys} - R_{2s}M_{xs} - k_{sw}M_{xs} + k_{ws}M_{xw}\,, \tag{3.11}$$

$$\frac{dM_{ys}}{dt} = \Delta\omega_s M_{xs} + \omega_1 M_{zs} - R_{2s}M_{ys} - k_{sw}M_{ys} + k_{ws}M_{yw}\,, \tag{3.12}$$

$$\frac{dM_{zs}}{dt} = -\omega_1 M_{ys} - R_{1s}(M_{zs} - M_{0s}) - k_{sw}M_{zs} + k_{ws}M_{zw}\,, \tag{3.13}$$

$$\frac{dM_{xw}}{dt} = -\Delta\omega_w M_{yw} - R_{2w}M_{xw} + k_{sw}M_{xs} - k_{ws}M_{xw}\,, \tag{3.14}$$

$$\frac{dM_{ys}}{dt} = \Delta\omega_w M_{xw} + \omega_1 M_{zw} - R_{2w}M_{yw} + k_{sw}M_{ys} - k_{ws}M_{yw}\,, \tag{3.15}$$

$$\frac{dM_{zs}}{dt} = -\omega_1 M_{yw} - R_{1w}(M_{zw} - M_{0w}) + k_{sw}M_{zs} - k_{ws}M_{zw}\,, \tag{3.16}$$

where $\omega_1 = \gamma B_1$ and $\Delta\omega = \omega$ defines the chemical shift. The subscripts x, y and z indicate the components in the x, y and z directions, respectively, while the subscripts s and w indicate the pool (solute or water), and *k* is the exchange rate. It should be noted, however, that pool *s* in the equations does not include the protons of semisolids for which the MT process is MTC. Furthermore, for MTC, it is common to exclude the transverse terms due to the fast dephasing of transverse magnetization.[1] It is, however, also worth noting that some studies have indicated potential transverse contribution.[68]

Because of the complexity of the equations, reasonable assumptions need to be made, including simplified approximations that yield closed-form solutions. A straightforward solution can be obtained using *the approximation of a weak saturation pulse and complete saturation,* wherein it is assumed that the saturation

pulse is on-resonance for the solute ($\Delta\omega_s = 0$), the protons of the solute are completely saturated, and no *spillover*[a] occurs. However, for complete saturation to be obtained, a very strong saturation pulse is required. Since this condition is not realistic, an assumption using *steady-state solutions under the weak saturation pulse approximation* is made (steady-state refers to the equilibrium of the resulting magnetization due to the transfer of saturation versus relaxation processes, in which the time-varying magnetization terms can be assumed to be zero). Another approximation is the *time-dependent solutions under the weak saturation pulse approximation*, wherein a two-step approximation is used, more reflective of the real scenario.[67] The saturation of the solute protons, which approaches steady-state, is instantly followed by transfer to the water pool, which in turn approaches steady state with a rate equal to $R_{1w} + k_{ws}$. From the solutions of the latter approximation, the expression for PTR given in Equation 3.7 is obtained for a two-pool model under the assumptions of no back exchange of saturated protons, no spillover effects and an exchange in the slow exchange regime.

3.3 The Z-spectrum

The MT effects are typically studied using the Z-spectrum, which, as previously mentioned, is the normalized water signal intensity as a function of irradiation frequency. Before discussing the fitting of Z-spectra, it is important to understand its features. When the RF irradiation occurs at the resonance frequency of water protons (DS), the largest signal reduction will be observed.[1,4] Furthermore, the offset ($\Delta\omega$ in radians per second) can also be expressed in absolute terms in Hz (commonly used in MTC experiments) or in relative terms in ppm (most commonly used in CEST experiments).[1] The DS peak is defined to be centered at 0 ppm (or 0 Hz), and it is well modeled by the Lorentzian function[60,69] at low B_1 (more details on this topic in Chapter 4). *The full width at half maximum* (FWHM) of the Lorentzian can be described as given in Equation. 3.17. Notice that the FWHM is commonly referred to simply as the *linewidth* (LW), a nomenclature that has been adopted in this thesis.

$$LW = \frac{\omega_1}{\pi}\sqrt{T_{1w}/T_{2w}}\,. \tag{3.17}$$

The absolute magnitude LW of the DS broadens with increasing field strength due to a reduction in T_2 and an increase in T_1, as can be inferred from Equation 3.17.

[a] Spillover refers to the unintended direct saturation of water protons caused by a saturation radiofrequency that is too close or with a too broad bandwidth relative to the water resonance. The effect becomes more pronounced at higher B_1-strength.

For the convention of using ppm, however, which implies normalization by the resonance frequency, the DS peak narrows with increasing field strength since the LW increase in Hz is small compared to the increase in resonance frequency with field strength. Another convention worth noting, which is taken from NMR spectroscopy, is that higher relative frequency (technically corresponding to a lower magnetic field and lower absolute frequency) is to the left due to the referencing of the protons' frequency in a particular chemical compound called *tetramethyl silane*. Using this convention, water protons in the NMR spectrum resonate at approximately 4.75 ppm, a number used in MR spectroscopy. However, since the Z-spectrum is dominated by the DS signal, and due to the frequent use of asymmetry analysis relative to that signal, the DS signal in Z-spectra is assigned to 0 ppm. Adding to the terminology, and understandable as explained above, frequencies higher than the water proton resonance frequency are sometimes referred to as *downfield* from water, whereas lower frequencies are *upfield* from water.

As a result of MT, the signal intensity of free water will be reduced at the resonance frequencies of the saturated protons in the Z-spectrum. When multiple target protons resonate in the same irradiated frequency range, MT effects occur simultaneously. The relative contribution from each component to the Z-spectrum depends on several factors, including the molecular concentration and type/number of interactions, the proton exchange rate, the B_0-field, the B_1-field (both B_1-strength and pulse shape) and the sequence settings (e.g., pulse and inter-pulse duration).[1] In consequence, an overlap and coalescence of the effects may occur, and improved separation may be possible with the correct settings of the experimental environment and acquisition parameters.

In particular, an increase in B_0 results in a larger *resonance dispersion* (since the resonance frequency is proportional to B_0), signifying a larger separation between the resonances of the spin pools for different molecules (Figure 3.5). This also results in the change of some spin pools to the slow exchange regime, which is more favorable for optimizing signal enhancement.[1,4] However, it should be mentioned that there are a few limiting factors when moving up in field strength. The *specific absorption rate* (SAR), which gives the rate at which energy is absorbed per unit mass in the object studied, is proportional to the square of field strength. This is especially problematic when using continuous wave saturation at higher B_1. However, SAR could be reduced in part by using smaller and more local RF-coils.[1] Another limitation is the amplification of any field inhomogeneities, which arises from the dependence of the Larmor frequency on field strength. This can, however, be alleviated by retrospective field correction methods.[70–73] Of course, a final limiting factor is the restricted availability of scanners with higher field strengths (≥7 T). The sensitivity of detecting fast-exchanging protons increases with B_1 (Equation 3.8). This, nevertheless, also increases the DS spillover effects and the MTC effects (Figure 3.5), both of which hamper the detectability of CEST and rNOE effects. It should be noted that some degree of coalescence is impossible to

overcome at lower B_0-field, especially for OH-groups close to the DS. One approach used to increase the detectability of these peaks is to alter their concentration by, for example, using contrast agents (such as glucose).[1]

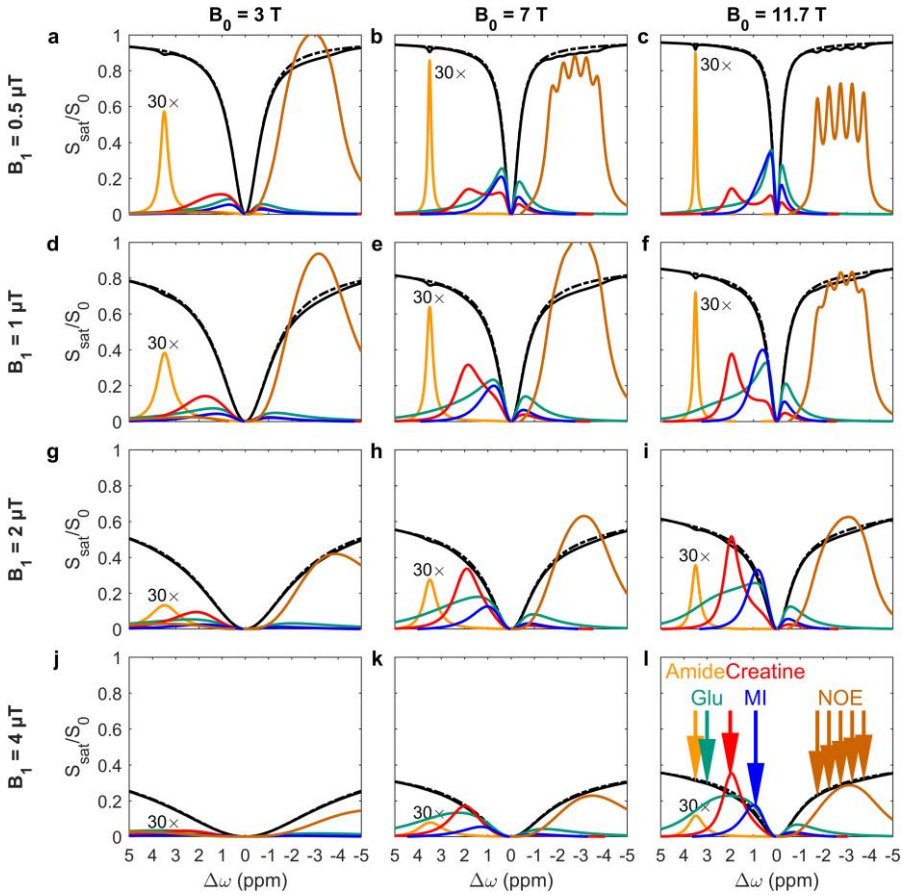


Figure 3.5. Simulation of the effect of B_0 and B_1 on Z-spectral components. A solid black line represents the Z-spectrum, while a dashed black line represents the resulting spectrum after removing the MTC and DS contributions. The colored lines represent the individual isolated components (indicated in Box l). Notice the increase in spillover and MTC effects with the increase in B_1-strength as well as the increased saturation transfer of fast exchanging protons (from, e.g., creatine), which gets reduced at very high B_1 due to the dominance of the stronger spillover effects of DS and MTC effect. Also, notice the increased dispersion of the components when moving towards higher B_0-field strength, e.g., reduced coalescence of the glutamate amine protons with water due to their shift towards the slower exchange regime. The increased dispersion is also reflected in the increased separation between the 5 rNOEs, however, the rNOE signal reduces with increased B_1 strength since they are already in the slow exchange regime. The figure is adapted from reference [1] with permission from the publisher.

3.3.1 Common components in the Z-spectrum

3.3.1.1 The DS pool

The DS is one of the most dominant contributions to the Z-spectrum (Figure 3.6) and the largest at low B_1. In some experiments, it is targeted and isolated by using a saturation pulse of low B_1-strength and a short duration (< 0.5 μT and < 1 s), resulting in what is often referred to as the *water saturation shift referencing* (WASSR) spectrum.[70] This term is based on the use of DS signals for voxel-by-voxel B_0-field shift correction.[70,71,73] At low B_1 it can also be used for estimating and removing small water-based spillover effects.[58,74] Other applications include the determination of local field shifts for whole-brain magnetic susceptibility mapping[75] and the study of relaxation effects in tissue[73] since the LW of the DS depends on both the longitudinal and transverse relaxation rates (R_1 and R_2).[73,76] As described in Paper III, it was recently shown that these spectra can be further utilized in DS *dynamic glucose enhanced* (DS-DGE) imaging. With DGE, glucose[b] uptake has usually been studied in a dynamic time series for a fixed saturation frequency that matches the average resonance frequency of the exchangeable OH groups of the administered glucose (typically occurring in a range from 1.2-2 ppm downfield).[77–80] Thus, the increase in OH-concentration after administration is reflected in the water signal reduction at the selected resonance frequency. Subsequently, a dynamic curve is calculated by taking the normalized difference between the averaged baseline signal (i.e., for a few images prior to glucose infusion) and the signal at different time points during and after infusion. The increased signal in the dynamic curve thus reflects an increased glucose uptake.[77,78] Contrary to conventional DGE, D-glucose uptake in the DS-DGE approach is studied via the LW-changes of the DS peak pre- and post-infusion (ΔLW). The study described in Paper III showed in both simulated and in vivo data that the DS-DGE method is promising in distinguishing different healthy tissues as well as different tumor types. In vivo results showed an increase in ΔLW of 1% for *gray matter* (GM) and *white matter* (WM), while the number was between 5-20% for various tumor types and 40% for *cerebrospinal fluid* (CSF).[81]

3.3.1.2 The MTC pool

In addition to the DS, several pools may be visible in the Z-spectrum. A relatively large contribution can usually be seen due to the MTC effects (Figure 3.6). As previously explained, these effects stretch far off-resonance beyond the proton spectral range of mobile macromolecules[1]; moreover, its total magnitude at higher B_1 can become larger than that of DS. In reality, semisolids (e.g., membranes having abundant protons or lipid chains with many side groups) will contribute with tens to

[b] Mainly D-glucose due to its wide availability, biodegradability, and FDA-approval for oral and intravenous administration.

hundreds of peaks on the Z-spectrum with center frequency values around 2-3 ppm. However, the peaks have very broad lines, given by $LW \sim 1/(\pi\sqrt{T_2})$, since semisolid protons have a very short T_2 (10-40 μs) compared to those of mobile macromolecules (ms range), which instead have finite LWs. An additional broadening of the MTC peaks is caused by so-called *chemical shift anisotropy* (CSA). Due to anisotropic (non-uniform) electronic environments around the nuclei in semisolids, the shielding effects will vary in different directions relative to the external magnetic field, thus resulting in different shifts on the Z-spectrum.[82] A third effect broadening the MTC signal is the occurrence of large dipolar couplings (kHz range) that split the resonance signals. Due to the superposition of several chemical shifts and dipolar splits, the resulting broadening will occur. This effect is not apparent for mobile macromolecules, for which the fast molecular tumbling averages out the anisotropic shielding and the dipolar splits. The combined effects of a short T_2, CSA, and dipolar splits result in a broad unresolved signal for semisolids. While not directly affecting the proton spectrum of mobile molecules (including the DS), the partial saturation of semisolid molecules leads to the eventual saturation of the water,[1,59,60] thus producing a broad background signal affecting the DS and other Z-spectral components.[83] It should be mentioned that MTC is a highly active field of research with a lot of unresolved controversy. A symmetrical approximation of the MTC contribution in the Z-spectrum has been and is still being used, especially in the case of low B_1-strength (where the MTC effect is smaller), and it is most likely a good approximation due to the dominance of the DS. However, several extensive studies[84–86] have experimentally shown an asymmetry of a few percent of the MTC component using a saturation power between 1 and 3 μT. This is expressed at offsets beyond the spectral range of the protons from mobile molecules, which signifies $|\Delta\omega| > 5$ ppm (Figure 3.8). More recent work[87] has indicated the occurrence of both fast and slow exchanging protons from semisolids. It is theorized that fast exchanging protons could arise from the direct irradiation of OH-groups of semisolids or bound water, which exchanges close to the free water resonance.[1,61] This results in a contribution with an approximately symmetric component, while the slower exchanging components are from rNOEs and expressed asymmetrically below the fine structure of the rNOE components.[1] It should be mentioned that the latter asymmetry is observed even for lower B_1 powers (1 μT). The different composition of macromolecules in tissue leverages the potential to differentiate tissue based on the study of the MTC component since it will be larger in tissues with a high concentration of semisolids.[60,88] This can, for example, be applied in the brain to differentiate between WM (with a high myelin content) and GM as well as CSF (with molecules mainly in a liquid environment).[89–92] It can also be utilized to detect demyelinated WM regions which, for example, occur with *multiple sclerosis* (MS)[65,66] and other diseases.[93–96]

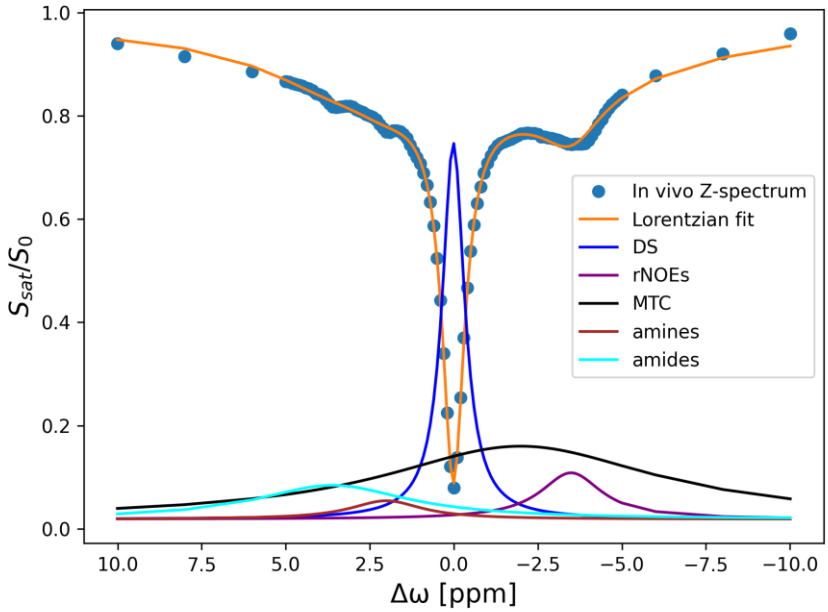


Figure 3.6. Z-spectrum with isolated components. A Z-spectrum (blue dots) from a rat brain acquired at a field strength of 11.7 T using a saturation pulse with a B_1 of 0.7 μT and a pulse duration of 4000 ms. The Lorentzian fits of 5 isolated components, DS (blue), amides (cyan), amines (brown), rNOE (purple), and MTC (black), add up to the orange fit of the Z-spectrum.

3.3.1.3 CEST components

Several CEST effects might be visible in the Z-spectrum downfield. The contributions will vary based on the studied environment and depending on whether it contains endogenous or exogenous compounds. Note that the examples that follow serve to reflect the intricacy and highly complex appearance of the Z-spectrum which result from the effects of chemical environments on the exchangeable protons.

3.3.1.3a: Amide protons

In the brain environment, some of the most common CEST signals are from primary *amides*, which are integral parts of the peptide bond in mobile proteins and peptides.[1,4] They have the general formula R-C(O)N(H)-R' wherein R and R' represent different *organyl* groups or hydrogen atoms in different amino acids, and CO is a *carbonyl group*. Together, the two groups form the so-called *acyl-group*. Furthermore, the NH is sometimes referred to as the *reduced amino group* (amino group if it holds two hydrogens). Hence, it can be inferred that a peptide-bond contains two functional groups, the acyl- and the (reduced) amino groups. This is because amides are formed through a chemical reaction between *carboxylic acid*

(RCOOH) and *ammonia* (NH_3) or amine groups of individual amino acids[c].[97] The exchangeable amide protons of mobile proteins and peptides occur over a range of several ppm in the spectrum, but multiple amides of the more floppy parts[d] of these macromolecules have an average resonance frequency centered around 3.5 ppm (Figure 3.6).[1,4,98] The total concentration from the many amides in mobile proteins and peptides in most tissues is between 5 – 8 mM,[99,100] while the exchange rate is ~30 Hz,[99] thus making it a very attractive candidate for CEST studies (Figure 3.7). It should be noted, however, that amides of semisolid structures have a much broader range of resonance frequencies and shorter T_1 and T_2 time constants and instead contribute to MTC effects.[16] Numerous successful applications for detecting pathologies have been implemented in clinical or preclinical settings in which the change in tissue environment affects the concentration of amides. The exchange of amide protons to water, for example, is base-catalyzed,[4,16,98] and a reduction in pH reduces the exchange rate and leads to reduced signal intensity for such protons in the slow exchange regime.[99] This has been exploited in several studies to detect ischemic stroke,[101–106,8,107] wherein the reduced blood flow causes an increase in the *adenosine triphosphate* hydrolysis and bicarbonate buffering, thus resulting in a decrease in pH. This was further utilized in a preclinical study to differentiate between hemorrhage and ischemic stroke by hyper- versus hypo-intensity.[108] The change of amide concentration in proteins and peptides has further been used to detect pathologies. For example, it has been reported that neurological disorders such as MS can be detected by the increased protein content in WM.[109,110] In addition, preclinical studies are indicative of the detection of Alzheimer's disease due to the reduction of mobile protein signals.[111] The accumulation of proteins in lymph nodes is another detectable change that has been utilized for the detection of breast cancer treatment-related lymphedema.[112] Moreover, a large field of application for amide signals regards tumor imaging. The increased concentrations of mobile proteins and peptides in tumors[100] can generate increased amide signals. This has been utilized in several studies to differentiate between low- and high-grade gliomas.[113–115] The study of the amide protons has resulted in what is usually referred to as *amide proton transfer weighted* (APT-w) imaging, wherein the weighting indicates that with this type of analysis there are mixing of multiple signal contributions from, for example, MTC and rNOE.

3.3.1.3b Amine and guanidinium protons

Other naturally occurring mobile macromolecules that may appear in the frequency range 2 – 3 ppm on the Z-spectrum are the amine and guanidinium proton groups. While both primary (RNH_2) and secondary (R_2NH) amines carry exchangeable

[c] It should be noted that both primary (RNH_2) and secondary (RNH) amines can participate in chemical reactions that produce amides.

[d] In chemistry floppy parts refer to molecules with large-amplitude, low-force constant vibrational modes that allow for significant structural changes, such as inversions or internal rotations.

protons, the latter usually has a slower exchange rate and requires a higher concentration to achieve a similar CEST effect (Figure 3.7c).[4] The amine protons typically have a fast exchange rate with water, for example, 5500 Hz[1] for the amine protons in *glutamate* under physiological conditions (pH = 7.3 and T = 310 K). The guanidinium protons have a moderate exchange rate of 1100 Hz (a recent report shows it even in the 200–300 Hz range).[116] Practically speaking, this indicates that neither is in the favorable slow exchange regime, leading to a coalescence between their signals and the water signal (Figure 3.6). This is especially true at lower field strengths, where the resonance dispersion is small. This means that the guanidinium protons only approach the slow exchange regime when the acquisition is performed at a very high field strength (Figure 3.5).[1] Amines can be found in amino acids such as glutamate (as previously mentioned) and *glutamine*. These are both prevalent in the brain[e] and have exchangeable amine protons that resonate at approximately 3 ppm.[117] Guanidinium, on the other hand, is an integral part of *creatine* (Cr), which is a naturally occurring compound in the brain and muscle.[1,118,119]

Some of the clinical and preclinical applications utilizing the amine protons include the early diagnosis of encephalitis by glutamate content. This is because areas of the brain with encephalitis exhibit an increased glutamate amine proton signal intensity that indicates an increased *microglial* activation which, in turn, reflects the pathology.[120] Similar to amides, the exchange of amines to water is base-catalyzed. The fast exchange rate of amines can thus be slowed down at a reduced pH leading to an increased CEST signal. This is opposite to the amides, which show a reduction in the CEST signal when slowed down, since they are already in the slow exchange regime at physiological conditions. This effect has been utilized in tumor imaging where the higher concentration of proteins enhance the visibility of acidic areas containing glutamine.[121–123] Moreover, guanidinium protons can be used to reflect creatine concentration changes, for example, in the calf muscle pre- and post-moderate exercise.[124–126] It should be mentioned that creatine is an amino acid derivative, so it also holds a primary amine group in addition to the guanidinium group.[127] However, because of *delocalized* bonds between the electrons of the two groups, the exchangeable protons have proximal resonances resulting in one indistinguishable peak at approximately 2 ppm. In contrast, *phosphocreatine* (PCr), which has exchangeable guanidinium protons with a secondary amine bound to the additional phosphate group, produces two separate peaks at approximately 2.5 and 2 ppm for the two groups.[118,119,128] Notice that the PCr's peak for the guanidinium protons is shifted downfield compared to that of Cr partly because the phosphate group is electron-withdrawing, which thus deshields the guanidinium group and results in the shift.[129]

[e] Glutamine in blood, extravascular extracellular space, and cells, and glutamate mainly in the cell.

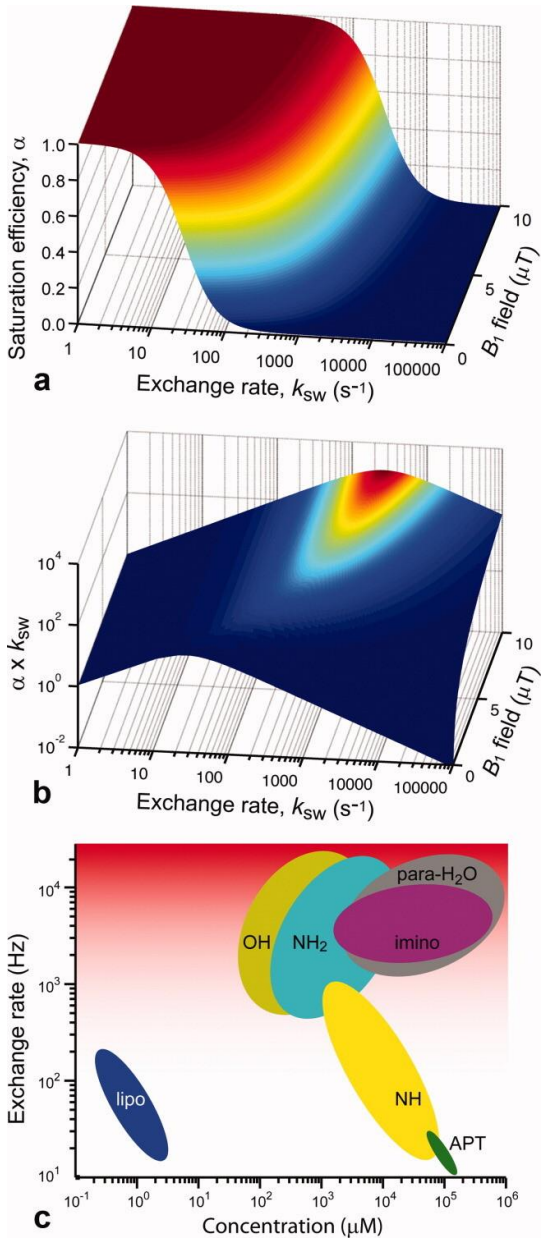


Figure 3.7. Saturation efficiency, exchange rate, B_1 power and concentrations. a) The effect on saturation efficiency with the change of exchange rate and B_1-field strength. Maximal saturation efficiency is rapidly achieved for low exchange rates as B_1 is increased. When the exchange rate surpasses the intermediate regime, even higher saturation powers are insufficient for the required saturation. b) An exchange rate that is too slow halts the saturation transfer, while one that is too fast results in insufficient time to saturate protons. Note, however, that the y-axis shows the product of the exchange rate and saturation efficiency, which is flatter over the exchange rate range. c) The concentration needed for different (common) CEST agents to obtain the same 5% CEST effect. The figure was adapted from reference [4] with permission from the publisher.

3.3.1.3c OH-groups

The hydroxyl group, common in but not exclusive to sugars, holds exchangeable protons which, depending on the chemical environment of the macromolecule, typically resonate between 0 and 3.5 ppm.[1] One naturally occurring sugar in the brain is *myoinositol,* which has OH protons with a resonance frequency of approximately 0.9 ppm.[130] However, this Z-spectral component is coalesced with water due to its protons' fast exchange rate of 2000 Hz. The CEST peak will then be inseparable from the water peak even at higher field strengths (Figure 3.6) since the exchange rate remains in the fast exchange regime. However, the DS peak yields an asymmetry that might be detectable.[1,130] Lactate, also with an exchangeable OH group, is another naturally occurring compound. It occurs, for example, in muscle as a product of *glycolysis* (the breakdown of glucose to *pyruvate* and adenosine triphosphate) under *anaerobic* (oxygen-poor) environmental conditions.[131] Similar to creatine, it increases in concentration in muscle after exercise, as shown by lactate-weighted CEST imaging.[132] In contrast to naturally induced changes of compound concentrations, such as in Cr and lactate during muscle activation, artificially altered concentrations can be achieved for low-concentration compounds (μM-mM). Examples of these are the biodegradable compounds that contribute with protons from OH groups with resonance frequencies close to the resonance frequency of the water pool. These can be administrated as CEST contrast agents. This signifies that their uptake or metabolism can be reflected by studying the change of concentration in a dynamic fashion, i.e. pre- and post-administration, as is done with the previously explained DGE technique[64,77,78,133–135] wherein D-glucose is used as a contrast agent. D-glucose holds five exchangeable OH protons with resonance frequencies at 0.66, 1.28, 2.08 and 2.88 ppm[f] having exchange rates in the intermediate to high exchange regime at clinical field strengths.[136,137] Thus, the pools are non-differentiable and instead contribute to a nonsymmetric broadening of the DS. In conventional DGE, a resonance frequency within the mentioned range is typically selected and studied.[77,78] DGE imaging has found various fields of application, the largest being cancer imaging where various studies have shown signal enhancement in tumorous areas after glucose infusion.[77,78,133–135]

3.3.1.4 rNOEs

Upfield in the Z-spectrum at approximately -1.6 ppm, a peak attributed to chlorin phospholipids can be found[138,139] as well as several rNOEs including primary, secondary, and tertiary aliphatic groups between -2 and -3.5 ppm. However, they are generally only detectable as one broad peak at clinical field strength due to the proximity of resonances (Figure 3.5). Moreover, they can only be visible at low B_1

[f] One OH-group at 0.66 ppm, three at 1.28 ppm, and one occurring at both 2.08 and 2.88 with a prevalence ratio of 9:16. This is because cyclic glucose comes as two *anomers*, α and β. Anomers differ in the orientation of the OH-group bound to the anomeric carbon atom, which is the carbon part of the carbonyl group in linear form.

due to increased interference from MTC effects with an increased B_1 strength (see Figure 3.6). Some studies have exploited the changes in rNOE contributions for detecting various brain tumors,[58] including astrocytoma[74] and glioblastoma[140], as well as for distinguishing between low- and high-grade glioma.[141]

3.2.3 Analysis of Z-spectral components

As indicated in the previous section, the Z-spectrum holds a plethora of biochemical information that has successfully been utilized in various clinical or preclinical applications. To obtain this information, a proper analysis of the Z-spectrum is required since the considerable reduction in the Z-spectral signal intensity due to DS might overshadow small CEST effects. Furthermore, another cause of interference in studies of in vivo CEST effects is the MTC effect from semisolid tissue components.[4,70] For improved detection of CEST effects, the symmetry around the DS is often utilized by removing the DS as well as symmetric MTC effects in the so-called *magnetization transfer ratio asymmetry*[1,70,142,143] (MTR_{asym}) spectrum (Figure 3.4C), which is defined as:

$$MTR_{asym} = \frac{S_{sat}(-\Delta\omega) - S_{sat}(\Delta\omega)}{S_0}, \tag{3.18}$$

where S_{sat} is the water signal intensity at the irradiation offset $\Delta\omega$ from the water proton resonance frequency and S_0 is the signal intensity without saturation.

While the MTR_{asym} is a quick and straightforward approach for Z-spectral analysis and proven to be useful in some applications,[103,105,106,144–146] it becomes problematic with the increased visibility of rNOEs from mobile macromolecules. This is because the downfield CEST- and upfield rNOE effects are the main contributions to the asymmetry of the Z-spectrum in the proton exchange range (5 to -5 ppm). Hence, with the MTR_{asym} approach, the effects on either side of the DS are superpositioned[g]. In addition, the assumption of a symmetric MTC contribution poses another problem, since an asymmetrical contribution can even occur at lower B_1-strengths as previously mentioned. Consequently, the MTR_{asym} approach might be rendered unsuccessful. For this reason, several alternatives have been proposed. A simple approach is the so-called *Lorentzian difference* (LD), where the DS is fitted with a Lorentzian and subsequently subtracted from the Z-spectrum to isolate both CEST and rNOE effects resulting from slow exchanging components.[74,83,146] While remedying the mixing of CEST and rNOE components caused by the MTR_{asym}, the approach is still highly sensitive to MTC contributions which might occur even

[g] This is also the reason why compounds targeted with asymmetry-analysis in the in vivo environment should be referred to as weighted, e.g., APT-weighted CEST.

at lower B_1 irradiation power. Therefore, a more accurate approach is to approximate each component by an individual Lorentzian, thus providing a labeled model for the spectrum that can be defined as:

$$Z_{lab}(\Delta\omega) = Z_{base} - \sum_{i=1}^{n} L_i(\Delta\omega)\,, \tag{3.19}$$

where Z_{base} is the background signal (close to 1), L_i is the Lorentzian for component i at an offset $\Delta\omega$, and n is the number of components in the model. Several further intricate approaches have been proposed, such as the MTR *exchange dependent relaxation contribution* (MTR_{Rex}),[147] which uses inverse Z-spectral analysis and compensates for the DS and MTC components through inverse addition. Another approach is the *water relaxation compensated exchange dependent relaxation contribution*,[148,149] which adds a correction for the T_{1w}-effect to the MTR_{Rex} through a simple normalization.[1,126,140] These approaches are widely used for Z-spectral analysis. However, it should be noted that these assume no interaction between the different pools and their success is thus restricted to the low B_1-setting. The assumption becomes flawed with the increase of B_1 and the coalescence of pools in the fast exchange regime, such as myoinositol and glutamate, with the water pool.

Nevertheless, all these approaches rely on one key component, which is fitting. Furthermore, prior to fitting there is a need to have a spectral model that approximates the Z-spectrum as accurately as possible. The Lorentzian spectral model was briefly mentioned, but there are a few other models that should be considered for potentially more accurate modeling, as will be discussed in the next chapter.

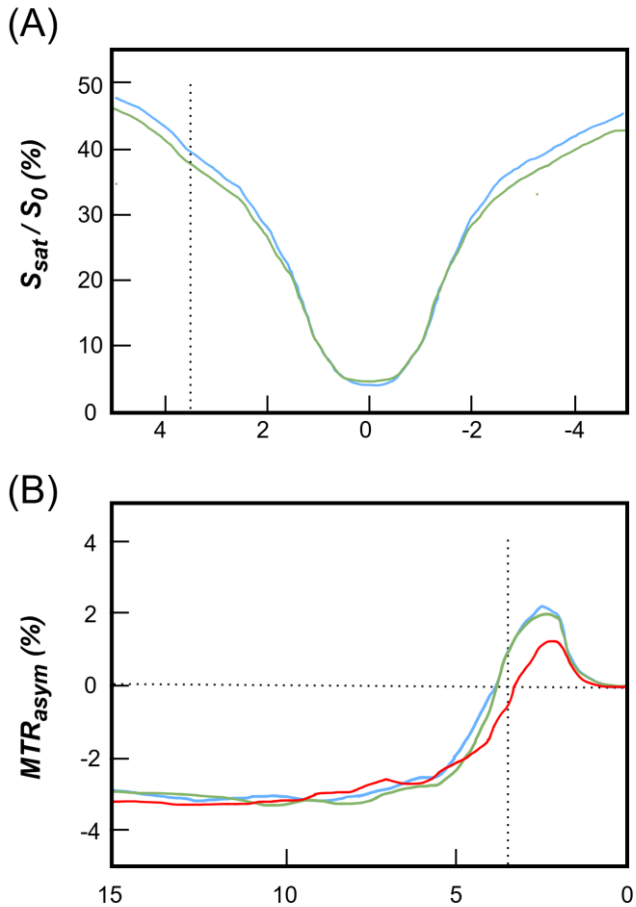


Figure 3.8. Nonsymmetrical MTC contributions in vivo. Signals from normocapnic (green), hypercapnic (blue) and postmortem (cardiac arrest, red) rat brains at 4.7 T. The signals are averaged over five regions of interest across the brains. (A) A Z-spectrum centered around the proton exchange regime. (B) CEST and rNOE effects are apparent, especially on the MTR_{asym} in the proton exchange regime (0-4 ppm). It is worth noting that nonzero contribution is apparent beyond the 4 ppm limit, which should not be the case if the MTC had been symmetrical around the water resonance frequency. Figure readapted from reference [1] with permission from the publisher.

4. Modeling of the Z-spectrum

4.1 Bloch McConnell modeling

A Z-spectrum can be described by the numerical solutions of the BMC equations under the assumption of a steady-state. To model a Z-spectrum using the BMC equations, several key parameters that describe the system's magnetization dynamics and exchange processes need to be defined. For a two-site model (e.g., a large water pool and a small solute pool as described in Chapter 3.2), the exchange rates (k_{sw} and k_{ws}), longitudinal (R_{1w} and R_{1s}) and transversal (R_{2w} and R_{2s}) relaxation rates as well as the chemical shift of the solute pool are needed.[51] Furthermore, the initial magnetizations of the two pools must be determined along with the population fractions. Saturation-specific parameters — including the pulse strength, duration, and the frequency offsets — must also be specified. With these parameters, the BMC equations for a two-site model (Equations 3.11 – 3.16) must be solved using a numerical algorithm for each offset frequency before obtaining a model for the full Z-spectrum.[149,150] This explains that the model's complexity and computational load scales with the applied RF protocol as well as the number of pools considered in the model. Therefore, the model needs to assume a limited number of pools. Another important consideration is that, while the BMC equations take into consideration the exchange between the individual pools and the water pool, the exchange between the pools and the MTC pool are not accounted for. Furthermore, rNOE can only be approximated as one or a few individual pools (with no interaction between pools). In addition, the BMC equations have a restricted number of observables for the separation of model parameters such as transversal relaxation, exchange rate, and the proton fraction of the studied components.[1]

Analytical solutions for the BMC equations have been proposed. A solution combining CEST and MTC highlighted the need to include the semisolid MTC pool size fraction for more accurate solutions.[149] In a more recently proposed analytical solution,[151] the computational complexity was reduced. It should be mentioned that both analytical and numerical solutions have been made available with open access.[149,152] Nevertheless, despite attempts to reduce the computational complexity of the numerical solutions, the analytical solutions are still currently limited to the two-site model (one solute and bulk water) while, in reality, the modeled spectra mostly contain several components. In combination with the inherent limitations, such as the restricted consideration for pool interactions and number of observables,

the high computational complexity might not make BMC modeling worthwhile. Therefore, several multi-pool spectral profiles can instead be used to estimate the observed Z-spectra. In the following sections, one traditional profile, the Lorentzian spectral model,[153,154] and a more recently used, the Voigt profile,[154,155] are defined and compared based on empirical findings and MT theory.

4.2 Lorentzian modeling

The most used spectral profile to model Z-spectral components is the Lorentzian.[153,154] It has been shown to accurately fit the DS within the narrow frequency interval between 5 and -5 ppm,[156] as well as various CEST peaks acquired using low irradiation field strength and short saturation duration (B_1 < 0.5 μT and t_{sat} < 1 s).[24,157–159] The common approach is to estimate each pool with a separate Lorentzian as described in Equation 3.19, wherein each Lorentzian spectral profile is defined as:

$$f_{Lorentzian}(\Delta\omega, A, \delta, \Gamma) = \frac{-A}{\pi}\left[\frac{\Gamma}{(\Delta\omega-\delta)^2+\Gamma^2}\right], \tag{4.1}$$

where $\Delta\omega$ is the saturation offset frequency from the DS saturation frequency, A is the amplitude of the peak or area when normalization is performed, δ is the center of the peak relative to the DS center value, and Γ is the *half-width at half maximum* (HWHM) of the peak. For comparability between different spectral models, as well as for more practical applications, derived parameters such as the height parameter (weighted amplitude) and the LW are often adopted. For the Lorentzian model, the height and LW are calculated as:

$$H_L = \frac{A}{\pi\cdot\Gamma}, \tag{4.2}$$

$$LW_L = 2\cdot\Gamma\,. \tag{4.3}$$

4.3 Voigt modeling

Recently, Voigt profiles[154,155] have been utilized for fitting Z-spectral components.[160] Similar to the multi-pool Lorentzian approach (Equation. 3.19) each pool of the Z-spectrum is approximated with a separate Voigt profile. Since the Voigt profile is the result of a convolution between a Gaussian and Lorentzian profile, the degrees of freedom are increased compared to the Lorentzian due to the

additional parameter from the Gaussian component when estimating the LW. The Voigt profile is defined as:

$$f_{Voigt}(\Delta\omega, A, \delta, \sigma_G, \Gamma) = \frac{-A \cdot \mathrm{real}[w(z)]}{\sigma_G \sqrt{2\pi}}, \tag{4.4}$$

with

$$w(z) = e^{-z^2}\left(1 + \frac{2i}{\sqrt{\pi}} \int_0^z e^{t^2} dt\right), \tag{4.5}$$

and with $z = (\Delta\omega - \delta + i \cdot \Gamma)/(\sigma_G\sqrt{2})$. σ_G is the standard deviation of the Gaussian which determines the Gaussian LW. Furthermore, the LW and height for the Voigt profile are given by:

$$LW_V = 1.1 \cdot \Gamma + \sqrt{0.87 \cdot \Gamma^2 + 5.5 \cdot \sigma_G^2}, \tag{4.6}$$

$$H_V = A \cdot \frac{\mathrm{real}\left[\mathrm{w}\left(i \frac{\Gamma}{\sigma_G\sqrt{2}}\right)\right]}{\sigma_G\sqrt{2\pi}}. \tag{4.7}$$

4.4 Comparison of spectral models

As briefly mentioned earlier, the Z-spectral components are well modeled by a Lorentzian when the saturation power and duration are kept low and short. Due to the minimal MTC contribution, this has been utilized to isolate the DS effect, e.g., in WASSR spectra.[70] Under the same acquisition conditions, the Lorentzian has also shown success in approximating other Z-spectral components with multi-pool Lorentzian models, including, for example, individual peaks for amides, amines, and a single peak for the aliphatic groups,[157,158] as well as a peak approximating the MTC contribution.[157] It has been suggested that the MTC can be approximated by a Lorentzian inside of the proton spectral range under the previously mentioned acquisition conditions.[156] However, a generally more accurate approximation for MTC is the super-Lorentzian,[2,161–163] which is a derivation of the Lorentzian having a wider profile. Moreover, even at lower B_1 irradiation powers, the Z-spectra constitute convolved contributions of various CEST, rNOE and potential MTC contributions (see Chapter 3.2.1), making it important to keep several considerations in mind. Firstly, rNOE effects from mobile macromolecules typically overlap, ultimately affecting the spectral shape. Secondly, for tissues that include strong MTC effects, the shape of the spectral components shifts toward a Gaussian profile

due to MTC-induced widening of both DS and the remaining peaks in the spectrum. Finally, fast-exchanging protons exhibit exchange-regime dependent lineshapes. All these effects, possibly even visible at lower B_1, become more pronounced with the increased irradiation power. Thus, depending on the saturation scheme and the spectral components studied, the shape of Z-spectral components can significantly deviate from the Lorentzian shape.

It has been experimentally shown that the Voigt lineshape significantly reduces the systemic fitting error for NMR ^{1}H-spectra compared to using the Gaussian or Lorentzian spectral lineshapes separately.[164] The Voigt profile, constituting both profiles tends to model the above-mentioned possible effects better. This holds true even at the low-to-intermediate saturation powers around 0.8 μT as empirically verified in the study reported in Paper II (Table 4.1). The improved fitting accuracy of the Voigt lineshape, even at lower saturation powers, could potentially be used to support the theorized symmetrical MTC contribution[1] at lower B_1 since, as previously stated, the lineshape transitions to become more Gaussian with the increased MTC contribution, a feature that the Voigt profile captures better. The overall improved goodness-of-fit of the Voigt profile compared to the Lorentzian is, of course, accredited to the increased degrees of freedom.

Another noteworthy aspect is the asymmetry of components, which is partly due to asymmetric MTC contributions but also the coalescence with the water peak for fast exchanging protons in the vicinity of the DS. In the previous chapter, two such mentioned examples were the myoinositol peak (centered around 0.9 ppm) and the rNOE peak from the aliphatic groups (centered around -3 ppm). To the best of my knowledge, no publications have yet reported the application of asymmetrical spectral models for fitting individual Z-spectral components. In this regard, several spectral models are good candidates, such as the *split Lorentzian*, *skewed Voigt,* or *Fano resonance,* all of which offer a direct or a derived parameter reflecting asymmetry. The asymmetry parameter could provide a new quantity for analyzing the Z-spectral components and, subsequently, new insights into the composition and structure of the Z-spectra, thus enhancing interpretability and potential clinical relevance. A more thorough discussion on this topic, together with a successful asymmetrical fit of rNOEs in an in vivo example, is presented in Chapter 7.

Table 4.1 Comparison of goodness-of-fit between the Voigt and Lorentzian spectral models. The mean absolute error (MAE) comparing the fits of the in vivo spectra using the Lorentzian and Voigt models on each peak together with the p-values from a Wilcoxon test to include the significance of the difference (level of significance $\alpha = 0.05$). The table was adapted from Paper II.

Peak	MAE Lorentzian	MAE Voigt	p-value
0 ppm	0.0184	0.0166	<0.005
3.6 ppm	0.00751	0.00716	<0.005
2 ppm	0.00871	0.00829	<0.005
-2.8 ppm	0.0101	0.00980	<0.005
Total	0.0979	0.00927	<0.005

The increased degrees of freedom for both the Voigt profile and the asymmetrical spectral models improve the goodness-of-fit compared to the Lorentzian, but they also entail some challenges and limitations. In particular, an increased time complexity adds to the already time-consuming conventional fitting approaches on the clinical timescale. Although limited to the interactions between CEST pools and the water pool, the BMC-modeling approach has the benefit of including pool interactions in the equations. However, its high model and time complexity make the choice of BMC fitting undesirable. This very same reason, in contrast, makes the multi-pool Lorentzian a preferred choice, especially during the acquisition conditions previously mentioned. Nevertheless, on the clinical timescale, multi-pool Lorentzian fitting with the fastest conventional algorithm (more details on this topic in Chapter 5) has been criticized for being far from feasible, thus prompting the search for innovative approaches to close the gap.[165] Since higher degrees of freedom in a spectral model yield a scalable increase in time consumption with each additional peak, more complex spectral models are rarely considered. This is despite their potential to significantly increase the goodness-of-fit as demonstrated in Paper II when using the Voigt profile instead of the Lorentzian. The scalable time increase was also demonstrated in Paper II, where the fitting time with the most efficient conventional algorithm for four peaks was approximately four versus eleven hours per brain for the Lorentzian and Voigt models, respectively. Hence, there is undoubtedly a need for groundbreaking approaches to accelerate the backbone of the fitting processes for any spectral model. This is conveniently the topic of the next chapter.

5.Z-spectral fitting

5.1 Conventional approaches

5.1.1 Fitting methods

Spectral fitting refers to optimizing a predefined model function, $f(x; p_1, \dots, p_n)$, where x denotes the discrete independent variable and p_i are the model parameters that best fit the observed discrete spectral data. This is achieved by determining the parameter set that optimizes a specified objective function. Several methods have traditionally been used as the foundation for fitting spectral models to discrete experimental data, with each having its benefits and drawbacks. Two methods worth mentioning are the *maximum likelihood estimation*[166] *(MLE)* and *Bayesian regression*.[167,168] The former relies on finding certain parameters, θ, that maximize the likelihood function, $P(D|\theta)$, all while observing the given data, D. The likelihood is defined as:

$$P(D|\theta) = \prod_{i=1}^{N} P(y_i|x_i, \theta)\,, \tag{5.1}$$

where y_i is the predicted intensity value by the spectral model $f(x_i, \theta)$ for the point i while N is the number of points in the observed spectrum. Consequently, the objective is to find the parameters that maximize the likelihood:

$$\theta_{MLE} = \operatorname{argmax} \sum_{i=1}^{N} \log P(y_i|x_i, \theta)\ . \tag{5.2}$$

Notice that the logarithm is taken to simplify the optimization process. Furthermore, it should be noted that it is a common assumption that errors are normally distributed since this simplifies the expression of Equation 5.1 to a product of Gaussian distributions, which in turn simplifies the expression of Equation 5.2 to a sum of Gaussians.

Similarly, Bayesian regression also utilizes the likelihood, $P(D|\theta)$, but it also incorporates prior distributions of the parameters $P(\theta)$ in accordance with *Bayes' theorem*,[169] which is given as:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \tag{5.3}$$

where $P(D)$ is a normalizing constant (marginal likelihood) ensuring a posterior distribution, $P(\theta|D)$, that adds up to 1. In terms of fitting a spectral model to experimental data, the expression of the posterior is calculated based on each point of the spectrum; hence, the expression becomes:

$$P(\theta|D) = \prod_{i=1}^{N} \frac{P(y_i|x_i, \theta)P(\theta)}{P(D)}. \tag{5.4}$$

Subsequently, the objective is to maximize the posterior, which can, for example, be done using *maximum a posteriori estimation* (MAP). This is given as:

$$\theta_{MAP} = \text{argmax}[\log P(D|\theta) + \log P(\theta)]\,. \tag{5.5}$$

Notice that the marginal likelihood is omitted in Equation 5.5 since it is constant with respect to θ and does not affect the optimization. Furthermore, it should be noted that with MAP, a fixed prior distribution (e.g., based on physiologically relevant ranges for the spectral model parameters) is used for Bayesian inference. Several other algorithms can be applied for more intricate Bayesian inference, but they are often associated with more complex applications than spectral fitting and are thus a topic outside the scope of this thesis; readers interested in this subject are referred to the literature on *full Bayesian inference* (such as reference [170]). Another noteworthy consideration is that, similar to MLE, Gaussian assumptions are common regarding the likelihood and the prior, which simplifies the computations of Equations 5.4 and 5.5.

One of the main objectives when fitting is to reduce the complexity of the method used while withholding the fitting accuracy, which is an arguably strong reason for *least squares* (LS) fitting being the most-used method. With LS, the sum of squared residuals is considered, which is given as:

$$S = \sum_{i=0}^{n} (y_i - f(x_i, \theta))^2\,. \tag{5.6}$$

The objective of the LS method is thus to minimize the sum of squared residuals, which can be expressed as:

$$\theta_{LS} = \text{argmin}\, S\,. \tag{5.7}$$

5.1.2 Optimization algorithms for the conventional approach

All the methods mentioned above rely on iterative optimization to either maximize or minimize the objectives (Equations 5.2, 5.5 or 5.7). For LS, one of the most common optimization algorithms is the *gradient descent* (GD) algorithm and its variants, as well as the *multivariate Newton's method* (MVN), *Gauss-Newton algorithm* (GNA) and the *Levenberg-Marquart algorithm* (LMA). All utilize first-order partial derivatives to guide the updating of parameters. GD (Equation 5.8) takes the gradients (partial derivative of the objective function with respect to each parameter) multiplied by a so-called *learning rate* ($0 < \eta < 1$), which reduces the step size as the extreme point (minimum for LS) is approached. While the first-order partial derivatives reflect the steepness and direction needed to reach the extreme point of the objective function, the curvature is reflected by the second-order partial derivatives, which are used instead of the learning rate by the other three algorithms for guiding the optimization process. However, the curvature is computed differently by the three algorithms. For multivariate applications (such as nonlinear spectral fitting of multi-parameter models), the MVN (Equation 5.9) relies on calculating both the gradient and the true *Hessian* (second-order partial derivative of each point's residual with respect to each parameter). In contrast, the GNA (Equation 5.10) is a simplification of Newton's method for LS problems wherein the Hessian is replaced by an approximation based on the (lower complexity) *Jacobian* (which instead is the first-order partial derivative of each point's residual with respect to each parameter). The approximation is valid close to the minimum point, where the residuals are small (since $H = J^T J + \sum_i r_i \cdot \nabla^2 r_i$ and the second-order term has a small contribution close to the minimum), but it becomes problematic far from the minimum. The LMA (Equation 5.10) addresses this limitation by incorporating a damping factor (κ) to balance between the GNA when the residuals are small, and the GD when the residuals have higher values. The update rules for the algorithms mentioned are given in the following equations:

$$\theta_{n+1} = \theta_n - \eta \nabla S(\theta_n)\,, \tag{5.8}$$

$$\theta_{n+1} = \theta_n - [H_S(\theta_n)]^{-1} \nabla S(\theta_n)\,, \tag{5.9}$$

$$\theta_{n+1} = \theta_n - (J^T J)^{-1} J^T \boldsymbol{r}\,, \tag{5.10}$$

$$\theta_{n+1} = \theta_n - (J^T J + \kappa I)^{-1} J^T \boldsymbol{r}\,, \tag{5.11}$$

where n is the current iteration, ∇ denotes the gradient while η is the learning rate, H_S is the Hessian of the sum of residuals while J is the Jacobian and $\boldsymbol{r}$ the matrix of the point-wise residuals. κ is the damping factor and I denotes the identity matrix.

A further comparison and more in-depth look into optimization functions is outside the scope of this thesis; however, an important consideration is that the combination of the method (e.g., MLE, Bayesian regression or LS) and the optimization algorithm (e.g., GD, MVN, GNA or LMA) is what comprises the complete fitting algorithm. Furthermore, the combination of LS and any of the optimization algorithms previously considered results in comparable computational and time complexity, as shown in Table 5.1. This is because while the complexity of MVN is the highest out of the four, it usually results in the fastest convergence requiring a lower number of iterations. This contrasts with GD which, despite having the lowest complexity, usually requires more iterations to converge. Furthermore, LS in combination with any of the optimization algorithms mentioned results in relatively low complexity compared to other traditional fitting algorithms. However, the need for major reductions in time and computational complexity (which will be addressed in terms of the proposed ML-based methods later in this chapter) is the reason for the indifference shown towards the particularity of the optimization algorithm. Consequently, the combination of the LS method with any of the mentioned optimization algorithms is referred to as the *LS-algorithm* for the remainder of this thesis.

Table 5.1 Comparing common optimization algorithms for LS. The complexity analysis is based on the following variables: M refers to the number of data points, p to the number of parameters and k the number of iterations.

Algorithm	Complexity	Preferred use
Gradient descent	$O(k \cdot M \cdot p)$	General nonlinear optimization
Multivariate Newton	$O(k \cdot M \cdot p^2 + p^3)$	General nonlinear optimization
Gauss-Newton	$O(k \cdot M \cdot p^2)$	Mainly for linear (or moderately nonlinear) LS
Levenberg-Marquardt	$O(k \cdot M \cdot p^2)$	Nonlinear LS

GD – complexity due to gradient calculations, MVN – complexity due to inversion of Hessian, GNA – complexity due to calculation of Jacobian, LMA – complexity due to calculation of Jacobian.

Despite being the most-used method for fitting due to its lower complexity relative to other conventional methods, LS remains limited in terms of clinical applications and on the clinical timescale. These limitations signify that it is sensitive to *signal-to-noise ratio* (SNR), the choice of starting values, and most importantly, that LS results in a high computational complexity relative to the clinical timescale, which also scales with the degrees of freedom of the spectral model considered. As was shown empirically in the results described in Paper II, fitting a 10-parameter model (e.g., the Lorentzian) to cover four Z-spectral peaks requires 0.27 s/spectrum[h]. To

[h] For LS-based fitting, the most relevant central processing unit (CPU) specifications are high single-thread performance (fast clock speed and a high number of instructions per cycle), 4-8 cores with 8-16 threads for parallel tasks, a large L3 cache (12-32 MB), and a support for vectorized instructions like AVX2. Most available CPUs fulfill these requirements and result in comparable fitting times.

put this into perspective, in clinical settings, a full brain typically contains approximately 50,000 spectra, which translates into a total fitting time of approximately 4.3 hours. By using a more precise model with higher degrees of freedom (e.g., the Voigt with 14 parameters) the fitting time scales accordingly to approximately 11 hours. Neither are feasible for clinical use; this has also been stated by Foo et al.[165], who concluded that a scientific leap is needed to reduce the gap. It is in this regard, and with the motivation to also remedy other LS-based shortcomings, that two ML-based solutions were developed. A solution based on artificial *neural networks (*NNs) was used in the work described in Paper I, and another solution described in Paper II was based on *gradient boosted decision trees* (GBDT). Each has its own benefits and limitations as will be further discussed later in this chapter.

5.2 Gradient boosted decision trees

5.2.1 Decision trees

To understand GBDT, an intuitive understanding of general *decision trees* (DTs) is needed. The concept of DTs was developed with contributions from various fields and the modern use in ML and data analysis can be traced back to the early- to mid-1980s with the development of several DT algorithms.[171,172] However, the idea of using tree-like structures for decision-making (sometimes referred to as *tree-diagrams*) has much older roots, with applications in *operations research* and *decision analysis*. In its basic form, a DT constitutes a *root node* wherein the first data-dividing prompt is stated. The root node subsequently branches to *decision nodes* (or *internal nodes*), which consider alternative paths to the root prompt. Moreover, the decision nodes proceed to branch recursively, presenting alternative pathways to the prompt of their parent node. The branching stops when there are no further splits, at the so-called *leaf nodes* (Figure 5.1).

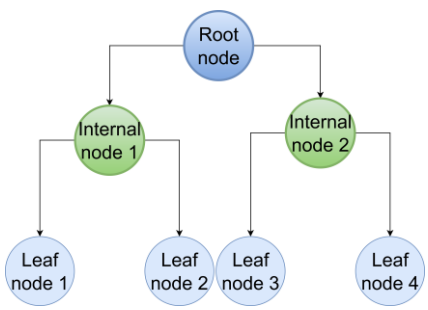


Figure 5.1 Components of a decision tree. The root node branches into decision (internal) nodes, which in turn branch to subsequent nodes. With no further splits, the final nodes are leaf nodes.

While the building blocks and the structure of DTs are straightforward, the way to construct a tree to achieve an intended task might not be. A good starting point for differentiating target groups within a dataset would be to analyze defining characteristics of the samples in a dataset, referred to as *features*, and select "splitting points" based on the features that produce desirable subgroups with respect to a predefined similarity metric. Even for the simplest of tasks, however, this would be a tedious assignment to complete manually, particularly with a growing dataset and especially when there are several features to adhere to. This calls for DT algorithms which, in simple terms, are procedures or methods for building the trees. To guide the splits and tree formation, the algorithms require a measurable quantity, a so-called *split criterion*. Before delving into the DT algorithms, it should be clarified that DTs constitute two *task modalities,* that is, *classification* and *regression* trees. While they are conceptually and algorithmically similar, the difference is that the former are used when the desired outputs are categorical classes, such as a label indicating if the input comes from a pathological or a healthy sample, whereas the latter refers to the case where the outputs are continuous (numerical) values, such as in the case of predicting the spectral model-parameters for a discrete valued experimental spectrum as input.

The foundation to DT algorithms was laid out in the early works of Breiman et al. and Quinlan et al. with the *classification and regression trees* (CART)[171] and the *Iterative Dichotomiser 3* (ID3),[172] respectively. The basic formulation of any DT algorithm (Algorithm 1) is based on processing the targeted dataset beginning from the root node by evaluating feature-and-split-point combinations and then splitting based on the combination that optimizes the split criterion. The process proceeds recursively at each decision node until a stopping condition is met, typically a predefined maximum depth of the tree or no further splits possible. The latter could be due to a predefined subset size, only one class remaining in the subset (for classification), no further improvement with regards to the split criterion, or only one sample remaining in the leaf node. For a trained tree, the most optimal feature-and-split-point combinations at each node, based on the training data, have been established. Consequently, during inference, the formed tree structure is used to determine the class/value of the input data as given in Algorithm 2.

Algorithm 1. General formulation of the DT algorithm.

```
# Pseudocode for the general formulation of DT algorithms
1. Initialize root node with full dataset D.
2. Define function f to determine best split(s):
  - For each feature F:
    - For each split point sp:
      - Split D into subsets D_1,..., D_n.
      - Calculate criterion c for all splits.
      - Select the split combination which optimizes c.
3. Define recursive function to build tree:
  - If stopping criterion met, return leaf node.
  - Else:
    - Find best split using f.
    - Create decision node based on best splits.
    - Recursively build subtrees with subsets D_1,..., D_n.
4. Build tree starting from root node using recursive function.
```

Algorithm 2. Inference using a trained DT.

```
# Pseudocode for the inference algorithm of a trained DT
1. To predict class/value(s) of new instance x:
  - Start at root node.
  - Traverse tree by comparing x feature values to decision nodes' split points.
  - Follow path to correct subtree based on comparison.
  - Return class label/averaged value of reached leaf node.
```

Beyond the fundamental skeleton (Algorithm 1), the intricate differences across DT algorithms arise from *architectural design choices* including (i) split criterion (impurity-, variance- or statistics-based), (ii) *split type* (binary or multiway), (iii) *search strategy*[i] (greedy, semi-global or global search), (iv) *handling of missing values*[j] (imputation, surrogate, fractional or missingness-aware splits), and (v) *pruning type*[k] (pre- or post-pruning). The taxonomy of architectural design choices and some of the classical DT algorithms are summarized in Figure 5.2. For the design choices, there are additional fine structured options which are considered *hyperparameters*. It should be noted that some design choices are strictly associated with task modality. Impurity-based split criteria (e.g., *Gini Index*,[171] *Information Gain*,[172] and *Gain Ratio*[173]) are used for classification trees since they are defined to measure the mixing of classes in a dataset. However, variance-based split criteria, such as, *mean squared error* (MSE) and *mean absolute error* (MAE), are used strictly for regression trees wherein the features have numerical values and can thus be evaluated based on variance. Furthermore, binary splits are mostly used in regression trees, while multiway splits are more common for classification trees since the features in the data used for classification typically have multiple discrete categories which might necessitate the multiway splits to optimize the split criterion considered. The remaining design choices, however, are non-specific for task modality and can be observed in both regression and classification tree algorithms. It should be noted that the development of DT algorithms is an active field of research, so additional modifications to the design choices and mechanisms have emerged and continue to do so. However, the algorithmic setting has been shifted to *ensemble-based algorithms,* which will be further explained in the next section.

---

[i] The predominant search strategy in DT algorithms is greedy search wherein the algorithm selects the feature-and-split-point combination that locally optimizes the split criterion at each node. In contrast, global search seeks the combination of splits that yields the optimal overall tree structure, but it is seldom applied due to its prohibitive computational complexity. Semi-global search strategies offer a compromise by incorporating limited foresight evaluating multiple levels or maintaining a subset of promising candidates to improve decision quality. However, these methods remain substantially more computationally intensive than greedy search, which explains the latter's widespread adoption in practice.

[j] Imputation replaces missing values with estimates (e.g., mean or median), often used when native handling is absent. Surrogate splits use alternative features that approximate the original split to guide decisions when data are missing. Fractional splitting distributes missing instances across branches proportionally to observed data. Missingness-aware methods treat missingness as informative and incorporates it directly into the split logic.

[k] Pruning techniques aim to reduce overfitting by simplifying the tree structure. Pre-pruning halts tree growth early based on predefined criteria (e.g., minimum samples per node or maximum depth), while post-pruning allows full tree construction followed by the removal of branches that do not improve generalization, which is typically validated on a separate dataset.

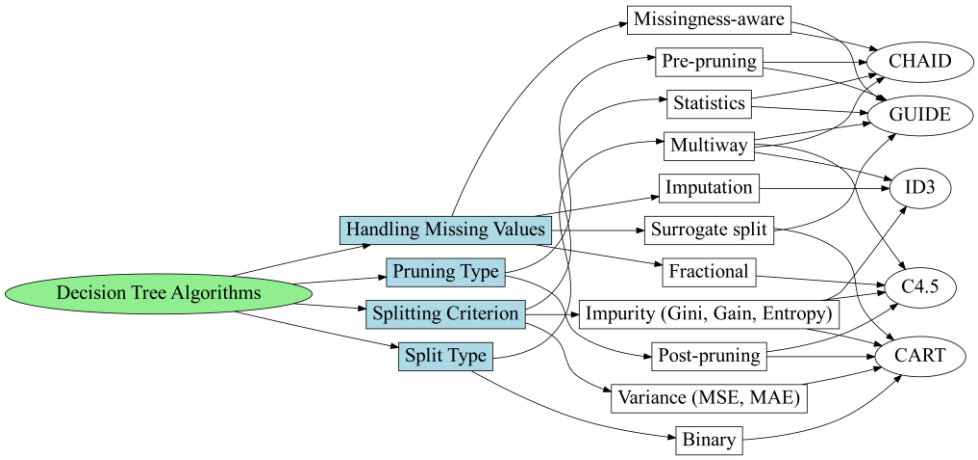


Figure 5.2 Taxonomy of DT algorithms. The four main design choices for classical DT algorithms, their intricate variants, and how the most common DT algorithms relate to these.

From Figure 5.2 it can be inferred that the CART algorithm constitutes greedy search, binary splits, a surrogate split-based handling of missing values, and post-pruning to reduce the risk of overfitting. It can also be seen that the algorithm could either have an impurity-based metric (Gini index) in the classification setting or a variance-based metric for regression, which is reflected by criteria such as MSE and MAE, the former defined in Equation 5.12.

$$MSE_{DT} = \frac{1}{N}\sum_{i=1}^{N}(y_i - c_m)^2\,, \tag{5.12}$$

where N is the number of samples in the node, y is the *ground truth* (GT) value for sample i, and c_m the average of the GT values for all samples in the node. When a split has been performed, based on a feature-threshold combination, the error (Equation 5.12) is calculated for both resulting nodes and the weighted error added for the total error:

$$Error_{tot} = \frac{n_L}{n}Error_L + \frac{n_R}{n}Error_R\,, \tag{5.13}$$

where n_L and n_R are the number of samples in the left and right nodes. respectively, n is the total number of samples, and $Error$ denotes the variance-based error. It should be noted that the right-hand side of Equation 5.13 has two terms due to the binary split of CART trees and would be replaced with an equal number of terms to the number of splits if it was a multiway split.

Despite being proposed as early as 1984[171] and being a relatively simple algorithm, CART has withstood the test of time due not only to its simplicity and, consequently, high interpretability, but also because of its strong baseline performance on both tabular and structured data originating from its nonlinear nature and high flexibility in the main setting. This signifies its suitability for both classification and regression, as well as its inherent handling of missing values and outliers. However, the CART algorithm has considerable limitations with regard to high variance and the risk of overfitting, especially on smaller datasets. Another even more important limitation for the regression setting is the piecewise constant predictions that are the result of each leaf node predicting a constant value, which is the average of the target values in the node. This leads to a step function-like prediction surface which can be non-smooth and sensitive to small changes. The simplicity of the CART algorithm is the main reason for its use as the backbone for ensemble methods. A cornerstone for ensemble-based methods was the *random forests* (RF) proposed in the paper by Breiman et al. in 2001.[174] The concept builds on several weak learners, i.e., individual CART trees, trained on a *bootstrap* (subset) from the original dataset in a *sampling with replacement*[1] fashion and with randomly sampled features. *Aggregation* is subsequently applied during inference, signifying that the majority vote for classification or the average of the predictions for regression is taken. This is given as:

$$pred_{RF} = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t(x)\,, \tag{5.14}$$

where T is the number of trees in the ensemble, and $\hat{y}_t$ is the prediction of the t-th tree. In this way, the limitations of individual CART are overcome since averaging smooths out the piecewise constant predictions, thus reducing variance and producing a more stable, continuous prediction surface.

5.2.2 Learning from the errors of the predecessors

In the last decade, another group of ensemble-based techniques has emerged in the form of gradient-boosted DTs. Unlike RF and other *bagging*-based methods, the trees of the GBDT ensemble are not trained independently. Instead, each tree learns from the errors of the preceding ensemble to strengthen the final learning. This is done by starting with an initial prediction, $F_0(x)$, for the considered dataset, $D = \{x_i, y_i\}_{i=1}^{N}$, where x represents the feature vectors, y is the target values and N is the number of samples in the dataset. In practice, $F_0(x)$ is given a constant value for all samples of D. The constant is determined based on the closed-form solution for

[1] The same sample can reoccur in the bootstrap

the global minimum of the *loss function*[m], $\mathcal{L}$, to ensure stable training and reduce the risk of overfitting. A common loss function used with GBDT for regression is MSE defined as:

$$\mathcal{L}_{MSE}\big(y, F(x)\big) = \frac{1}{2}\big(y - F(x)\big)^2 . \tag{5.15}$$

For MSE, the closed-form solution is thus the mean of y. Using the predictions $F(x)$, the *pseudo-residuals* can be computed as given in Equation 5.16 to represent the direction and magnitude in which the current model needs to be adjusted to reduce the loss.

$$r_i^{(M)} = -\left[\frac{\delta\mathcal{L}\big(y_i, F(x_i)\big)}{\delta F(x_i)}\right]_{F(x_i)=F_{M-1}} , \tag{5.16}$$

where M is the current iteration. It should be noted that for $\mathcal{L}_{MSE}$, the pseudo-residuals are the actual residuals given as $y - F(x)$. With the calculated pseudo-residuals used as targets, a new DT, h_M, is trained resulting in $h_M(x) \approx r_i$. During the formation of the individual CART-like trees of the ensemble, the splits are guided not by the standard variance-based split criteria but instead by the loss-aware *gradient-based gain* defined as:

$$Gain = \frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda_{reg}} - \frac{G^2}{H+\lambda_{reg}} - \lambda_{split} , \tag{5.17}$$

where G is the sum of pseudo-residuals, H is the corresponding sum of second-order derivatives, λ_{reg} is a regularization term for the leaf weights, and λ_{split} is the minimum gain threshold. Notice that the second-order derivative for $\mathcal{L}_{MSE}$ is equal to 1. The process of training new trees using the previous pseudo-residuals as targets is repeated recursively until the loss of a validation dataset has converged or a predefined number or iterations, M, has been reached. For a trained GBDT ensemble, the prediction given a feature vector x is:

$$F_M(x) = F_0(x) + \eta \sum_{m=1}^{M} h_m(x) , \tag{5.18}$$

[m] The terminology in ML is very delicate; in terms of the individual trees, the term "split criterion" is used for the metric that guides the formation of the tree. However, in terms of GBDT, the algorithm is guided by the loss function, which is incorporated into the split criterion of the induvial trees. It should also be noted that the term "loss function" is used in terms of the neural networks that are described in the next section.

where $h_m(x)$ is the individual tree's predicted pseudo-residuals and η is the learning rate. The process described above is, of course, the general formulation of the GBDT algorithm, and more fine-tuned differences have emerged since, the most notable being *XGBoost*,[175] *LightGBM*,[176] and *CatBoost.*[177] These have gained prominence for their predictive accuracy and flexibility, although they offer trade-offs in speed, regularization, and categorical feature handling.

5.2.2.1 XGBoost, LightGBM, and CatBoost

XGBoost utilizes level-wise tree construction similar to CART, meaning it works on all the nodes at the same level before proceeding in depth. However, in contrast to CART, it integrates loss-aware second-order optimization, regularization (L1 and L2), and sometimes histogram-based binning of the threshold values to enhance generalization and control overfitting.[175] LightGBM uses a leaf-wise growth strategy that aggressively splits the leaf with the largest potential gain, thus accelerating training via histogram-based feature binning. *Gradient-based One-Side Sampling* selectively holds high-gradient samples while downsampling well-fitted points to preserve distributional integrity.[176] This allows significant computational efficiency without compromising accuracy. CatBoost, in contrast, constructs so-called oblivious (symmetric) trees using a consistent split at each depth level and integrates ordered boosting, a permutation-driven approach that prevents target leakage and stabilizes training on small datasets. Its native handling of categorical variables through target-aware encodings minimizes preprocessing and hyperparameter tuning.[177] Consequently, the algorithms have their strengths and limitations. While CatBoost often achieves superior accuracy with minimal tuning, LightGBM is superior in training speed and scalability. XGBoost, on the other hand, remains a robust, interpretable baseline with strong regularization capabilities. Collectively, these frameworks exemplify diverse optimization strategies in the GBDT category, each designed to address different primary concerns.

5.2.3 Gradient boosted decision trees for spectral fitting

In the context of spectral fitting, the problem can naturally be framed as a regression task, wherein the outputs are continuous-valued parameters that optimally fit the spectral model to the discrete input spectrum. The formulation is also well aligned with *supervised learning*, which is training an ML model with a paired dataset $D = \{x_i, y_i\}_{i=1}^{N}$ constituting sample-and-GT pairs. GBDT, known for their robustness, interpretability, and ability to model complex nonlinear relationships, present desirable alternatives to conventional optimization-based methods (Chapter 5.1). In the simplest form, the feature vectors, i.e. inputs, are taken to be the intensity values of the discrete points. Additionally, the possibility of *feature engineering*, which is extracting features from the raw spectra (e.g., descriptive statistics) could provide

further insight for strengthening the performance of the trained model. However, this approach requires careful design of features with a strong implicit relationship to the model parameters, which in turn requires an intricate empirical analysis. In the work described in Paper II, it was shown that the raw intensities as features were sufficient for the high accuracy mapping of the parameter values for both Lorentzian and Voigt spectral models including four Z-spectral peaks.

As previously mentioned, the output in spectral fitting consists of multiple model parameters. Moreover, the number of parameters typically scales with both the number of peaks present in the fitted spectrum and the degrees of freedom associated with the spectral model considered. The problem thus needs to be treated as a multi-output regression task, an aspect which thus far has not been addressed for the GBDT algorithms since all previous theory was focused on single-output formulations. On that note, one straightforward approach is to train separate GBDT ensembles for each model parameter using the same feature vectors as inputs. While this method has demonstrated success in various applications, the approach inherently neglects potential correlations among the model parameters that could be leveraged for improved efficiency and accuracy of mapping, particularly for more complex spectral models. An alternative approach, which is natively supported by the CatBoost library, involves extending the MSE to a multi-target version, referred to as multi-target MSE. This is defined as**:**

$$multiMSE = \frac{1}{N}\sum_{i=1}^{N} \| y_i - \hat{y}_i \|^2 , \tag{5.19}$$

where y_i and $\hat{y}_i$ denote the true and predicted vectors of model parameters for sample i, respectively, and N is the number of training samples.

Analogous to feature-engineering, the output could also be adjusted. In Chapter 4, it was explained that derived parameters such as H (Equations 4.2 and 4.6 for Lorentzian and Voigt respectively) and LW (Equations 4.3 and 4.5 for Lorentzian and Voigt respectively) are commonly used for inter-model comparisons and for practical applications. In this regard, targeting the derived parameters streamlines direct analysis. In the work of Paper II, it was empirically shown that both approaches are feasible.

5.3 Densely connected neural networks

5.3.1 Theory

Inspired by the biological neuron, the *perceptron* was invented by Frank Rosenblatt in 1958.[178] The algorithm operates by computing a weighted sum of its inputs and applying a nonlinear transformation via a so-called *activation function* to obtain the output. The activation function is essential for enabling the model to map nonlinear relationships in the data. Mathematically, the perceptron can be expressed as:

$$\hat{y} = \phi\left(\sum_{i=1}^{N} x_i \cdot w_i + b\right) = \phi(\boldsymbol{w}^T \boldsymbol{x} + b)\,, \tag{5.20}$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$ is the input vector, $\boldsymbol{w} = [w_1, w_2, \ldots, w_N]^T$ is the weight vector, b is the bias term, $\phi(\cdot)$ is the activation function, and $\hat{y}$ is the output of the perceptron. By connecting multiple perceptrons, the *multilayer perceptron* (MLP), better known as an artificial NN, is constructed. Figure 5.3 illustrates an NN with its main building blocks.

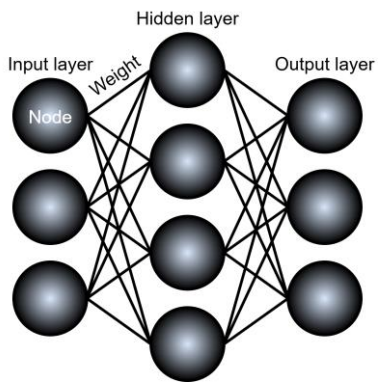


Figure 5.3 General overview of the building blocks of a NN. A NN with the three main layers: input, hidden and output. Each purple circle is a node (perceptron), and the links between the nodes are the weights (the trainable parameters).

In this architecture, each perceptron is referred to as a node or neuron. The first and last layers are termed as the input and output layers, respectively, while any intermediate layers are denoted as *hidden layers.* When an artificial NN constitutes more than one hidden layer, it is referred to as a *deep neural network* (DNN). The weights and biases in a NN are known as trainable parameters and are iteratively adjusted during training to minimize a loss function. The total number of trainable parameters, and thus the model's complexity, is determined by the number of layers

and nodes per layer. A highly complex model can effectively approximate complex mappings of inputs to outputs, which is analogous to using a spectral model with a high number of degrees of freedom to fit a discrete spectrum. However, increasing model complexity also increases the risk of overfitting, wherein the model learns noise or specific patterns in the training data that do not generalize to unseen data. The trade-off between model complexity and generalization capabilities is a central consideration in NN design and training and will be addressed further in the coming paragraphs.

In supervised learning, the training of an NN constitutes three fundamental steps: (i) *forward propagation*, (ii) *error calculation and gradient evaluation*, and (iii) *backpropagation for parameter updates*. The forward propagation for a given input $\boldsymbol{x} = \boldsymbol{a}^0$ results in the network computing an output $\hat{\boldsymbol{y}}$ by sequentially applying affine transformations and activation functions across all the layers, which can be expressed in a general form as follows:

$$\boldsymbol{a}^{(l)} = \phi^{(l)}\left(\boldsymbol{w}^{(l)}\boldsymbol{a}^{(l-1)} + b^{(l)}\right), \quad l = 1, \dots, L\,, \tag{5.21}$$

where $\boldsymbol{a}^{(l)}$ denotes the activation of layer l, and L is the total number of layers in the network. In the error calculation step, the discrepancy between the predicted output $\hat{\boldsymbol{y}}$ and the true target $\boldsymbol{y}$ is quantified using a predefined loss function $\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})$, which, as with GBDT, should be variance-based, e.g., multiMSE for a multi-regression task (Equation 5.19). The gradients of the loss with respect to the model parameters $\delta\mathcal{L}/\delta\boldsymbol{w}$ and $\delta\mathcal{L}/\delta b$ are also computed, utilizing the chain rule due to the sequential nature of the forward propagation. Finally, for the backpropagation and parameter update step, the computed gradients are propagated backward through the network and used to update the parameters in the direction that minimizes the loss. This is done with the guidance of an optimization function, which in most cases is GD (Equation 5.8) or one of its variants.

It follows that the selection of the number of trainable parameters, along with the choice of activation functions, loss function and optimization algorithm constitute the hyperparameter configuration of the network. The hyperparameters are not learned during training but are instead predefined, and their choice can highly influence the trained model's performance. Consequently, finding a balanced hyperparameter combination reflects the previously mentioned trade-off between model complexity and generalization capabilities. To systematically find an optimal hyperparameter configuration, *hyperparameter optimization* (HPO) is employed. The process consists of three main steps, which are (i) defining the *search space*, (ii) defining the *objective function,* and (iii) choosing the *search strategy*. The first step encapsulates the variability allowed during the search by defining the ranges or choices possible for each hyperparameter. The second step determines the

performance criterion to evaluate improvements in generalizability as a function of the hyperparameter configuration, which is evaluated based on the performance on a validation dataset. The final step determines the method by which the search space is explored. For a relatively small search space *grid search,* which exhaustively evaluates all possible combinations may be computationally feasible. However, for higher dimensional or continuous search spaces, this search method becomes computationally unfeasible. *Random search* is thus preferred in these cases since it randomly samples configurations from a uniform distribution and is more efficient in high-dimensional settings.[179] More involved strategies such as *Bayesian optimization*,[180] *Hyperband*,[181] and *population based training*[182] offer a trade-off between computational complexity and efficiency.

5.3.1 The single Lorentzian fitting neural network

The *single Lorentzian fitting neural network* (sLoFNet) is a DNN developed in the work described in Paper I to fit discrete DS spectra which adhere to Lorentzian shapes. Consequently, the outputs are the Lorentzian shape parameters (Equation 4.1). The architectural design was obtained via HPO, and the defined search space and resulting hyperparameter configuration are given in Table 5.2. During the HPO process, candidate configurations were evaluated on a held-out validation dataset using a composite search objective that incorporated both MAE and root-MSE. The adopted search strategy was random search. The inclusion of both metrics in the objective function was intended to balance sensitivity to large errors (captured by root-MSE) with robustness and interpretability (offered by MAE). While root-MSE is widely used, it has been shown to be less reliable for inter-model comparisons due to its dependence on the average error (i.e., MAE), the number of samples, and the variance within the samples.[183] In contrast, MAE provides a scale- and variance-invariant measure of performance. Consequently, by combining both metrics, a selection of models with strong generalization performances across a range of error profiles is favored.

Table 5.2 sLoFNet architecture and search space for its HPO. The defined search space and the resulting hyperparameter configuration from the HPO for sLoFNet. This table is adapted from Paper I.

Hyperparameter	Choices	Results from HPO
Number of hidden layers	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	5
Nodes per hidden layer	[128,256, 512, 1028, 2048]	2048
Loss function	[MAE, Root-MSE]	Root-MSE
Final activation	[Tanh, Linear, None]	None
Batch size	[32, 64,128, 256]	32

The optimization algorithm for the model was chosen explicitly to be the *adaptive moment estimation*[184] (Adam), because it is particularly effective for large datasets and high-dimensional parameter spaces. It is given as:

$$\theta_{t+1} = \theta_t - \eta \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon}, \tag{5.22}$$

where $\widehat{m}_t = m_t/(1-\beta_1^t)$ and $\hat{v}_t = v_t/(1-\beta_2^t)$ are the bias-corrected estimates of the first moment estimate $m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$ (mean of gradients) and second moment estimate $v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$ (uncentered variance), respectively, and where $g_t = \nabla_\theta \mathcal{L}_t$ is the gradient of the loss function with respect to the trainable parameters at time point t. β_1 and β_2 are the decay rates for the moment estimates, and ϵ is a constant (typically 10^{-8}) to prevent division by zero.

It should be noted that repeating the HPO would likely yield different optimized configurations since the obtained configuration is one of many solutions to the objective function. Using an interpolation-based approach, Goodfellow and Vinyals[185] empirically showed that there is a smooth plateau region of the objective function landscape in which all model configurations perform comparably. Thus, the focus should be on ensuring that the obtained configuration is one that generalizes well.

5.4 Comparison of fitting approaches

5.4.1 Theoretical foundation

While Big O notation[186] provides a theoretical framework for understanding how different models scale with input size and model complexity, it does not always fully explain the empirical performance. The inference of DNNs has a complexity $O(L \cdot N \cdot M)$, where L is the number of layers, N the number of nodes per layer, and M the number of datapoints. In contrast, GBDT scale as $O(D \cdot M)$, where D is the maximum depth of the trees in the ensemble. It should be noted that for training, both methods have additional computational cost. For GBDT, the complexity becomes $O(T \cdot D \cdot M)$, where T is the number of trees in the ensemble. For DNNs, training complexity scales as $O(E \cdot L \cdot N \cdot M)$, where E is the number of training epochs. In comparison, the nonlinear LS-algorithm has an inference-time complexity of approximately $O(k \cdot M \cdot p^2)$. This formulation highlights that LS-based methods scale quadratically with the number of parameters p while the inference of both ML-methods is independent of the parameter count. This difference becomes especially significant with high-dimensional models since the

additional dependence on data size renders LS fitting infeasible for clinical applications.

5.4.2 Practical factors

The empirical results described in Papers I and II confirm that both ML-based approaches significantly outperform LS in terms of inference time. For instance, the inference when using the ML-models (on the average available graphical processing units, GPUs) required approximately 1 second per brain for four-pool Z-spectra compared to several hours for the LS-algorithm. The latter aligns with theoretical expectations that LS complexity scales with the number of parameters. This was further validated in Paper II, wherein the difference in inference time between a 10-parameter model (four-pool Lorentzian) and a 14-parameter model (four-pool Voigt) was negligible in the ML setting but increased tangibly from 4 to 11 hours in the LS setting. On this basis, the ML-based approaches close the gap for clinical application.

In addition to the theoretical foundation, several practical factors are of importance. GBDT rely on fast, shallow tree traversal and benefit from highly optimized implementations (e.g., LightGBM and CatBoost), thus requiring significantly fewer *floating-point operations* compared to both DNNs and the LS-algorithm. While DNNs are more computationally intensive due to matrix operations and nonlinear activations, they leverage efficient tensor operations and are well-suited for GPU acceleration. Both ML-approaches, in fact, benefit from GPU-level parallelism, which further increases the performance advantage. In contrast, LS-methods often involve computationally expensive function evaluations that depend on the formulation of the spectral model (e.g., Voigt or Lorentzian), and they are typically implemented in general-purpose solvers that lack hardware-specific optimizations. Taken together, these theoretical and practical considerations explain why GBDT are empirically the fastest, followed by DNNs, with the LS-algorithm being the slowest.

5.4.3 Deep neural networks versus gradient boosted decision trees

It is worth noting that while GBDT models exhibit only marginally faster inference times compared to DNNs, they are significantly more efficient in terms of training speed. This difference arises primarily from the algorithmic structure of DNNs, which involve iterative forward propagation, error calculation, gradient computation, and parameter updates across multiple layers (Chapter 5.3.1), all operations that scale with the number of trainable parameters and the depth of the network. In contrast, GBDT are trained sequentially by fitting shallow DTs to residuals, a process that is inherently less computationally intensive and more

flexible to parallelization. It should be mentioned, however, that the computational burden of DNNs such as sLoFNet can be reduced slightly by decreasing the number of trainable parameters without necessarily compromising performance. As suggested by Goodfellow and Vinyals,[185] the loss landscape of NNs often contains a broad plateau region in which many configurations yield comparable performance. Consequently, identifying a model within this region that also minimizes the number of trainable parameters is therefore desirable since it offers a favorable trade-off between efficiency and generalization. Nonetheless, due to their inherently lower algorithmic complexity and the simplicity of their inference mechanism, GBDT maintain a computational advantage over DNNs especially in the training phase. This makes them especially attractive for applications in which reduced time and computational complexity is of primary concern. The increased algorithmic complexity of DNNs does, however, offer advantages. It enables modeling of highly nonlinear and intricate relationships, thus making DNNs more suitable for complex tasks. When properly regularized, DNNs also tend to have greater robustness in terms of variance and noise in the data.

It follows that any choice between GBDT and DNNs should be guided by the specific requirements of the intended application. If models need to be retrained frequently, such as when acquisition protocols for the Z-spectra change or due to other changes to the imaging conditions, GBDT are preferred due to their significantly faster training times and lower computational load. In contrast, when robustness against data variability is a primary concern, DNNs offer a powerful choice for fitting. Their increased model complexity and strong capacity for generalization enable them to better tolerate variations in the input data, such as noise fluctuations, reduction of sampling density, or moderate inconsistencies. In this context, robustness thus refers to the model's ability to maintain stable performance in the presence of such well-represented perturbations in the training data. However, it is important to recognize that this robustness is inherently limited since excessive variability in the training data might obstruct convergence and hinder the model's ability to learn the intended mapping. For instance, a model trained to predict spectral fitting parameters such as sLoFNet should ideally be invariant to moderate noise, as shown to be the case in Paper I, which indicated a marginal increase of fitting error with added noise up to 4%. However, the DNN-based fitting model should not be expected to perform denoising as an additional task. If denoising is required, it should be addressed through dedicated models.

Furthermore, the lower algorithmic complexity of GBDT provides an additional benefit beyond the reduced time complexity, that of enhanced interpretability. The decision paths within individual trees can be readily analyzed to understand the influence of specific features on predictions. In this way, targeted feature engineering is facilitated, which can further improve model performance.

5.4.4 Training data and mapping

Beyond the differences originating in the algorithmic design, another significant difference between sLoFNet (Paper I) and the GBDT-based fitting approach (Paper II) lies in the nature of the training data. In the main configuration, sLoFNet was trained on simulated spectra, where each input-output pair consisted of a discrete spectrum simulated over a predefined range of the invariant variable x, with the corresponding Lorentzian shape parameters (Equation 4.1) sampled from uniform distributions. The parameter ranges were selected based on values observed in vivo, as reported in the literature. Since only a single spectral peak was modeled, the outputs were the shape parameters for that peak. In contrast, the GBDT-based fitting model was trained on in vivo spectra that had been pre-fitted using the LS-algorithm to extract the shape parameters for four peaks modeled using either the Lorentzian (Equation 4.1) or Voigt (Equation 4.4) functions. In the work described in Paper I, a similar approach was adopted with a DNN resulting in the $\text{sLoFNet}_{\text{Invivo}}$ variant, which is algorithmically identical to sLoFNet but trained on a smaller size LS-fitted in vivo dataset instead of the simulated dataset used for sLoFNet to enable a direct comparison. This comparison revealed that $\text{sLoFNet}_{\text{Invivo}}$ demonstrated potential but lacked full generalizability due to the small training data size of approximately 50,000 spectra. Specifically, empirical results indicated that between 500,000 and 1,000,000 simulated samples were needed to ensure generalizability of sLoFNet, as validated on both simulated and in vivo test sets. On this note, it is important to consider the cost of increased training time which, as previously discussed, scales linearly with the number of training samples. The preprocessing step for the in vivo training data presents an additional practical challenge since generating large-scale in vivo training datasets requires extensive LS-based fitting, which is computationally expensive and scales poorly with both the number of samples and parameters of the spectral model considered. Conversely, the GBDT-based model achieved convergence and generalization with a relatively small dataset of approximately 5,000 samples, significantly reducing the LS burden. Nonetheless, the reliance on LS to generate GT labels remains a bottleneck for the approach in which in vivo data are used for training.

In theory, one could simulate multi-peak spectra using parameter ranges observed in vivo, similar to the sLoFNet approach. However, as the number of modeled pools increases, the complexity of accurately capturing inter-parameter dependencies also increases. The dependencies, such as correlations between peak amplitudes, widths and positions, are difficult to model explicitly. Preliminary analyses using covariance matrices and kernel-based methods suggested the presence of nontrivial correlations, further complicating the simulation process. Moreover, training on real in vivo data implicitly captures such dependencies, including potential interactions between Z-spectral pools, which are otherwise difficult to encode in the simulated data.

A successfully trained ML-based fitting model learns to approximate the mapping between the spectral input and model parameters, effectively learning the underlying spectral model (e.g., Lorentzian or Voigt) while tolerating the variability observed in the experimental data. This includes both *reducible* and *irreducible* sources of error. The latter, which originate from inherent noise and biological variability, places a fundamental limit on model performance regardless of training strategy. However, the former can be suppressed with the proper representation in the training data. Consequently, while simulated data offer scalability and control, they may fail to capture the full complexity of in vivo spectra, particularly in multi-pool models. The success of sLoFNet is largely attributable to the simplicity of the single-pool model. For more complex models, training on LS-fitted in vivo data remains the most effective strategy, provided that sufficient data are available. As such, the use of experimental data should be prioritized whenever feasible, since it offers the most faithful representation of the underlying biological and physical processes.

6. Z-spectral denoising

6.1 Available denoising approaches

6.1.1 Principal component analysis

Principal component analysis (PCA) is a statistical method that leverages a linear and orthogonal transformation to reorient high-dimensional data to a new coordinate system in order to find the uncorrelated directions (*principal components*, PCs) that describe the data variance the most.[187] Consequently, the largest data variability can be captured in just a few PCs. This allows a reduction in dimensionality to facilitate interpretation, and it has several potential applications.

Before computing the PCs, the feature matrix, $X \in \mathbb{R}^{N \cdot n}$ is standardized, where N is the number of samples and n the number of features. This process involves centering and scaling the data to ensure that all features contribute equally to the analysis. This can be mathematically expressed as subtracting the mean of each feature μ_j and dividing by its standard deviation, σ_j, given as[187]:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}. \tag{6.1}$$

Therefore, the resulting standardized data matrix can be denoted $\tilde{X} \in \mathbb{R}^{N \cdot n}$. The next step is to compute the covariance matrix $C \in \mathbb{R}^{n \cdot n}$, which captures the pairwise linear correlations between the features[187]:

$$C = \frac{1}{N-1} \tilde{X}^T \tilde{X}. \tag{6.2}$$

It should be noted that the denominator is Bessel's correction, which serves to remove bias from the estimation of the population covariance of a sample. Furthermore, each entry $c_{ik} \in C$ represents the covariance between features i and k. This signifies that each entry on the diagonal of C (i.e., $i = j$) holds the (uncorrelated) variances of the features while each off-diagonal entry ($i \neq j$) represents how the pairs of features vary together. Consequently, the next step

involves eigen-decomposition of the covariance matrix to obtain the eigenvalue matrix, Λ, (explained variance) and the eigenvectors matrix, Q, (PCs)[187]:

$$C = Q\Lambda Q^T . \tag{6.3}$$

This is solved for the individual eigenvalues λ as $\det(C - \lambda I) = 0$, where the eigenvalues form the diagonal of Λ. For each eigenvalue, the corresponding eigenvector $\vec{v}$ is solved through $(C - \lambda I)\vec{v} = 0$; these eigenvectors form the columns of the matrix Q.[187]

The eigenvectors define the direction of maximum variance (PCs), and the corresponding eigenvalues quantify the amount of variance explained in these directions. Eigenvectors are thus sorted in descending order of their associated eigenvalues. Selecting the top k components results in a so-called projection matrix $W_k \in \mathbb{R}^{n \times k}$ which can be used to transform the data, and $\tilde{X}_{PCA} = \tilde{X}W_k$ where $\tilde{X}_{PCA} \in \mathbb{R}^{n \times k}$ is the reduced dimensionality representation. Due to the stochastic nature of noise, it often manifests in the directions of low variance, so discarding components with low variance can serve as an effective denoising approach (also leveraged for compression). However, it should be noted that the choice of threshold highly affects the denoising performance. PCA has, nevertheless, been used successfully for denoising Z-spectra in the work by Breitling et al.[26]

6.1.2 Deep learning-based denoising

Several DL-based denoising approaches have been proposed over the years, with different algorithmic constructions and learning mechanisms.

6.1.2.1 Autoencoders

Autoencoders (AE) are a class of NNs designed to learn a compressed latent representation of the input. This is accomplished with the help of the two-component design that includes an *encoder* and a *decoder*. The encoder is typically designed in a manner that reduces the number of nodes per layer successively until reaching the end of the encoder, at which point the number of nodes represents the dimension of the latent representation. The final layer of the encoder connects to the first layer of the decoder, which is commonly designed to mirror the encoder. In other words, each layer of the decoder is followed by a layer with an increased number of nodes in order to reconstruct the input from the latent representation. Hence, an AE is usually symmetrical in its architectural design. When reducing the number of hidden layers to a single layer and using a linear activation for all the nodes, an AE is reduced to the PCA transformation,[187,188] with the number of nodes in latent space corresponding to the number of PCs used. The objective of an AE is to minimize the reconstruction error by typically using a loss function such as MSE.

AEs have been successfully applied for denoising[189] in what is referred to as *denoising autoencoders* (DAEs).

6.1.2.2 Convolutional neural networks

When the weights in a conventional NN (Chapter 5.3.1) are replaced with *convolutional kernels*, the resulting architecture is referred to as a *convolutional neural network* (CNN). Consequently, the forward pass of the individual perceptrons (Equation 5.20) is replaced by a discrete convolution between the input, x, and a kernel, w, which for a 1D input can be defined as:

$$y[n] = \phi\left(\sum_{k=0}^{K-1} x[n-k] \cdot w[k]\right) \text{ for } n = 0,1,\dots,N-K-2\,, \qquad (6.4)$$

where $y[n]$ are the point-wise values of the output vector y, N is the length of the input x, and K is the length of the kernel w. Notice that just as in the conventional NN case, the output is passed through an activation function ϕ. Thus, in CNNs, the trainable parameters are convolutional kernels referred to as filters, which are iteratively updated during training. The filters will form during the training process based on highly prevalent local spatial dependencies in the training data. Consequently, if the input contains the learned structures of the filters, at each layer during both forward propagation and inference, the convolutions will produce a feature map with highlighted regions that can subsequently be leveraged for the intended task. It should be noted that a CNN can have several filters at each layer, which is analogous with the possibility of several nodes per layer in conventional NNs. As with conventional NNs, in the supervised setting, CNNs are trained using the previously described three-steps learning algorithm (Chapter 5.3.1).

It is also common for CNNs to utilize dimensionality reductions by applying so-called *pooling layers*. The two most common are *max pooling* and *average pooling*, which are defined in Equations 6.5 and 6.6, respectively.

$$y[n] = \max\{x_{s\cdot n}, x_{s\cdot n+1}, \dots, x_{s\cdot n+K-1}\}\,, \qquad (6.5)$$

$$y[n] = \frac{1}{K}\sum_{j=0}^{K-1} x_{s\cdot n+j}\,, \qquad (6.6)$$

where n denotes the position in input x, s is the stride (step size) of the pooling window, and K is the window size. By using pooling layers, the computational load can be alleviated due to the reduction in data size. Furthermore, the projection of a higher dimension to a lower one will emphasize prominent features, either by keeping max values or averaging them with max and average pooling respectively. This results in an inherent translation invariance and noise suppression.

When convolutional layers are used in both the encoder and decoder paths of an AE, the resulting architecture is referred to as a *convolutional autoencoder* (CAE). As clarified in the previous section, the AEs aim to transform the input to a lower dimensional representation. Therefore, pooling layers are essential building blocks for achieving dimensionality reduction in CAEs. However, it was also clarified that AEs aim to reconstruct the input from the latent dimension. Consequently, the counterpart to pooling layers, i.e., upsampling layers, are necessary for regaining the input size. A common upsampling process is based on *nearest neighbor linear interpolation*, which is defined as:

$$y[n] = x_{n/s} + \left(\frac{ns}{s}\right)\left(x_{n/s+1} - x_{n/s}\right), \tag{6.7}$$

where $s = 2$ is common for doubling the dimensionality. Alternatively, *transpose convolution* with a $stride > 1$ (commonly equal to 2) is used to have a *learned upsampling* instead. It should be noted that CAEs used for denoising are referred to as *denoising convolutional autoencoders* (DCAEs).

6.1.2.3 State-of-the-art approaches

In the context of CEST, the most recent DL-based denoiser applied to individual Z-spectra is that of Kurmi et al.,[28] who developed a DCAE and incorporated a deterministic *Kullback-Leibler* (KL) divergence in a *curriculum learning-based approach*. In their work, the authors built a symmetric AE with a four-block encoder, and thus a four-block mirrored decoder (Figure 6.1A). Each block of the encoder constituted a 1D convolutional layer followed by the *exponential linear unit* (ELU) activation function and a subsequent max-pooling layer. The number of filters used at the convolutional layers was doubled block-wise starting from 32 and ending at 256 filters, with the kernel size for all convolutions set to 3. The output of the encoder was flattened and followed by a dense layer with 32 nodes making up the latent space of the architecture. To mirror the encoder, the decoder started with a dense layer whose size was equal to that of the flattened layer, which was then followed by the four decoder blocks. Each decoder block contained a *transposed convolutional layer* with a $stride = 2$ serving to upsample the input, followed by a convolutional layer. The number of filters per convolutional layer mirrored the block-wise progression of the encoder, i.e., starting with 256 and ending with 32 filters and the kernel size was consistently set to 3. A final convolutional layer with one filter followed, and a subsequent cropping layer was applied to regain the input shape. It should be noted that each layer in the decoder was followed by a linear activation function.

The curriculum learning-based approach (Figure 6.1B) that Kurmi et al. used comprised two steps. In the first step, a baseline model was trained on noisy z-spectra with varying noise levels (1 to 5 %) and two different saturation powers. In

the second step, two models were fine-tuned starting from the baseline, and they were focused on one saturation power each. At this training step, a deterministic KL-divergence was applied between the output of the fine-tuned model and the baseline to restrict tangible divergence and thus ensure stability. Furthermore, *context learning* was employed for the inference. This signifies that, for each input to be inferred, a temporary model was further fine-tuned by applying a deterministic KL-divergence between the model output and a PCA denoised version. Again, this was done to restrict the model output from deviating tangibly and to ensure stability.

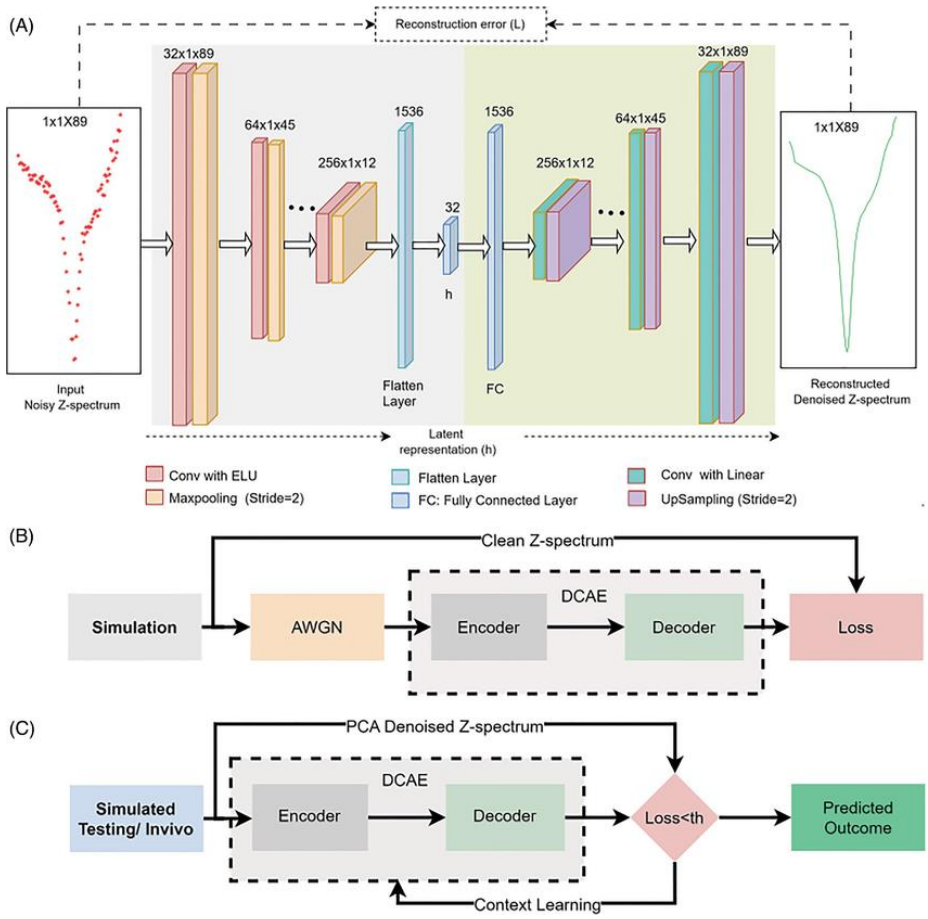


Figure 6.1. Overview of the denoiser by Kurmi et al. (A) The architecture, (B) the training process setup, and (C) the structure of the context learning inference. The figure was adapted from the original work[28] with permission from the publisher.

6.2 Constrained loss autoencoder residual denoiser

In the work described in Paper IV, a novel DL-based denoiser was proposed with an asymmetric CAE architecture including several innovative key components (Figure 6.2), two of which are particularly noteworthy. The first is a three-way encoder wherein the first and second paths target short- and long-range dependencies, respectively, while the third path targets noise explicitly. The second is a subtractive layer that aims at removing the mapped latent noise from the latent representation of the noisy input. Additionally, a custom loss was utilized combining the reconstruction loss of the full spectrum $\mathcal{L}_{Full}$ with the reconstruction losses constrained to spectral regions of interest ($\mathcal{L}_{rNOE}$, $\mathcal{L}_{guanidinium}$ and $\mathcal{L}_{amides}$), with a weighting factor, ζ, defined by:

$$\mathcal{L}_{Comb} = \mathcal{L}_{Full} + \zeta\left(\mathcal{L}_{rNOE} + \mathcal{L}_{guan} + \mathcal{L}_{amides}\right). \tag{6.8}$$

Consequently, the denoiser was named the *constrained loss autoencoder residual denoiser* (CLAERD). Each of the encoder's paths contains four blocks wherein the fourth block of the first and second paths was shared since the outputs of their respective third blocks were merged via a concatenation layer. Furthermore, each block of the encoder's paths constituted a convolutional layer and a subsequent max-pooling layer according to the common hierarchical CAE feature extraction strategy. This reduces the spatial dimension (with pooling) while deepening the feature representation with a block-wise increase of the filters, in this case, according to 16 →32→64→256 filters. All convolutional layers used a kernel size equal to 3, except the convolutions of the first three blocks of the second path (long-range dependencies path), which used a kernel size equal to 9. Moreover, the output from the merged feature maps of noisy signals were given as input to the subtraction layer together with the feature maps from the noise path. A one-path decoder followed, constituting an upsampling followed by two convolutional layers with 256 and 16 filters, respectively and a kernel size equal to 3 for both. Finally, a flattening layer was applied and connected to a dense layer with the number of nodes set equal to the input size to obtain the reconstructed denoised spectra.

The activation functions used after the convolutions were carefully chosen to achieve the intended mechanism for each architectural component. Firstly, the activation function after each convolution in the first and second paths was the *leaky-rectified linear unit* (leaky-ReLu), except for the final block, which used the *hyperbolic tangent* (tanh) function. The leaky-ReLu was used to extract both signal and noise features since noise might have subtle negative components that will be preserved with the Leaky-ReLu as opposed to the standard *ReLu*. This is because the leaky-ReLu permits a small nonzero gradient for negative values while maintaining the feature-enhancing behavior of ReLu through linear growth of

positive inputs, an important property for the amplification of prominent signal features. The activation functions in the noise path were carefully chosen to refine the noise representation. The leaky-ReLu was used for the first block to preserve small noise variations. This was followed by the ELU in the second block to enhance feature expression with smoother transitions by reducing bias shift, while still preserving negative noise components. The swish activation function, given as; $x \cdot 1/(1 + \exp(-x))$, was applied to the output of Block 3 due to its smooth non-monotonic nature, which is advantageous for capturing complex noise structures. The tanh was used for the final block of both the noisy signal and noise paths to align their outputs by centering it around zero before the subtraction between the two. It should be noted that all layers of the decoder used a linear activation function to preserve the obtained distribution of the denoised latent representation.

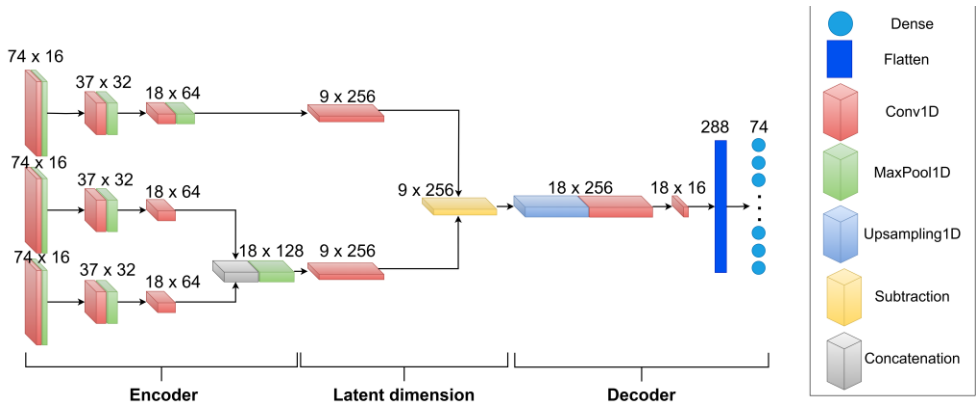


Figure 6.2 The architecture of CLAERD. The input is passed to a three-way encoder (the leftmost block): the first path (bottom left) focuses on short-range dependencies, the second (middle left) on long-range dependencies, and the third (top left) on noise features. The "residual subtraction" layer (yellow) preforms the denoising in latent space and the following decoder reconstructs the denoised spectra. The Figure is adapted from Paper IV.

6.2.1 Latent space analysis

When subtracting two biased magnitude signals (Rician distributions), the bias can be cancelled leaving a difference which is well explained by a Gaussian-like distribution.[47] This was empirically confirmed in Paper IV, as can be seen in Figure 6.3A, where a Rician governed noise-free signal was subtracted from the corresponding noisy signal spectra. The theoretical expectation is that the latent mapping of noise should have a Gaussian-like distribution, suggesting that successful noise mapping should be indicated by a Gaussian-like distribution in the final feature maps of the noise path. Furthermore, the noise maps of the blocks in the noise path should successively move towards a Gaussian-like distribution. This was indeed verified in the results presented in Paper IV, wherein a nonsignificant

D'Agostino's K^2 test (i.e., for a Gaussian-like distribution) was only the case for the final block of the noise path as shown in Figure 6.3B-E.

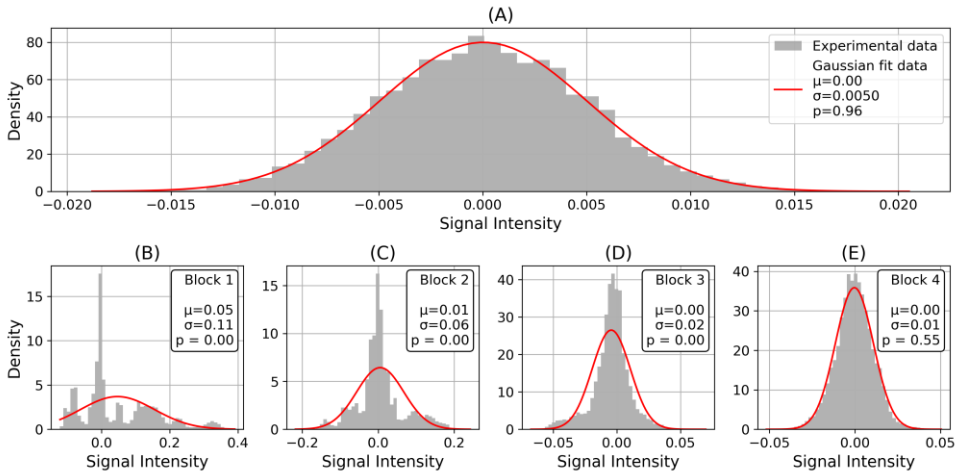


Figure 6.3. Histogram analysis of latent space. (A) The histogram and a corresponding Gaussian fit for the difference between two biased distributions (noisy signal and noise-free signal). The histograms of the feature maps for the noise and a corresponding attempted Gaussian fit for the maps from (B) Block 1, (C) Block 2, (D) Block 3, and (E) Block 4. The figure is adapted from Paper IV.

Furthermore, Figure 6.4 shows *t-distributed stochastic neighbor embedding* maps which project the high-dimensional latent spaces of CLEARD for noisy signal and noise, respectively, onto two dimensions. The plots revealed one major cluster for the latent space of noise while two major clusters could be observed for the noisy signal latent space. This is likely to be explained as noise and signal components compared to only noise components in the noise latent space projection (Figure 6.4B). Furthermore, the fine-grained clusters of the noisy signal projection were better delineated compared to the projection of only noise. This is again indicative of successful mapping since the signal components are many, e.g., from different peaks of the Z-spectra, while the variation in noise was primarily based on noise levels.

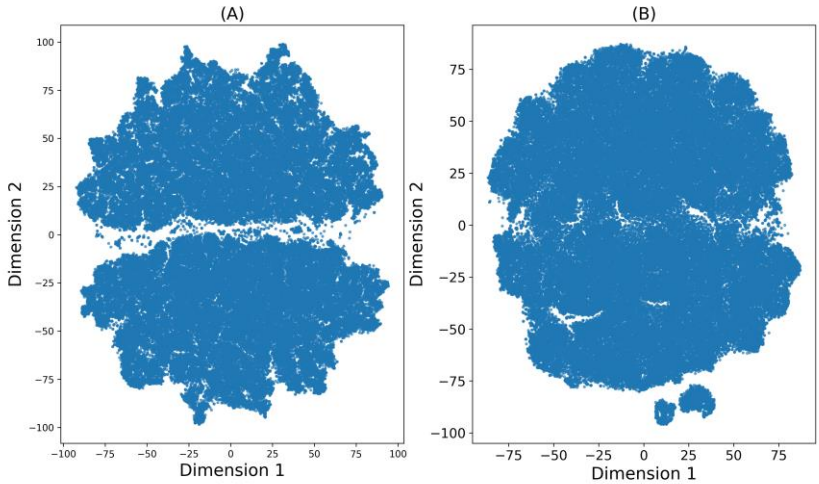


Figure 6.4 t-distributed stochastic neighbor embedding maps of latent dimensions. Maps originating from (A) the latent representation of the noisy signal and (B) from the latent representation of noise. Notice the two major clusters for the noisy signal representation. Furthermore, one can observe the stronger delineation of the fine-structured clusters for the noisy signal representation. The Figure is adapted from Paper IV.

6.3 Comparison of denoising methods

In terms of quantitative denoising performance, it was shown in Paper IV that CLAERD outperformed both PCA and DCAE with a statistically significant difference. The analysis included standard metrics such as MAE, the *peak signal-to-noise ratio*, *structural similarity index measure*, and *coefficient of determination* (R^2), all of which serve to reflect general denoising performance. Moreover, the quantitative analysis described in Paper IV also included metrics focusing on the capability of recovering Z-spectral peaks. This particularly includes the peak recovery score (PRS), which is defined as:

$$PRS = 1 - \frac{|AUC_{GT} - AUC_{denoised}|}{AUC_{GT}}, \tag{6.9}$$

where AUC_{gt} denotes the area under the curve for the GT spectra, and $AUC_{denoised}$ is the corresponding AUC for the denoised spectra. Additionally, ROC analysis was included in the comparison with emphasis on sensitivity and specificity. The analysis revealed that CLAERD provided a statistically significant improvement for all peaks considered in terms of PRS. As for the ROC analysis, an elevated AUC for the ROC curve was observed, and the close-to-perfect sensitivity was similar for

both CLAERD and DCAE, whereas the specificity of CLAERD was elevated for all peaks. A final quantitative metric called the *total variation* (TV) was used to measure the smoothness of the denoised spectra, with lower values implying increased smoothness. An average TV score of approximately 1.80 was observed for all denoisers, which was lower than the value of the noisy spectra, equal to 1.87, as should be expected. Despite the comparable TV scores of the three denoisers, interestingly, DCAE showed a trend of marginally lower TV scores. However, the quantitative results unambiguously indicated an enhancement by CLAERD, and Z-spectra contain contributions from various components, which should thus inherently increase the TV value even in a noise-free case. Consequently, the trend of lower values by DCAE could potentially be explained by over-smoothing.

It should be noted that PCA-based denoising remains simple and fast but is limited by its strong dependency on the choice of PCs for the denoising performance. This can be remedied with cross-validation to find the optimal number of PCs to include. The main limitation remains, however, since the linearity of the transformation makes the distinction between noise and fine structure components in the low variance PCs close to impossible. On the other hand, DCAE leverages enhanced DL-based denoising by using the innovative technique of combining a curriculum-based training strategy with context learning for inference. Both are guided by the deterministic KL-divergence, which was constrained by the PCA reference in the inference setting. Although this increases the approach's stability, the PCA's limitations (particularly the inability to separate noise from fine structured details) will be partially inherited.

This, in part, explains the elevated performance by CLAERD compared to DCAE in terms of the peak recovery metrics. The specificity was shown to be significantly lower using PCA compared to both CLAERD and DCAE. Even if DCAE produced slightly higher specificity values compared to PCA, the approach lagged behind CLAERD due to its KL-based constraint on PCA during inference. It should be noted that an increase in the specificity of detected Z-spectral peaks can add great value. This means that in certain applications, such as APT-weighted imaging for detecting ischemic strokes, tumorous regions, or neurodegenerative diseases where subtle biochemical changes carry high clinical importance, the ability to distinguish true spectral signals from noise ensures accurate diagnosis, treatment planning, and monitoring.

7. Conclusions, reflections, and future perspectives

7.1 Conclusions from the completed work

Despite the many promising applications that take advantage of MT mechanisms, several limitations still need to be overcome to unlock the full potential of its imaging techniques. The work of this thesis revolved around finding ML-based solutions to some of the current limitations affecting MT imaging techniques. As a result, two Z-spectral fitting approaches and one denoising approach were developed and thoroughly evaluated. It was previously explained in Chapter 5.4 that these two fitting approaches differ in algorithmic design. Based on the results presented in Papers I and II, however, it can be concluded that by replacing the conventional backbone of spectral fitting (the LS algorithm) with any of the ML solutions, the sought-after leap towards clinically feasible fitting times can be achieved. Furthermore, based on the analysis described in Chapter 5.4, it can be concluded that the choice between the two ML fitting solutions should be based on their intended use. If robustness is essential, such as when the targeted data reflect high variance or uncertainty, the higher complexity of DNNs should be prioritized. If the aim is flexibility, such as being able to redesign the acquisition protocols or even change the spectral model, the GBDT solution should be leveraged. Moreover, robustness refers to the trained model's ability to become insensitive to common and moderate variations in the data such as noise fluctuations and other perturbations (as explained in Chapter 5.4). Since the ML-fitting approaches are not actual denoisers, more extensive denoising requires a dedicated model such as the one proposed in Paper IV. It is concluded that this denoising model can recover overshadowed Z-spectral components due to its intricate algorithmic design targeting noise in latent space.

On that note, organizing the processing and analytical methods in a pipeline would greatly benefit standardization. Several pipelines have, of course, already been created for CEST data, and to this end, the developed solutions described in this thesis could be integrated in a straightforward manner. In the early steps of analyzing and processing CEST data, denoising might be necessary, and CLAERD can be integrated here. A subsequent step in the pipeline is the B_0-field correction, commonly performed through fitting DS spectra, and sLoFNet can be used for this.

Furthermore, implementing GBDT as a replacement for the LS-backbone for subsequent fitting steps is also straightforward. These additions would be a great step towards increasing efficiency and standardization when analyzing and processing CEST data.

7.2 A critical approach to a brilliant tool

When using AI-based solutions, a good awareness of the tool used is essential since many risks arise along with the great opportunities. An understanding of the application of a developed model and its limitations is thus of great importance. With sLoFNet, for example, the intended use is to fit DS Z-spectra, which adhere to a Lorentzian shape while providing accelerated fitting times and robustness against noise and sampling density. However, knowing this tool's limitations means understanding that this robustness is not infinite. Reducing the sampling density below a certain point increases the fitting error tangibly, as does having noise levels beyond the levels reflected in the invariance of the model (since the model is not an actual denoiser). Furthermore, a non-Lorentzian spectrum would, of course, also result in a very poor fitting.

Lack of awareness of these two aspects is a topic that many AI critics have used as leverage to pinpoint the risks of AI. In the work by Antun et al.,[190] for example, it was shown that DL-models trained to reconstruct MRI brain images from undersampled k-space data resulted in major artifacts when introducing perturbations in the form of either changes in k-space or in the image domain (heart-shapes were drawn into the brain images to reflect unfamiliar structures in the brain). However, it was not clarified in that work that the models used were not trained on these particular variations. Therefore, the poor performance should not be surprising when the training data did not include a remotely close reflection of the variations. This is because an AI model's generalizability and invariance reflect what has been included in the training data. A model that specializes in denoising brain MR images, for example, will have a problem denoising knee MR images, regardless of how well it performs on brains if invariance to anatomical structure had not been included in the training. As an analogy, a skilled car mechanic would have problems fixing a microwave oven unless he had explicitly been trained to do so. The main difference is that an AI model will always give an output when given an input, which necessitates a high degree awareness.

The black-box nature of many AI models makes it even more important to understand their intended applications and limitations. While some insight into their mechanisms is possible by applying statistical and projection-based analyses on the internal representations (as done with CLAERD), the interpretability reduces with the increased complexity of the models. Thus, interpretability represents another

argument for using a lower complexity algorithm whenever possible. For the GBDT used in the Paper II project, for example, interpretability is more straightforward since the decision making can be evaluated by analyzing the resulting trees after training, which is not as easy to do with the DNNs used for sLoFNet.

7.3 Future work

7.3.1 Asymmetrical profiles for Z-spectral components

The topic of using various spectral models for fitting Z-spectral components was discussed in Chapter 4, where it was mentioned that while BMC-based spectral modeling takes into consideration some pool interactions, it is restricted by the computational complexity tied to its closed form numerical solutions and the assumptions made about these. This was further explained to be the reason for the popularity of the multi-pool Lorentzian approach under acquisition conditions with short durations and low power saturation pulses. In Paper II, it was empirically shown in this regard that there are spectral models such as the Voigt profile which, due to its increased degrees of freedom, manages to model the Z-spectral components with significantly improved accuracy, albeit with increased time complexity. By replacing the LS-based backbone of fitting with the GBDT solution, however, clinical feasibility could be possible not only for the Lorentzian but also for any spectral model, since the time complexity depends on the algorithmic complexity instead of the number of parameters. This subsequently opens a new realm of opportunities for exploring several other spectral models for fitting experimental data, particularly asymmetrical ones. It is on this note that a new interesting research question emerges.

Some Z-spectral components have asymmetric peak shapes due partly to asymmetric MTC contributions and the fast-exchanging components causing coalescence with other peaks in their vicinity (such as the DS peak). Two such examples are the myoinositol peak (approximately 0.9 ppm) and the rNOE peak of the aliphatic groups (approximately -3.6 ppm).[1] Both the MTC contribution and the saturation of the mobile compounds' protons depend on the saturation pulse shape, composition, and timing. As such, the shape of the signal is expected to vary depending on the saturation power (and duration). The examples in Figure 7.2 show the pool for the aliphatic groups (rNOE peak) from an in vivo mouse brain acquired at a field strength of 11.7 T and a saturation pulse power of 0.7 μT, fitted using the conventional Lorentzian in Figure 7.2A and using an asymmetrical spectral model in Figure 7.2B (the *split Lorentzian* in this case, although a few other interesting candidates such as the *skewed Voigt* or *Fano resonance* could also be explored). It can be observed that the fitted component has an asymmetrical peak, as further

verified in Table 7.1, which presents a nonzero value for the asymmetry parameter of the split-Lorentzian model. Table 7.1 also includes the fitting statistics, indicating a significant improvement both in terms of R^2 and χ^2 metrics for the asymmetric model. The asymmetry parameter can potentially provide an additional, previously unexplored, quantity for Z-spectral component analysis since it could be reflective of the MTC contribution under carefully designed experimental conditions. One such study design could address the change of the asymmetry parameter as a function of saturation pulse power.

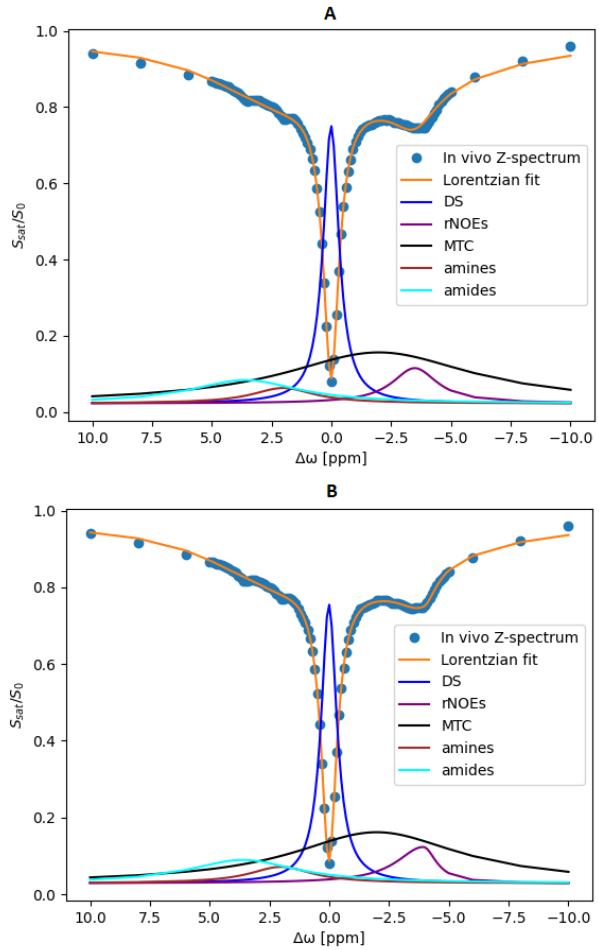


Figure 7.2. Visual comparison of symmetric to asymmetric Z-spectral fitting. Comparing (A) the conventional Lorentzian to (B) an asymmetrical spectral model, namely the split-Lorentzian for fitting the peak for the aliphatic groups at approximately -3.6 ppm. Note that the remaining components are fitted with a conventional Lorentzian in both the top and bottom plot.

Table 7.1 Quantitative comparison of symmetric to asymmetric Z-spectral fitting. The fitting statistics of the Lorentzian compared to the split Lorentzian on the same in vivo spectrum as shown in Figure 1, as well as the value for the asymmetry parameter.

Metric	Lorentzian	Split Lorentzian
χ^2	0.00441	**0.00232**
reduced χ^2	4.65e-05	**2.47e-05**
R^2	0.9984	**0.9992**
Asymmetry [ppm]	0.0	**0.70**

7.3.2 AE-based overcomplete transformation

In data science, various transformations are commonly applied to increase the class-dependent features and thus the interpretability of the data being analyzed. Transformation to higher dimensions means adding features that could provide meaningful class-characteristic features, subsequently streamlining the distinction between classes (see general example in Figure 7.3). Note that classes in the data could refer to signal components, noise, etc.

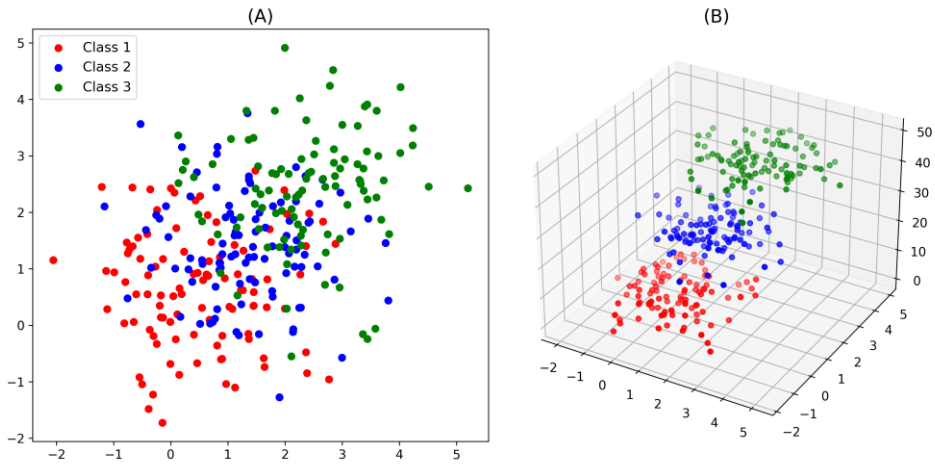


Figure 7.3. Transforming data to higher dimensions. (A) A plot including data from three different classes. Note that the classes are almost indistinguishable based on the plotted features (x and y axes). (B) The same data has been transformed, resulting in the addition of a new feature (dimension), and the distinction is subsequently straightforward.

It should be noted that adding features does not always provide straightforward distinctions between the classes, especially when analyzed with linear methods; instead, it could result in what is referred to as *overcomplete* representations. Therefore, it should be mentioned that the opposite is also performed, i.e., dimensionality reduction, wherein the overcomplete representation of data is reduced to get a more focused representation of the most significant features. Such

a transformation can be performed via PCA[191] as was explained in Chapter 6. In the realms of AI, not surprisingly, dimensionality reduction approaches have also been developed. In particular, AE networks have been applied diligently to transform data and map the most important features in latent space by using lower dimensions in the hidden layers compared to the input layer. The potential of AE to transform data nonlinearly also provides the potential of finding otherwise non-detectable dependencies when relying on linear transformations such as PCA. The common view is that the nonlinear transformations of any DNN are, in fact, the main characteristics that allow patterns to be found in the data that otherwise would not be possible to find with simple solutions; however, this is done in a black-box fashion due to the high complexity of DNNs (which instead rely on learning from data).

Instead of using AEs to reduce dimensionality, future work in this regard could perform transformations to a higher dimension (to increase the number of features as previously explained) by designing AEs with larger-size hidden layers than the size of the input data. To produce meaningful features, however, the AE must be presented with auxiliary input wherein potential, albeit unknown, dependencies exist (e.g., the Z-spectrum, B_0- and B_1-maps, a noise-scan or baseline-scan for dynamic images etc.). The transformation to a higher dimension with the AE would allow these dependencies to be found and connected in new features, which can subsequently be used for easier class distinction. Such class distinctions could, for example, be between signal and noise. Hence, the proposed idea of an *AE-based overcomplete transform* (AEbOT) can be integrated into the developed DL-based denoiser described in Paper IV, thus allowing for denoising with even higher fidelity. Another example is to utilize the AEbOT to target and extract specific Z-spectral components. Yet another application could be to make the AEbOT an integral part of an NN with recurrent modules (e.g., *long-short term memory* modules targeting the sequential nature of dynamic data) or *transformer networks* (targeting long range dependencies) in order to distinguish motion-induced artifacts from the actual signal. The handcrafted AE-based transformation to an "overcomplete representation" effectively means utilizing the black-box nature under controlled conditions, because implicit dependencies will form the new features.

Acknowledgements

Our lives are shaped by the eyes we choose to view our surroundings through. Every individual can, upon request, point out some misfortune in their life. Equally, the same individuals could, upon request, point out some joy in their life. I deeply believe that we could actively choose to be happy and content by shifting focus to the set of eyes viewing joy. While easier said than done, by knowing this, as soon as the eyes of misfortune are being used one can make effort to switch to the eyes of joy. Suddenly, the "misfortune" will turn out to be a blessing in disguise.

Six years ago, I was finishing my master's thesis in the field of synchrotron-based imaging (a choice which had been very hard since I had to put aside MRI which I had resonated with and found a deep interest in, as a result of the comprehensive course with Professor Ronnie Wirestam). As I was at the end of the master's project, my supervisor at the time got an unexpected change of circumstances forcing a move abroad. This meant a full stop to the potential PhD position we had spoken about. A blessing in disguise indeed! The following 6 months after graduation would be spent unsuccessfully applying to various positions. Then one evening, while visiting a friend, I got an email from non-other than Professor Linda Knutsson – someone who until then was known to me as "the happy pulse-sequence lady from the MRI-course". She mentioned that she had been informed by Ronnie that "an ambitious student" was sitting idle and that she potentially had something to offer. And so, my journey started, including a reconciliation with my previous love interest -MRI- but with the beautiful addition of CEST, and even more beautiful AI.

So, I'd like to start off by expressing my outermost gratitude to Ronnie Wirestam, first and foremost for being and remaining a mentor, but also for keeping me in mind and for being a great co-supervisor. In the correct chronological order, I'd like to extend my outermost gratitude to "the happy pulse-sequence lady form the MRI-course" who later grew to become my work-mom. She has provided me with deep trust, support, and truly contributed to my professional and personal growth. To Professor Peter van Zijl for encouragement, humbleness, putting up with my endless questions and his inspiring love for science. I would also like to express my gratitude to Dr. Anina Seidemo and Patrick Lehmann, with whom I've shared many fruitful conversations, Dr. Arthur Chakwizira for being an inspiring and kind-hearted brother, Professor Pia Sundgren for supporting my PhD studies, Associate Professor Hye-Young Heo for mentorship and Associate Professor Nirbhay Yadav for being a supporting and inspiring mentor and friend. I am grateful to the people of the MRI

department and to everyone at MSF for providing inclusive environments. Special thanks to Professor Michael Ljungberg who has been a supportive co-supervisor, Professor Katarina Sjögreen for encouragement, Professor Crister Ceberg for his kindness and unwavering willingness to help, Titti Owman for securing the venue for the defense, Liselotte Eriksson for administrative assistance and Associate Professor Mikael Gunnarson for his kindness and encouragement. I also want to thank all my co-authors and everyone who has contributed to and supported my learning. Special thanks to Eva Lundin, Erika Malmqvist, Olof Hylén and the kind group I was privileged to study the master's years with.

Finally, to the people that I love, you are an endless source of support, and I could not thank you enough even if I wrote another thesis especially for that. Most of all, I thank God for the infinite blessings, and particularly the blessing of a curious and critical mind. It is truly the key to unveiling and understanding the truth.

الحمد لله رب العالمين

References

1. van Zijl PCM, Lam WW, Xu J, Knutsson L, Stainsz GJ. Magnetization transfer contrast and chemical exchange saturation transfer MRI: Features and analysis of the field-dependent saturation spectrum. *Neuroimage*. 2018;168:222-241. doi: 10.1016/j.neuroimage.2017.04.045
2. Henkelman RM, Stanisz GJ, Graham SJ. Magnetization transfer in MRI: A review. *NMR Biomed*. 2001;14(2):57-64. doi:10.1002/nbm.683
3. Kogan F, Hariharan H, Reddy R. Chemical exchange saturation transfer (CEST) imaging: Description of technique and potential clinical applications. *Curr Radiol Rep*. 2013;1(2):102-114. doi:10.1007/s40134-013-0010-3
4. van Zijl PCM, Yadav NN. Chemical exchange saturation transfer (CEST): What is in a name and what isn't? *Magn Reson Med*. 2011;65(4):927-948. doi:10.1002/mrm.22761
5. Ward KM, Aletras AH, Balaban RS. A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J Magn Reson*. 2000;143(1):79-87. doi:10.1006/jmre.1999.1956
6. Hancu I, Dixon WT, Woods M, Vinogradov E, Sherry AD, Lenkinski RE. CEST and PARACEST MR contrast agents. *Acta Radiol*. 2010;51(8):910-923. doi:10.3109/02841851.2010.502126
7. Msayib Y, Harston GWJ, Tee YK, et al. Quantitative CEST imaging of amide proton transfer in acute ischaemic stroke. *Neuroimage Clin*. 2019;23:101833. doi:10.1016/j.nicl.2019.101833
8. Yu L, Chen Y, Chen M, et al. Amide proton transfer MRI signal as a surrogate biomarker of ischemic stroke recovery in patients with supportive treatment. *Front Neurol*. 2019;10:104. doi:10.3389/fneur.2019.00104
9. Zhou J. Amide proton transfer imaging of the human brain. *Methods Mol Biol*. 2011;711:227-237. doi:10.1007/978-1-61737-992-5_10
10. Zhou J, Lal B, Wilson DA, Laterra J, van Zijl PCM. Amide proton transfer (APT) contrast for imaging of brain tumors. *Magn Reson Med*. 2003;50(6):1120-1126. doi:10.1002/mrm.10651

11. Consolino L, Anemone A, Capozza M, et al. Non-invasive investigation of tumor metabolism and acidosis by MRI-CEST imaging. *Front Oncol*. 2020;10:161. doi:10.3389/fonc.2020.00161

12. Huang J, Chen Z, Park SW, Lai JHC, Chan KWY. Molecular imaging of brain tumors and drug delivery using CEST MRI: Promises and challenges. *Pharmaceutics*. 2022;14(2). doi:10.3390/pharmaceutics14020451

13. Loi L, Zimmermann F, Goerke S, et al. Relaxation-compensated CEST (chemical exchange saturation transfer) imaging in breast cancer diagnostics at 7T. *Eur J Radiol*. 2020;129:109068. doi:10.1016/j.ejrad.2020.109068

14. Evans VS, Torrealdea F, Rega M, et al. Optimization and repeatability of multipool chemical exchange saturation transfer MRI of the prostate at 3.0 T. *J Magn Reson Imaging*. 2019;50(4):1238-1250. doi:10.1002/jmri.26690

15. Tang Y, Xiao G, Shen Z, et al. Noninvasive detection of extracellular pH in human benign and malignant liver tumors using CEST MRI. *Front Oncol*. 2020;10:578985. doi:10.3389/fonc.2020.578985

16. Jones KM, Pollard AC, Pagel MD. Clinical applications of chemical exchange saturation transfer (CEST) MRI. *J Magn Reson Imaging*. 2018;47(1):11-27. doi:10.1002/jmri.25838

17. Zhao R, Luk W, Niu X, Shi H, Wang H. Hardware acceleration for machine learning. In: *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. 2017:645-650. doi:10.1109/ISVLSI.2017.127

18. Strigl D, Kofler K, Podlipnig S. Performance and scalability of GPU-based convolutional neural networks. In: *2010 18th Euromicro Conference on Parallel, Distributed and Network-Based Processing*. 2010:324. doi:10.1109/PDP.2010.43

19. Berggren K, Xia Q, Likharev KK, et al. Roadmap on emerging hardware and technology for machine learning. *Nanotechnology*. 2020;32(1):012002. doi:10.1088/1361-6528/aba70f

20. Welser J, Pitera JW, Goldberg C. Future computing hardware for AI. In: *2018 IEEE International Electron Devices Meeting (IEDM)*. 2018:1.3.1-1.3.6. doi:10.1109/IEDM.2018.8614482

21. Demigha S. The impact of big data on AI. In: *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*. 2020:1395-1400. doi:10.1109/CSCI51800.2020.00259

22. Jain A, Patel H, Nagalapatti L, et al. Overview and importance of data quality for machine learning tasks. In: *Proceedings of the 26th ACM*

SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '20. Association for Computing Machinery; 2020:3561-3562. doi:10.1145/3394486.3406477

23. Growth in AI and robotics research accelerates. *Nature News*. Published October 12, 2022. Accessed September 4, 2025. https://www.nature.com/articles/d41586-022-03210-9
24. Glang F, Deshmane A, Prokudin S, et al. DeepCEST 3T: Robust MRI parameter determination and uncertainty quantification with neural networks—application to CEST imaging of the human brain at 3T. *Magn Reson Med*. 2020;84(1):450-466. doi:10.1002/mrm.28117
25. Pemmasani Prabakaran RS, Park SW, Lai JHC, et al. Deep-learning-based super-resolution for accelerating chemical exchange saturation transfer MRI. *NMR Biomed*. 2024;37(8):e5130. doi:10.1002/nbm.5130
26. Breitling J, Deshmane A, Goerke S, et al. Adaptive denoising for chemical exchange saturation transfer MR imaging. *NMR Biomed*. 2019;32(11):e4133. doi:10.1002/nbm.4133
27. Chen H, Chen X, Lin L, et al. Learned spatiotemporal correlation priors for CEST image denoising using incorporated global-spectral convolution neural network. *Magn Reson Med*. 2023;90(5):2071-2088. doi:10.1002/mrm.29763
28. Kurmi Y, Viswanathan M, Zu Z. Enhancing SNR in CEST imaging: A deep learning approach with a denoising convolutional autoencoder. *Magn Reson Med*. 2024;92(6):2404-2419. doi:10.1002/mrm.30228
29. Bloch F. Nuclear induction. *Phys Rev*. 1946;70(7-8):460-474. doi:10.1103/PhysRev.70.460
30. Purcell EM, Torrey HC, Pound RV. Resonance absorption by nuclear magnetic moments in a solid. *Phys Rev*. 1946;69(1-2):37-38. doi:10.1103/PhysRev.69.37
31. Vlaardingerbroek MT, Luiten A, Boer JA, Knoet F. *Magnetic Resonance Imaging: Theory and Practice*. Springer; 2003.
32. Bjørnerud A. *The Physics of Magnetic Resonance Imaging*. Oslo, Norway: Department of Physics, University of Oslo; 2008.
33. Levitt MH. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2nd ed. Wiley; 2008.
34. Landau LD, Lifshitz EM. *Statistical Physics*. Vol 5. 3rd ed. Elsevier; 2013.

35. U.S. Food and Drug Administration. FDA clears first 7T magnetic resonance imaging device. Published October 12, 2017. Accessed September 4, 2025. Available from FDA

36. Griffiths DJ. *Introduction to Electrodynamics*. 4th ed. Boston, MA: Pearson; 2013.

37. Sijbers J, den Dekker AJ, Scheunders P, Van Dyck D. Maximum-likelihood estimation of Rician distribution parameters. *IEEE Trans Med Imaging*. 1998;17(3):357-361. doi:10.1109/42.712125

38. Cárdenas-Blanco A, Tejos C, Irarrazaval P, Cameron I. Noise in magnitude magnetic resonance images. *Concepts Magn Reson Part A*. 2008;32A(6):409-416. doi:10.1002/cmr.a.20124

39. Johnson JB. Thermal agitation of electricity in conductors. *Phys Rev*. 1928;32(1):97-109. doi:10.1103/PhysRev.32.97

40. Nyquist H. Thermal agitation of electric charge in conductors. *Phys Rev*. 1928;32(1):110-113. doi:10.1103/PhysRev.32.110

41. Redpath TW. Signal-to-noise ratio in MRI. *Br J Radiol*. 1998;71(847):704-707. doi:10.1259/bjr.71.847.9771379

42. Voss RF, Clarke J. Flicker (1/f) noise: equilibrium temperature and resistance fluctuations. *Phys Rev B*. 1976;13(2):556-573. doi:10.1103/PhysRevB.13.556

43. den Dekker AJ, Sijbers J. Data distributions in magnetic resonance images: a review. *Phys Med*. 2014;30(7):725-741. doi:10.1016/j.ejmp.2014.05.002

44. Chaudhari A, Kulkarni J. Noise estimation in single coil MR images. *Biomed Eng Adv*. 2021;2:100017. doi:10.1016/j.bea.2021.100017

45. Aja-Fernández S, Tristán-Vega A, Alberola-López C. Noise estimation in single- and multiple-coil magnetic resonance data based on statistical models. *Magn Reson Imaging*. 2009;27(10):1397-1409. doi:10.1016/j.mri.2009.05.025

46. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med*. 1995;34(6):910-914. doi:10.1002/mrm.1910340618

47. Wink AM, Roerdink JBTM. BOLD noise assumptions in fMRI. *Int J Biomed Imaging*. 2006;2006:12014. doi:10.1155/IJBI/2006/12014

48. Marintchev A, Frueh D, Wagner G. NMR methods for studying protein-protein interactions involved in translation initiation. *Methods Enzymol*. 2007;430:283-331. doi:10.1016/S0076-6879(07)30012-8

49. Scalar Coupling. In: *Fundamentals of Protein NMR Spectroscopy*. Springer Netherlands; 2006:135-151. doi:10.1007/1-4020-3500-4_7

50. Huenges M, Kessler H. Nuclear magnetic resonance spectroscopy | structural chemistry of peptides. In: Worsfold P, Poole C, Townshend A, Miró M, eds. *Encyclopedia of Analytical Science*. 3rd ed. Academic Press; 2017:438-452. doi:10.1016/B978-0-08-101983-2.00295-4

51. Zhou J, Wilson DA, Sun PZ, Klaus JA, Van Zijl PCM. Quantitative description of proton exchange processes between water and endogenous and exogenous agents for WEX, CEST, and APT experiments. *Magn Reson Med*. 2004;51(5):945-952. doi:10.1002/mrm.20048

52. Jin T, Autio J, Obata T, Kim SG. Spin-locking versus chemical exchange saturation transfer MRI for investigating chemical exchange process between water and labile metabolite protons. *Magn Reson Med*. 2011;65(5):1448-1460. doi:10.1002/mrm.22721

53. Davis DG, Perlman ME, London RE. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1 rho and T2 (CPMG) methods. *J Magn Reson B*. 1994;104(3):266-275. doi:10.1006/jmrb.1994.1084

54. Fischer MWF, Majumdar A, Zuiderweg ERP. Protein NMR relaxation: theory, applications and outlook. *Prog Nucl Magn Reson Spectrosc*. 1998;33(3-4):207-272. doi:10.1016/S0079-6565(98)00023-5

55. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*. 1982;104(17):4546-4559. doi:10.1021/ja00381a009

56. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc*. 1982;104(17):4559-4570. doi:10.1021/ja00381a010

57. Solomon I. Relaxation processes in a system of two spins. *Phys Rev*. 1955;99(2):559-565. doi:10.1103/PhysRev.99.559

58. Zhou Y, Bie C, van Zijl PCM, Yadav NN. The relayed nuclear Overhauser effect in magnetization transfer and chemical exchange saturation transfer MRI. *NMR Biomed*. 2023;36(6):e4778. doi:10.1002/nbm.4778

59. Edzes HT, Samulski ET. Cross relaxation and spin diffusion in the proton NMR of hydrated collagen. *Nature*. 1977;265(5594):521-523. doi:10.1038/265521a0

60. Wolff SD, Balaban RS. Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo. *Magn Reson Med*. 1989;10(1):135-144. doi:10.1002/mrm.1910100113

61. Liepinsh E, Otting G. Proton exchange rates from amino acid side chains—implications for image contrast. *Magn Reson Med*. 1996;35(1):30-42. doi:10.1002/mrm.1910350106

62. Vinogradov E, Keupp J, Dimitrov IE, Seiler S, Pedrosa I. CEST-MRI for body oncologic imaging: are we there yet? *NMR Biomed*. 2023;36(6):e4906. doi:10.1002/nbm.4906

63. Cember ATJ, Nanga RPR, Reddy R. Glutamate-weighted CEST (gluCEST) imaging for mapping neurometabolism: an update on the state of the art and emerging findings from in vivo applications. *NMR Biomed*. 2023;36(6):e4780. doi:10.1002/nbm.4780

64. Knutsson L, Xu X, van Zijl PCM, Chan KWY. Imaging of sugar-based contrast agents using their hydroxyl proton exchange properties. *NMR Biomed*. 2023;36(6):e4784. doi:10.1002/nbm.4784

65. Igarashi T, Kim H, Sun PZ. Detection of tissue pH with quantitative chemical exchange saturation transfer magnetic resonance imaging. *NMR Biomed*. 2023;36(6):e4711. doi:10.1002/nbm.4711

66. McConnell HM. Reaction rates by nuclear magnetic resonance. *J Chem Phys*. 1958;28(3):430-431. doi:10.1063/1.1744152

67. Zhou J, van Zijl PCM. Chemical exchange saturation transfer imaging and spectroscopy. *Prog Nucl Magn Reson Spectrosc*. 2006;48(2):109-136. doi:10.1016/j.pnmrs.2006.01.001

68. Wu XZ, Listinsky JJ. Effects of transverse cross relaxation on magnetization transfer. *J Magn Reson B*. 1994;105(1):73-76. doi:10.1006/jmrb.1994.1103

69. Mulkern RV, Williams ML. The general solution to the Bloch equation with constant rf and relaxation terms: application to saturation and slice selection. *Med Phys*. 1993;20(1):5-13. doi:10.1118/1.597063

70. Kim M, Gillen J, Landman BA, Zhou J, van Zijl PCM. Water saturation shift referencing (WASSR) for chemical exchange saturation transfer (CEST) experiments. *Magn Reson Med*. 2009;61(6):1441-1450. doi:10.1002/mrm.21873

71. Liu G, Qin Q, Chan KWY, et al. Non-invasive temperature mapping using temperature-responsive water saturation shift referencing (T-WASSR) MRI. *NMR Biomed*. 2014;27(3):320-331. doi:10.1002/nbm.3066

72. Müller-Lutz A, Matuschke F, Schleich C, et al. Improvement of water saturation shift referencing by sequence and analysis optimization to enhance chemical exchange saturation transfer imaging. *Magn Reson Imaging*. 2016;34(6):771-778. doi:10.1016/j.mri.2016.03.013

73. Smith SA, Bulte JWM, van Zijl PCM. Direct saturation MRI: theory and application to imaging brain iron. *Magn Reson Med*. 2009;62(2):384-393. doi:10.1002/mrm.21980

74. Jones CK, Huang A, Xu J, et al. Nuclear Overhauser enhancement (NOE) imaging in the human brain at 7T. *Neuroimage*. 2013;77:114-124. doi:10.1016/j.neuroimage.2013.03.047

75. Lim IAL, Li X, Jones CK, Farrell JAD, Vikram DS, van Zijl PCM. Quantitative magnetic susceptibility mapping without phase unwrapping using WASSR. *Neuroimage*. 2014;86:265-279. doi:10.1016/j.neuroimage.2013.09.072

76. Ying X. An overview of overfitting and its solutions. *J Phys Conf Ser*. 2019;1168(2):022022. doi:10.1088/1742-6596/1168/2/022022

77. Xu X, Yadav NN, Knutsson L, et al. Dynamic glucose-enhanced (DGE) MRI: translation to human scanning and first results in glioma patients. *Tomogr*. 2015;1(2):105-114. doi:10.18383/j.tom.2015.00175

78. Xu X, Chan KWY, Knutsson L, et al. Dynamic glucose-enhanced (DGE) MRI for combined imaging of blood-brain barrier breakdown and increased blood volume in brain cancer. *Magn Reson Med*. 2015;74(6):1556-1563. doi:10.1002/mrm.25995

79. Wu Y, Derks SHAE, Wood TC, et al. Improved postprocessing of dynamic glucose-enhanced CEST MRI for imaging brain metastases at 3 T. *Eur Radiol Exp*. 2023;7(1):78. doi:10.1186/s41747-023-00390-5

80. Seidemo A, Lehmann PM, Rydhög A, et al. Towards robust glucose chemical exchange saturation transfer imaging in humans at 3 T: arterial input function measurements and the effects of infusion time. *NMR Biomed*. 2022;35(2):e4624. doi:10.1002/nbm.4624

81. Knutsson L, Yadav NN, Mohammed Ali S, et al. Dynamic glucose enhanced imaging using direct water saturation. *Magn Reson Med*. 2025;94:15-27. doi:10.1002/mrm.30447

82. Kraus J, Sarkar S, Quinn CM, Polenova T. Chapter Two – Solid-state NMR spectroscopy of microcrystalline proteins. In: Webb GA, ed. *Annual Reports on NMR Spectroscopy*. Vol 102. Academic Press; 2021:81-151. doi:10.1016/bs.arnmr.2020.10.002

83. Zhang XY, Wang F, Li H, et al. Accuracy in the quantification of chemical exchange saturation transfer (CEST) and relayed nuclear Overhauser enhancement (rNOE) saturation transfer effects. *NMR Biomed*. 2017;30(7). doi:10.1002/nbm.3716

84. Hua J, Jones CK, Blakeley J, Smith SA, van Zijl PCM, Zhou J. Quantitative description of the asymmetry in magnetization transfer effects around the water resonance in the human brain. *Magn Reson Med*. 2007;58(4):786-793. doi:10.1002/mrm.21387

85. Swanson SD. Protein mediated magnetic coupling between lactate and water protons. *J Magn Reson*. 1998;135(1):248-255. doi:10.1006/jmre.1998.1535

86. Swanson SD, Malyarenko DI, Fabiilli ML, Welsh RC, Nielsen JF, Srinivasan A. Molecular, dynamic, and structural origin of inhomogeneous magnetization transfer in lipid membranes. *Magn Reson Med*. 2017;77(3):1318-1328. doi:10.1002/mrm.26210

87. Xu J, Chan KWY, Xu X, Yadav N, Liu G, van Zijl PCM. On-resonance variable delay multipulse scheme for imaging of fast-exchanging protons and semisolid macromolecules. *Magn Reson Med*. 2017;77(2):730-739. doi:10.1002/mrm.26165

88. Balaban RS, Ceckler TL. Magnetization transfer contrast in magnetic resonance imaging. *Magn Reson Q*. 1992;8(2):116-137.

89. Jiang X, van Gelderen P, Duyn JH. Spectral characteristics of semisolid protons in human brain white matter at 7 T. *Magn Reson Med*. 2017;78(5):1950-1958. doi:10.1002/mrm.26594

90. Yarnykh VL. Pulsed Z-spectroscopic imaging of cross-relaxation parameters in tissues for human MRI: theory and clinical applications. *Magn Reson Med*. 2002;47(5):929-939. doi:10.1002/mrm.10120

91. Sled JG, Pike GB. Quantitative imaging of magnetization transfer exchange and relaxation properties in vivo using MRI. *Magn Reson Med*. 2001;46(5):923-931. doi:10.1002/mrm.1278

92. Ramani A, Dalton C, Miller DH, Tofts PS, Barker GJ. Precise estimate of fundamental in-vivo MT parameters in human brain in clinically feasible times. *Magn Reson Imaging*. 2002;20(10):721-731. doi:10.1016/s0730-725x(02)00598-2

93. Thomas JD. Magnetization transfer in magnetic resonance imaging. *Radiol Technol*. 1996;67(4):297-306.

94. Alizadeh A, Dyck SM, Karimi-Abdolrezaee S. Myelin damage and repair in pathologic CNS: challenges and prospects. *Front Mol Neurosci*. 2015;8:35. doi:10.3389/fnmol.2015.00035

95. Love S. Demyelinating diseases. *J Clin Pathol*. 2006;59(11):1151-1159. doi:10.1136/jcp.2005.031195

96. Popescu BFG, Lucchinetti CF. Pathology of demyelinating diseases. *Annu Rev Pathol*. 2012;7:185-217. doi:10.1146/annurev-pathol-011811-132443

97. Gotor V, Gotor-Fernández V, Busto E. 7.6 Hydrolysis and reverse hydrolysis: hydrolysis and formation of amides. In: Carreira EM, Yamamoto H, eds. *Comprehensive Chirality*. Elsevier; 2012:101-121. doi:10.1016/B978-0-08-095167-6.00707-2

98. Elvidge JA, Jones JR, O'Brien C, Evans EA, Sheppard HC. Base-catalyzed hydrogen exchange. In: Katritzky AR, Boulton AJ, eds. *Advances in Heterocyclic Chemistry*. Vol 16. Academic Press; 1974:1-31. doi:10.1016/S0065-2725(08)60458-4

99. Zhou J, Payen JF, Wilson DA, Traystman RJ, van Zijl PCM. Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI. *Nat Med*. 2003;9(8):1085-1090. doi:10.1038/nm907

100. Yan K, Fu Z, Yang C, et al. Assessing amide proton transfer (APT) MRI contrast origins in 9L gliosarcoma in the rat brain using proteomic analysis. *Mol Imaging Biol*. 2015;17(4):479-487. doi:10.1007/s11307-015-0828-6

101. Zhao X, Wen Z, Huang F, et al. Saturation power dependence of amide proton transfer image contrasts in human brain tumors and strokes at 3 T. *Magn Reson Med*. 2011;66(4):1033-1041. doi:10.1002/mrm.22891

102. Tee YK, Harston GWJ, Blockley N, et al. Comparing different analysis methods for quantifying the MRI amide proton transfer (APT) effect in hyperacute stroke patients. *NMR Biomed*. 2014;27(9):1019-1029. doi:10.1002/nbm.3147

103. Harston GWJ, Tee YK, Blockley N, et al. Identifying the ischaemic penumbra using pH-weighted magnetic resonance imaging. *Brain*. 2015;138(Pt 1):36-42. doi:10.1093/brain/awu374

104. Heo HY, Zhang Y, Burton TM, et al. Improving the detection sensitivity of pH-weighted amide proton transfer MRI in acute stroke patients using extrapolated semisolid magnetization transfer reference signals. *Magn Reson Med*. 2017;78(3):871-880. doi:10.1002/mrm.26799

105. Song G, Li C, Luo X, et al. Evolution of cerebral ischemia assessed by amide proton transfer-weighted MRI. *Front Neurol*. 2017;8:67. doi:10.3389/fneur.2017.00067

106. Lin G, Zhuang C, Shen Z, et al. APT weighted MRI as an effective imaging protocol to predict clinical outcome after acute ischemic stroke. *Front Neurol*. 2018;9. doi:10.3389/fneur.2018.00901

107. Foo LS, Harston G, Mehndiratta A, et al. Clinical translation of amide proton transfer (APT) MRI for ischemic stroke: a systematic review (2003–2020). *Quant Imaging Med Surg*. 2021;11(8):3797-3811. doi:10.21037/qims-20-1339

108. Wang M, Hong X, Chang CF, et al. Simultaneous detection and separation of hyperacute intracerebral hemorrhage and cerebral ischemia using amide proton transfer MRI. *Magn Reson Med*. 2015;74(1):42-50. doi:10.1002/mrm.25690

109. Dula AN, Asche EM, Landman BA, et al. Development of chemical exchange saturation transfer at 7 T. *Magn Reson Med*. 2011;66(3):831-838. doi:10.1002/mrm.22862

110. Dula AN, Pawate S, Dethrage LM, et al. Chemical exchange saturation transfer of the cervical spinal cord at 7 T. *NMR Biomed*. 2016;29(9):1249-1257. doi:10.1002/nbm.3581

111. Wells JA, O'Callaghan JM, Holmes HE, et al. In vivo imaging of tau pathology using multi-parametric quantitative MRI. *Neuroimage*. 2015;111:369-378. doi:10.1016/j.neuroimage.2015.02.023

112. Donahue MJ, Donahue PCM, Rane S, et al. Assessment of lymphatic impairment and interstitial protein accumulation in patients with breast cancer treatment-related lymphedema using CEST MRI. *Magn Reson Med*. 2016;75(1):345-355. doi:10.1002/mrm.25649

113. Jiang S, Eberhart CG, Zhang Y, et al. Amide proton transfer-weighted magnetic resonance image-guided stereotactic biopsy in patients with newly diagnosed gliomas. *Eur J Cancer*. 2017;83:9-18. doi:10.1016/j.ejca.2017.06.009

114. Zhang H, Yong X, Ma X, et al. Differentiation of low- and high-grade pediatric gliomas with amide proton transfer imaging: added value beyond quantitative relaxation times. *Eur Radiol*. 2021;31(12):9110-9119. doi:10.1007/s00330-021-08039-w

115. Zou T, Yu H, Jiang C, et al. Differentiating the histologic grades of gliomas preoperatively using amide proton transfer-weighted (APTW) and

intravoxel incoherent motion MRI. *NMR Biomed*. 2018;31(1):e3850. doi:10.1002/nbm.3850

116. Zhang Z, Wang K, Park S, et al. The exchange rate of creatine CEST in mouse brain. *Magn Reson Med*. 2023;90(2):373-384. doi:10.1002/mrm.29662

117. Cai K, Haris M, Singh A, et al. Magnetic resonance imaging of glutamate. *Nat Med*. 2012;18(2):302-306. doi:10.1038/nm.2615

118. Wang K, Huang J, Ju L, et al. Creatine mapping of the brain at 3T by CEST MRI. *Magn Reson Med*. 2024;91(1):51-60. doi:10.1002/mrm.29876

119. Chen L, Schär M, Chan KWY, et al. In vivo imaging of phosphocreatine with artificial neural networks. *Nat Commun*. 2020;11(1):1072. doi:10.1038/s41467-020-14874-0

120. Jia Y, Chen Y, Geng K, et al. Glutamate chemical exchange saturation transfer (GluCEST) magnetic resonance imaging in pre-clinical and clinical applications for encephalitis. *Front Neurosci*. 2020;14:750. doi:10.3389/fnins.2020.00750

121. Harris RJ, Cloughesy TF, Liau LM, et al. pH-weighted molecular imaging of gliomas using amine chemical exchange saturation transfer MRI. *Neuro-Oncol*. 2015;17(11):1514-1524. doi:10.1093/neuonc/nov106

122. Harris RJ, Cloughesy TF, Liau LM, et al. Simulation, phantom validation, and clinical evaluation of fast pH-weighted molecular imaging using amine chemical exchange saturation transfer echo planar imaging (CEST-EPI) in glioma at 3 T. *NMR Biomed*. 2016;29(11):1563-1576. doi:10.1002/nbm.3611

123. Marathe K, McVicar N, Li A, Bellyou M, Meakin S, Bartha R. Topiramate induces acute intracellular acidification in glioblastoma. *J Neurooncol*. 2016;130(3):465-472. doi:10.1007/s11060-016-2258-y

124. Kogan F, Haris M, Debrosse C, et al. In vivo chemical exchange saturation transfer imaging of creatine (CrCEST) in skeletal muscle at 3T. *J Magn Reson Imaging*. 2014;40(3):596-602. doi:10.1002/jmri.24412

125. Kogan F, Haris M, Singh A, et al. Method for high-resolution imaging of creatine in vivo using chemical exchange saturation transfer. *Magn Reson Med*. 2014;71(1):164-172. doi:10.1002/mrm.24641

126. Rerich E, Zaiss M, Korzowski A, Ladd ME, Bachert P. Relaxation-compensated CEST-MRI at 7 T for mapping of creatine content and pH—preliminary application in human muscle tissue in vivo. *NMR Biomed*. 2015;28(11):1402-1412. doi:10.1002/nbm.3367

127. Pischel I, Gastner T. Creatine – its chemical synthesis, chemistry, and legal status. In: Salomons GS, Wyss M, eds. *Creatine and Creatine Kinase in Health and Disease*. Springer Netherlands; 2007:291-307. doi:10.1007/978-1-4020-6486-9_15

128. Xu J, Chung JJ, Jin T. Chemical exchange saturation transfer imaging of creatine, phosphocreatine, and protein arginine residue in tissues. *NMR Biomed*. 2023;36(6):e4671. doi:10.1002/nbm.4671

129. Carey FA. *Organic Chemistry*. McGraw-Hill Higher Education; 2000.

130. Haris M, Cai K, Singh A, Hariharan H, Reddy R. In vivo mapping of brain myo-inositol. NeuroImage. 2011;54(3):2079-2085. doi:10.1016/j.neuroimage.2010.10.017

131. Li X, Yang Y, Zhang B, et al. Lactate metabolism in human health and disease. Signal Transduct Target Ther. 2022;7(1):305. doi:10.1038/s41392-022-01151-3

132. DeBrosse C, Nanga RPR, Bagga P, et al. Lactate chemical exchange saturation transfer (LATEST) imaging in vivo: a biomarker for LDH activity. Sci Rep. 2016;6(1):19517. doi:10.1038/srep19517

133. Boyd PS, Breitling J, Zimmermann F, et al. Dynamic glucose-enhanced (DGE) MRI in the human brain at 7 T with reduced motion-induced artifacts based on quantitative R1ρ mapping. Magn Reson Med. 2020;84(1):182-191. doi:10.1002/mrm.28112

134. Herz K, Lindig T, Deshmane A, et al. T1ρ-based dynamic glucose-enhanced (DGEρ) MRI at 3 T: method development and early clinical experience in the human brain. Magn Reson Med. 2019;82(5):1832-1847. doi:10.1002/mrm.27857

135. Xu X, Sehgal AA, Yadav NN, et al. D-glucose weighted chemical exchange saturation transfer (glucoCEST)-based dynamic glucose enhanced (DGE) MRI at 3T: early experience in healthy volunteers and brain tumor patients. Magn Reson Med. 2020;84(1):247-262. doi:10.1002/mrm.28124

136. Yadav NN, Xu J, Bar-Shir A, et al. Natural D-glucose as a biodegradable MRI relaxation agent. Magn Reson Med. 2014;72(3):823-828. doi:10.1002/mrm.25329

137. Zaiss M, Anemone A, Goerke S, et al. Quantification of hydroxyl exchange of D-glucose at physiological conditions for optimization of glucoCEST MRI at 3, 7 and 9.4 Tesla. NMR Biomed. 2019;32(9):e4113. doi:10.1002/nbm.4113

138. Zhang XY, Wang F, Afzal A, et al. A new NOE-mediated MT signal at around −1.6 ppm for detecting ischemic stroke in rat brain. Magn Reson Imaging. 2016;34(8):1100-1106. doi:10.1016/j.mri.2016.05.002

139. Zhang XY, Wang F, Jin T, et al. MR imaging of a novel NOE-mediated magnetization transfer with water in rat brain at 9.4 T. Magn Reson Med. 2017;78(2):588-597. doi:10.1002/mrm.26396

140. Zaiss M, Windschuh J, Paech D, et al. Relaxation-compensated CEST-MRI of the human brain at 7T: unbiased insight into NOE and amide signal changes in human glioblastoma. NeuroImage. 2015;112:180-188. doi:10.1016/j.neuroimage.2015.02.040

141. Heo HY, Jones CK, Hua J, et al. Whole-brain amide proton transfer (APT) and nuclear Overhauser enhancement (NOE) imaging in glioma patients using low-power steady-state pulsed chemical exchange saturation transfer (CEST) imaging at 7T. J Magn Reson Imaging. 2016;44(1):41-50. doi:10.1002/jmri.25108

142. Vinogradov E, Sherry AD, Lenkinski RE. CEST: from basic principles to applications, challenges and opportunities. J Magn Reson. 2013;229:155-172. doi:10.1016/j.jmr.2012.11.024

143. Guivel-Scharen V, Sinnwell T, Wolff SD, Balaban RS. Detection of proton chemical exchange between metabolites and water in biological tissues. J Magn Reson. 1998;133(1):36-45. doi:10.1006/jmre.1998.1440

144. Heo HY, Tee YK, Harston G, Leigh R, Chappell MA. Amide proton transfer imaging in stroke. NMR Biomed. 2023;36(6):e4734. doi:10.1002/nbm.4734

145. Sun PZ, Cheung JS, Wang E, Lo EH. Association between pH-weighted endogenous amide proton chemical exchange saturation transfer MRI and tissue lactic acidosis during acute ischemic stroke. J Cereb Blood Flow Metab. 2011;31(8):1743-1750. doi:10.1038/jcbfm.2011.23

146. Tietze A, Blicher J, Mikkelsen IK, et al. Assessment of ischemic penumbra in patients with hyperacute stroke using amide proton transfer (APT) chemical exchange saturation transfer (CEST) MRI. NMR Biomed. 2014;27(2):163-174. doi:10.1002/nbm.3048

147. Zaiss M, Bachert P. Exchange-dependent relaxation in the rotating frame for slow and intermediate exchange – modeling off-resonant spin-lock and chemical exchange saturation transfer. NMR Biomed. 2013;26(5):507-518. doi:10.1002/nbm.2887

148. Zaiss M, Xu J, Goerke S, et al. Inverse Z-spectrum analysis for spillover-, MT-, and T1-corrected steady-state pulsed CEST-MRI – application to

pH-weighted MRI of acute stroke. NMR Biomed. 2014;27(3):240-252. doi:10.1002/nbm.3054

149. Zaiss M, Zu Z, Xu J, et al. A combined analytical solution for chemical exchange saturation transfer and semi-solid magnetization transfer. NMR Biomed. 2015;28(2):217-230. doi:10.1002/nbm.3237

150. Woessner DE, Zhang S, Merritt ME, Sherry AD. Numerical solution of the Bloch equations provides insights into the optimum design of PARACEST agents for MRI. Magn Reson Med. 2005;53(4):790-799. doi:10.1002/mrm.20408

151. Shaghaghi M, Cai K. Analytical solution of the Bloch-McConnell equations for steady-state CEST Z-spectra. Magn Reson Imaging. 2024;109:74-82. doi:10.1016/j.mri.2024.02.015

152. Schuenke P. BMCTool. https://github.com/schuenke/BMCTool

153. Haken H, Wolf HC, Brewer WD, eds. Modern methods of optical spectroscopy. In: The Physics of Atoms and Quanta: Introduction to Experiments and Theory. Springer Berlin Heidelberg; 2005:399-416. doi:10.1007/3-540-29281-0_22

154. Reilly JT, Walsh JM, Greenfield ML, Donohue MD. Analysis of FT-IR spectroscopic data: the Voigt profile. Spectrochim Acta Part Mol Spectrosc. 1992;48(10):1459-1479. doi:10.1016/0584-8539(92)80154-O

155. Whiting EE. An empirical approximation to the Voigt profile. J Quant Spectrosc Radiat Transf. 1968;8(6):1379-1384. doi:10.1016/0022-4073(68)90081-2

156. Zaiss M, Schmitt B, Bachert P. Quantitative separation of CEST effect from magnetization transfer and spillover effects by Lorentzian-line-fit analysis of Z-spectra. J Magn Reson. 2011;211(2):149-155. doi:10.1016/j.jmr.2011.05.001

157. Cai K, Singh A, Poptani H, et al. CEST signal at 2 ppm (CEST@2ppm) from Z-spectral fitting correlates with creatine distribution in brain tumor. NMR Biomed. 2015;28(1):1-8. doi:10.1002/nbm.3216

158. Desmond KL, Moosvi F, Stanisz GJ. Mapping of amide, amine, and aliphatic peaks in the CEST spectra of murine xenografts at 7 T. Magn Reson Med. 2014;71(5):1841-1853. doi:10.1002/mrm.24822

159. Mohammed Ali S, Yadav NN, Wirestam R, et al. Deep learning–based Lorentzian fitting of water saturation shift referencing spectra in MRI. Magn Reson Med. 2023;90(4):1610-1624. doi:10.1002/mrm.29718

160. Zeng Q, Machado M, Bie C, et al. In vivo characterization of glycogen storage disease type III in a mouse model using glycoNOE MRI. Magn Reson Med. 2024;91(3):1115-1121. doi:10.1002/mrm.29923

161. Ceckler T, Maneval J, Melkowits B. Modeling magnetization transfer using a three-pool model and physically meaningful constraints on the fitting parameters. J Magn Reson. 2001;151(1):9-27. doi:10.1006/jmre.2001.2326

162. Li AX, Hudson RHE, Barrett JW, Jones CK, Pasternak SH, Bartha R. Four-pool modeling of proton exchange processes in biological systems in the presence of MRI–paramagnetic chemical exchange saturation transfer (PARACEST) agents. Magn Reson Med. 2008;60(5):1197-1206. doi:10.1002/mrm.21752

163. Morrison C, Henkelman RM. A model for magnetization transfer in tissues. Magn Reson Med. 1995;33(4):475-482. doi:10.1002/mrm.1910330404

164. Marshall I, Higinbotham J, Bruce S, Freise A. Use of Voigt lineshape for quantification of in vivo 1H spectra. Magn Reson Med. 1997;37(5):651-657. doi:10.1002/mrm.1910370504

165. Foo LS, Yap WS, Tee YK. Determination of computationally efficient multi-pool model fitting approach for pulsed chemical exchange saturation transfer MRI. In: Proceedings of the 2019 8th International Conference on Bioinformatics and Biomedical Science. ICBBS '19. Association for Computing Machinery; 2020:32-39. doi:10.1145/3369166.3369175

166. Fisher RA, Russell EJ. On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond A. 1997;222(594-604):309-368. doi:10.1098/rsta.1922.0009

167. Savage LJ. The foundations of statistics. Nav Res Logist Q. 1954;1(3):236-236. doi:10.1002/nav.3800010316

168. Wakefield JC, Smith AFM, Racine-Poon A, Gelfand AE. Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. J R Stat Soc C Appl Stat. 1994;43(1):201-221. doi:10.2307/2986121

169. Bayes T, Price null. LII. An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc Lond. 1997;53:370-418. doi:10.1098/rstl.1763.0053

170. Liu H, Wasserman L. Bayesian inference. In: Statistical Machine Learning. Carnegie Mellon University; 2014.

171. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Wadsworth & Brooks/Cole; 1984.

172. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1(1):81-106. doi:10.1007/BF00116251

173. Quinlan JR. C4.5: Programs for Machine Learning. Published online 1993.

174. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. doi:10.1023/A:1010933404324

175. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM; 2016:785-794. doi:10.1145/2939672.2939785

176. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS '17. Curran Associates Inc.; 2017:3149-3157.

177. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS '18. Curran Associates Inc.; 2018:6639-6649.

178. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386-408. doi:10.1037/h0042519

179. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13(10):281-305.

180. Frazier PI. A tutorial on Bayesian optimization. Published online 2018. Accessed September 4, 2025. https://arxiv.org/abs/1807.02811

181. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: a novel bandit-based approach to hyperparameter optimization. Published online 2018. Accessed September 4, 2025. https://arxiv.org/abs/1603.06560

182. Jaderberg M, Dalibard V, Osindero S, et al. Population based training of neural networks. Published online 2017. Accessed September 4, 2025. https://arxiv.org/abs/1711.09846

183. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 2005;30(1):79-82.

184. Kingma DP, Ba J. Adam: a method for stochastic optimization. Published online 2017. Accessed September 4, 2025. https://arxiv.org/abs/1412.6980

185. Goodfellow IJ, Vinyals O, Saxe AM. Qualitatively characterizing neural network optimization problems. Published online 2015. Accessed September 4, 2025. https://arxiv.org/abs/1412.6544

186. Knuth DE. Big Omicron and big Omega and big Theta. SIGACT News. 1976;8(2):18-24. doi:10.1145/1008328.1008329

187. Jolliffe I. Principal component analysis. In: Wiley StatsRef: Statistics Reference Online. 2014. doi:10.1002/9781118445112.stat06472

188. Ladjal S, Newson A, Pham CH. A PCA-like autoencoder. Published online 2019. Accessed September 4, 2025. https://arxiv.org/abs/1904.01277

189. Casella M, Dolce P, Ponticorvo M, Marocco D. From principal component analysis to autoencoders: a comparison on simulated data from psychometric models. In: 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE). 2022:377-381. doi:10.1109/MetroXRAINE54828.2022.9967686

190. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. 2008:1103. doi:10.1145/1390156.1390294

191. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc Natl Acad Sci. 2020;117(48):30088-30095. doi:10.1073/pnas.1907377117

Printed by Tryckeriet i E-huset, Lund 2025

Magnetization transfer imaging techniques, such as *chemical exchange saturation transfer (CEST)* have shown promise in both clinical and preclinical applications. The available methods for processing and analyzing CEST data, however, still carry many limitations. This thesis provides a solid background to magnetization transfer imaging and further introduces, for processing, a machine learning-based denoiser and, for analysis, two machine learning-based solutions for spectral fitting of CEST data. The proposed solutions tackle some of the current limitations in their respective categories. By solving these limitations, the proposed methods provide beneficial implications on the clinical feasibility of the magnetization transfer techniques.

LUND UNIVERSITY

Medical Radiation Physics, Lund
Faculty of Science
ISBN 978-91-8104-634-2 (print)