$$u^b$$

Graduate School for Cellular and Biomedical Sciences

University of Bern

Multiparametric MR Spectroscopy: evaluation of quantitative frameworks based on modeling and deep learning

Ph.D. Thesis submitted by

Rudy Rizzo

for the degree of

Ph.D. in Biomedical Engineering

Supervisor
Prof. Dr. sc. nat. Roland Kreis
Magnetic Resonance Methodology
Institute of Diagnostic and Interventional Neuroradiology
University of Bern

Co-advisor
PD Dr. Johannes Slotboom
Support Center for Advanced Neuroimaging
Institute of Diagnostic and Interventional Neuroradiology
University of Bern

Date of dissertation: 22 May 2023

Date of publication: 1 February 2025

License Statement

This thesis, titled *Multiparametric MR Spectroscopy: evaluation of quantitative frameworks based on modeling and deep learning*, is submitted in partial fulfillment of the requirements for the degree of *Ph.D. in Biomedical Engineering* at the *University of Bern, Bern, Switzerland*. This work is licensed under *Creative Commons Attribution −NonCommercial 4.0 International License* **CC BY-NC 4.0** *except:*

pages 147-156 reproduced with permission from Springer Nature;

pages 157-236 modified with permission from John Wiley and Sons;

Figures 1.1-1.10, 2.2 (B,C), 2.3, 4.2, 4.3, and 4.5 reproduced with permission from John Wiley and Sons under License Agreement 5924810669017

Figures 2.5-2.8, 3.2 and 3.3 reproduced with permission from John Wiley and Sons and Figure 3.5 reproduced with permission from W.B./Saunders Co. under License Agreement 1554261

Terms of Use

Accepted by the Faculty of Medicine, the Faculty of Science and the Vetsuisse Faculty of the University of Bern at the request of the Graduate School for Cellular and Biomedical Sciences

Bern,   Dean of the Faculty of Medicine

Bern,   Dean of the Faculty of Science

Bern,   Dean of the Vetsuisse Faculty Bern

Abstract

Magnetic Resonance Spectroscopy (MRS) is a superb technique for the diagnosis and treatment monitoring of many diseases. It relies on the non-invasive acquisition and evaluation of the metabolic content in a selected body area. In fact, levels of metabolites' concentration can provide cell-type-specific insights into the function and pathophysiology of various organs.

Furthermore, the chemistry of the human body can be explored through sensitization of the Nuclear Magnetic Resonance (NMR) signal to a variety of properties beyond plain metabolite concentrations. This is the realm of multiparametric MRS. T_2 relaxation rates contain essential information on the cellular microenvironment, acting as potential biomarkers for abnormalities and carrying a crucial role in absolute quantification. In contrast, metabolic diffusion is driven by particle features and the geometry of intracellular compartments. It may provide inference on the chemical exchange and bonding, untangle interactions between metabolites and other compounds, or discern behaviors across different tissue compartments. Conventional techniques for measuring these quantities, such as Diffusion-Weighted MRS (DW-MRS) or Multiple Echo Times (MTE) acquisitions, are time-consuming, inefficient, and rarely used within clinical timeframes. Instead, to solve the issues for absolute quantification, clinical MRS relies on regionally T_2 tabulated values, which do not account for inter-subject variability and are often lacking, particularly in pathology. Besides ignoring a relevant MR quantity, such assumptions introduce quantification errors and biases.

To speed up the quantification tasks, recent developments in machine learning algorithms, explicitly concerning deep learning (DL) architectures, have found an increasing interest within the scientific community. DL toolboxes come as pre-trained models where run-time metabolic concentrations are provided on-the-fly. However, despite the uprising research throughput, a low acceptance rate for such tools is currently found in the clinics. On the one hand, that is due to the high complexity of such architectures, which translates into a low level of interpretability still conveyed to black-box assumptions. But on the other hand, DL does not yet provide reliable and well-established uncertainty measures of its predictions, which leaves the average user unaware of potential intrinsic errors.

The current work starts with an introduction given in Chapter 1, where a brief overview of the relevant properties of NMR, from classic and quantum physic perspectives, is treated. It follows, in Chapter 2, a quick dive into a practical MRS experiment where the fundamentals of signal excitation and recording, as well as the crucial properties of MRS signals, are discussed in a synthetic yet realistic context. Chapter 3 wants to disclose the MR properties and challenges of two human organs (brain and prostate) that drove the focus of the various investigations encountered during this work. It follows an overview of the methods, clinical interests, and potentials in the context of multiparametric MRS, mainly focusing on estimating metabolite-specific T_2 rates and diffusion properties, which is disclosed in Chapter 4. The background part is concluded in Chapter 5 with an overview of the principles of DL, exploring in more detail the current challenges on interpretability and uncertainty measures as well as the state-of-the-art designs deployed for MRS quantification.

The first part of the main contribution of this work is given in Chapter 6. DL is deployed across many architectural designs and is twinned to tailored MRS processing aiming to enhance features in the data that are more or less prone to DL computation. Results are further analyzed concerning dataset biases and possible training strategies to overcome such limitations, namely, ensemble of models and data augmentation. A window into interpretability and uncertainty measures is also explored, offering a first method to integrate MRS predictions of concentrations with their reliability. An analysis that compares these measures to traditional Cramer Rao Lower Bounds (CRLBs) in fitting follows.

The second part of the main contributions of this work is explored in Chapter 7 with a focus on different aspects of multiparametric MRS. First, the urge of speed in multiparametric MRS to simultaneously and accurately produce metabolite concentrations and T_2 rates is explored, introducing a novel acquisition method that combines bi-dimensional fitting and truncated multi-echo acquisitions. The benefits and limitations of the methodology are disclosed both in a single-voxel experimental fashion and in a 2D MRSI setup targeting the human brain. Second, diffusion-weighted MRS is deployed for the first time in the human prostate to untangle and explore MRS properties. The results offer an alternative viewpoint to the complex chemical bonding of proteins with some of the main prostatic metabolites. Eventually, an initial investigation of DW-MRS for pathological cases is outlined. Although severely limited by a small cohort of patients, it promises exciting potential currently interpreted based solely on the prostate microstructure.

Contents

Introduction & Background

1 | Principles of NMR

The beauty of in-vivo Nuclear Magnetic Resonance (NMR) lies in mimicking the fundamental behavior that features human beings over other living species: articulated dialogues. In fact, from the wide pool of medical imaging technologies available to a radiologist, NMR represents the safer and most complex dialogue between the operator, or the developer, and the tissue in hand, at a (sub-) millimeter scale. The level of interaction with the physics is indeed such complex that any limited manipulation of a delivered message, upon excitation, twinned to any fiddling with the communication channel, upon recording, can end up in an entirely different reply from the inquired tissue, generally called MR contrast. What a beauty.

The study of the interaction between matter and electromagnetic radiation is known as spectroscopy. Atoms and molecules contain distinct energy levels corresponding to different quantified electronic, vibrational, or rotational states. The absorption and emission of photons with an energy that perfectly matches the energy level difference between two states describe the interaction between atoms and electromagnetic radiation. Since the energy of a photon is proportional to its frequency, different forms of spectroscopy are typically identified by the frequencies involved. NMR spectroscopy, or MR Spectroscopy (MRS) employs radio-frequency (RF) in the 10-1000 MHz range.

NMR is the study of nuclei's magnetic properties and associated energy. Radio-frequency energy is absorbed when the nuclei are put in a (strong) external magnetic field. The induced energy level difference of nuclei in an external magnetic field is relatively modest compared to thermal energy at ambient temperature, resulting in almost equally populated energy levels. As a result, photon absorption is relatively low, making NMR a relatively insensitive technique in comparison to other types of spectroscopy. However, because of the low energy absorption, NMR is a noninvasive and nondestructive technology well suited for in-vivo studies.

This chapter is strongly inspired by [1]

1.1. Nuclear spin and magnetization

The birth of Magnetic Resonance (MR) is the consequence of the definition of the concept of spin. Any rotating object is characterized by angular momentum, which describes the object's tendency to continue spinning. Subatomic particles like electrons, neutrons, and protons have an intrinsic angular moment, or spin, which exists even when the particle is not spinning. Electron spin is a product of relativistic quantum physics and has no classical equivalent explanation. However, this can be conceptualized as a magnetic field formed by rotating currents within a spinning particle. It is typically well represented as a nucleus (or ball) rotating around its axis.

1.1.1. Classical description

A spin-bearing particle is depicted as a ball with a given charge rotating around an axis. The angular momentum $\boldsymbol{L}$ featuring a rotating object with mass m and distance from the rotational axis r is described by $\boldsymbol{L} = mr\boldsymbol{v}$, where $\boldsymbol{v}$ indicates the spinning velocity. The rotation (or variation) of the charged particle generates a current that induces a magnetic moment, which features the particle and is referred to as microscopic magnetization: $\boldsymbol{\mu} = \gamma\boldsymbol{L}$ where γ is the gyromagnetic ratio that characterizes the particle (i.e., $\gamma = e/2m$ where e is the charge of the particle). As suggested by the headlight *Nuclear* MR, the particles of interest are nuclei. Even though various nuclei can be targeted and are of interest for *in-vivo* NMR, the purpose of the current work will focus on proton NMR (or ^{1}H-NMR), which refers to the isotope ^{1}H, which is typically observed in water and other compounds, such as metabolites, lipids, and different macromolecules. Moreover, (1) protons are present with high intrinsic abundance in all mammalian tissues in the form of water or lipids (e.g., the human brain contains around 6×10^{19} spins/mm^3) and (2) protons feature the highest γ among the commonly encountered MR-visible isotopes.

Molecules are present in solution and experience molecular tumbling with various rotations, translations, and collisions. As a result, the amplitude and direction of one proton's magnetic field at the site of another proton vary with time. The spin orientation has no preference in the absence of an external magnetic field, and the spins are oriented randomly across the sample, Fig.1.1(A). Summation over all orientations results in (near) complete cancellation of magnetic moments, and consequently, a macroscopic magnetization vector is absent. When the protons are exposed to an external magnetic field $\boldsymbol{B_0}$ will experience a torque $\boldsymbol{T}$ which is depicted by $\boldsymbol{T} = \boldsymbol{\mu} \times \boldsymbol{B_0}$. A torque is mathematically described as a change of angular momentum over time, $\boldsymbol{T} = \partial\boldsymbol{L}/\partial t$, which features the precession of the spin's magnetic moment, described as $\partial\boldsymbol{\mu}/\partial t = \boldsymbol{\mu} \times \gamma\boldsymbol{B_0}$. This effect is referred to as Larmor precession, where all the spins will begin to precess around the external magnetic field at the same Larmor frequency $\omega_0 = \gamma B_0$, Fig.1.1(B).

The spin orientation is further perturbed by local field fluctuations caused by molecular tum-
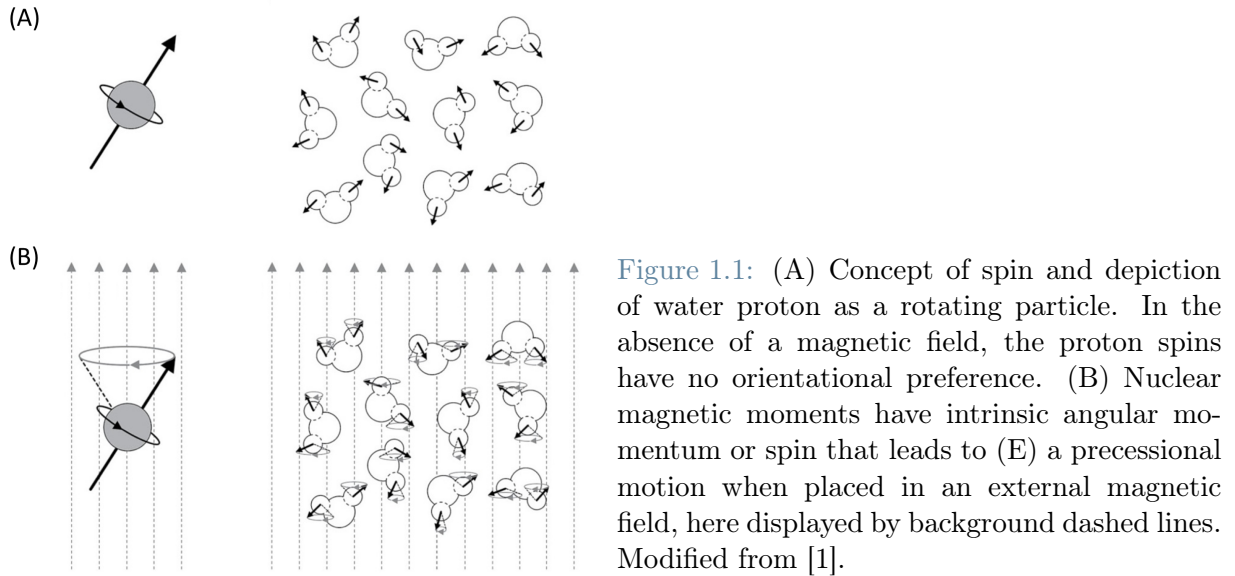
(A)

(B)

Figure 1.1: (A) Concept of spin and depiction of water proton as a rotating particle. In the absence of a magnetic field, the proton spins have no orientational preference. (B) Nuclear magnetic moments have intrinsic angular momentum or spin that leads to (E) a precessional motion when placed in an external magnetic field, here displayed by background dashed lines. Modified from [1].

bling. These disturbances are mostly but not entirely random. A strong external magnetic field promotes parallel spin orientation according to the concept of minimal energy. The corresponding magnetic energy can be computed by $E = \boldsymbol{\mu} \cdot \boldsymbol{B_0} = -\mu B_0 cos\theta$ where θ is the angle between the magnetic moment and the external magnetic field. Therefore, the magnetic energy can be maximized (μB_0) when spins are aligned anti-parallel with $\boldsymbol{B_0}$ ($\theta = \pi$) and minimized ($-\mu B_0$) when spins are aligned parallel with $\boldsymbol{B_0}$ ($\theta = 0$). As a result, the entirely random spin orientation distribution evolves over time into one that is somewhat biased toward parallel spin orientation. However, the net number of spins biased towards parallel orientation is minuscule (on the order of one in a million) but adding the magnetic moments, on the other hand, results in a macroscopic magnetization vector parallel to the external magnetic field. The magnitude of the longitudinal equilibrium magnetization, and hence the strength of the induced NMR signal, is proportional to the number of spins biased parallel to the main magnetic field. The longitudinal equilibrium magnetization M_0 can be calculated from the Boltzmann distribution, and results proportional to

$$M_0 \propto \frac{N\gamma^2 B_0}{T} \tag{1.1}$$

Which reveals several important features of the signals detected by NMR.

- the thermal equilibrium magnetization M_0 is directly proportional to the number of spins N in the sample. This feature makes NMR a quantitative method in which the detected signals are, in principle, proportional to the concentration.

- The quadratic dependence of M_0 on the gyromagnetic ratio γ implies that nuclei resonating at high frequency generate the strongest NMR signals. The linear dependence of M_0 on the magnetic field strength B_0 implies that higher magnetic fields improve the sensitivity.

- the inverse proportionality of M_0 to the temperature T indicates that sensitivity can be

enhanced at lower sample temperatures. However unrealistic for in-vivo applications.

It needs to be noted that the actual experimental sensitivity is determined by many additional factors, like RF coil characteristics, pulse sequence details, sample volume, natural abundance of the studied nucleus, sample noise, relaxation parameters, and spectral resolution.

The current thesis project considered standard clinical magnetic field strength at 3T and hydrogen as nuclei (^{1}H-NMR or proton NMR). Hydrogen has the highest NMR relative sensitivity, given the already discussed intrinsic properties.

1.1.2.   Quantum physics description

One of the fundamentals of quantum mechanics is that energy, momentum, and other properties are quantified: restricted to specific, discrete values. Different vibrational modes are characterized by quantified energy levels given by $E = (n + 1/2)hv$, where n is an integer quantum number ($n = 0, 1, 2, \dots$) corresponding to the separate levels, h is the Plank constant, and v is the frequency of vibration.

Energy sampling between quantified levels can happen by absorption or by emission. Molecules that vibrate at a lower energy level with quantum number n can move to a higher energy level with a quantum number $n + 1$ by absorption of electromagnetic energy with a frequency v that exactly matches the energy-level difference ΔE, according to $\Delta E = hv$. Conversely, a molecule with high vibrational energy can drop down to a lower vibrational energy level through the emission of electromagnetic waves of specific frequency v, Fig.1.2(A).

Intrinsic angular momentum, or spin, is also quantified both in amplitude and orientation. In the quantum mechanical picture of NMR, spins are either in the parallel (α) or anti-parallel (β) orientation concerning the external magnetic field. The spin states are characterized by energy $E = -mhv$ with m representing the spin quantum number associated with α ($m = 1/2$) and β ($m = -1/2$) spin orientations. NMR transitions are only allowed between energy levels for which the spin quantum number m changes by ± 1, Fig.1.2(B).

It may appear that all spins are either in the α or β spin state and that no other orientations are allowed. The quantum mechanical view and the classical view can be reconciled to a certain degree by knowing that spins can also be in so-called superposition states besides the pure α and β quantum states. However, since NMR does not observe single spins, the collapse into pure spin states never happens, and the spins within the sample remain in a superposition state throughout the experiment.
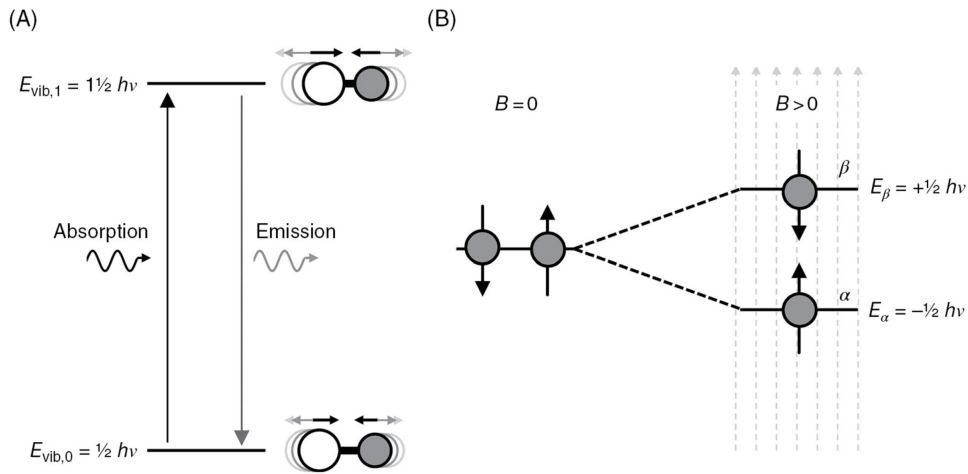
Figure 1.2: Quantum energy levels. (A) Absorption and emission of electromagnetic radiation with a frequency that matches the energy-level difference between two vibrational states allow the transition from one to the other for a given vibrating molecule. (B) Split of NMR transition between α and β states upon external magnetic field $B > 0$. Figure from [1].

Density matrix formalism

While the classical description helps to gain a general understanding of the NMR physics and can be deployed to fully describe isolated (i.e., non-interacting) spin systems (e.g., like a first-level approximation of water spins in conventional MRI, where despite j-coupling interaction present for water protons, its effect is not seen), quantum mechanics (QM) is fundamental to understanding interacting spin systems and phenomenon like line splitting due to spin-spin coupling (details in Section 1.3). The QM formal description, known as *density matrix* (or *operator*) *formalism*, covers a fundamental role in the simulation of spin-system evolution upon delivery of NMR manipulation (RF pulses, gradients, etc.), particularly crucial for the description of MR spectra and fundamental for quantification purposes, see Section 2.6.2.

The density matrix formalism is not directly concerned with magnetization but rather deals with the energy states of the spin system under investigation. A density matrix calculation typically starts with the creation of a $2^N \times 2^N$ thermal equilibrium density matrix $\boldsymbol{\sigma}(0)$, where N represents the number of different spins in a spin system. The effects of RF pulses, delays, and magnetic field gradients are described by time-dependent operators $\boldsymbol{\sigma}(t)$ created by Hamiltonians $\boldsymbol{H}$. They describe the total energy of the spin system and can be seen as a generalization of rotation matrices for larger spin systems. The evolution of the density matrix under the influence of RF pulses and delays is governed by the *Liouville–von Neumann equation*, operating on the initial spin state $\boldsymbol{\sigma}(0)$ at time $t = 0$ and transforming it into a new density matrix $\boldsymbol{\sigma}(t)$ at time t:

$$\boldsymbol{\sigma}(t) = e^{-i\boldsymbol{H}t}\boldsymbol{\sigma}(0)e^{+i\boldsymbol{H}t} \tag{1.2}$$

The propagation of the spin system throughout multiple steps of the NMR manipulation is followed by the time of signal acquisition. An NMR spectrum can be extracted from the final density matrix and displayed following Fourier transform by extraction of the quadrature elements (i.e., trace) of the density matrix. An in-depth description of the quantum state function and the derivation of the solution that characterizes a quantum measurement can be found in [2]. Visual aid is supported in Fig.1.3, where a two-spin system with four energy levels and their respective spin population yields a 4×4 thermal equilibrium density matrix $\boldsymbol{\sigma}(0)$. The energy-level populations are contained in the non-zero element (gray) on the matrix diagonal. The Lioville-von Neumann equation manipulates the density matrix according to the desired NMR experiment. At the time of signal recording, the off-diagonal element of the matrix can be nonzero and correspond to the various coherences allowed in the two-spin system.
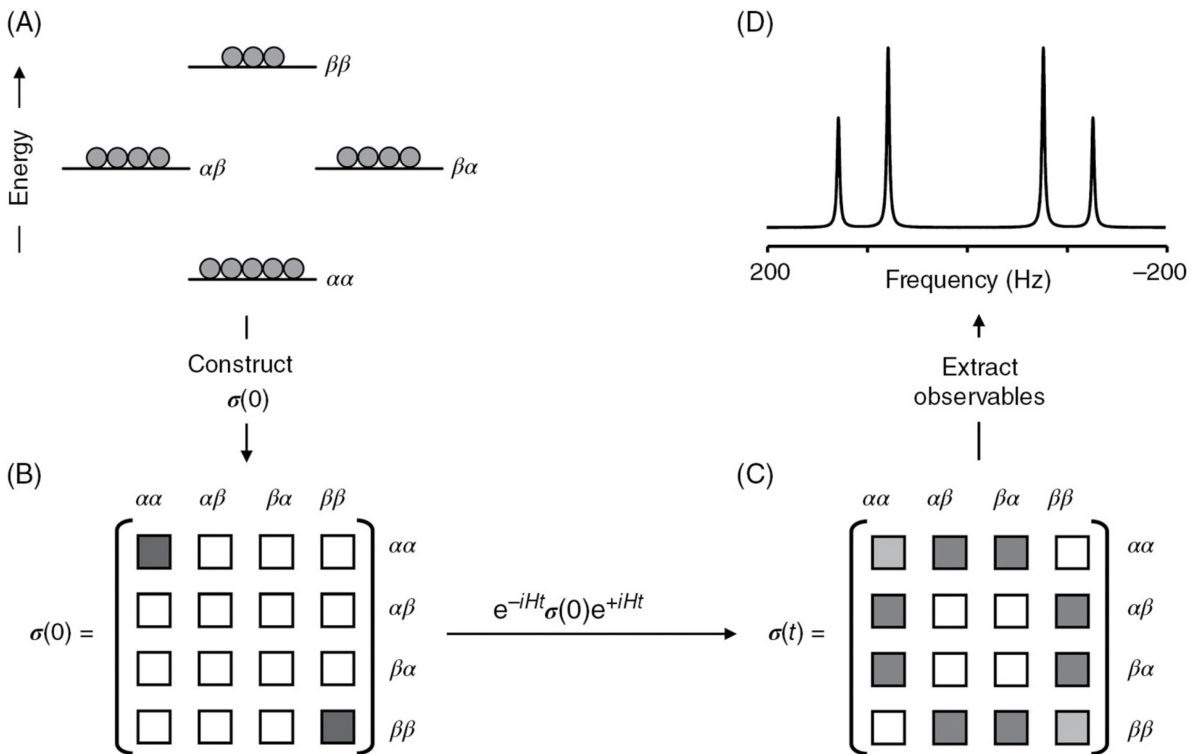
Figure 1.3: Quantum-mechanical NMR description using density matrix. The description of (A) two-spin system energy levels and the associated nuclear spin population is provided by (B) the initial density matrix. (C) Evolution of the density matrix under NMR manipulation. Gray areas correspond to nonzero elements, while white areas correspond to nulled elements. (D) At signal recording, the observable populations (or coherences) are extracted as a spectrum. Figure from [1].

For a weakly-coupled two-spin system, the general density matrix can be decomposed into a linear sum of matrixes in which each matrix corresponds to a particular spin coherence (or population). For example, the diagonal elements correspond to spin populations which give rise to longitudinal magnetization, whereas off-diagonal elements give rise to single-quantum

coherences or transverse magnetization, Fig.1.4. Different combinations of matrix elements give rise to different and mixed contributions of longitudinal, transversal, or multi-quantum coherences.
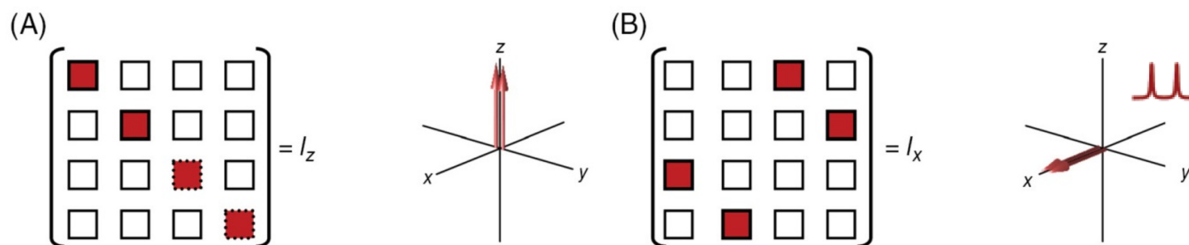
Figure 1.4: Density matrix decomposition: a visual aid. Sub-elements of a two-spin density matrix are selected in red: real (solid) and imaginary (dotted) components. A correspondent visualization of the macroscopic components inspired by the correlated vector model [1] is added. (A) The diagonal elements correspond to spin populations which give rise to longitudinal magnetization. (B) Off-diagonal elements give rise to single-quantum coherences, detectable by signal quadrature. Modified from [1].

1.2. RF excitation and NMR signal recording

The longitudinal magnetization vector gives the signal detected in an NMR experiment. On the other hand, the static, longitudinal magnetization is never directly observed because its minor contribution would be dominated by much larger contributions from magnetic properties associated with electron currents within atoms and molecules. Instead, the longitudinal magnetization is moved into the transverse plane, where the precessing magnetization may trigger a signal in a receiver coil at a precise Larmor frequency. To alter the orientation of the longitudinal magnetization, M_0, an additional time-varying (i.e., frequency- or phase-modulated) magnetic field B_1, with frequency matching the Larmor frequency and perpendicular to the magnetic moment B_0, is deployed. See Fig.1.5(A). The intensity of B_1 is much smaller than B_0, but is capable of influencing M_0 given its matching resonance frequency. The modulation of nuclear spin and their macroscopic magnetization exploiting the concept of *resonance* explains the letter R in *NMR*.

The net macroscopic magnetization, also referred to as *longitudinal magnetization* and indicated by M_z, experiences a coherent rotation (i.e., a torque) towards the transverse plane since the angle between the single spin magnetic moments and the magnetic field B_1 is constant. As a result, the spins have been excited. The B_1 field is referred to as an *excitation* pulse when the amplitude and duration of the B_1 field are modulated to produce a 90° (or $\pi/2$) rotation of the magnetization from the longitudinal axis into the transverse plane, Fig.1.5(B). This is referred to as *transverse magnetization* and indicated by M_{xy}. When the amplitude or duration of the B_1 field is doubled, the initial thermal equilibrium magnetization rotates 180° (or π), which is accomplished by an *inversion* pulse. Following excitation, the B_1 field is removed, leaving the

magnetization in the transverse plane to be detected through electromagnetic induction in a nearby receiver coil. As a result, the free magnetization in the transverse plane evolves into a signal that is called Free-Induction-Decay (FID), Fig.1.5(C).
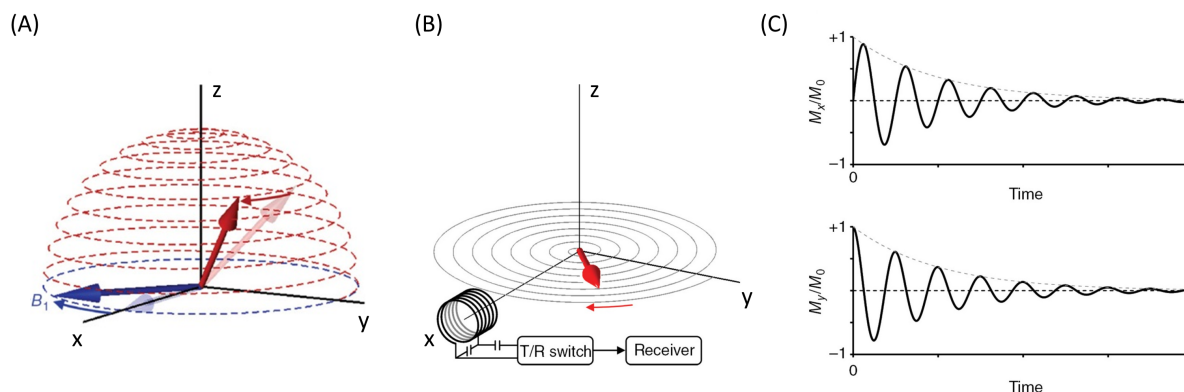
Figure 1.5: Excitation and recording. (A) Rotation of the macroscopic magnetization vector (red) in a non-rotating coordinate system under the influence of B_0 along the z axis and the oscillating magnetic field B_1 in the xy plane. (B) The macroscopic magnetization has been flipped about 90° and fully on the xy plane, which oscillation induces a current on the indicated coil generating the (C) FID signal with components M_x and M_y. Modified from [1].

1.2.1.   Relaxometry

Relaxation can be qualitatively described as the process by which the macroscopic magnetization vector returns to the thermal equilibrium M_0 state following a perturbation. In other words, random local field fluctuations lead the Larmor frequencies of different spins to run out of sync over time. For spins in solution, the randomly fluctuating magnetic fields are predominantly caused by the magnetic moment of other nearby spins. The magnetic moment of one spin affects the local field of another spin randomly (both in amplitude and orientation) due to Brownian motion and molecular tumbling. Fig.1.6 shows a typical distribution of the local magnetic field at a nucleus as a function of time.

The restoration of the longitudinal equilibrium magnetization M_0 is characterized by the longitudinal or spin–lattice relaxation time constant T_1, while the disappearance of transverse magnetization is described by the transverse or spin–spin relaxation time constant T_2. Both relaxation times are time constants used to characterize what are assumed to be first-order rate processes. Relaxation rates $R_{1(2)} = 1/T_{1(2)}$ can alternatively be used instead of relaxation times. For ^{1}H-NMR in biological tissues, the T_2 relaxation time constants (10–500 ms) are typically much shorter than the T_1 relaxation time constants (500–3000 ms). Furthermore, the relaxation mechanism depends on the rate of molecular motion. Thus, it is temperature and B_0 dependent.
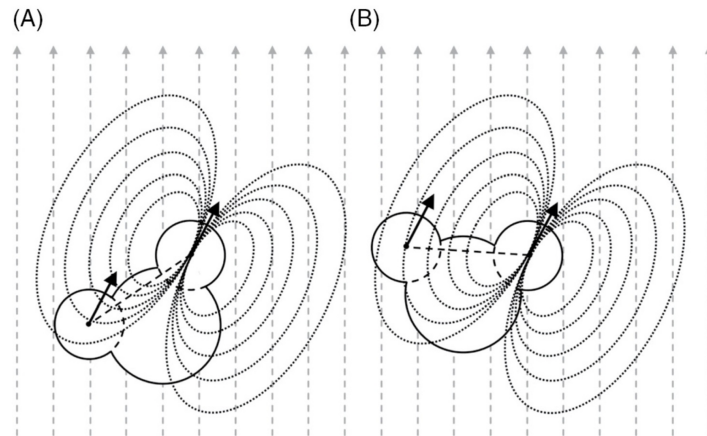
Figure 1.6: Interaction of local magnetic fields within a water molecule. (A) In gray is depicted the magnetic moment of one proton in an external magnetic field that perturbs the magnetic field at the position of another proton. (B) As the water molecule rotates and translates randomly, the direction and magnitude of the perceived local magnetic field change over time, giving rise to time-varying magnetic fields required for relaxation. Figure from [1].

T_1 relaxation

The spin-lattice (or longitudinal) relaxation time T_1 quantifies the rate of energy transfer from the nuclear spin system to the neighboring system of molecules (the lattice). For any given NMR energy transition ΔE experienced by an excited nuclear spin, there will be some possible change within the lattice involving the same quantity of energy, inducing translations, rotations, and internal motions of molecules. For instance, a downward flip of a nuclear spin will cause an acceleration of some motion of the molecule in which the flip occurred. Such energy transfer requires a fluctuating magnetic field at the site of the nuclear spin involved, which must have a component at the Larmor frequency of the nucleus under consideration. Only x and y components of the local field can cause T_1-relaxation (i.e., the motion in the lattice). The most common source of the such local fluctuating field is direct dipolar interaction. Other mechanisms for spin-lattice relaxation are interaction with unpaired electrons (e.g., paramagnetic substances) or interactions with electric field gradients for quadrupolar nuclei. The spin-lattice relaxation time determines what recycle delay between pulses should be used. The effects of T_1 relaxation on MR signal detection can be minimized or eliminated by choosing a sufficiently long repetition time. This relaxation in the z-direction leads to the restoration of Boltzmann equilibrium. The recovery of longitudinal magnetization is described by

$$\frac{\partial M_Z}{\partial t} = \frac{M_0 - M_Z}{T_1} \tag{1.3}$$

The standard method for measuring T_1 is known as *inversion-recovery*. First, a 180° pulse inverts the magnetization along the z-axis. A time period T_I (inversion time) is allowed, during

which spin-lattice relaxation occurs, causing M_z to go from the value of $-M_0$ through zero to its equilibrium value of M_0. A 90° pulse is then applied, and the FID is recorded. The experiment is repeated with different T_I delays, allowing the determination of the T_1 value, Fig.1.7. Quantitatively, the decay of M_z is given by the Bloch equation Eq.1.3. Integration of Eq.1.3 with $M_z = -M_0$ at $t = 0$ gives: $M_z = M_0(1 - 2e^{-t/T_1})$.
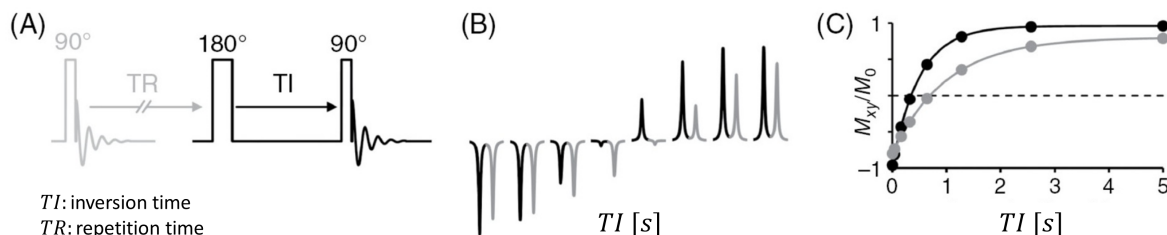
Figure 1.7: Measurement of T_1 relaxation via *inversion-recovery*. (A) pulse sequence diagram. (B) The intensity of two compounds upon different inversion times and (C) signal recovery over time. Modified from [1].

T_2 relaxation

Spin-spin (or transverse) relaxation time T_2 is used to quantify the rate of magnetization decay within the xy plane. In a perfectly homogeneous magnetic field, after a 90° pulse, the nuclear spins are aligned in one direction (and said to be phase coherent). Still, this arrangement is gradually lost due to intrinsic direct interaction between the spins, without energy transfer to the lattice. The disappearance of macroscopic magnetization (i.e., M_{XY} assumed as the module of the signal) from the transverse plane is described by

$$\frac{\partial M_{XY}}{\partial t} = -\frac{M_{XY}}{T_2} \tag{1.4}$$

T_2 relaxation does not affect the total amount of z-magnetization, but the degree of synchronization of the transverse magnetization components. However, a microscopic and macroscopic inhomogeneity combination leads to additional signal dephasing and a more rapid signal decrease, characterized by T_2^* relaxation. Under most in-vivo conditions, the inhomogeneity contribution to T_2^* relaxation dominates. The causes of macroscopic inhomogeneities are caused by B_0 inhomogeneity and local field variations, whereas microscopic inhomogeneities can be related to the unwanted diffusion of molecules.

The measurement of T_2 refers to *spin-echo* or *stimulated echo* experiments, where the contribution of macroscopic magnetic field inhomogeneity to the observed T_2^* relaxation can be eliminated. In such a design, the phase variation acquired by individual spins before the 180° pulse due to magnetic field inhomogeneity is inverted after the refocusing pulse. During the second $TE/2$ period, the spins continue to accumulate phase due to the macroscopic magnetic field

inhomogeneity that is part of T_2^* relaxation. However, since the acquired phase was inverted, the phase accumulation during the second $TE/2$ period leads to signal rephasing and spin echo formation at the echo-time TE. As a result, the random magnetic field variations underlying T_2 relaxation cannot be refocused by a spin-echo, leading to a reduced signal intensity at the echo-time TE. Visual aid is provided in Fig.1.8.
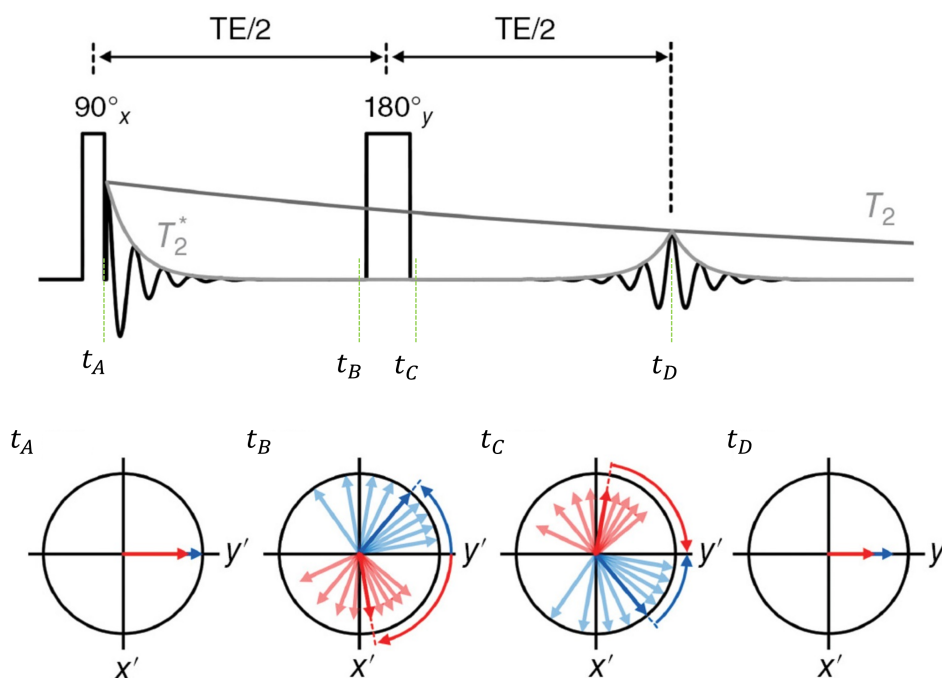
Figure 1.8: T_2 and T_2^* relaxation in a *spin-echo* experiment. (Top) A spin-echo pulse sequence shows the rapid signal decay due to T_2^* relaxation immediately following excitation (t_A). Echo formation is displayed at echo time TE (t_D), upon deployment of a refocusing pulse ($t_B - t_C$) whereby the echo intensity is reduced by T_2 relaxation. (Bottom) Evolution of transverse magnetization of two compounds (red and blue) with different chemical shifts and magnetic field inhomogeneity yields phase dispersion across the sample. The 180° pulse allows inversion of the acquired phases during the first $TE/2$ period. Modified from [1].

1.3. Chemical shift and scalar coupling

The Larmor frequency of a given nucleus does not only depend on the external magnetic field strength, but it deviates from its expected value given the shielding of the magnetic field at the nucleus by the surrounding electron cloud (i.e., the chemical environment surrounding the nucleus given the electronic configuration) which alters the nucleus' perceived external field leading to a lower Larmor resonance frequency. Such a phenomenon is known as *chemical shift*. In Fig.1.9, a visual example is given. The electron cloud can be seen as a current rotating around the nucleus, generating a local magnetic field opposite the external magnetic field B_0. As the electron cloud density surrounding the proton decreases, so does the electron-induced magnetic field B_e. The total magnetic field at the nucleus B_n will therefore be larger. The Larmor

resonance frequency can become independent of the magnetic field strength when expressed in parts-per-million (ppm) relative to a reference compound. Thus, the NMR resonance frequencies, or chemical shifts, give direct information about the chemical environment of nuclei, thereby yielding the unambiguous detection and assignment of compounds.

(A)

(B)
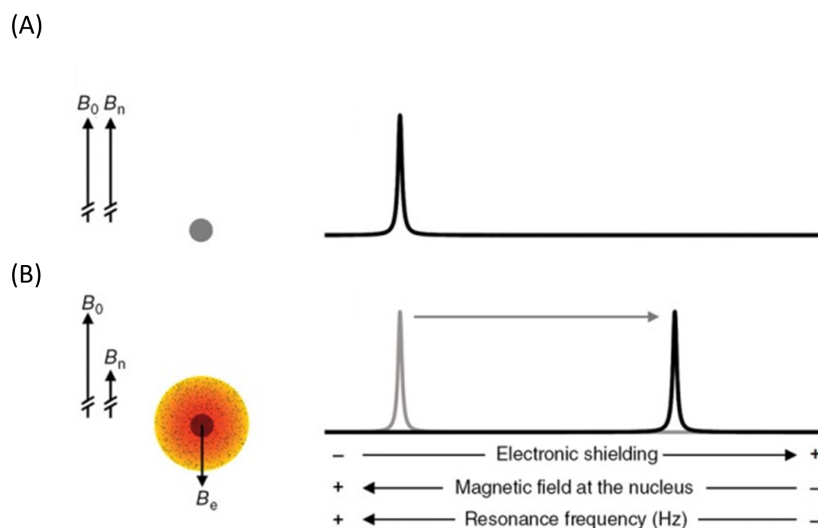
Figure 1.9: Chemical shift. (A) The magnetic field B_n at the nucleus of a single proton is equal to the external magnetic field B_0 leading to a spectrum with a signal at Larmor frequency $\omega_n = \frac{\gamma}{2\pi} B_n$. (B) The electron cloud in a given atom can be modeled as small currents that generate a magnetic field B_e that opposes B_0. The magnetic field at the nucleus is reduced, leading to a lower Larmor frequency. Modified from [1].

Scalar coupling, also known as *J-coupling*, is a type of spin-spin coupling that occurs as a result of the coupling of the magnetic moments of two or more nuclei within a molecule. In other words, the magnetic fields of the nuclei interact with each other, causing the energy levels of the nuclei to split into different energy states. Such splits into multiple energy levels are observed in NMR spectra by splitting resonances into several smaller lines. Scalar coupling is a quantum effect that originates from the fact that nuclei with magnetic moments can influence each other directly through electrons in chemical bonds.

Considering the two isolated nuclei in Fig.1.10(A): the blue one (spin H) resonates at a higher frequency v_H, whereas the red one (spin C) resonates at a lower v_C frequency. They display in frequency via two singlets at different energy levels and thus different resonating frequencies, Fig.1.10(D-top). The Fermi contact governs the interaction between the nuclear and electron spins within the nucleus and energetically favors an antiparallel over a parallel arrangement. As soon as the two nuclei are combined, see Fig.1.10(B), but chemically isolated, the resulting interaction affects the energy levels, which now represent the sum of the single energy levels. However, if these two nuclei are now supposed to be covalently bound, see Fig.1.10(C), the interaction between the two electrons spins inside the chemical bond must follow an antiparallel

orientation, as demanded by the Pauli exclusion principle. When both nuclear spins are parallel or antiparallel to the external magnetic field B_0, respectively $\alpha\alpha$ or $\beta\beta$ state, the two bonding electrons cannot both be antiparallel to the nuclear spins, leading to an energetically less favorable state. However, for the $\alpha\beta$ or $\beta\alpha$ conditions, both electron spins can be antiparallel to the nuclear spins leading to an energetically more favorable situation. The equivalent gain or drop in energy is proportional to the J-coupling constant, which measures the strength of the chemical bond between the two nuclei and therefore defines the energy level transition and the correspondent resonating frequencies. J-coupling yields unique spectral and temporal fingerprints for chemically coupled compounds, Fig.1.10(D-bottom). Such unique time evolutions can provide information about the structure and dynamics of a molecule.
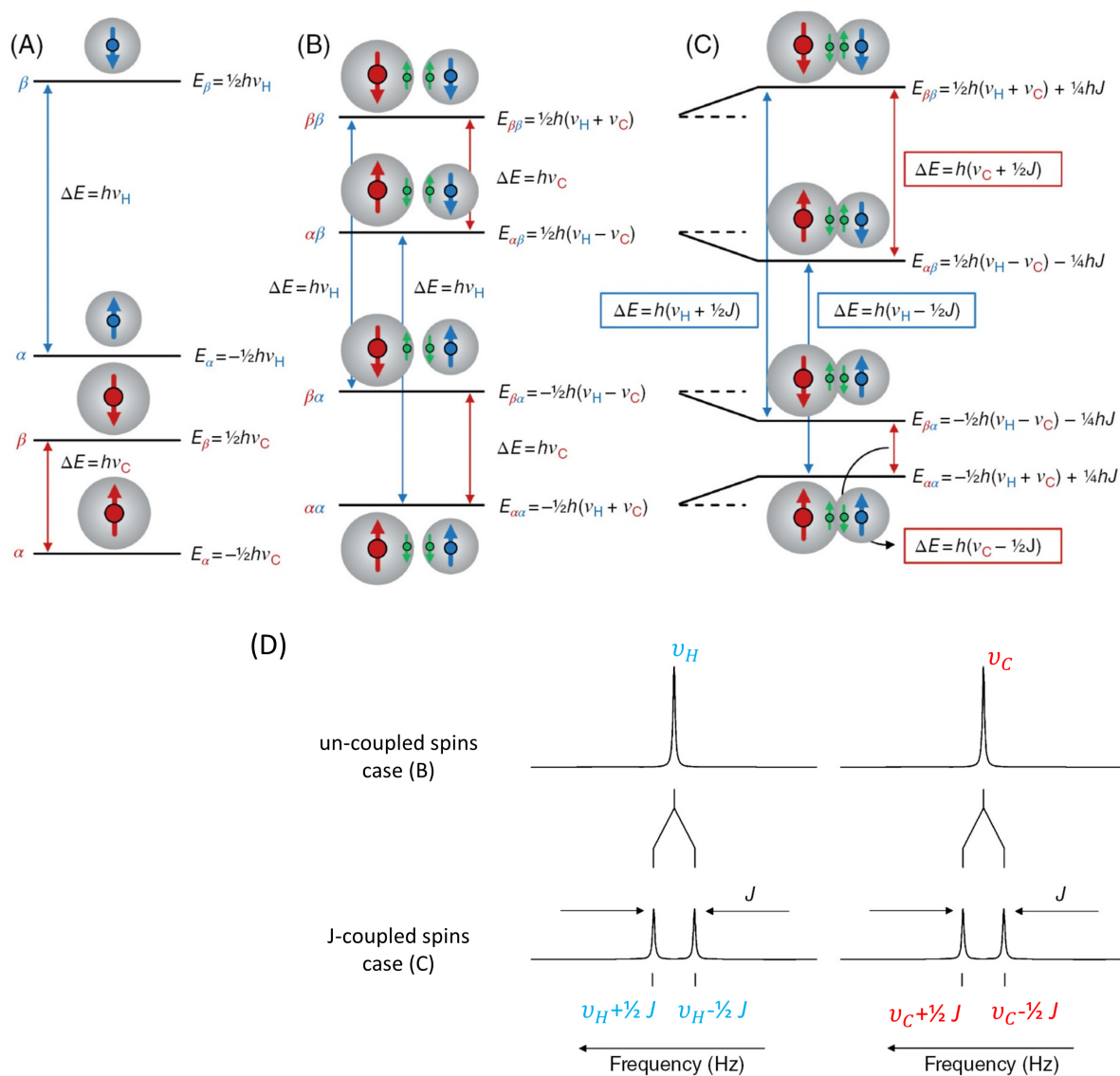
Figure 1.10: The effect of scalar coupling on the energy level of a two-spin system. Energy levels for (A) isolated or (B) combined spins H and C. In the combined mode, the energy levels represent the sum of the isolated energy levels. Transition frequencies v_H (blue) and v_C (red) yield the NMR resonating frequency (D, top). (C) Shows energy levels for spins H and C when chemically bonded. Spin states follow less or more energetically favorable states yielding splitting of the energy levels and consequence splitting of the NMR visible frequencies (D, bottom). Modified from [1].

2 | An illustrative NMR experiment

Alike grammar in any natural language, NMR follows a set of complex rules, which defines the meaning of the message at the end of this intriguing communication channel, and that is carried out between the two operators: the user, or radiologist, and the tissue of interest, or patient. Such rules are rather well defined by physics, as explored in Chapter 1, but need engineering to be applied in the real world.

An ideal phantom containing three compounds in three different partitions is placed at the isocenter of an MR scanner. The MR scanner generates a strong external magnetic field B_0, which homogeneity is perfect only in a given surrounding space around the magnet's isocenter. The three compounds feature (1) a different chemical structure that results in different Larmor frequencies (ω_1, ω_2 and ω_3), and (2) different concentrations in solution ($C_3 > C_2 > C_1$).

2.1. Signal localization

An NMR measurement assumes spatial localization to restrict the signal detection to a well-defined region of interest (ROI), isolating the acquired spectra from unwanted signals coming from outside the ROI (e.g., lipids). Moreover, it helps manage tissue heterogeneity, minimizing partial volume effects and giving a handler to B_0 and B_1 variations. MRS localization takes care of excitation and selection of spin pools of interest and is followed by the recording phase, where the receiver coils will detect the relaxation phenomenon. Overall, such a system results in an ordered set of RF pulses, gradients, and switches that are actioned according to a tailored design, which aims to control the spin pool of interest to encode any desired feature. This system is known as the *MR pulse sequence*.

For the purpose of the thesis, we will introduce the reader to the *semi-LASER* localization technique in both single-volume (Single-Voxel MRS), and multi-volume (MR Spectroscopic Imaging, MRSI) approaches. A semi-LASER localization consists of a single-scan 3D localization method

---

This chapter is partly inspired by [1]

that combines a frequency selective excitation pulse with four adiabatic refocusing pulses. In Fig.2.1, the MR sequence diagram is reported. First, a frequency-selective pulse is made slice selective by twinning to a selective gradient along the x direction, which yields a 3D slice selection (i.e., a slab).
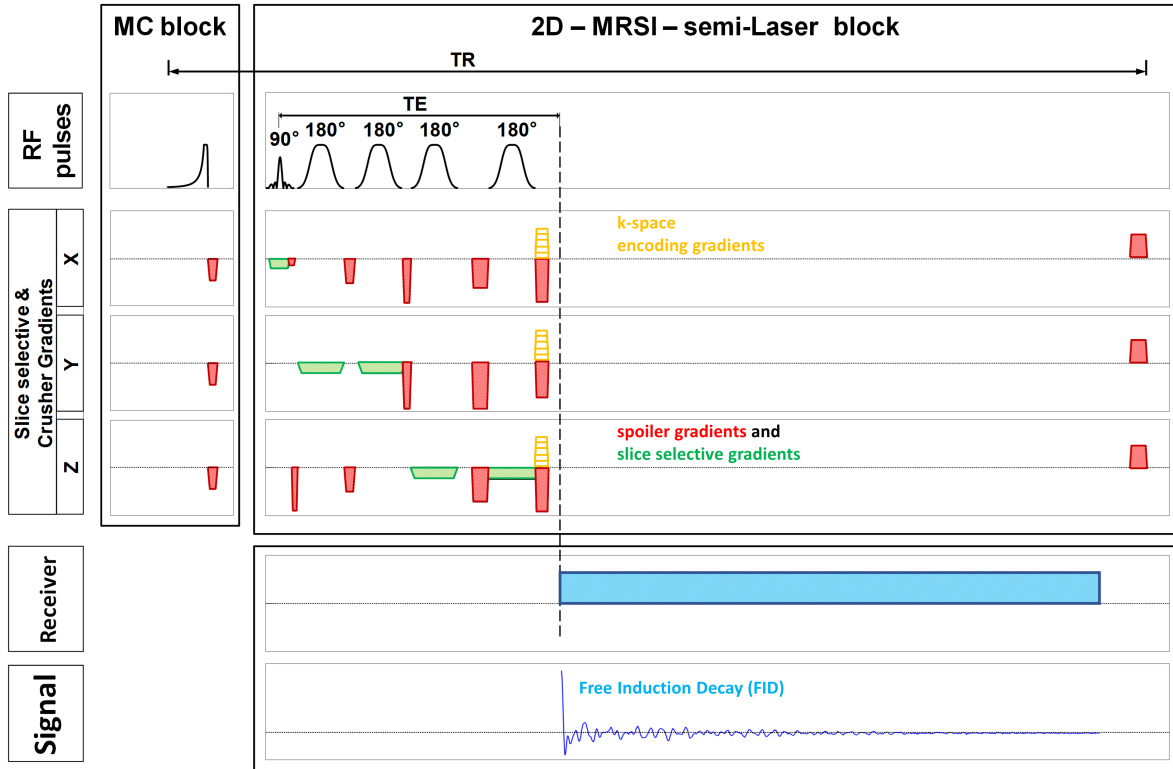
Figure 2.1: Display of a semi-LASER acquisition scheme.

Slice selective pulses

The logic behind slice selective pulses is the twinning of a frequency-modulated wave (i.e., the RF pulse) to a magnetic field gradient. Magnetic field gradients $G(\boldsymbol{r})$ describe a linearly dependent magnetic field $B(\boldsymbol{r})$ that vary according to $\boldsymbol{r}$, the position along a given axis

$$G(\boldsymbol{r}) = \frac{\partial B(\boldsymbol{r})}{\partial \boldsymbol{r}} = G \tag{2.1}$$

They are created by two looping currents with opposite directions across two coils orthogonally arranged to the desired gradient direction. The static magnetic field B_0 upon application of a gradient displays a linear position-dependent intensity $B(\boldsymbol{r}) = B_0 + \boldsymbol{G} \cdot \boldsymbol{r}$ which results in a position-dependent Larmor frequency: $\omega(\boldsymbol{r}) = \omega_0 + \gamma \boldsymbol{G} \cdot \boldsymbol{r}$. A selective RF pulse with a specific transmitter frequency offset ω_{RF} from ω_0, and a given frequency bandwidth $\Delta\omega$ is applied simultaneously with such magnetic field gradient, thus only the selected frequencies,

or inversely, a selected spatial position can be excited. The RF pulse amplitude and duration determine the resulting nutation angle in the net magnetization. The position of the slice is defined by the transmitter frequency of the RF pulse, and the RF pulse bandwidth adjusts the slice thickness and the gradient field amplitude G according to the following equation

$$\Delta L = \frac{\Delta\omega}{\gamma G} \tag{2.2}$$

An example of different slice selections and locations is given in Fig.2.2(A). A head 3D model reports the variation of the magnetic field along z directions, Fig.2.2(B), and a corresponding possible slab selection, Fig.2.2(C).
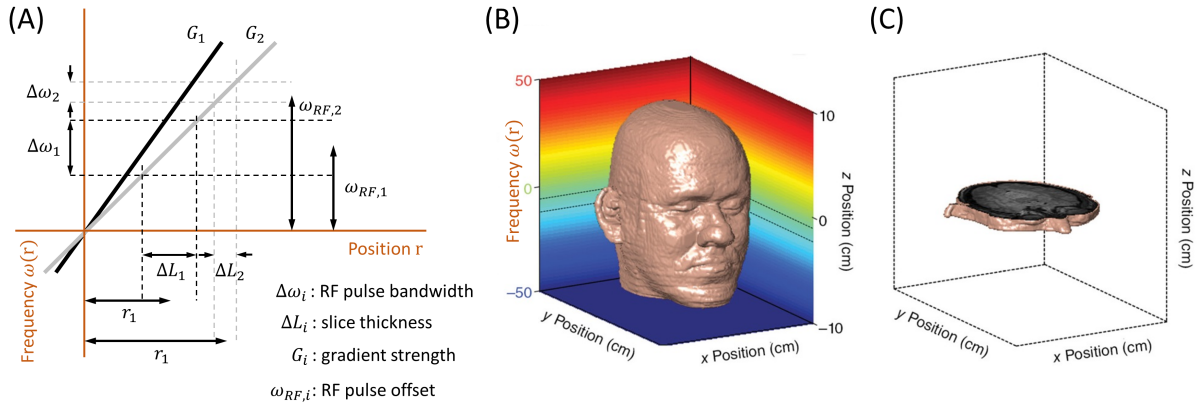
Figure 2.2: Slice selection via frequency-modulated RF and gradient strength modulation. (A) configuration of gradients and RF pulses, (B) frequency encoding, and (C) slice selection of spin visualized for a head 3D model. Extended from [1].

Adiabatic pulses

The second part of the ROI selection follows the principle of frequency-selective refocusing pairs of Adiabatic Full Passage (AFP) pulses. In short, AFP pulses are an exclusive class of RF pulses that allow uniform inversion (or refocusing) of magnetization, by employing frequency (or phase) and amplitude modulations. Adjusting frequency from off-resonance to on-resonance makes adiabatic pulses insensitive to the nonuniformity of the B_1 field and variation in frequency offset. However, the frequency modulation of a single adiabatic pulse induces a nonlinear B_1 and position-dependent phase across the slice, leading to severe signal cancellation. A second, identical AFP pulse can refocus the nonlinear phase such that perfect refocusing can be achieved. Two pairs of AFP pulses allow a 2D slice selection that intersects with the already selected slab, yielding the selection of a 3D volume of interest (or voxel), Fig.2.3. The advantages of the LASER technique over alternative short echo-time techniques, like STEAM and PRESS, are twofold. Namely (1) the method is entirely adiabatic and (2) by employing high-bandwidth AFP pulses, the localization can be exceptionally well defined, both in terms of minimal chemical shift

displacement, as well as sharpness of the localization edges.
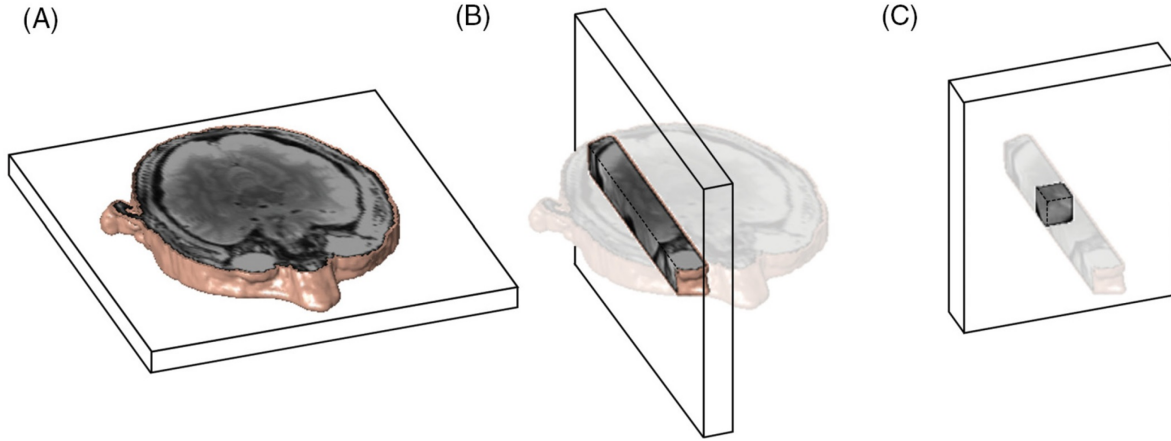
Figure 2.3: Principle of 3D voxel selection. Single voxel selection based on magnetic field gradients on consecutive directions: (A) 1D plane (or slab), (B) 2D column (or beam), and finally, (C) 3D voxel. Figure from [1].

2.2. Time and frequency domain

The result of localization is the spatial identification of a rotating transverse magnetization, as introduced in Section 1.2. Such a rotating magnetic field induces an oscillating electric current into a nearby receiving coil, formally known as FID. In short, by solving the Bloch equations describing the motion of the macroscopic magnetization, the resulting FID signal within the selected sample containing three compounds with different Larmor frequencies can be described with a complex function $s(t) = R(t) + j \cdot I(t)$ where the real $R(t)$ and imaginary $I(t)$ components are given by

$$R(t) = M_x(t) = \sum_{i=1}^{3} M_{0,i} cos(\omega_i t + \phi) e^{-t/T_{2,i}^*}$$

$$I(t) = M_y(t) = -\sum_{i=1}^{3} M_{0,i} \ sin(\omega_i t + \phi) \ e^{-t/T_{2,i}^*}$$

(2.3)

The FID is represented by the summation of the contribution of single compounds, each one described by four independent parameters:

- $M_{0,i}$: contribution of compound i to overall magnetization at thermal equilibrium;

- ω_i: Larmor frequency of compound i;

- ϕ: phase of the transverse magnetization relative to the x axis immediately after excitation and assumed common to all substances;

- $T_{2,i}^*$: transverse relaxation rate characteristic of compound i.

All four parameters describing the FID can be easily recognized in the case of a single Larmor frequency (single compound), Fig.2.4(A-C). The M_0 parameter is critical in many NMR applications as it is directly related to the concentration. For a single-frequency signal, M_0 is proportional to the intensity of the FID immediately after excitation, Fig.2.4(A-C, FID side).

The Fourier transform of the FID comes in aid, allowing visual distinction of the three different resonating frequencies, otherwise overlapping and proving hard or even impossible to untangle and identify observing the time domain FID, Fig.2.4(D). The correspondent Fourier transform can also be evaluated in its real and imaginary components, respectively, known as the absorption and dispersion part of the spectrum, Fig.2.4(A-D, spectrum side). The four parameters describing the time domain single compound FID can also be recognized in the absorption spectrum. The first point of the FID, M_0, equals the area under the curve of the absorption spectrum. The frequency ω_i of the signal can be recognized as the position on the frequency axis. $T_{2,i}^*$ relaxation time can be obtained from the frequency width at half maximum (FWHM) according to $FWHM = \frac{1}{\pi T_{2,i}^*}$.

The assumption on $T_{2,i}^*$ characterizes the nature of the line depicted in Fig.2.4. Typically, the contribution to $T_{2,i}^*$ is represented by *Voigt* line shapes. The summation of a Lorentzian and a Gaussian contribution gives a *Voigt* line. Variations of field inhomogeneities that are expected not to change over time are supposed to be symmetric to the echo-time TE. Thus, following a Gaussian distribution and modeled by the Gaussian term. The Lorentzian contribution can be further displayed as the sum of relaxation rates $1/T_{2,i}^* = 1/T_2' + 1/T_{2,i}$, where $1/T_2' = \int_r \Delta B(r)$ is the relaxation rate contribution attributable to field inhomogeneities ΔB across the voxel, here described as integral in a spherical volume of radius r. On the other hand, $1/T_{2,i}$ is the *effective* T_2 as the spin-spin relaxation property of the molecule i, see Section 1.2.1. The Lorentzian assumption for effective T_2 comes from the solution of the Bloch equation Eq.1.4. Under most *in-vivo* conditions, the inhomogeneity contribution to T_2^* relaxation is dominant, such that resonance line widths are approximately equal despite minor differences in T_2 relaxation.
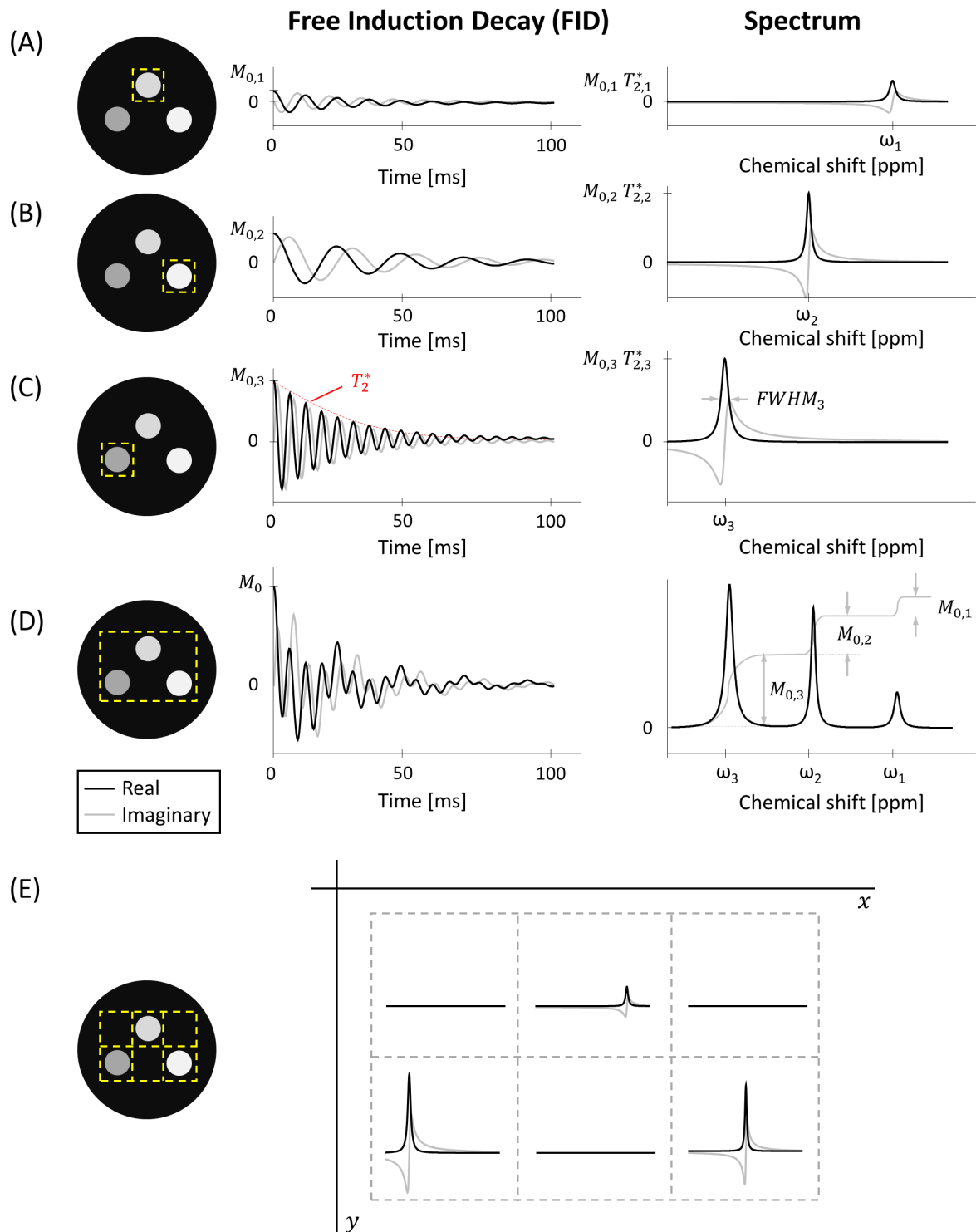
Figure 2.4: Acquisition of MR Spectra. A phantom with three distinct compartments is acquired with SVS selecting either single compartments (A,B,C) or all compartments (D). Free Induction Decay and correspondent spectrum are reported for each selection. (E) Shows a correspondent MRSI acquisition where the location of each compartment is encoded in the spatial coordinates x, y. Perfect water suppression is assumed.

Averaging

It must be mentioned that such acquisition of good quality spectra in Fig.2.4 may be feasible in a phantom, where the nuisance contributors are reduced. However, in an in-vivo setup, signal averaging is required to overcome the higher noise contribution, which hinders the detection of the metabolite fingerprint [3]. In signal averaging, the FID's signal-to-noise ratio (SNR) is improved by adding or averaging the FID signals of several consecutive, identical experiments (called signal transients or shots). Averaging N FID signals leads to an improvement in SNR of a factor $\sqrt{N}$. Consequently, an in-vivo NMR experiment is a compromise between sufficient SNR and the allowable duration of the experiment [1].
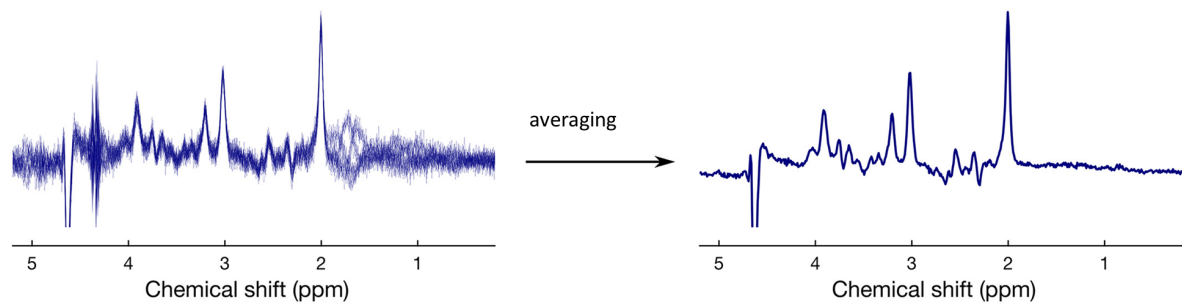
Figure 2.5: Effect of averaging for in-vivo MRS (3 T human brain MRS data, TE=68 ms). Improved SNR is observed after averaging single transients. Single transients are affected by minor spurious echoes around 1.8 ppm and 4.3 ppm. However, following averaging of the phase cycled scans, these are effectively removed so that this spectrum can be safely analyzed. Modified from [3].

2.3. Water and lipid suppression

Fig.2.4 may fool the reader if one assumes the synthetic phantom, where the compounds differ from water but are in solution, assuming water as a solvent (typically the case). In general, one should consider that the water content in mammals amounts to 60% circa, making it the most abundant compound in mammalian tissues. Thus, the water concentration level in tissue is more than 10^4 higher than the metabolite concentrations. Therefore, care in treating the water signal must limit unwanted contributions to the acquired spectra that could make metabolites indistinguishable. The water signal in a ^{1}H-MR spectrum resides at 4.65 ppm and is typically affected by shorter T_2 relaxation; hence it appears broader than typically well-acquired metabolite spectra. The presence of a vast water signal leads to baseline distortions in the MR spectrum, hampering metabolite signal investigations, Fig.2.6(A). Suppression of the water signal or clever handling of its contribution is required to eliminate baseline distortions, yielding a reliable and consistent detection of metabolite resonances in tissues [3]. Conventional approaches to removing the water contribution rely on water pre-saturation, deploying pulse trains before signal localization [4].

In Fig.2.1, a metabolite-cycling pulse (i.e., MC block) is played out before the semi-LASER localization. This non-water suppression method relies on the interleaved acquisition of isolated water and metabolite spectrum. An MC pulse is a frequency-selective pulse that, in two consecutive scans, generates alternating signal polarization of metabolite resonances. Either upfield (chemical shift < 4.4 ppm) or downfield (chemical shift > 5.0 ppm) are inverted. Thus, the summation of two consecutive shots yields the water signal only, whereas subtraction results in the metabolite spectrum. This method yields interleaved water signals that can be exploited for robust post-processing steps, like phase/frequency alignment and eddy current correction.

Moreover, lipid resonances within 1-2 ppm may become dominant in the proton MR spectrum and contribute to baseline distortion in organs like the prostate, which are embedded in fatty tissues. The spectral region of Citrate is regularly contaminated with lipid signals, especially because periprostatic lipids are often inside the ROI [5]. In such cases, signal suppression (for both water and lipids) can be well achieved by exploiting frequency-selective suppression pulses, like MEGA pulses, Fig.2.6(B). MEGA suppression relies on a pair of inversion pulses placed within the localization block in an element whose function is to refocus transverse magnetization. The advantages of MEGA suppression over other suppression techniques are minimizing T_1 recovery of the suppressed signals, adding minimal spectral distortion, and its robustness to field inhomogeneities or incorrect RF calibration [6]. On the other hand, care on signal interpretation must follow since the MEGA pulses also influence all other signals resonating within their selected bandwidth. This may also impact the signal intensity and shape of other protons in these molecules (i.e., signal editing).

(A)

(B)
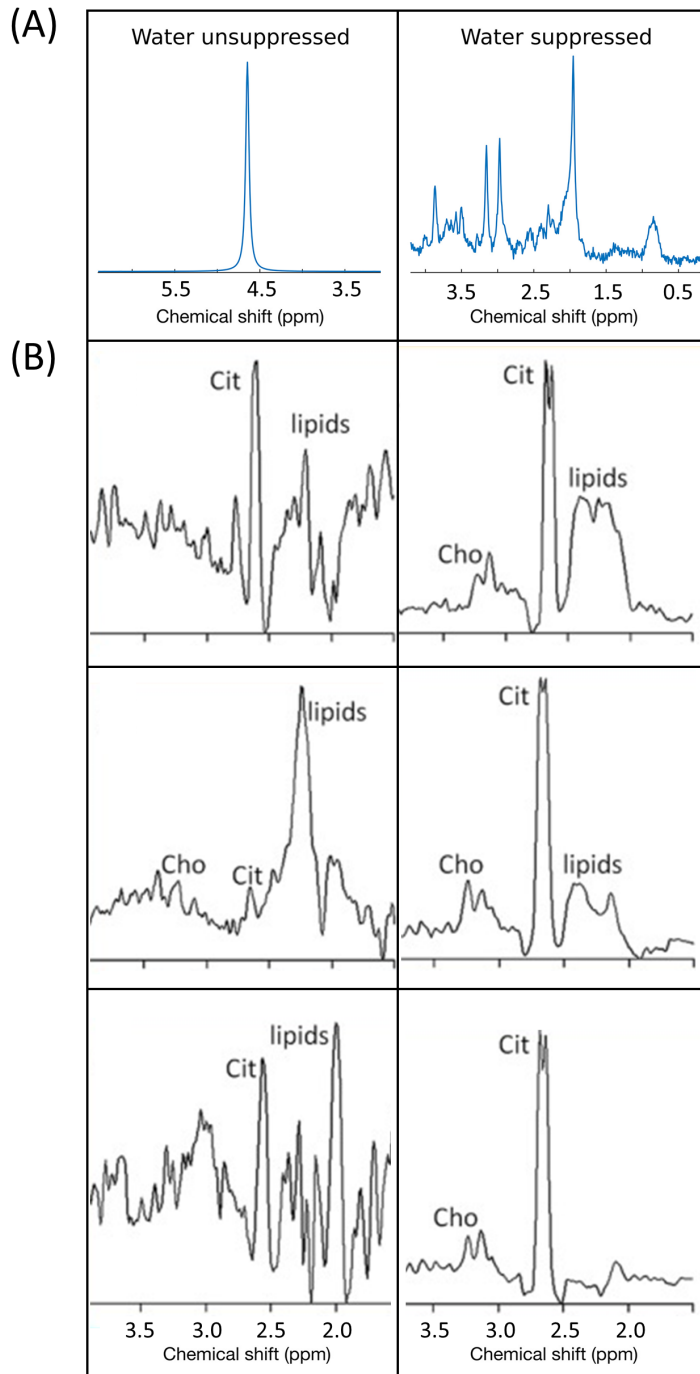
Figure 2.6: (A) effect of water suppression and brain metabolite spectra visibility. (B) Effect of lipid suppression on prostate spectra. (left) poor lipid suppression and substantial contamination, (right) good lipid handling, and good spectral quality. Metabolite of interests: Cho - Choline, Cit - Citrate. Figure modified from [3, 5].

2.4.   MR Spectroscopic Imaging

The advantages of single-voxel MRS are

- the volume is typically well-defined with minimal contamination (e.g., extracranial lipids in brain MRS);

- the magnetic field homogeneity across the volume can be readily optimized, leading to

- improved water suppression and spectral resolution.

However, the main disadvantage is seen in Fig.2.4(A-C): if the single volume focuses on a small part of the object (like compartment A), the signal coming from other parts (like compartments B and C) will be missed. On the other hand, multi-volume or multi-voxel localization, known as magnetic resonance spectroscopic imaging (MRSI), allows the encoding and detection of metabolic profiles from multiple spatial positions, thereby offering a complete characterization of the entire object under investigation, Fig.2.4(E). On certain aspects, however, MRSI is technically more challenging due to

- significant magnetic field inhomogeneity across the entire object,

- spectral degradation due to intervoxel contamination,

- typically longer data acquisition times, and

- the processing of large multidimensional datasets.

The single voxel semi-LASER described will now be extended to process MRSI data and will also provide an illustrative example of MRSI concepts. However, advanced sequence design and optimization will not be discussed since they are outside the purpose of the current work. In short, MRSI requires the extension of phase and frequency encoding from MR Imaging (MRI) to display spatial information through *k-space* encoding. However, in traditional MRSI, frequency encoding is avoided since adding time-varying carrier frequency will strongly alter the intrinsic Larmor frequency of the single compounds of interest, yielding unreadable spectra. However, advanced reconstruction techniques based on fast MRI can nowadays rely on frequency encoding as well for MRSI.

The concept of k-space can be quickly introduced considering the overall description of the signal $S(t)$ coming from an ROI selecting all the phantom's compartments A, B, and C in Fig.2.4. $S(t)$ represents, as depicted in Fig.2.4(D), the sum of signals from the elementary volumes $s(\boldsymbol{r}, t)$, from each point $\boldsymbol{r}$ in the sample (each compartment can be assumed to have x, y coordinates as indicated in fig Fig.2.4(E)). Mathematically, it can be described as

$$S(t) = \int_{-\infty}^{\infty} s(\boldsymbol{r}, t) d\boldsymbol{r} \tag{2.4}$$

The overall spectrum $F(\omega)$ is calculated via Fourier transform of $S(t)$ and equals the sum of spectra from elementary volume elements $f(\omega, \boldsymbol{r})$:

$$F(\omega) = \int_{-\infty}^{\infty} S(t) e^{-j\omega t} dt = \int_{-\infty}^{\infty} f(\omega, \boldsymbol{r}) d\boldsymbol{r} \tag{2.5}$$

The application of phase encoding gradient will induce a phase shift on each elementary volume according to

$$f(\omega, \boldsymbol{r}) = f(\omega, \boldsymbol{r}) e^{-j\gamma \boldsymbol{r} G t} \tag{2.6}$$

And such that the whole spectrum can now be described as a function of ω and G:

$$F(\omega, G) = \int_{-\infty}^{\infty} f(\omega, \boldsymbol{r}) e^{-j\gamma \boldsymbol{r} G t} d\boldsymbol{r} \tag{2.7}$$

With the introduction of $k = \gamma G t$ (valid for rectangular gradient waveforms G), the *k-formalism* is in place.

$$F(\omega, k) = \int_{-\infty}^{\infty} f(\omega, \boldsymbol{r}) e^{-jk\boldsymbol{r}} d\boldsymbol{r} \tag{2.8}$$

The phase-modulated spectra of the entire sample $F(k, \omega)$ represent the inverse Fourier transformation of the spectra $f(r, \omega)$ from the individual volume elements.

$$f(\omega, \boldsymbol{r}) = \int_{-\infty}^{\infty} F(\omega, k) e^{-jk\boldsymbol{r}} d\boldsymbol{r} \tag{2.9}$$

Where $\boldsymbol{r}$ represents any direction in a 3D volume and can be extended to Cartesian directions by applying orthogonal gradients G_x, G_y, and G_z. For a 2D MRSI readout, one would need two phase-encoding gradients G_x and G_y to yield spectra $f(\omega, x, y)$. Such gradients are indicated in Fig.2.1 and predisposed right before readout acquisition. It is noted that a more general definition of $k(t) = \gamma \int_0^t G(t) \, dt$ applies when time-varying gradient waveforms are deployed.

Although SVS is appropriate for the investigation of a focal lesion, a specific anatomical region, or diffuse brain disease, MRSI is preferred when the location of interest is uncertain or multiple areas need to be evaluated simultaneously, such as when investigating metabolite distributions across heterogeneous lesions due to a tumor [7].

2.5.   Digitalization

All the equations that have introduced the concepts of FID and spectra assume continuous sampling of signal either in time or k-space domain. However, in reality, signals are digitally sampled at discrete n-time or k-space positions. Therefore, their definition impacts the quality of acquired data through different definitions of resolution.

Spectral resolution

The minimum sampling rate to untangle and correctly digitize oscillating frequency (and avoid aliasing) in the interval $[-F_N, F_N]$ is defined by Nyquist theory, for which the spectral bandwidth SW must be equal to $2F_N$. Such sampling is run over a given number of N data points spaced by dwell time $\Delta t = 1/SW$. Following the Fourier transformation of the FID, the spectrum will also contain N points spread out over the spectral width SW. Therefore, the digital spectral resolution is SW/N, which is equivalent to the reciprocal of the total acquisition time, i.e., $1/T_{acq}$. The digital resolution can be improved by increasing the total acquisition time T_{acq}. A smaller SW limits the minimum width required to avoid aliasing, whereas increasing the number of acquisition points N comes at the cost of a decreased SNR as more noise and less signal is acquired. Zero-filling is the post-processing step that simulates the extension of the acquisition time by adding a string of points with zero amplitude to the FID. While it can significantly improve the visual appearance of spectra, it does not increase the information content of data. Still, it impacts signal modeling and the following quantification approaches, especially regarding assumptions on noise distribution.

Field-Of-View (FOV) resolution

The increments in gradient area ΔG to encode $k-$locations determine the digitization rate in the spatial frequency or $k-$space domain. Nyquist sampling yields the maximal phase shift difference between two increments over the entire FOV to avoid aliasing:

$$FOV = \frac{2\pi}{\gamma \Delta G t} \tag{2.10}$$

The nominal voxel size is directly related to the FOV and the number of phase encoding increments N (number of acquired $k-$point along a given direction): $\Delta V = FOV/N$. In practice, the minimum voxel size is, besides the FOV, determined by the allowable measurement time and sensitivity. For conventional MRSI, which is approached in this work, the encoding of $N_x \times N_y \times N_z$ volume elements (voxels) requires $N_x \times N_y \times N_z$ separate acquisitions, each with a repetition time TR. Temporal resolution can be increased conventionally, combining (1) elliptical (or spherical) acquisition and (2) $k-$space apodization during acquisition. The elliptical acquisition consists of the acquisition of $k-$space low-spatial-frequency coordinates only. The high-spatial

coordinates, which contribute less to the bulk of the observed signal, are ignored, reducing the total measurement time by about 21.5% for 2D MRSI. $K-$space apodization during acquisition can be performed by running fewer signal averages at high $k-$space coordinates. This optimizes the acquisition time and yields a gaussian-weighting on the Point-Spread-Function (PSF), which significantly minimizes inter-voxel signal contamination.

2.6. Analysis of MR spectra

An in-vivo MRS experiment aims to estimate the relative or absolute concentrations of tissue metabolites within a specific anatomical region of interest. Once the time domain MRS data have been acquired, several steps are needed to obtain meaningful and reliable concentration estimates. First, a series of preprocessing steps should be applied to prepare the spectrum for analysis. Next, analysis of the processed dataset is performed, often by peak fitting, to estimate the metabolite signal intensities. Finally, the unitless signal intensity measures are converted into scaled concentration estimates to enable meaningful interpretation and comparisons of tissue metabolite levels between subjects and groups, regardless of the acquisition site or other measurement conditions.

2.6.1. Preprocessing

Preprocessing describes any series of operations applied to the acquired raw MRS data to prepare them for analysis. There are three main reasons for preprocessing in MRS.

- Remove/correct spectral imperfections: MRS data are unavoidably degraded by experimental imperfections (e.g., eddy currents, scanner drift, subject motion) [8]. Since spectral fitting models generally do not consider all of these imperfections, some preprocessing operations are needed to remove the imperfections, to the extent possible, in advance.

- Reduce data dimensionality: raw data are almost always multi-dimensional, with multiple signal averages acquired by multiple coil channels from parallel receive array coils. Thus, some preprocessing is needed to combine these signals and reduce the data dimensionality.

- Other preprocessing operations: such as Fourier transformation, phasing, apodization, and zero-filling, are not strictly related to data quality but can be used to aid in visual interpretation or peak fitting performance.

This work will only briefly describe techniques for removing or correcting spectral imperfections. A glimpse at some of the other possible preprocessing steps has already been introduced. Furthermore, it is noted that only traditional processing strategies will be mentioned in this section. As in many different fields, novel preprocessing methods relying on Machine Learning (ML) are being implemented and introduced to the MRS community. However, these methods are still

relatively new and may require further development before being deployed widely. Therefore they are not yet acknowledged in routines despite their State of the Art (SOTA) performances. These strategies will be mentioned in Section 5.3.

Eddy currents

Rapid gradient switching gives rise to unwanted short-lived (i.e., up to hundreds of milliseconds) fluctuations of the B_0 field, called *eddy currents*. For example, suppose the acquisition window occurs close to the end of a gradient pulse. In that case, a multi-exponential decaying B_0-field component may exist during the early part of the acquisition window, giving rise to a time dependence in the resonance frequencies of the acquired FID. This unwanted effect severely distorts spectral line shapes, Fig.2.7. The acquisition of an unsuppressed water spectrum using identical gradient strength and timing configuration aids the identification of such an effect. Any time dependence in the frequency of the water signal is monitored by non-linearity in phase. Subtracting such phase variation from water and metabolite signals yields eddy current correction [9].
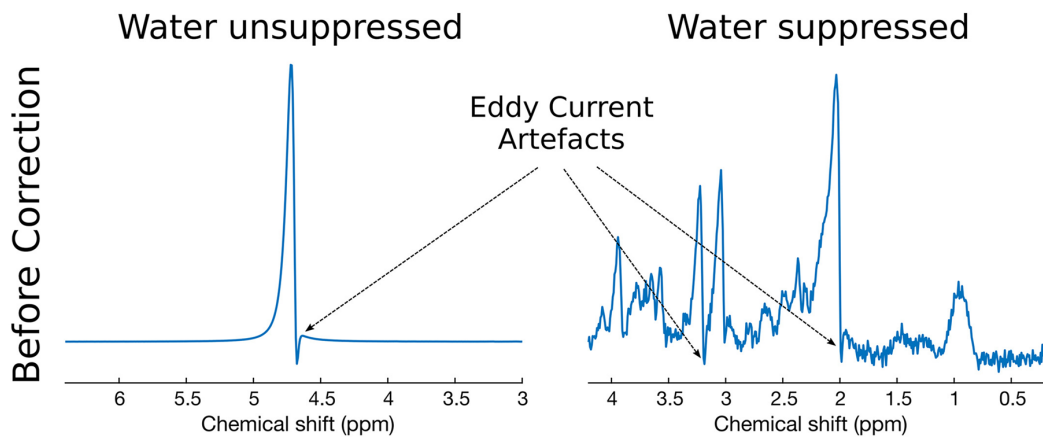
Figure 2.7: Display of eddy current artifacts in water-unsuppressed and metabolite spectra. Modified from [3].

Motion

Despite all efforts to control subject motion, a small amount of motion is practically inevitable in the timeframe of an MRS scan. Tiny amounts of motion, for example, due to normal physiological motion (breathing, cardiac pulsation, swallowing) or small bulk movements of less than a few millimeters, result in small changes in the frequency and phase of the individual transients. A retrospective processing algorithm can easily correct for such drifts. However, in the case of large amounts of motion, such as the gross motion of the head or limbs, much larger spectral distortions are observed. In addition, severe gross motion can lead to unwanted tissue sampling outside the region of interest. In this case, it may be necessary to remove the transients that

are most severely affected by motion or discard the dataset altogether. Although highly promising, different prospective motion correction strategies for ^{1}H-MRS (e.g., navigator images or optical tracking) are not yet in mainstream use, and thus retrospective correction methods are commonly employed [10].

Frequency and phase drifts

The main magnetic field of an MRI scanner, B_0, is subject to subtle temporal drift due to the heating and cooling of the ferromagnetic passive shim elements, which are in thermal contact with the gradient coils. As a result, frequency drifts are observed during most MRS experiments. The magnitude of these drifts can vary significantly from scanner to scanner, day time, and is also affected by the use of gradient-intensive pulse sequences before the MRS scan, but ranges typically between 1 and 10 Hz of total drift during a typical MRS acquisition. Moreover, as mentioned above, physiological or small bulk motion during the MRS scan leads to additional frequency and phase offsets, independent of scanner drift. If not corrected, these frequency and phase drifts will lead to broader spectral peaks, signal-to-noise ratio (SNR) reduction, and line shape distortion. Several methods exist for retrospective correction of frequency and phase drifts. The one used during the developments of the current work involved either (1) tracking the frequency and phase of the residual water peak to estimate and correct the frequency and phase drifts or (2) relying on the water-unsuppressed signal deploying metabolite-cycling schemes.

Nuisance peak removal

In-vivo ^{1}H-MRS acquisitions are specifically designed to suppress nuisance signals such as water and outer volume signals, as briefly mentioned in Section 2.3. However, perfect suppression is challenging, as the signals to be suppressed are usually orders of magnitude larger than the signals of interest. As a result, noticeable contamination of spectra is relatively common. It varies with the region of interest, and minimizing signal contamination is an essential aspect of data preprocessing.

Poor water suppression can be handled in two ways:

- residual water signal can be removed before spectral analysis by fitting the peak to a series of line shape components via singular value decomposition (SVD) and then subtracting the resulting fit from the spectrum (i.e., the state-of-the-art algorithm is known as Hankel-Lanczos SVD, HLSVD [11]);

- residual water is not removed, but the analysis is performed using a fitting model incorporating water peak modeling.

Lipid contamination, like poor water suppression, can impact spectral quantification and be dealt with similarly. The main difference is that contaminating lipid peaks are generally much

broader than residual water peaks and often overlap with metabolites of interest, making lipid contamination more challenging to correct in preprocessing. Suppression during signal recording or inclusion of lipid basis set in modeling are currently the best practices. See Section 2.3.

Spurious echoes (also referred to as *ghosts*) are another commonly observed nuisance signal in in-vivo ^{1}H-MRS, Fig.2.8. These are typically caused by unwanted coherence pathways and often originate from tissues outside the region of interest. The issue of unspoiled coherences is best dealt with by modifications to the acquisition, such as changing the timing or amplitudes of the spoiler gradients, improving B_0 homogeneity, or improving the phase cycling scheme. If spurious echoes occur near the end of an FID, apodization can significantly reduce their appearance, but requires care in the follow-up analysis routines.
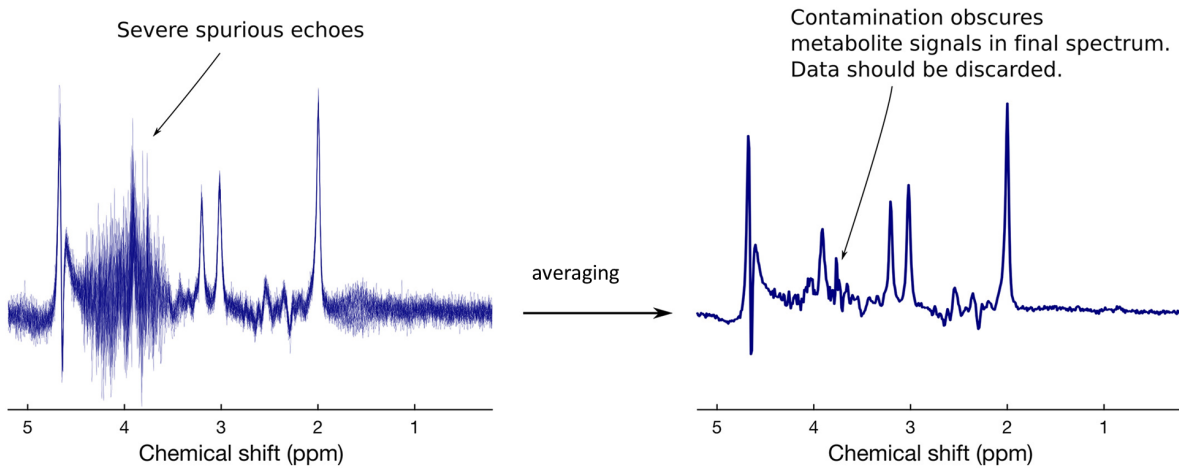
Figure 2.8: Visualization of spurious echoes and signal contamination. Modified from [3].

2.6.2. Quantification

After preprocessing, the spectrum is ready for quantification. The goal of quantification is, first, to estimate the spectral peak areas (or equivalently, time-domain signal amplitudes) of the various metabolites of interest in the spectrum, as well as that of some reference signals. Second, these quantities will then be converted to meaningful concentration units.

Linear Combination Model (LCM) fitting

In linear combination model fitting, each metabolite's contribution to the overall spectrum is modeled as a single response function called *basis spectrum*. The basis spectrum describes an individual metabolite's full spectral contribution and can be obtained either by phantom experiment or by numerical simulation. For this work, numerical quantum mechanical simulations for each metabolite k were used to generate the basis spectrum $\xi_k^{num}(t)$ [12, 13]. The input for such calculations is the spin system (i.e., chemical shift and j-coupling between all spins of

the $k-$compound), the pulse sequence, and the strength of the external magnetic field. The simulation of complex interacting spin systems follows the quantum mechanics *density matrix formalism*, introduced in Section 1.1.2 and is currently implemented in various publicly available NMR/MRS software packages. Sometimes, ideal pulses are used for simulations to the extent that the real pulse profile is superfluous (i.e., yielding minor inaccuracies)[14]. When simulations are used as a basis set, signal relaxation is typically excluded and fully modeled as parameters in the fit [15, 16].

Once the full set of metabolite basis spectra (the basis set) is produced, a constrained non-linear least-squares analysis can be used to fit a linear combination of the basis spectra to the acquired MR spectrum, by adjusting their individual parameters: typically amplitudes A_k and frequencies ω_k. Additional metabolite-specific or global terms, such as spectral phase ϕ and linewidth (either modeled as Gaussian broadening β and/or Lorentzian broadening α), are normally included to improve the fit. The relative amplitudes (or weights) of the various metabolite basis spectra in the best fit correspond to the estimated relative signal intensities. The time-domain mathematical expression of LCM for a given metabolite spectrum $\xi_{met}(t)$ follows

$$\xi_{met}(t) = \sum_k A_k \cdot e^{j(\omega_k t + \phi)} \cdot e^{-\alpha_k t - \beta^2 t^2} \cdot \xi_k^{num}(t) \tag{2.11}$$

where line broadening is characterized by Voigt line shape (i.e., the contribution of both Gaussian and Lorentzian decay).

The formalism of LCM models the acquired signal $s(t)$ by

$$s(t) = \xi(t, \boldsymbol{\theta}) + w(t) \tag{2.12}$$

where ξ is intentionally used instead of ξ_{met} alluding to a more general model, for example, including baseline contributions, and is parameterized by $\boldsymbol{\theta}$, the parameter vector space. $w(t)$ represents the measurement noise, typically assumed white Gaussian distributed. A constrained non-linear least-squares approach yields the estimated parameter in the following minimization process called χ^2 minimization:

$$\hat{\boldsymbol{\theta}} = argmin_{\boldsymbol{\theta}}\Big(\xi(t, \boldsymbol{\theta} - s(t)\Big)^2 \tag{2.13}$$

The fit residual (i.e., the difference between the fit and the data) indicates the goodness of the fit or the presence of unmodeled peaks in the spectrum. Visually, suppose the model sufficiently explains the data. In that case, residues should distribute as the measurement noise, yielding a white Gaussian process (i.e., no bias, and residue variance equal to the measurement noise power). Quantitatively, this can be expressed using the fit quality number (FQN), which is the ratio of the variance in the fit residual divided by the variance in the pure spectral noise. The

FQN should be close to 1 for an ideal fit.

To reduce the grade of the complexity of such models, *prior-knowledge* is applied. That consists of enforcing fixed relationships between the model parameters, reducing their degree of freedom. Linear combination model fitting allows for the inclusion of broad macromolecular components (characteristic of brain spectra, see Chapter 3.1, as well as baseline components (often either *model-free* spline functions or Voigt-lines) to account for any remaining broad, unmodeled background signal contributions. However, since the unmodeled baseline components are poorly characterized by definition, these estimates often represent the most significant source of uncertainty in fitting models. An example of a fitted spectrum with insights through single fitted basis set components is displayed in Fig.2.9.

Converting ^{1}H-MRS signal amplitudes A_k into metabolite concentrations C_k entails comparing the metabolite signals with a chemical concentration reference. For this work, only the internal water signal is used as the reference, but alternatives are available [3]. Metabolite levels can be reported simply as a ratio to the reference metabolite and not corrected for relaxation or partial volume effects. When referenced to a known concentration standard and corrected for partial volume and relaxation effects, concentrations are reported in terms of *absolute* concentrations, using either *molar* (moles per L of solution) or *molal* units (moles per kg of tissue).

Partial volume correction accounts for the sources of the estimated signal. For instance, when using water referencing in the brain, one should consider from which compartment water is coming: gray matter - GM, white matter- WM, or cerebrospinal fluid - CSF. In fact, each compartment is characterized by different tissue water densities. If the volume fraction of CSF is minimal, metabolite contribution coming from this tissue is typically negligible. Likewise, correction for relaxation must include the fact that water proton signals relax at different rates in different tissue types, as do metabolite protons. Accounting for both partial volume and tissue-specific water relaxation leads to the following expression for *molal* concentrations $[C_k]$:

$$[C_k] = \frac{A_k}{A_{H_2O}} \frac{f_{GM}^{H_2O} R_{GM}^{H_2O} + f_{WM}^{H_2O} R_{WM}^{H_2O} + f_{CSF}^{H_2O} R_{CSF}^{H_2O}}{1 - f_{CSF}^{H_2O} R_k} \frac{2}{N_k} [H_2O] \qquad (2.14)$$

where $[H_2O]$ is the molal concentration of pure water; $f_x^{H_2O}$ and $R_x^{H_2O}$ model the water volume fraction or water relaxation rate of tissue type x, respectively, while R_k is the metabolite-specific relaxation rate. Such rates depend on the acquisition protocol. For a typical spin echo experiment (like semi-LASER or STEAM), they are approximated by $R = e^{-TE/T_2}(1 - e^{-TR/T_1})$ when $TR >> TE$. The conversion to *molar* units needs the inclusion of water tissue densities in Eq.2.14, which allows conversion from tissue volume to water mole fractions.
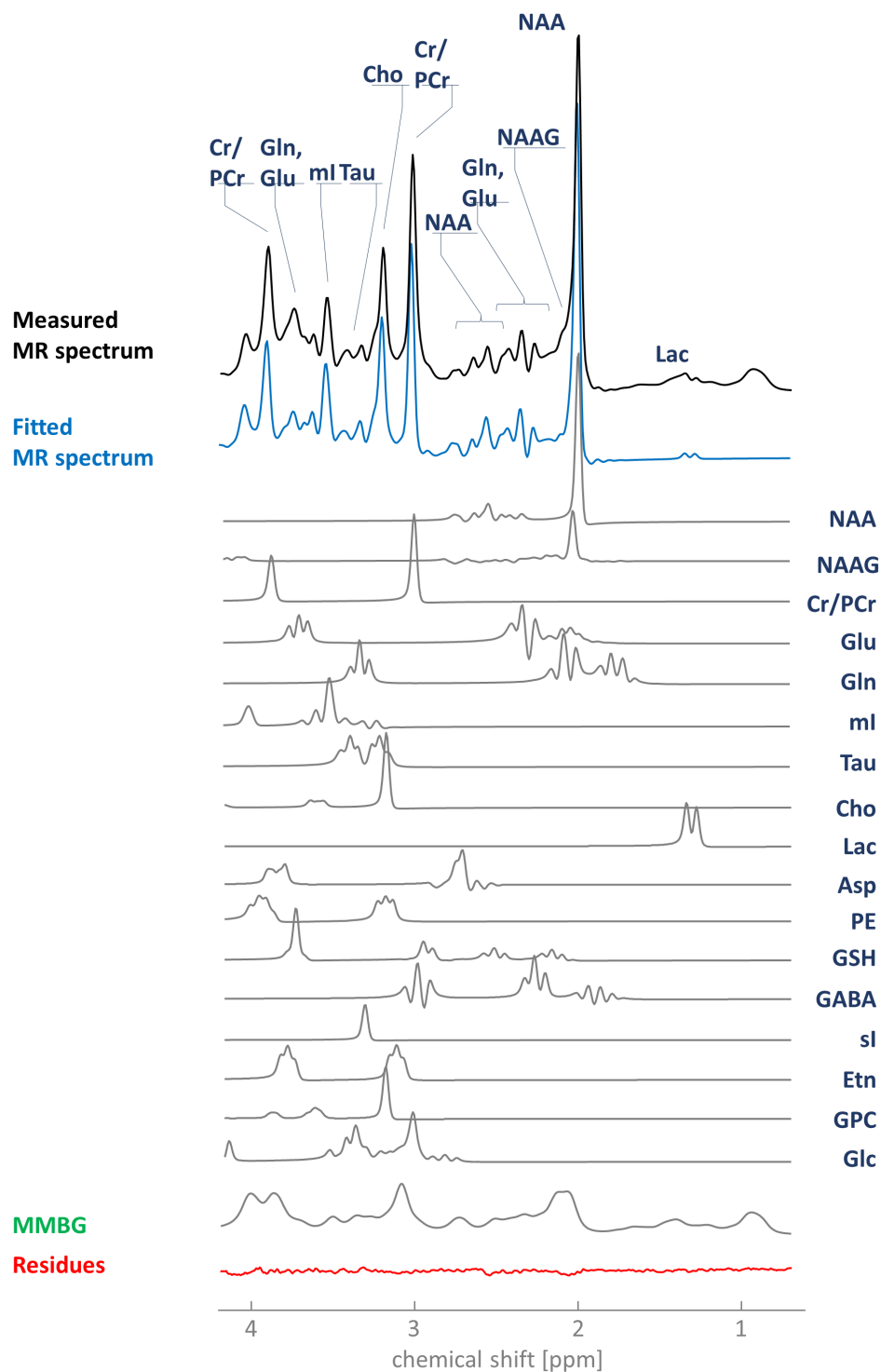
Figure 2.9: LCM fitting on an OCC brain spectrum (3T, semi-LASER). Spectrum (black) and fitted model (blue). The fitted model results from the superimposition of the single metabolite basis set fitted to the data. The macromolecular background (MMBG) is modeled separately (green), and residues display like white noise (red). Metabolite nomenclature and detail on MMBG are reported in Chapter 3.1.

Cramér-Rao Lower Bounds

To estimate its expected spread, every quantified value should be presented alongside a meaningful error metric, such as its standard deviation (indicated as STD, SD, or σ) [17]. However, as MRS is a time-limited modality, it is not typically feasible to perform multiple repetitions of the same experiment to calculate the standard errors. As such, the field has adopted the theoretical Cramér-Rao lower bound (CRLB) as the alternative to circumvent this limitation [18].

CRLBs are lower bounds for the standard deviation of an estimated parameter. For example, from information theory, given a maximum likelihood estimator, like the one in Eq.2.13, exists a lower bound on the statistical errors $\sigma_{\boldsymbol{\theta}} = \sqrt{E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2]}$ that, irrespectively of the model used to estimate the values of the parameters, can not be superseded [19]. $E[\cdot]$ indicates the statistical expected value.

The CRLB of the parameter $\theta_i \in \boldsymbol{\theta}$ is defined as

$$CRLB_{\theta_i} = \sqrt{(F_{ii})^{-1}}, \tag{2.15}$$

where F is the Fisher information matrix $F = \Re(D^H D)$ where $\Re$ indicates the real part of a complex number, H denotes the Hermitian conjugation, and D the partial derivatives of the model function ξ, with closed formula calculation provided in Eq.2.12 upon assumption of gaussian noise:

$$D = \frac{1}{\sigma_i} \frac{\partial \xi(t, \theta_i)}{\partial \theta_i}. \tag{2.16}$$

It can be show that the inequality $\sigma_{\theta_i} \geq CRLB_{\theta_i}$ holds [20]. A visual aid to the definition of CRLBs is given in Fig.2.10 where the calculation is simplified on a scalar parameter space θ and displays the *likelihood* function L for the whole set of continuous estimated parameters [21].

Likelihood function: $L(\theta|x) = P_\theta(X = x)$

$\hat{\theta}_{ML}$

$L(x|\theta)$

Set of data #1
Set of data #2

$\dfrac{\partial L}{\partial \theta}$

$\dfrac{\partial^2 L}{\partial \theta^2}$

$\dfrac{\partial^2 L}{\partial \theta^2}$

$\dfrac{\partial^2 L}{\partial \theta^2}$

$$\sigma^2(\hat{\theta}_{ML}) \geq I^{-1}(\theta) = -\left[E \frac{\partial^2 L}{\partial \theta^2}\right]^{-1}$$
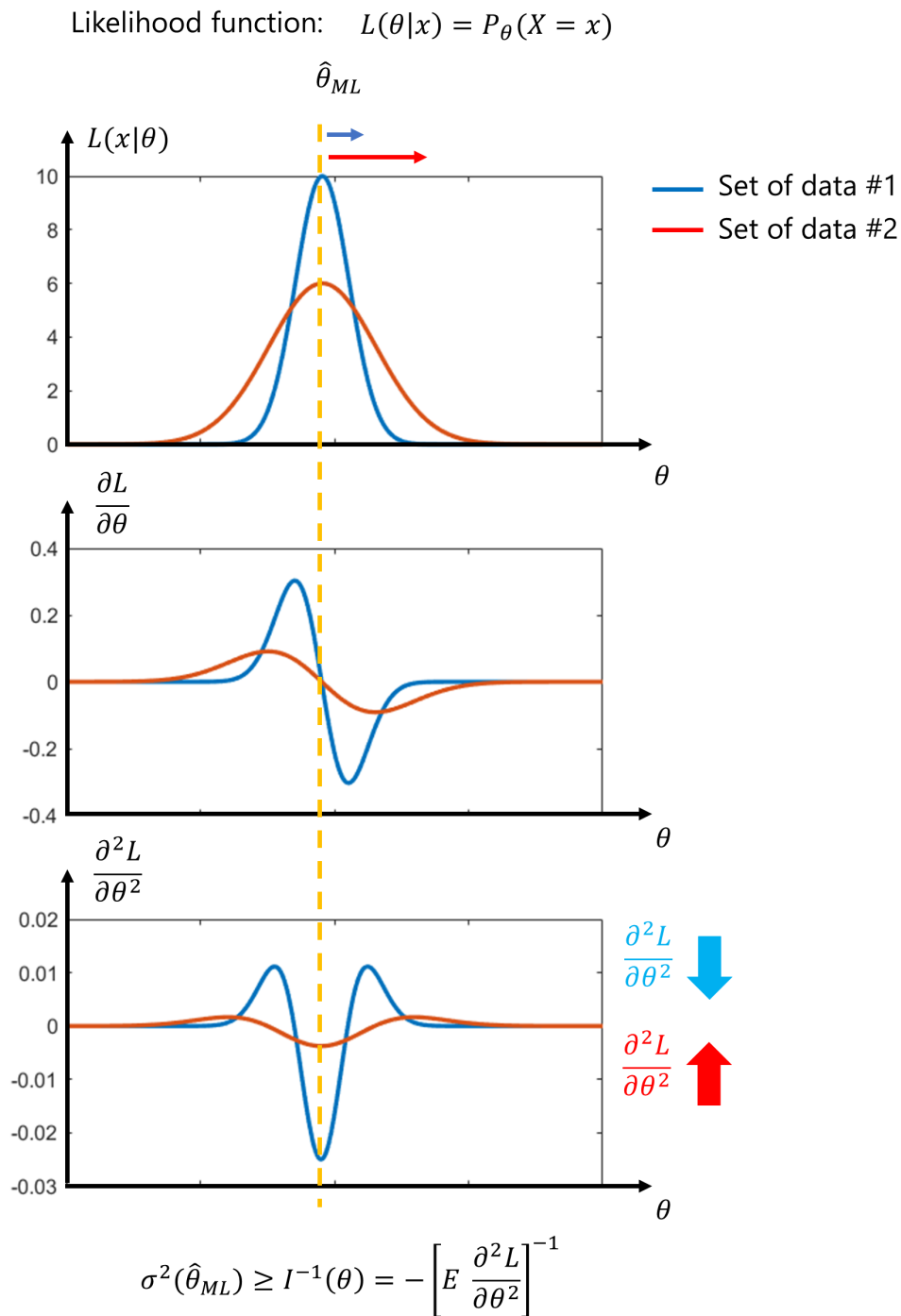
Figure 2.10: Two sets of data x (blu and red) are provided as input to an unbiased estimator with likelihood function L. The model describes the data and is parameterized on θ. The calculation of the likelihood L, first ∂L and second $\partial^2 L$ order derivative, is shown over the whole set of possible θ. The estimated parameter $\hat{\theta}_{ML}$ is the best estimate for both datasets but suits better the blue data sets (i.e., maximum likelihood). Visually, its likelihood spread less, and the CRLB evaluation as the second $\partial^2 L$ order derivative calculated for $\hat{\theta}_{ML}$ gives the measure of it. The inequality equation is reported at the bottom.

However, there are practical limitations to the theory and conditions that ensure or limit its validity.

- CRLBs require the knowledge of the *true* parameter values $\boldsymbol{\theta}$ as well as the true noise power (or standard deviation, σ), which are experimentally not known but only estimated: $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$. Thus, conveying *estimated* to *true* parameter space requires the assumption the model function is an unbiased estimator: $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$. It is noted that this proves evidence that the CRLBs upon which the MRS community rely on are to be defined as *estimated* CRLBs and not *true* CRLBs [17].

- *true* CRLBs for a model defined as in Eq.2.12 do not equal the standard deviation of its parameter. In fact, given the exponential terms in Eq.2.11, such an estimator cannot be considered an *efficient* estimator. Therefore CRLBs in the case of MRS must be regarded as strict lower bound [17].

- The solution of Eq.2.13 cannot be cast in a closed formula. Implementing the minimization algorithm that iteratively searches for such a solution (e.g., Levenberg-Marquardt) typically introduces a regularization term (or, in other words, a bias) that would help decrease the variance of the fitted parameters to avoid inflation of the parameter estimates. This prevents the condition of unbiased estimators even when full knowledge of *true* $\boldsymbol{\theta}$ is supposed [22].

The reader may then deduce that CRLB theory does not formally apply to the MRS case. However, it is current practice to compute and provide CRLBs as a rough estimate of the STD of the fitted parameters because its calculation is straightforward and can be cast in an analytical formula. Nevertheless, from the viewpoint of the underlying statistics of the problem, such estimates are incorrect.

Advanced modeling: a 2D frame

Many advanced MRS paradigms, such as dynamic, functional, or multiparametric MRS, produce 2D *spectral-π* datasets. This refers to a series of 1D spectra acquired sequentially, often while varying a sequence parameter or administering an external time-dependent stimulus or manipulation. Such designs include diffusion-weighted (DW-MRS), T_1 or T_2 subject-specific and metabolite-specific estimation, or functional experiments. Thus, such a dataset can be visualized in a 2D space where one dimension is either the spectral or temporal domain. In contrast, the second dimension is the π dimension, which explores the π-parameter of interest in the given paradigm (e.g., $b - value$s for a DW-MRS study or TEs in an MRS study with assessment of specific T_2s).

The analysis of the 2D *spectral-π* datasets produced by dynamic MRS experiments is conventionally done piecewise in two stages. First, each spectrum is fit using a linear combination of

basis functions to extract the temporal dependence of each metabolite's amplitudes, shown in Fig.2.9. Then, the π-series for each metabolite's amplitude fits the dynamic model, which describes the π behavior to extract the relevant parameters (e.g., T_2 rates in a multi-TE dataset). This is referred to as 1D-fitting, sequential, or piecewise fitting, Fig.2.11(top). Nevertheless, it has been suggested that multiple spectra comprising a dynamic dataset should be analyzed and fitted in tandem rather than sequentially using a model that combines the spectral and π degrees of freedom [16, 23, 24]. Such an approach utilizes the π-correlations inherent in the data to benefit the spectral estimations of metabolite amplitudes and provide more precise and accurate estimates of the π-parameters and concentrations altogether. This is referred to as 2D-fitting, dynamic, or simultaneous fitting, Fig.2.11(bottom).
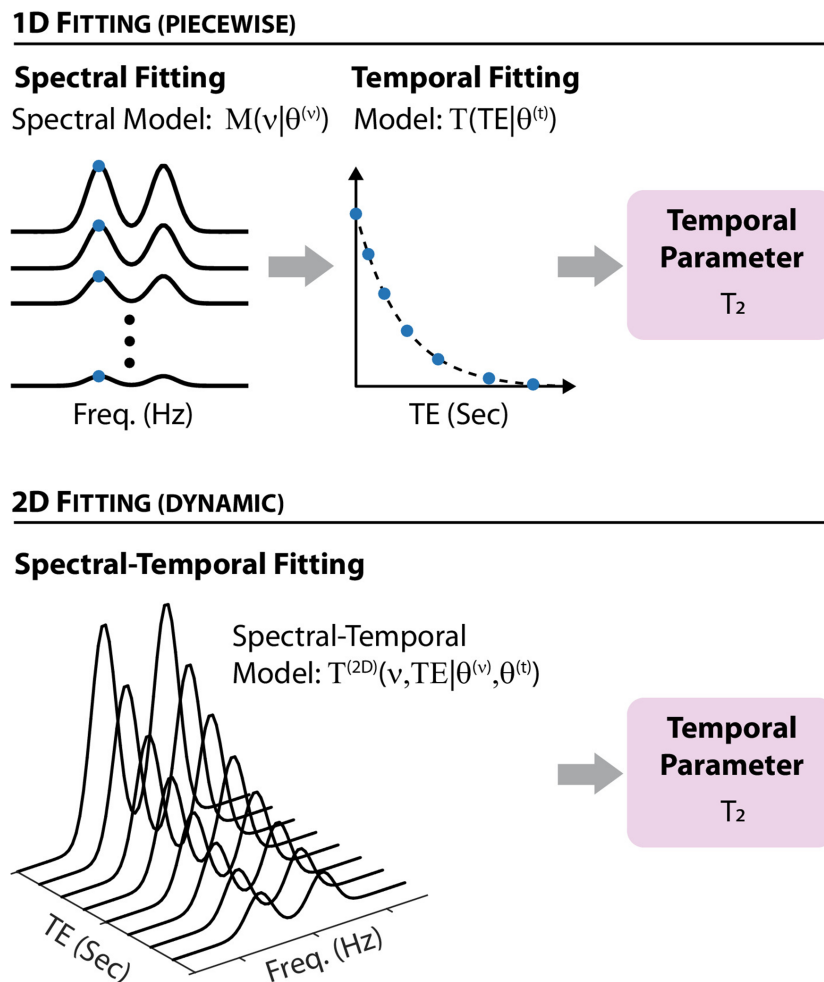
Figure 2.11: A schematic overview of 1D (piecewise) versus 2D (dynamic) fitting schemes for a fictitious multi-echo MRS experiment designed to estimate T_2 rates. 1D-fitting deploys two sequential models. First, lineshapes are fitted in the frequency domain (ν) independently for each response. Second, their amplitude values (visually depicted in blue dots) are cast into a second model describing the π-parameter space (TE). 2D-fitting deploys a model exploiting both domains at once (ν and TE). Figure from [24].

3 | Target organs

In the process of untangling the mysteries of such dialogue, I was introduced to the brain first. Despite its exciting tissue conformation and plenty of contrasts, I discovered it to be a relatively easy organ to access: quite stable and reliable. I dare to say, an almost passive interlocutor. In our "dialogue" figure of speech, I could relate the brain to a highly fascinating but shy person. Nevertheless, neurologists would say it is the most exciting and challenging organ because you can neither take it out to look at it better, solve any health issue, nor replace it. What I happily discovered during this journey is the prostate. Rather more attractive and active organ, which talks and lively delivers the most unpredictable messages. That dialogue was a real challenge, but a partly won one!

3.1. Brain

The brain is the central organ of the human nervous system, and with the spinal cord makes up the central nervous system. The brain consists of the cerebrum, the brainstem, and the cerebellum. It controls most of the body's activities, processing, integrating, and coordinating the information it receives from the sense organs and making decisions about the instructions sent to the rest of the body. The brain is contained in and protected by the skull bones of the head.

MRS of the human brain, first reported more than 30 years ago, is a mature methodology used clinically in many medical centers worldwide to evaluate brain metabolism. Its maturity is driven by the many anatomical advantages, such as the relatively close location to the body surface and the intrinsic anatomical stiffness and staticity, that make brain MRS relatively easy to perform, robust and reliable. In addition, its complementary use with MR imaging sequences provides valuable insights into several common neurologic diseases, including brain neoplasms, inherited metabolic disorders, demyelinating disorders, and infective focal lesions.

The clinical usefulness of ^{1}H MR spectroscopy has been established for brain neoplasms (i.e., brain tumor characteristics, progression and response to treatment [25]), neonatal and pediatric disorders (hypoxia-ischemia, inherited metabolic diseases, and traumatic brain injury), demyelinating disorders, and infectious brain lesions. The growing list of conditions for which ^{1}H MR spectroscopy may contribute to patient management, for its sensitivity to brain dysfunction

in the presence of apparently normal structural imaging, extends to its use as a biomarker of neurodegenerative disorders such as Alzheimer's disease, neurodegenerative diseases, epilepsy, and stroke [26]. However, despite its widespread usage in medical practice and although most modern clinical MR scanners support MRS capabilities, its ultimate clinical value continues to be a source of discussion, especially considering its routine deployment is primarily restricted to specialized centers with good access to MR research support [27, 7]. To facilitate expanded clinical acceptance and standardization of MR spectroscopy methodology, guidelines are provided for data acquisition and analysis, quality assessment, and interpretation [7]. Two examples of MRS clinical value are shown in Fig.3.1 for non- and neoplastic lesions.
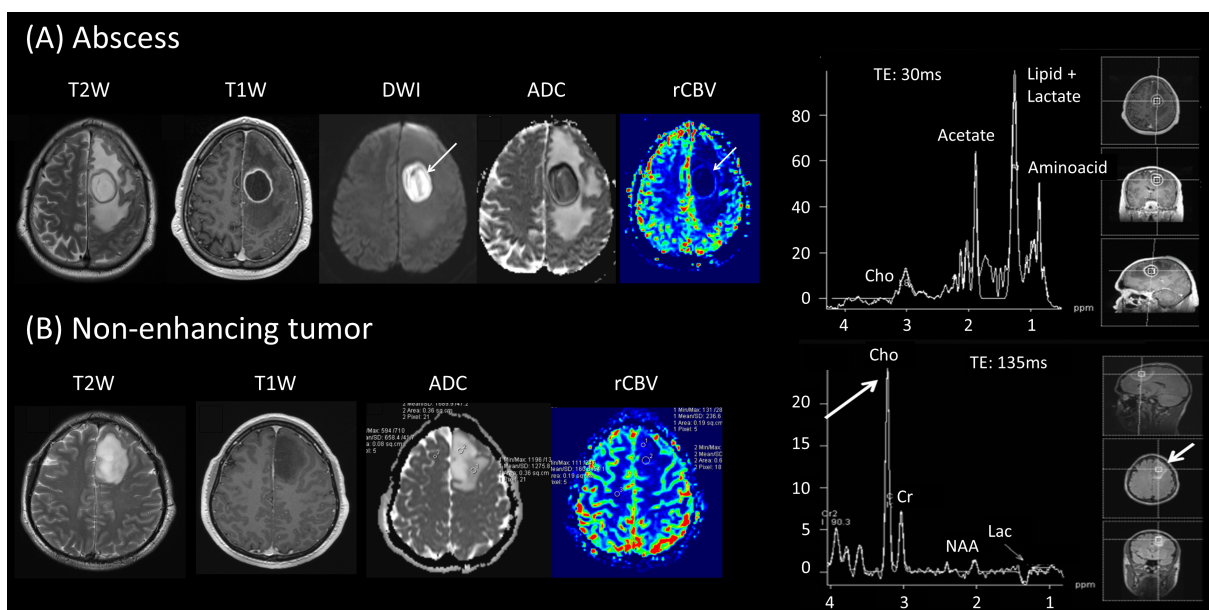
Figure 3.1: Examples of the clinical value of MRS twinned to conventional MRI. (A) Cerebral abscess. (B) Non-enhancing cerebral tumor. T2W and T1W: T2- or T1-weighted images, DWI and ADC: diffusion-weighted and Apparent Diffusion Coefficient mages, rCBV: relative Cerebral Blood Volume measured by perfusion imaging. Modified from [28].

In Fig.3.1(A), a patient with cerebral abscess has undergone a multi-parametric MR examination. Conventional MRI fully describes the morphology of the lesion. It is interesting to see how the ADC and calculated high b-value map complements the insufficient morphological information supplied by T1w and T2w imaging, yielding proper differentiation between tissues within the lesion boundaries (i.e., due to the higher viscosity of fluid) and bleeding brain tissue surrounding the lesion. MRS comes into play with localization within the lesion. MRS features of abscess show predominantly protein breakdown products on the right side of the ppm scale, including amino acid, acetate, and succinate peaks, as well as the presence of a lactate peak. In Fig.3.1(B), a case of multi-parametric MRI-guided biopsy for a non-enhancing tumor is displayed. To optimize biopsies and avoid tissue sampling error for non-enhancing tumors, the biopsy target can be selected from high choline, high rCBV (relative Cerebral Blood Volume), or

low ADC location. The case demonstrates homogeneous ADC and rCBV maps, and the stricking utility of a choline mapping produced by multiple single-voxel spectroscopy measurements to choose the highest choline peak and target the location for biopsy.

3.1.1.   Metabolites of interest

Because of its relatively low sensitivity, only small, mobile molecules in millimolar quantities are generally detectable in an in vivo ^1H MR spectrum. At commonly used field strengths such as 1.5 or 3.0 T, only signals from choline (Cho), creatine (Cr), and N-acetylaspartate (NAA) are observed in the normal brain at long echo times (e.g., 140 or 280ms). In contrast, compounds such as lactate, alanine, or others may be detectable in pathological conditions, which increases their concentration. At short echo times (e.g., 35 ms or less), other compounds such as glutamate, glutamine, myo-inositol, lipids, and macromolecular resonances are detectable [27]. Healthy and pathological spectra with the variation of metabolic content are reported in Fig.3.2. A healthy spectrum and detailed insight on the basis set of fitted metabolites for quantification is displayed in Fig.2.9. The biological significance of the major detectable compounds follows.
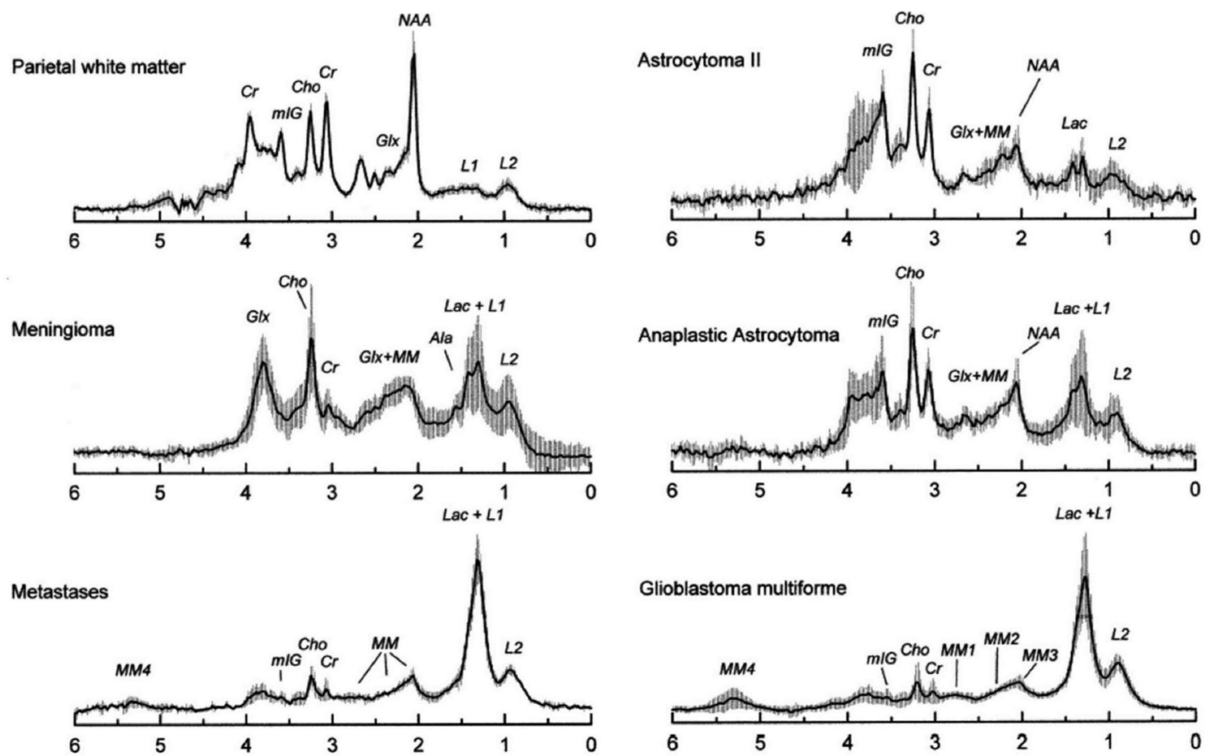
Figure 3.2: Group average spectra from normal and neoplastic pathological parietal white matter, recorded at 1.5T using single voxel spectroscopy (STEAM, TE = 30 ms, 4–8 cm^3 voxel size). All lesions have reduced NAA and also lower Cr than normal white matter. Spectra from the necrotic core of metastases and GBM show reduced levels of all metabolites and elevated lipids. Grade II and III astrocytomas show elevated Cho and lactate signals; grade II astrocytomas also have a prominent myo-inositol (mI) signal. Figure from [29].

N-Acetylaspartate (NAA) is the largest signal in the normal adult brain spectrum, which the most prominent resonance originating at 2.01 ppm, with a small and usually unresolved contribution from *N-acetylaspartylglutamate* (NAAG) at 2.04 ppm. NAA is one of the most abundant amino acids in the central nervous system. NAA is often referred to as a *neuronal marker* since studies have suggested that it is predominantly restricted to neurons, axons, and dendrites within the central nervous system. Overall, NAA does appear to be an excellent surrogate marker of neuronal health, but (as with all surrogate markers), it may sometimes change independent of neuron cell density or function. Nonuniform concentrations for NAA over the entire brain are reported with higher concentrations in gray matter (8-11 mM) compared to white matter (6-9 mM). Unlike NAA, NAAG is suggested to be involved in excitatory neurotransmission as well as a source of glutamate [27, 1].

The *total-Choline* (tCho) is a composite signal consisting of contributions from the trimethylamine groups of glycerophosphocholine (GPC), phosphocholine (PC), and a small amount of free choline itself, with the most prominent resonating peak at 3.20 ppm. A healthy human brain's concentration is approximately 1-2 mM with a nonuniform distribution. These compounds are involved in membrane synthesis and degradation and are often elevated in disease states where increased membrane turnover is involved (e.g., tumors). Glial cells have also been reported to have high levels of tCho [27, 1]. Sometimes tCho is wrongly referenced as *Choline* (Cho) only.

The *total-Creatine* (tCr) is a composite signal consisting of both creatine and phosphocreatine (typically assumed to contribute 50%-50% to the overall intensity). Two resonances are visible: the biggest methyl-ammonium (CH_3) resonance at 3.03 ppm and the methylene (CH_2) group at 3.93 ppm. This compound is involved in energy metabolism via the creatine kinase reaction, generating ATP. In vitro, glial cells contain a two-to four-fold higher concentration of creatine than neurons. tCr also shows quite significant regional variations, with lower levels in white matter (5.2-5.7 mM) than gray matter (6.4-9.7 mM) in normal brains [27, 1]. Sometimes tCr is wrongly referenced as *Creatine* (Cr) only.

One of the larger signals in short echo time spectra occurs from *myo-inositol* (mI) at 3.5–3.6 ppm. mI is a sugar with a suggested role as part of the inositol triphosphate intracellular second messenger system. Glial cells in vitro have been shown to contain higher levels of mI than neurons. mI has been reported to be reduced in hepatic encephalopathy and increased in Alzheimer's dementia and demyelinating diseases.

Glutamate (Glu) is the most abundant amino acid in the brain and is the dominant neurotransmitter. At 1.5 T, there is an almost complete overlap of Glu and *glutamine* (Gln), and they are detected as a composite "Glx" peak. At higher fields (3.0 T and above), Glu and Gln become better resolved and can be quantified individually with good accuracy using appropriate spectral analysis techniques. Glu presents an average concentration of 8-12 mM with significant differences between gray and white matter [27, 1].

3.1.2. Challenges

A large body of research demonstrates that (1) robust, high-quality MRS data may be acquired with the hardware available on current clinical MR systems; and (2) many technical challenges of performing clinically useful SVS and MRSI can be overcome with software improvements applied to current scanner hardware [7]. General improvements require (1) the implementation of a robust semi-LASER protocol to improve localization of MRS measurements, (2) the incorporation of simulated metabolite basis sets in spectral analysis, and (3) the deployment of time-efficient, robust, and high-quality automated shimming routines [7]. Trickier pitfalls in brain MRS are reported in the following sections.

Macromolecule Background

[1]H-MR spectra obtained at short TEs provide the best SNR for quantifying brain metabolites. However, broad peak-like signals from mobile macromolecules (*MMs* or *MMBG*) that overlap with the narrower peak-like small molecular metabolite signals and are not predefined by knowledge about their composition and concentrations complicate quantification at short TE. Therefore, MMs presence complicates metabolite quantification. Thus, separating broad MM signals from low molecular weight metabolites enables accurate determination of metabolite concentrations and is of primary interest in many studies [30, 31]. In Fig.2.9, an example of MM inclusion in the basis set is displayed. In Fig.3.3, measured and fitted MM spectra at different echo times are shown.

MM spectra are composite signals composed of multiple overlapping and closely spaced multiplets (due to scalar couplings) that originate from different amino acids. Moreover, spectral patterns of the same amino acids also differ slightly concerning their chemical shifts across different proteins. Thus, based on current knowledge, MM spectra in vivo are addressed to represent distributions of overlapping multiplets from different amino acids within other proteins, contributing to the apparent broad linewidths of the various peaks [30, 31].

For the most accurate and robust quantification of metabolite tissue contents, it is, thus, recommended to use an experimental MM spectrum as part of the basis set for linear combination model fitting. Such MM spectrum can be patient-specifically measured via T_1- or T_2-weighting sequences (hence exploiting MMs' considerably shorter T_1 and T_2 relaxation times) or via diffusion-weighting approaches (exploiting MMs' significantly shorter ADCs). Alternatively, simulated MM shapes can be modeled and precisely tuned to a study cohort [30, 31, 23].
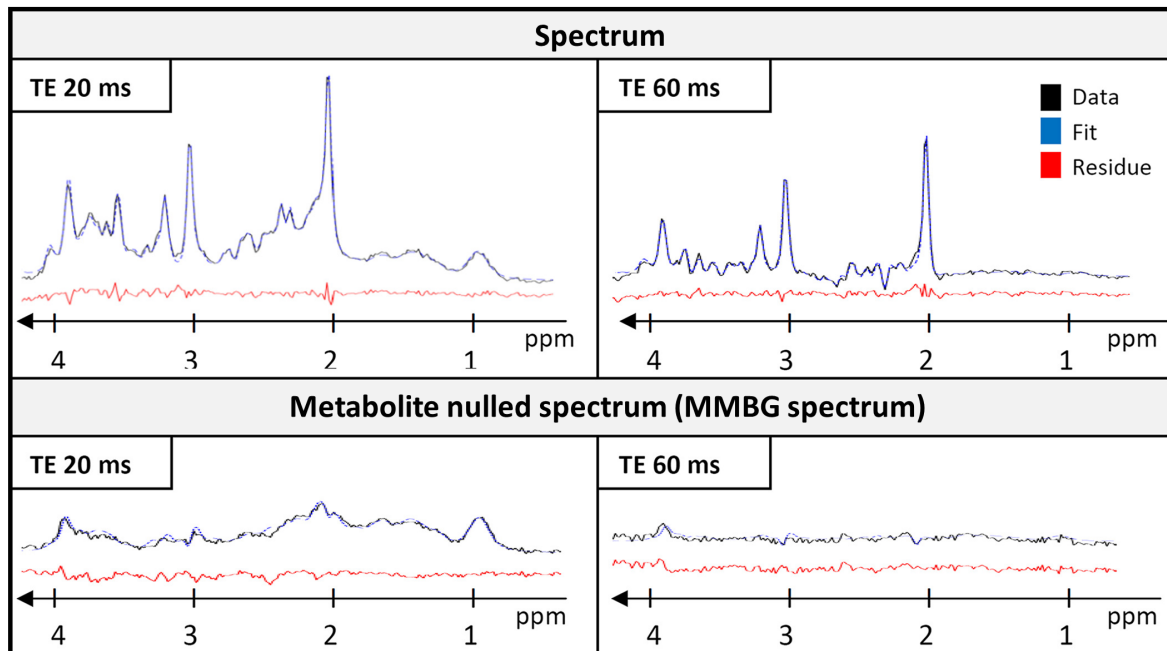
Figure 3.3: Illustration of the spectrum (top) and MM measured spectrum (bottom) at different TE: 20 and 60 ms. The MM spectrum is obtained via metabolite nulling with an inversion recovery measurement (TI=900ms). Measured data are displayed in black, fit in dotted blue lines, and residues in red. Acquisition-centered midline in the occipito-parietal cortex in gray matter. Modified from [23].

Signal contamination

One particular type of artifact is the *ghost* or *spurious echo*. In spectroscopy, spurious echoes result from insufficient spoiling gradient power in combination with local susceptibility variations. In-vivo susceptibility-induced magnetic field distortions arise from the presence of air/tissue and tissue/bone interfaces. In the brain, field distortions are most apparent in regions close to sinuses, such as the prefrontal cortex and temporal lobes. Because susceptibility differences scale with the static field strength, these distortions are stronger at higher fields. Ghosting artifacts are problematic because they superimpose with metabolite peaks at varying frequencies and may thus preclude reliable area estimation. Less troublesome are ghosts that appeared after complete FID relaxation in the time domain (e.g., somewhat after 400 ms at 3T) since postprocessing pipelines (truncation or apodization of the FID) can overcome such contamination (but with appropriate considerations and potential drawbacks that must be accounted for in follow up spectral analysis). Ghosting artifacts should be monitored in the acquisition phase and can be limited by applying OVS bands over nasal cavities, Fig.3.4. Automatic detection of spurious echoes in spectra — and even more so the restoration of the affected spectra — has proven to be possible with modern deep learning strategies [32].
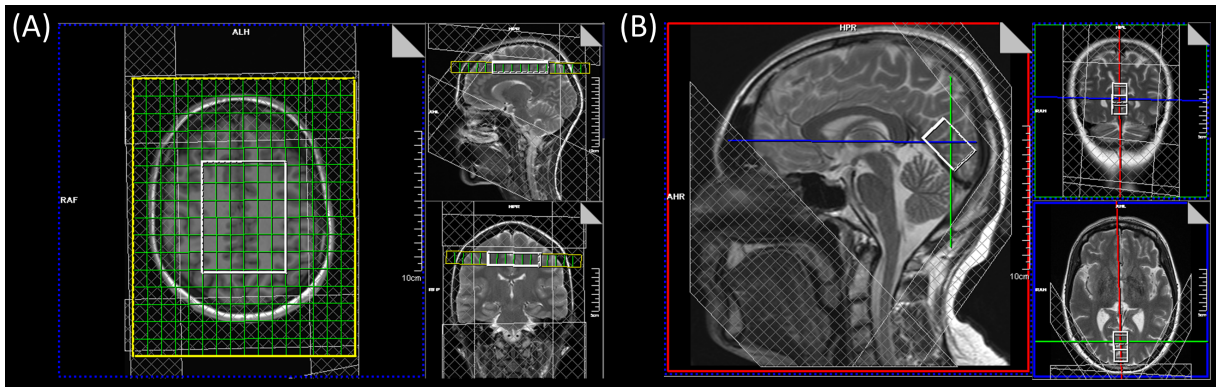
Figure 3.4: VOI and OVS bands placement on (A) MRSI in the supraventricular region and (B) SVS in the occipital lobe. OVS over lipid areas surrounding the brain are twinned to a thick OVS suppressing spurious echo that could possibly generate through the nasal cavity.

From Fig.3.4, it can be seen that the acquired VOI is placed in preferential areas far away from air/tissue interfaces to prevent shimming difficulties and away from the scalp to avoid spurious out-of-volume lipid signals. The subcutaneous lipid contamination should be constantly checked in the acquisition phase by monitoring the 0-2 ppm interval. As for macromolecules, when lipid contamination is unavoidable, a proper modeling integration via basis set is recommended for a reliable spectral analysis [7].

3.2. Prostate

The prostate is an accessory gland of the male reproductive system and a muscle-driven mechanical switch between urination and ejaculation. In adults, it is about the size of a walnut. Anatomically, the prostate is found in the pelvis. It sits below the urinary bladder, with the urethra passing through it and the rectum closing in from behind. It is surrounded by an elastic, fibromuscular capsule and contains glandular and connective tissue. The internal structure of the prostate is predominantly described by zones. Specifically, the prostate features three zones:

- Peripheral zone: the back of the gland that surrounds the distal urethra and lies beneath the capsule. It covers an average 70% fraction volume of an adult gland.

- Central zone: the zone that surrounds the ejaculatory ducts. It amounts to 20% of an adult gland fraction volume.

- Transition zone: it surrounds the proximal urethra and measures about 10% of an adult gland fraction volume.

Visualization of the prostate and its surrounding structures is reported in Fig.3.5 as in anatomy books and as MRI display in Fig.3.6.
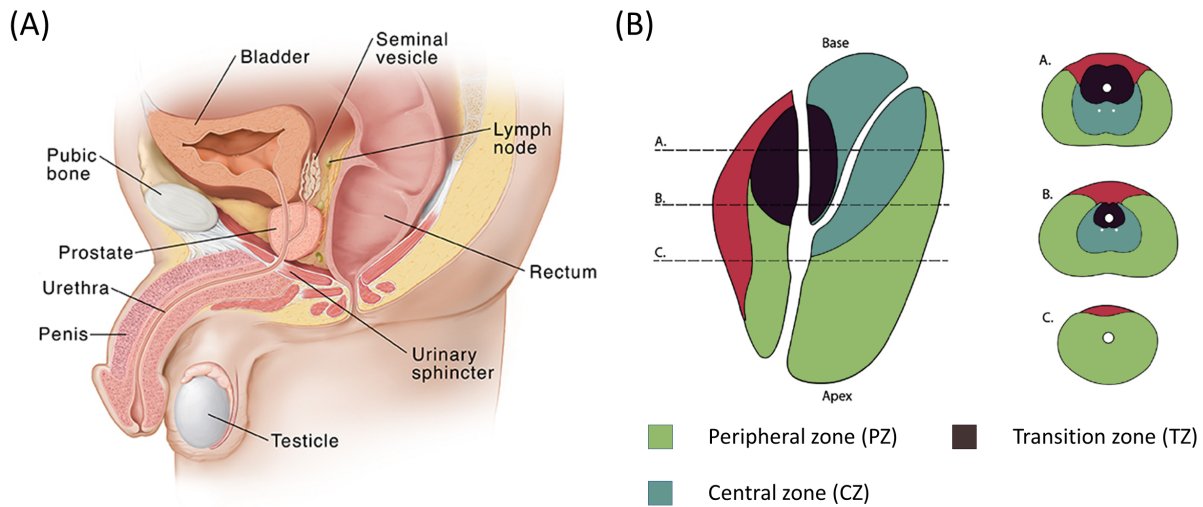
Figure 3.5: (A) Prostate anatomy and surrounding structures. (B) Prostate zones. Figure from [33]
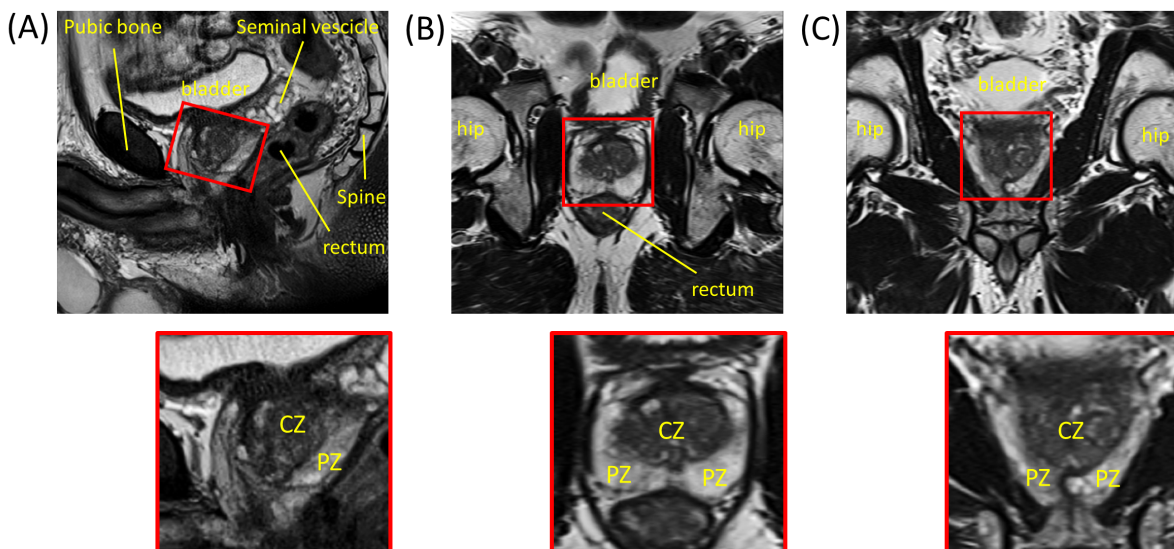
Figure 3.6: Prostate anatomy and surrounding structures depicted by MR Imaging. T2-weighted (A) sagittal, (B) transversal, and (C) coronal slices through a healthy prostate volunteer. (Top) General localization with surrounding structures. (Bottom) Zoom-in on the prostate with labeled prostate zones.

The clinical relevance of the prostate is given mostly by three pathologies:

- Enlarged prostate: the most common cause is benign prostatic hyperplasia (BPH) due to an increase in the number of cells that make up the prostate from a non-malignant reason. It is widespread in older men.

- Inflammation: called prostatitis, it can be caused by bacteria or other noninfective causes. It causes an enlargement of the prostate and related urination pain.

- Cancer: prostate cancer (PCa) is the second most commonly occurring cancer in men and the fifth leading cause of cancer death, with an estimated 1.4 million new cases and 375,000 deaths worldwide in 2020 [34].

The latter represents *the* most relevant clinical situation. For example, prostate cancer that is only present in the prostate is often treated with either surgical removal of the prostate or radiotherapy. Sometimes, the decision may be made not to treat prostate cancer. For example, suppose a tumor is small and localized, or the patient's life expectancy is less than ten years. In that case, the decision may be made to monitor for cancer activity at intervals (*active surveillance*) and defer treatment.

Autopsy studies of men not diagnosed with PCa have shown a PCa incidence of 60% in men over 80 years old, so screening for the disease with PSA testing also finds many cancers that would probably never need any treatment. Therefore, a significant issue in PCa management is to distinguish between potentially aggressive cancers that are clinically significant requiring treatment and those that will not need immediate treatment. In histopathology of biopsies, the aggressiveness of tumor lesions is characterized by Gleason grades on a scale from 1 to 5, determined at two locations, combined in a Gleason score (GS). Often lesions with a $GS \leq 3+3$ are defined as low risk, with $3+4$ as intermediate and $\geq 4+3$ as high risk [34].

Recently, multi-parametric MRI (mpMRI) and MRI-guided targeted biopsy have emerged as essential tools in detecting, grading, and staging PCa. Specifically, diffusion-weighted imaging (DTI) and the calculation of the water Apparent Diffusion Coefficient (ADC) are the game changers in such protocols, which visual power is reported in two examples in Fig.3.7. While MRI parameters can assess anatomical, morphological, and some physiological abnormalities associated with cancer development, complementary information on molecular aspects of this development can be derived from metabolic readouts, some of which may underlie earlier or more specific phases of disease progression. Tissue metabolites can be assessed non-invasively by ^{1}H Magnetic Resonance Spectroscopic Imaging (MRSI). As ^{1}H MRSI can be added seamlessly to MRI procedures, it may reinforce mpMRI in the non-invasive diagnosis of PCa. In particular, mpMRI currently suffers from a low pooled specificity, and low inter-reader reproducibility [34].
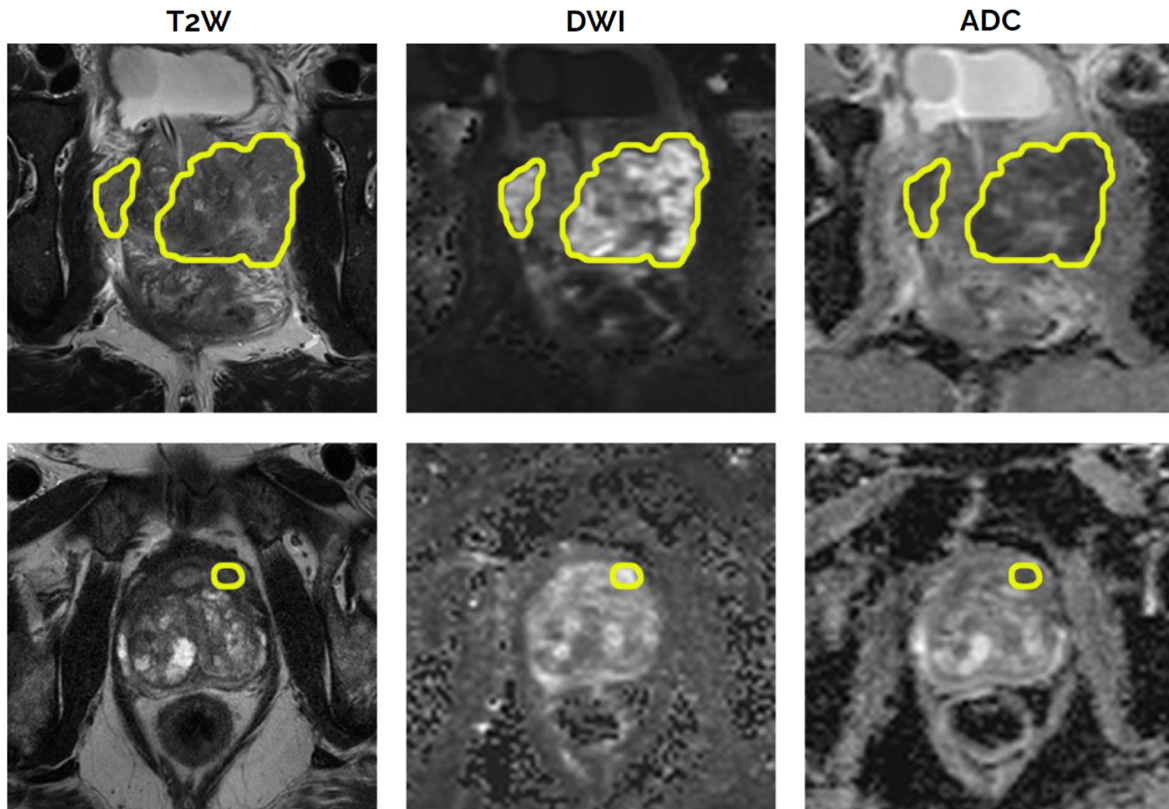
Figure 3.7: Display of mpMRI and PCa morphological heterogeneity. Transversal view of (Left) T2-weighted (T2W) imaging, (Middle) diffusion-weighted imaging (DWI), and (Right) apparent diffusion coefficient (ADC) maps constituting the prostate mpMRI scans for two different patients. Yellow contours indicate PCa lesions. Modified from [35].

Microstructure

The human prostate is composed of epithelial acini (i.e., a round cluster of cells) arranged in a fibromuscular stromal network, Fig.3.8(A). The epithelia are highly organized into a contiguous basal layer containing four major cell sub-types: stem cells, transit amplifying (TA) cells, committed basal (CB) cells, and a layer of columnar secretory luminal cells that make up the rest of the epithelium. The relative content of different epithelial cells in the normal prostate are luminal (60%) and basal (40%), with the stem cells constituting $\sim 1\%$ of total epithelia. The cellular composition of a cancerous acinus is characterized by luminal hyperproliferation, loss of the basal layer, basement membrane breakdown, immune cell infiltration, and stromal reactivity. In addition, cancer skews the epithelial cell percentages; the luminal cells make up $> 99\%$ of tumors, and basal CSCs are estimated to constitute $< 0.1\%$ of tumor epithelial cells. Visually, a consequent squeezing of the lumen is shown, Fig.3.8(B) [36].
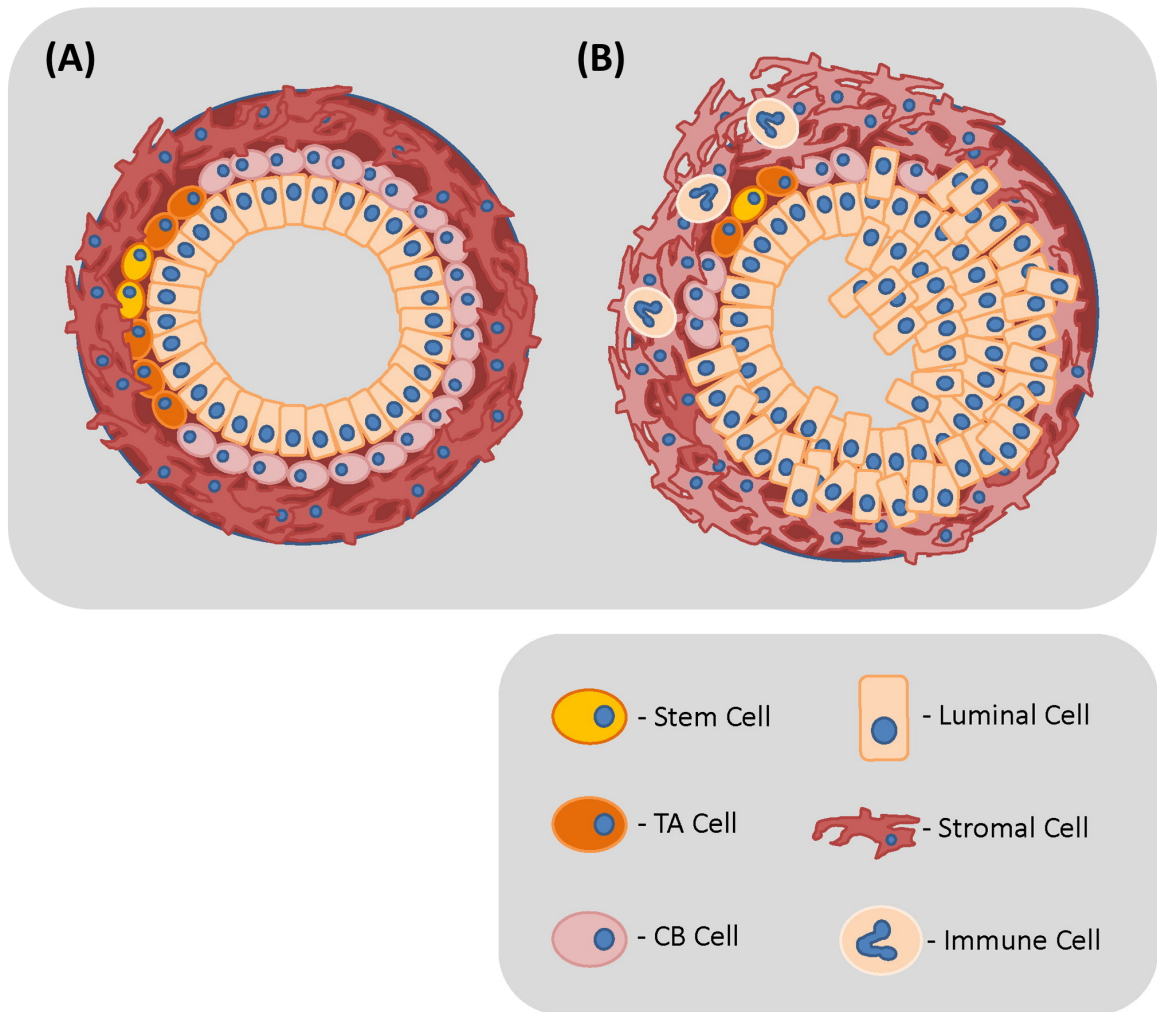
Figure 3.8: Epithelial cell constitution of (A) healthy prostate and (B) prostate cancer. Modified from [36].

3.2.1.   Metabolites of interest

The dominant metabolite peaks observed in MR spectra acquired from the prostate include protons from citrate (Cit), Choline compounds (tCho), (phospho-)creatine compounds (tCr), and signals of polyamines, mostly spermine (Spe). Due to recent development in MRS acquisition, protons of myo-inositol (mI) and taurine (Tau) may be less often but visible. Cit, mI, and Spe are considered dominant compounds in luminal space, whereas tCho and tCr are those prevalent in prostate cells. Fig.3.9 shows two spectra from a ^{1}H-MRSI acquisition of a PCa patient with correspondent histopathology [34].
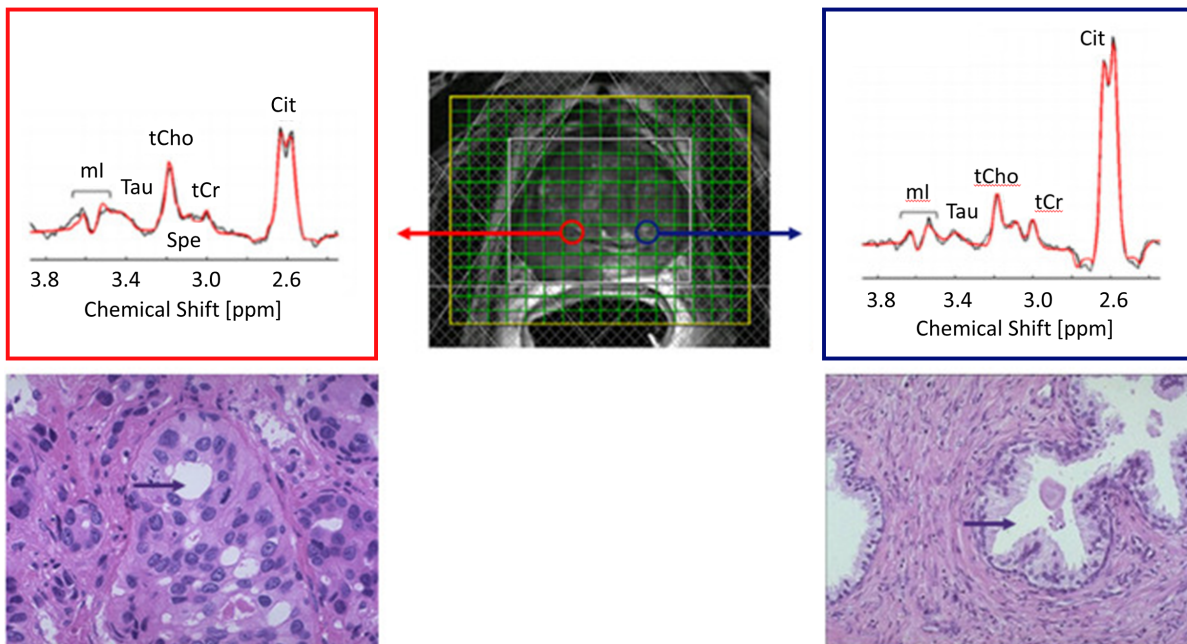
Figure 3.9: Spectra of healthy (right) and prostate cancer (left) tissues measured at 3T with a 3D ^{1}H-MRSI protocol. Region of interest and FOV sampling reported overlayed the MRI image. The tumor lesion reports decreased Cit and Spe and increased tCho levels. Under the panels, histopathology slides are shown, illustrating the reduced luminal space in the cancer lesion compared to healthy tissue (purple arrows). Modified from [35].

Citrate is the most intriguing and observable metabolite for ^{1}H-MRS of the prostate. Its spectral appearance varies substantially at different magnetic field strengths, pulse sequence timing, pH, and cation concentration. Such variability can be observed comparing the example spectra in the healthy case from Fig.3.9 and Fig.3.10. At the lower field, the two middle peaks of the quartet dominate and resonate close together at about 2.6 ppm. The healthy prostate accumulates high levels of Cit, in particular in the peripheral zone. The excess Cit is secreted in the prostatic fluid of the lumen and contributes to favorable conditions for sperm maturation and motility in seminal fluid. In vivo MRS assessments estimate average prostate tissue Cit concentrations to vary considerably between 30 and 70 mM. Citrate levels are decreased in prostate cancer tissue due to reduced production and secretion as modulated by the TCA cycle and reduced luminal space given cancer growth [34].

In healthy prostate tissues, high concentrations of polyamines mainly represent spermine. Like citrate, they are secreted by specialized ductal cells in the prostate and accumulate in the luminal space. In in-vivo MRS studies, the tissue concentration of spermine is estimated to be 7–18 mM. A linear relationship between the concentration of polyamines (spermine) and citrate has been observed, and a transient association between citrate, spermine, and zinc and the binding of this complex to proteins has been deduced from in vitro T_2 relaxation studies. Less spermine synthesis and decreased luminal space explain lower signal detection for spermine in prostate

cancer as well [34]. Precise and reliable quantification of Spe is nevertheless challenging, given a substantial overlap with the frequencies of tCr and tCho in the 3.12-3.18 ppm area. Resonating groups below 2.11 ppm area are typically suppressed or highly contaminated by lipid presence, Fig.3.10.

Choline and creatine compounds are present at relatively high concentrations in prostate cells, either epithelial or stromal cells. tCho tissue levels estimated via in-vivo MRS experiment were between 2 and 5 mM, and those of tCr between 4 and 9 mM. In PCa tissues, the levels of tCho are increased, mainly due to a higher phosphocholine and glycerophosphocholine content. This involves increased choline transport into tumor cells, increased choline kinase α and *phospholipase A2* expression and activity in tumors. Rising choline levels in tumors are often associated with increased cell density and tumor hypoxia, although the latter is limited for the prostate. Many MRS studies have observed a correlation between tCho levels or tCho signal ratios such as tCho/Cit or (tCho+Spm+tCr)/Cit and Gleason score [34].
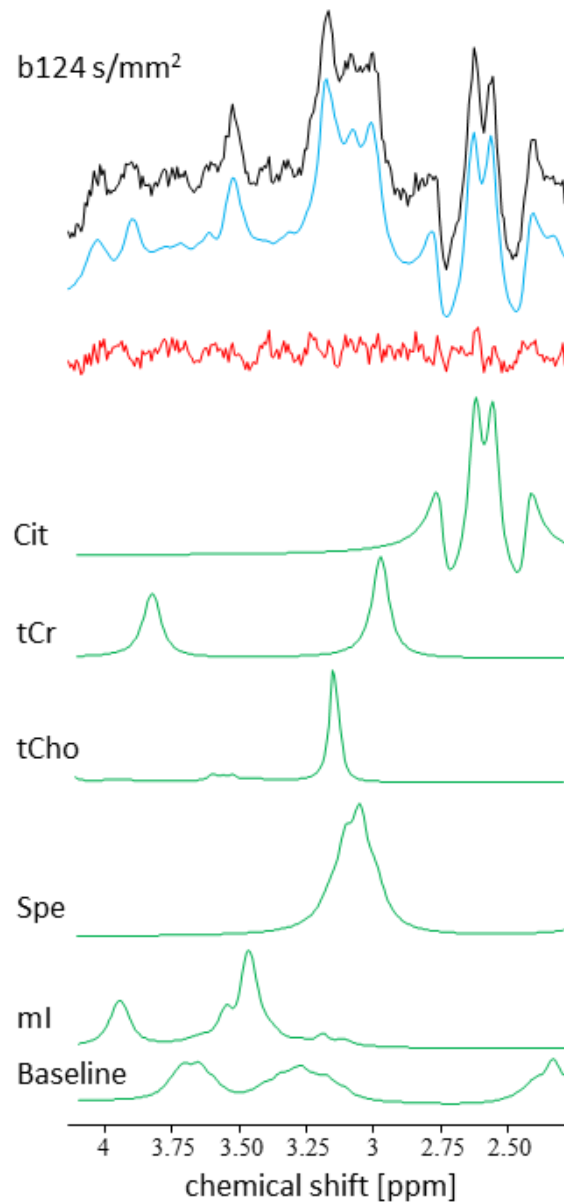
Figure 3.10: Spectrum of Single-Voxel MRS which VOI covered the whole prostate (black). Acquisition with a STEAM protocol with low diffusion weighting (b-value 124 s/mm^2). The fitted model (blue) and residues (red) are reported together with the fitted basis set (green).

3.2.2.   Challenges

Common challenges for prostate MRS and MRSI measurements are introduced in the following chapter.

Shimming

Optimizing B_0 field homogeneity (shimming) enormously improves spectral quality, reducing the decay of the apparent transverse magnetization in a voxel and giving a handle on lineshape broadening. That can significantly reduce possible severe spectral overlap between tCho, Spe, and tCr, which resonate relatively close to each other, Fig.3.10. Good B_0 homogeneity is also crucial for adequate water and lipid signal suppression since they may escape the frequency-selective pulses needed for their suppression when severely broadened or shifted. Aside from optimization via (1) dedicated acquisition or (2) post-processing pipelines to restore or reduce signal broadening [37, 34], a practical step to improve B_0 homogeneity is the preparation of the rectum with a cleansing enema and an endorectal gel filling or the prescription of light dietary and fasting six hours prior the MR examination.

Pulse sequence design

Because of prostate cancer's multi-focal and heterogeneous nature, volume selection methods are primarily employed with 3D MR spectroscopic imaging readouts to cover the whole prostate.

Gradient-modulated adiabatic pulses are considered the gold standard for volume localization. Adiabatic pulses have better slice profiles, reduce outer volume signal contamination, and are less sensitive to RF transmit field inhomogeneities. To lower RF power deposition, gradient-modulated offset independent adiabaticity (GOIA) pulses have been implemented for prostate MRSI. The application of GOIA-sLASER to prostate MRSI considerably reduces the contamination of spectra with lipid signals of fat surrounding the prostate. It hence improves the quality of the spectra and robustness of the measurement. That positively impacts the choice of shorter TR, with the consequence of (1) mitigating the relatively shorter (compared to the brain) T_1 relaxation effect of proton spins and (2) shortening the experiment time [34].

Echo timing covers a crucial role in the MRS of the prostate. On the one hand, for spin-echo type localization sequences, the shortest possible echo time TE is commonly used to minimize T_2 relaxation losses. On the other hand, longer TEs in most prostate MRS(I) experiments are chosen to decrease lipid nuisance. Moreover, the signal of the strongly coupled spin system of citrate dominates in prostate MR spectra and shows considerable variations as a function of interpulse timing (including TE). Therefore, for each pulse sequence timing and field strength, a TE is selected with a high citrate absorption signal intensity ($TE =$88ms is the choice for the State of the Art optimized semi-LASER 3D MRSI setup at 3T [5]).

Movement artifacts

Prostate MR measurements may suffer from movement artifacts due to the location of the prostate near the bladder and the bowels and relatively long acquisition times. Preparation techniques can be used to limit bowel movement, including anti-peristaltic drugs and the application of microenema to evacuate the rectum if necessary. Dietary restrictions, where the patients are instructed to fast six hours before the exam and consume water solely, are also applied. Potential acquisition techniques to reduce motion artifacts include applying a navigator or using rapid acquisition methods. Finally, in water signal unsuppressed MRS(I), the water signal can mitigate movement artifacts.

Lipid contamination

As the prostate is embedded in lipid tissue, substantial contamination by lipid signals may occur in MR spectra of prostate voxels. It is common to apply outer volume saturation (OVS) bands positioned around the prostate to reduce extraprostatic lipid signals. All spins in these bands are excited and then dephased by crusher gradients. OVS pulses were developed to compensate for poor edge profiles, B_1 field inhomogeneity, and chemical shift errors. A widely applied approach is spectroscopic signal suppression with double band-selective inversion with Mescher–Garwood (MEGA) pulses, in which dual-frequency pulses surrounded by crusher gradients selectively invert and dephase both the lipid and water signals. Due to the side-lobes of the spatial response function (SRF), signal contamination between neighboring voxels can be significant when peri-prostatic areas with high lipids are included in the VOI. These sidebands are commonly attenuated with a Hamming apodization filter in k-space. Another effective method to prevent lipid contamination is the application of more accurate localization sequences (i.e., semi-LASER).

4 | Multiparametric MRS

The beauty of a dialogue goes through the interplay between the two interlocutors. An interesting dialogue may require a good topic, but its true beauty resides in the articulation of motives, justifications, and arguments. Such dynamics, when proper, may trigger or draw any attention even to the drowsiest listener. This I relate to multiparametric NMR measurement: the capacity of arranging such a dialogue with the MR partner, capable of going beyond any static topic, but extending it further, into side argumentations that may complement the overall picture or substantially strengthen whatever theme carried out in the conversation.

The detection and quantification of a wide range of metabolite concentrations has led to the characterization of disease progression, allows the study of intervention by medication or surgery, and allows the identification or categorization of diseases by observing specific metabolic markers. However, despite its great importance, detecting static metabolite concentrations alone provides only a partial description of metabolism. In vivo metabolism is largely characterized by dynamic processes, like an enzyme-catalyzed chemical exchange, transfer of chemical groups through metabolic pathways, and, specifically for NMR, relaxation processes.

Multiparametric MR spectroscopy deploys appropriate experimental techniques that sensitize the NMR physics to various dynamic processes, yielding simultaneous and efficient quantification of multiple NMR parameters. For the purpose of this work, T_2 relaxation and diffusion weighting will be introduced.

4.1. Metabolite-specific T_2 relaxation

In Chapter 1.2, the overall NMR relaxation phenomenon was described as the process by which the macroscopic magnetization vector returns to the thermal equilibrium state following a perturbation. Specifically, the disappearance of transverse magnetization is described by the transverse or spin–spin relaxation time constant T_2. The orientation of nuclear spins can be changed by applying an oscillating magnetic field in the transverse plane near the Larmor frequency of the spins. A coherent perturbation of the spins can be achieved with an external coherent magnetic field B_1+. In contrast, incoherent, random magnetic fields internal to the sample lead to incoherent, random perturbations of the nuclear spin orientation. For spins in solution, the randomly fluctuating magnetic fields are predominantly caused by the magnetic moment of other nearby

spins. The magnetic moment of one spin affects the local field of another spin randomly (both in amplitude and orientation), due to Brownian motion and molecular tumbling. Therefore, since relaxation will be induced when the frequency of the local magnetic field is close to the Larmor frequency, a quantitative description of relaxation will characterize the frequency components variation of such local magnetic field over time. In other words, changes in T_2 relate well with changes in the spin surrounding environment. It must be noted that, in principle, any interaction that causes fluctuating magnetic fields can induce relaxation. The interactions which can generate the appropriate conditions for relaxation are (1) magnetic dipole–dipole interactions, (2) electric quadrupole interactions, (3) chemical shift anisotropy, (4) spin rotation interactions, (5) scalar coupling interactions and (6) interactions with unpaired electrons.

Relaxation in biological tissues is complicated even further due to compartmentalization and the exchange of molecules between different molecular environments. For example, assuming isotropic molecular motion for water in tissue is generally wrong. In fact, with a first degree of approximation, tissue water can be divided into *bulk* or *free water* and *protein-associated water*. Free water is assumed rotationally mobile, indicating long T_2 relaxation time constants. When water forms a hydration layer around proteins or other hydrophilic structures, it is called structured protein-associated water. Such water is characterized by restricted mobility and consequently hints at significantly reduced T_2 relaxation times. The structured and bulk water pools are in fast exchange such that the relaxation characteristics of the immobile pools can significantly shorten the observed relaxation times. The two-compartment model can be extended to more compartments, which could also extend to metabolites. As a matter of fact, this interplay of water with proteins relates well with the spermine compound in the prostate, where the characterization of spermine interactions with protein bonding is still yet under investigation, through T_2 measurements or analysis of its diffusivity (see Section 7.2).

4.1.1. Measurement of metabolite-specific T_2 times

The conventional technique to measure T_2 times is via a series of repeated spin-echo experiments at different TEs. Such an approach is sometimes referred to as Multiple Echo Times (MTE) experiment or, in this work, as Multi-Echo Multi-Shot (MEMS) approach, Fig.4.1.

Spin-echo experiments intrinsically reduce the T_2^* artifactual component in the measure, as discussed in Sections 1.2 and 2.2. The signal intensity acquired at different echo times decay exponentially, according to Eq.1.3. Deviations from the simple, exponential decay curve can be seen for water, as is the case of the separation between brain water ($T_2 < 200$ ms) and CSF water ($T_2 > 200$ ms). Larger or strongly-coupled spin systems are characterized by complicated modulation functions that can only be obtained through quantum-mechanical simulations. Such modulation due to scalar coupling needs to be quantitatively known to extract the metabolite-specific T_2 relaxation time constant. Current practice relies on simulating the metabolite basis sets for each different TE_i in the investigated batch $TE_{i=1}^n$. Subsequently, a sequential or bidi-

mensional fit extracts T_2 metabolite-specific times and concentrations. In Fig.4.1 an exemplary batch of brain spectra acquired for six TEs is arranged in a 2D frame. Along the 1st dimension is displayed the spectral contribution in frequency (chemical shift), and along the 2nd dimension is displayed the spectral variation upon increments of TE (i.e., T_2-weighting). The model that describes the pool of spectra can be written in the time domain as follow:

$$s(t, TE) = \sum_{n=1}^{N} A_n \xi_n^{TE}(t) e^{j(\omega_{TE}t + \phi_{TE})} e^{-\alpha_{TE}t^2} e^{-TE/T_{2,m}} \tag{4.1}$$

where a bidimensional inter-dependency on time t and echo time TE is evident. The overall contribution result is the sum of the simulated basis set ξ_n^{TE} for a specific metabolite n and a given sequence timing TE. Overall frequency and phase distortion, as well as field inhomogeneities α (i.e., the T_2^* modeled contribution), are assumed independent and free for each TE. Such flexibility can be discussed and limited according to the specific experimental design and/or necessity. The metabolite-specific relaxation rate $T_{2,m}$ is estimated as a parameter varying over the 2nd dimension and is assumed linked throughout different TEs for any given metabolite $m \in [1, M]$. Given limitations such as SNR and degree of freedom of the model, the range of independent fitted T_2 is typically smaller than the overall number of metabolites (i.e., $M < N$). Therefore T_2 may be assumed equal for a given pool of weakly represented metabolites. Such pool may vary according to the intensity of the external magnetic field B_0 [38].
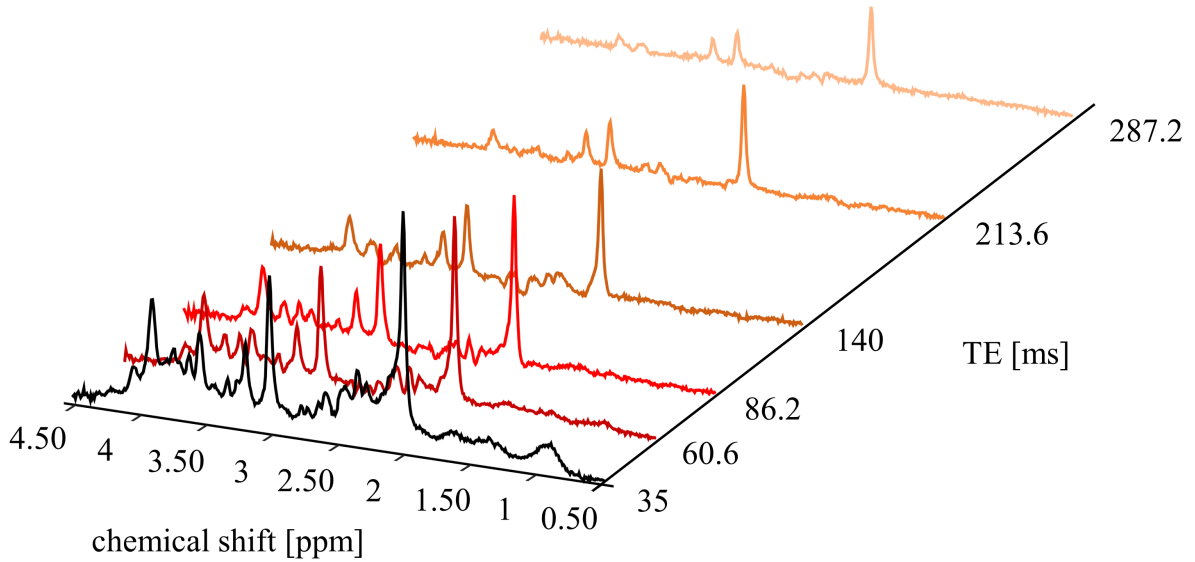
Figure 4.1: Multi-Echo Multi-Shot dataset of brain spectra acquired at 3T with a semi-LASER protocol and exploring a batch of six echo times. T_2-weighting is evident along the TE dimension.

A sequential fit approach would untangle the term $e^{-TE/T_{2,m}}$ from Eq.4.1 and enforce freedom onto estimating metabolite contribution (or intensity) A_n^{TE} for each different TE and therefore generating six independent fits modeled by $s_{TE}(t)$ instead of the bi-dimensional model $s(t, TE)$.

After sequential χ^2 minimization for each TE, the six estimated A_n^{TE} are fitted with the mono-exponential decay function $x(TE) = A_m^{TE=0}e^{-TE/T_{2,m}}$ to determine the specific T_2 for each metabolite m. A bidimensional (or simultaneous) fitting approach would take the whole Eq.4.1 and fit both areas A_n and $T_{2,m}$ simultaneously in one-step χ^2 minimization, where metabolite areas A_n do not vary between TEs but are constant and corrected by simultaneous T_2 weighting. It must be mentioned that such correction is only partial, and T_1 correction will still be needed when calculating absolute concentration level deploying Eq.2.14. The last observation suggests how profitable the estimate of measurement-specific (or, in other words, subject-specific) relaxation times is for in-line calculation of metabolic content.

4.1.2. Impact of T_2 for quantification

Absolute quantification in MRS requires knowledge of sequence properties, such as echo time TE, repetition time TR, and pulse inter-timings, as well as metabolites' relaxation times (T_1, T_2) and transmit inhomogeneity (B_1+). Conventional techniques for measuring these quantities, such as Inversion Recovery (IR) or Multiple Echo Times (MTE), are time-consuming, inefficient, and rarely used within clinical timeframes. Instead, clinical MRS relies on regionally tabulated values, which do not account for inter-subject variability and are often lacking, particularly in pathology. Such assumptions introduce quantification errors, and biases [1].

Moreover, under pathological conditions such as stroke or tumors, various metabolites may be present in different cellular micro-environments with different relaxation characteristics. In these cases, the measurement of T_2s is first meant to be embedded into the appropriate calculation of concentration levels and, second, could hold additional valuable information as biomarkers per se. For example, the relaxation times of the neuronal marker NAA reflect the neuronal microenvironment and may operate as an independent marker of neurodegeneration or inflammation [39]. So far, there is clear evidence for age-dependence of metabolite relaxation times [40] but also altered values in pathologies such as multiple sclerosis [41], Alzheimer's disease [42], and cancer [43, 44].

4.2. Diffusion-Weighted MRS (DW-MRS)

Diffusion is the random translational (or Brownian) motion of molecules or ions that is driven by internal thermal energy. Diffusion in the absence of internal concentration gradients, like diffusion in pure water, is intuitively described by a probability function that gives the probability of a particle having moved over a certain distance and time, Fig.4.2(A). For isotropic diffusion, such probability function predicts a *Gaussian* dependence on the displacement. As a random Brownian motion, the average displacement is zero. However, the average square displacement λ^2 associated with 3D diffusion can be calculated as $\lambda^2 = 6\Delta t$, where D is the diffusion coefficient (in $\mu m^2/ms$) and t is the diffusion time (ms), Fig.4.2(B). For freely diffusing water at room

temperature $D \sim 2.2\mu\mathrm{m}^2/\mathrm{ms}$, which means that the water molecules travel an average or root mean square distance of circa 25 μm in 50 ms. Deviations from the Gaussian distribution function will arise when geometrical constraints restrict the translational, Brownian displacements.

Diffusion of particles or molecules in biological tissues is driven by the molecule or particle size and any spatial or chemical constrain on molecules' motility. Therefore, diffusion may reflect insights on tissue micro-structures and compartmentalization (e.g., intra- or extra- cellular environment), chemical interactions across molecules and metabolic pathways (e.g., protein bonding and complexation) as well as description of the nature of molecule themselves (e.g., size of a molecule).
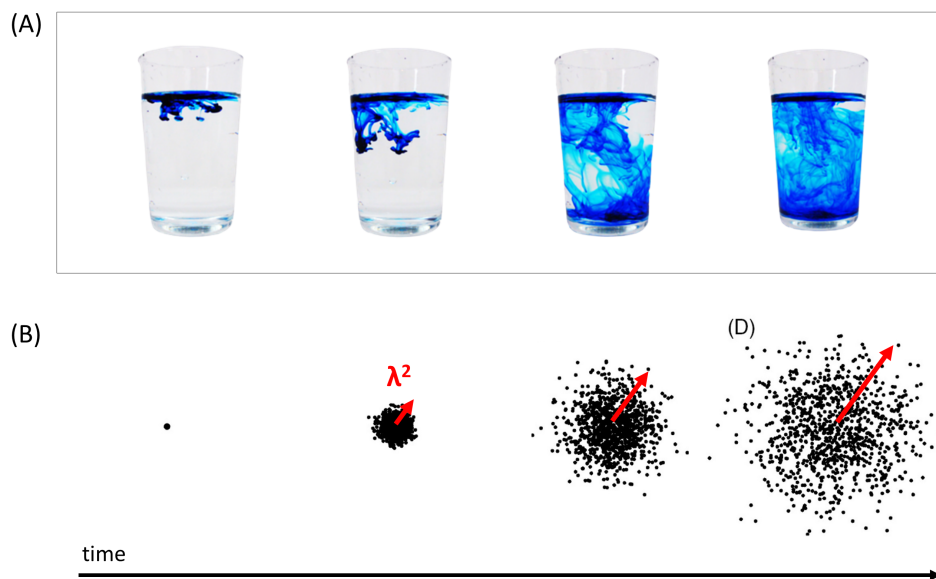
Figure 4.2: Diffusion and spatial displacement. (A) *Almost* free diffusion of blue ink in water. (B) Starting from a single point in space, the mean square displacement of diffusing particles mimicking ink molecules increases linearly with time. Extended from [1].

4.2.1. Introduction to Diffusion in NMR

Diffusion-weighted MR measurements rely on the signal attenuation arising from phase dispersal. To quantify translational motion, magnetic field gradients are employed such that spins at different spatial positions r will acquire spatially dependent phase shifts ϕ as a consequence of gradient-driven Larmor frequency $\omega(r)$: concept similar to the position-dependent Larmor frequency $\omega(\boldsymbol{r}) = \omega_0 + \gamma G r$, where ω_0 indicates the Larmor frequency induced by B_0, as illustrated in Chapter 2.1 for the definition of slice selective pulses. A visual example is given in Fig.4.3. The phase exerted on the spins by the magnetic field gradients is provided by $\phi = \gamma \delta G r$, where r is the position of the spin along the axis where the gradient with intensity G is played out for a duration δ (transition from time steps t_1 to t_2 in Fig.4.3). For stationary spins, the generated phase dispersal is refocused by a secondary magnetic field gradient applied with inverted

magnitude, and after the *diffusion time* Δ (i.e., the inter-time played out between two diffusion encoding gradients, see the transition from $t = 3$ to $t = 4$ in Fig.4.3-*first line*). Contrarily, suppose there is a macroscopic motion (e.g., flow) along the diffusion encoding gradient. In that case, all spins will undergo a displacement Δr that will produce a net phase shift over all spins $\Delta \phi \propto \gamma \delta G \Delta r$ (transition from t_3 to t_4 in Fig.4.3-*second line*, phase distortion represented at t_5). In the presence of diffusion (without macroscopic motion), the spins randomly move to different spatial positions. Since the translational motion has perturbed the linear dependence of phase on spatial position, the application of the second magnetic field gradient can no longer lead to a complete reversal of the phase acquired by the first magnetic field gradient. Therefore, the presence of diffusion leads to phase dispersal across the sample, which in turn will lead to phase cancellation and hence signal loss (transition from t_3 to t_4 in Fig.4.3-*third line*, signal loss represented at t_5). Signal attenuation depends on the diffusion coefficient, the area of the magnetic field gradient (i.e., amplitude and duration), and the separation between the magnetic field gradients.
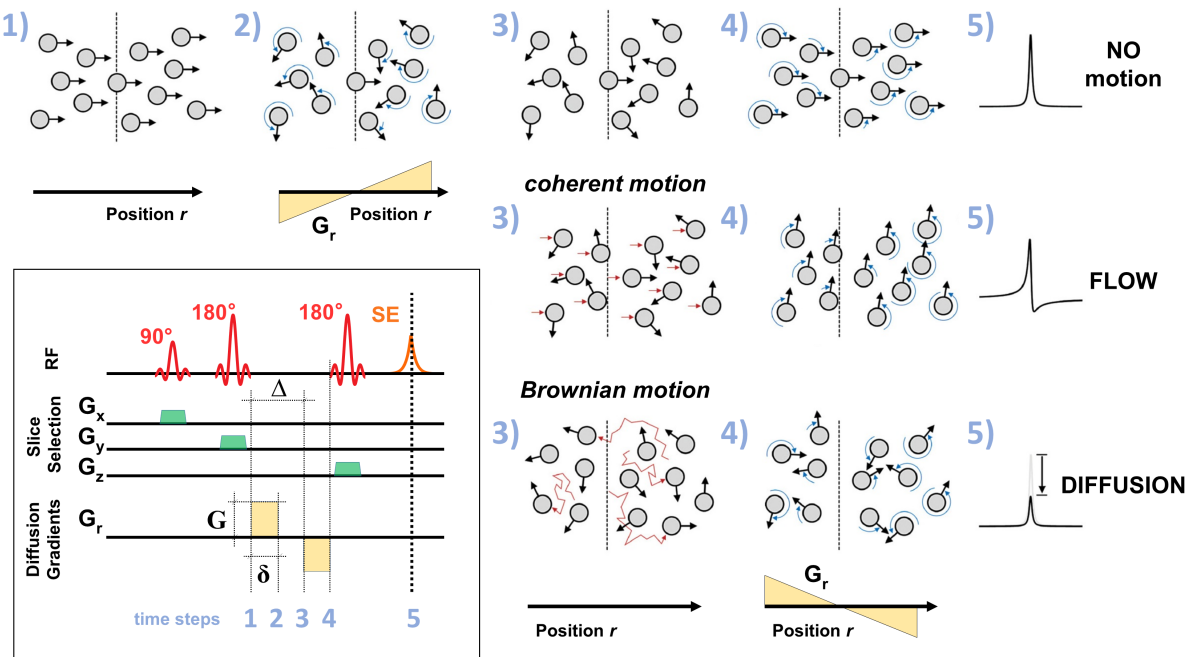
Figure 4.3: Diffusion, motion, and flow in the presence of magnetic field gradients. (box) A 3D localization experiment is completed by bipolar gradients of duration δ, intensity G and separation Δ. Spins' phases are reported for different time steps (1-5) following excitation, refocusing, and application of diffusion gradients. (right) Over three different lines, three scenarios are reported with consequent phase nulling, phase gain, and signal cancellation. Extended from [1].

Diffusion in the presence of time-varying magnetic field gradients is quantitatively described by the solution of the Torrey-Bloch equations, given by

$$M_{xy}(b) = M_{xy}(0)e^{-bD} \quad \text{with} \quad b = \gamma^2 \int_0^t \left[\int_0^{t'} G(t'')dt'' \right]^2 dt' \quad (4.2)$$

Which predicts a simple, exponential decay of the detected signal due to diffusion-related signal loss. The diffusion coefficient D can be quantitatively measured from the detected signal $M_{xy}(b)$ as long as the *b-value* is quantitatively known. The *b-value* represents the amount of gradient-induced dephasing and can be quantitatively calculated for any MR sequence with any magnetic field gradient combination. Usually, the diffusion experiment is repeated with different b-values (either by changing G, ϵ, or Δ) to achieve a reasonable span of signal intensities. In Fig.4.4, a visualization of such an effect is given.
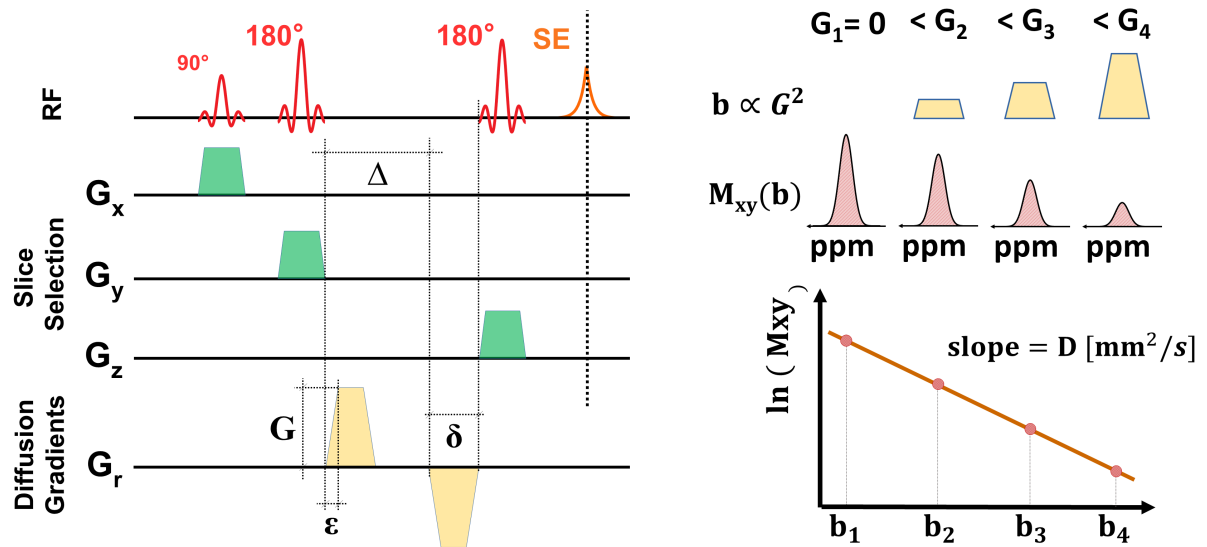
Figure 4.4: (left) Scheme of a 3D spin-echo MRS localization with the addition of bipolar trapezoidal diffusion gradients: Δ diffusion time, G gradient amplitude, ϵ gradient ramp-up time and δ gradient duration. (right) Acquisition of transversal magnetization M_{xy} as a function of different b-values. Four different gradient intensities encode four b-values. Signal loss is exponentially proportional to the diffusion gradient strength. D, the diffusion coefficient, describes the logarithmic slope.

For trapezoidal gradients, G, ϵ, and Δ yield closed formula calculation of the b-value. However, different gradient forms (e.g., half-sine, oscillating gradients, etc.) can be deployed alternatively, with respective pros and cons.

4.2.2. The value of DW-MRS

Diffusion-weighted magnetic resonance imaging (DWI or DW-MRI) generates images with encoded diffusion weighting on water protons. Molecular diffusion in tissues is not random but reflects interactions with many obstacles, such as macromolecules, fibers, and membranes. Water molecule diffusion patterns can reveal microscopic details about tissue architecture, either normal or diseased. The uniquity of DW-MR Spectroscopy (DW-MRS), which, contrarily to DWI, focuses on metabolite protons, resides in its cellular specificity. The intrinsic location of the metabolites, which are *in-vivo* typically present in higher concentrations within the cellular environment, contrasted to the unspecificity of water, found in mammalian tissue both in intra- and extra-cellular compartments. For example, DW-MRS of the brain may return insights on neuronal microstructure investigating diffusion for N-Acetylaspartate and glutamate, which are both considered neuronal biomarkers. Vice versa, it may provide insights into the microstructure of astrocytes via investigation of the diffusivity of myo-inositol and choline compounds, if acknowledged as glial biomarkers. In Fig.4.5, an example of an MRS experiment focused on brain spectra is reported. Six *b-values* are acquired. Diffusion weighting is visually evident from signal loss while increasing *b-values*. To be noted is the higher velocity of the water disappearance compared to metabolites. The intuition that water could be diffusing faster is confirmed by the signal intensity M_{xy} reported on the right either in absolute scale Fig.4.5(B) or logarithmic scale Fig.4.5(C). Water is characterized by bi-exponential decay whereas, in this context, metabolites are well explained by mono-exponential decay and distinct by different decay velocities. When the diffusion coefficient D is estimated using such an experiment, its measure is referred to as *Apparent Diffusion Coefficient* (ADC) to reflect (1) methodological uncertainties that are intrinsic in the MR experiment (e.g., artifactual signal loss from cytoplasmic streaming, blood and lymphatic flow in the microcirculation, bulk tissue motion from cardiac pulsations or respiration, and using dispersion due to susceptibility effects) and (2) the heterogeneity of human tissues.
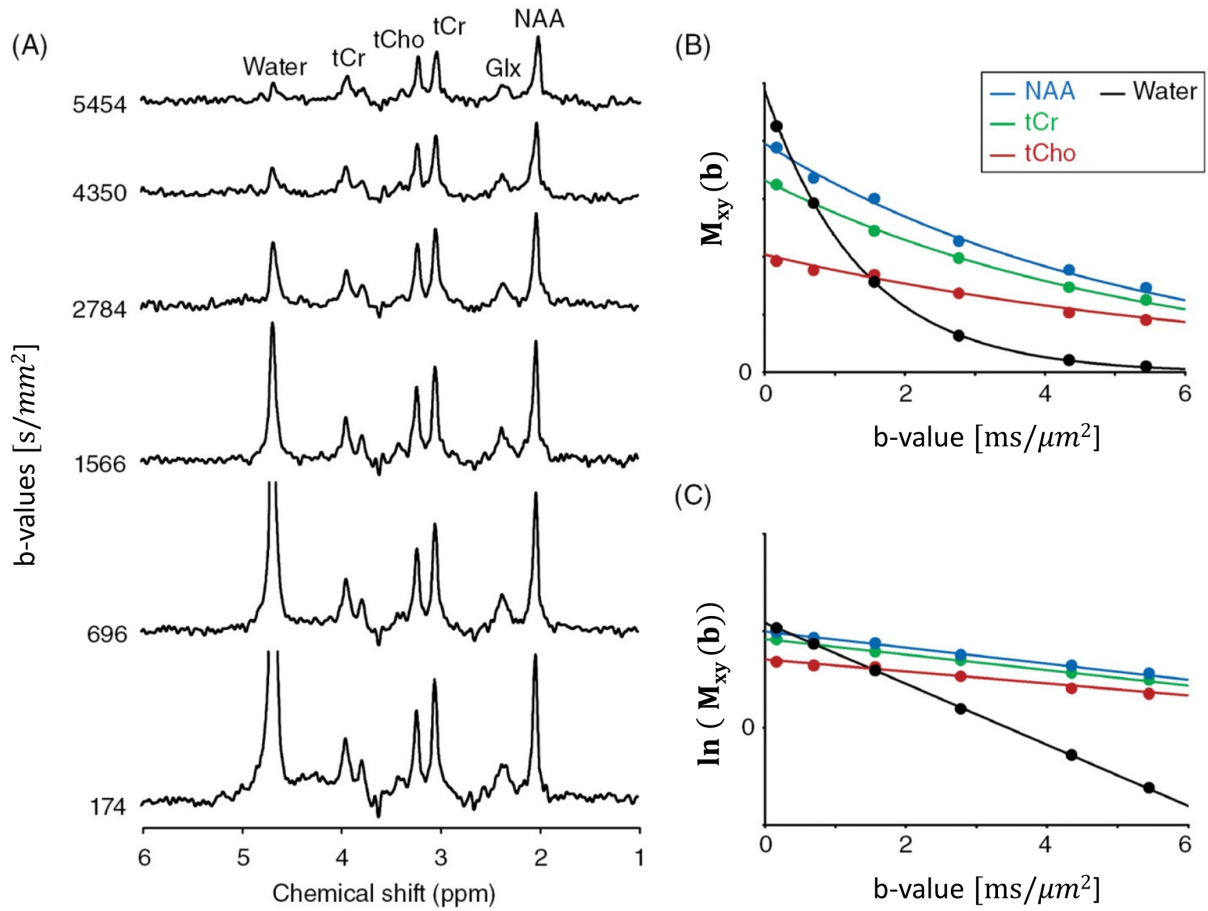
Figure 4.5: Diffusion-weighted ^{1}H MRS experiment. (A) ^{1}H MR spectra of rat brain acquired at 4.7T (PRESS, TR/TE = 3000/144 ms, CHESS water suppression). Different *b-values* are indicated with diffusion-weighting increasing from bottom to top. Observed metabolites are labeled. (B) The signal intensity of the major resonances and exponential fitting. (C) Natural logarithmic of the signal intensity and linear fitting. The diffusion coefficient equals the slope. Figure from [1].

Given a diffusion-weighted MR design, the access to micro-structure estimation is driven by the size of the compartment's explorable distances. Δ defines the extension of the micro-structural explainability of a diffusion experiment. It encodes the maximum diffusion time for which particles experiencing the diffusion encoding gradients are left to travel in the medium. That consequently gives information about the maximum allowed travel distance. For example, Fig.4.6 display different possible explorable travel distance for a neuronal molecule under diffusion weighting. The conformation of such explored space can be monitored by non-linear changes of the particle's diffusivity (i.e., ADC) over increments on diffusion time Δ.
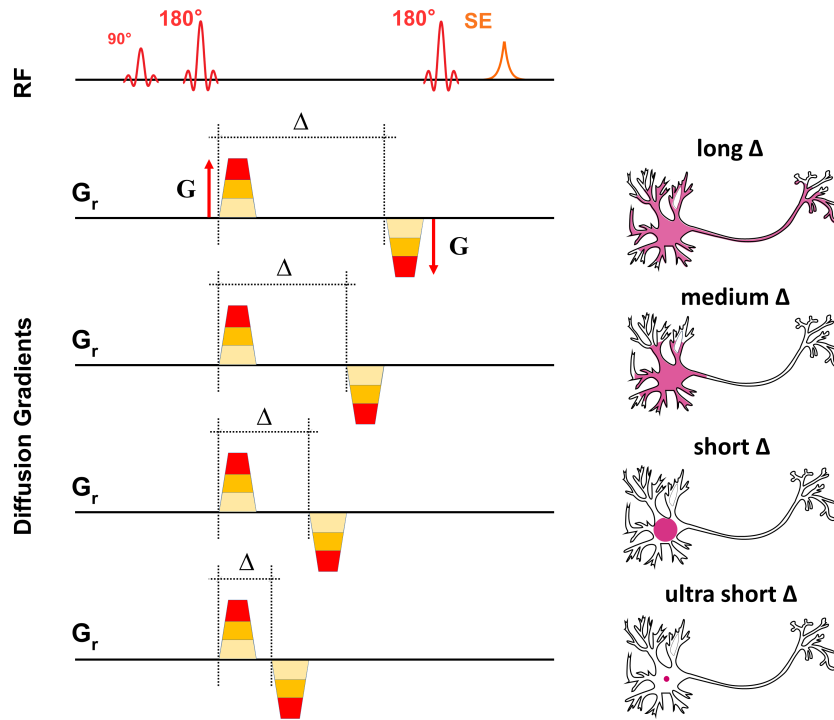
Figure 4.6: (right) Spin-echo localization with diffusion-weighting explored under variation of diffusion inter-time Δ. (left) Explorable volume or diffusion distance is indicated in pink, assuming the diffusing compound is located within a neuronal membrane. The diffusion encoding direction r is here assumed to be along the axon, left-to-right.

For this work, DW-MRS of the prostate investigated creatine, phosphocreatine, and choline as cellular biomarkers. In contrast, citrate and spermine as luminal biomarkers, where the lumen defines the space of a tubular structure characteristic of the prostate tissue arrangement and where such metabolites, when in excess, are secreted, see Chapter 3.2.

4.2.3. Challenges

DW-MRS measurements are sensitive to errors leading to non-negligible artifacts, such as eddy currents, b-value calculation inaccuracies, and motion-related artifacts.

Eddy currents

Significant errors in diffusivity estimations may result from distorted phase and shape of the spectral pattern as a consequence of eddy currents (see Chapter 2.6.1), which artefactual contribution is enhanced by the strong gradients required in a DW-MRS experiment. Bipolar diffusion gradient waveforms can eliminate spectral pattern distortions caused by eddy currents compared to single polarity gradient profiles. Furthermore, additional non-water suppressed or metabolite-cycled acquisitions for each diffusion-weighted scan can be employed.

Motion

It should be noted that linear non-accelerated motion does not lead to signal loss per se but does lead to a global phase offset. When the motion changes from scan to scan (e.g., macroscopic motion like patient motion, accelerated, decelerated, or pulsating blood flow, etc.), the net phase $\Delta\phi$ will also change. This scan-to-scan phase variation during signal averaging in MRS can lead to signal loss. This effect can be eliminated by storing and phase-correcting each transient separately before summation for MRS. Physiological motions can be minimized by employing triggering or motion tracking while measuring. For this work, a compensation scheme in post-processing based on water signals has been used to correct motion-related artifactual losses [45].

Dimensionality

DW-MRS relies on signal weighting, which worsens the already low intrinsic SNR of MR Spectroscopy. For the time being, the current more significant limitation is the display of such high cellular specificity in only rather extensive single-voxel experiments. Therefore, advancements toward encoding DW-MRS Imaging yielding micro-structural specificity and its spatial variation will be of absolute value.

5 | Deep Learning and its application to *in-vivo* MRS

MR spectroscopy, a technique so fine,
for studying the chemical make-up, so divine.
it is a type of medical scan,
a way to look inside the body of a man.
Magnetic fields and radio waves,
make it so we can see what's inside our caves.
From metabolism to disease,
MR spectroscopy makes it a breeze.
A non-invasive way to diagnose,
a powerful tool in the physician's notes.
With its high resolution and sensitivity,
MR spectroscopy is a technology of great versatility.
It's the future of medicine, that's for sure,
a powerful tool for finding the cure.
So let's give a cheer for MR spectroscopy,
it's like a magic trick, you'll see,
a technique that's revolutionizing the field,
making diagnoses more accurate, and more reveal.
It's like a detective solving a crime,
MR spectroscopy is the ultimate find!

"poem on MR Spectroscopy" by ChatGPT-3

This chapter is partly inspired by [46] and [47]

5.1. The charm of Artificial Intelligence

Artificial Intelligence (AI) is nowadays striking society with the charm of generative models. A well-known example is Deepfake, capable of generating synthetic yet realistic portrait images or videos of famous faces adapted to reproduce any short talk with any text of your liking. Worth mentioning are as well Vista simulator, a synthetic photo-realistic generator of environments for training autonomous vehicles or *OpenAI*'s works on natural language processing like image (DALL-E-2) or code/language (ChatGPT-3) generators.

AI starts pushing the limit of any *Turing's test* and lets the average user be unaware of small details and uncertainty, which are clear examples of AI's pitfalls and where human capabilities are yet fundamental and needed. We are definitely diving into a world where most likely, AI will not replace human beings, but a person *using AI* may do. I guess we all agree, for example, that the marvelous "poem on MR Spectroscopy" by ChatGPT-3 is affected by at least one striking ambiguous statement: *with its high [...] sensitivity.* Isn't it quite the opposite?

The recent *(November 2022)* performances of AI in its ability to generate fluent language have hit hard the scientific community worldwide, reflecting dark shadows and threatening the transparency of science itself. From its earliest times, science has operated by being open and transparent about methods and evidence, regardless of which technology has been in vogue. Researchers should ask themselves how the transparency and trustworthiness that the process of generating knowledge relies on can be maintained if they or their colleagues use software that works in a fundamentally opaque manner. Interestingly and tempestively, I must say, both *Science* and *Nature* have already commented with the utmost intent of laying down ground rules and principles on ethical AI usage in regards to authorship of AI-related bots or proper documentation of the usage of such tools as a research assistant in the process of doing science [48, 49].

5.2. Deep Learning (DL)

The term AI defines whatever system is capable, to a certain extent, of imitating an intelligent human behavior. Therefore, it generally carries a sheer broad and philosophically oriented meaning. On the other hand, Machine Learning (ML) is tuned on a more practical view: it defines any set of algorithms capable of automatically learning and improving from experience. ML deploys math and statistics twinned to hand-engineered features that help to tailor the processing of any raw input to a desired output. Sometimes, the feature engineering process can be time-consuming, brittle, and practically not scalable. As soon as underlying features are assumed to be intrinsically represented within a vast amount of raw data and therefore automatically learnable, it comes to the definition of Deep Learning (DL). The surge of DL since 2010 is a direct consequence of (1) Big Data science (i.e., availability of big datasets and their

easy collection and storage), advancements in (2) hardware (i.e., among all, graphics processing units - GPUs - were the game changer) and (3) software, especially plug&play toolboxes. All of this just brought back fundamentals of statistics (in fact, I dare to say, just modernly renamed into ML) known from the late '50s.

5.2.1. Supervised DL

Supervised deep learning is a class of DL algorithms that uses multiple layers to progressively extract higher-level features by matching raw input to the desired output. Therefore, it relies on the availability of labeled datasets: a vast amount of data in which solutions for a desired task (i.e., the output) are known and given. It assumes that a deep architecture can learn features by matching input-output in multiple iterations via a network-update training called *back-propagation*. After training, the algorithm should be capable of handling any *reasonable* unlabeled input predicting its most likely output according to the rules it has learned in training. This is also known as *learning by examples*.

Neural Networks

The fundamental ingredient of DL is a Neural Network (NN). NNs are computing systems inspired by the biological neural networks that constitute animal brains. NNs are based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain, Fig.5.1.
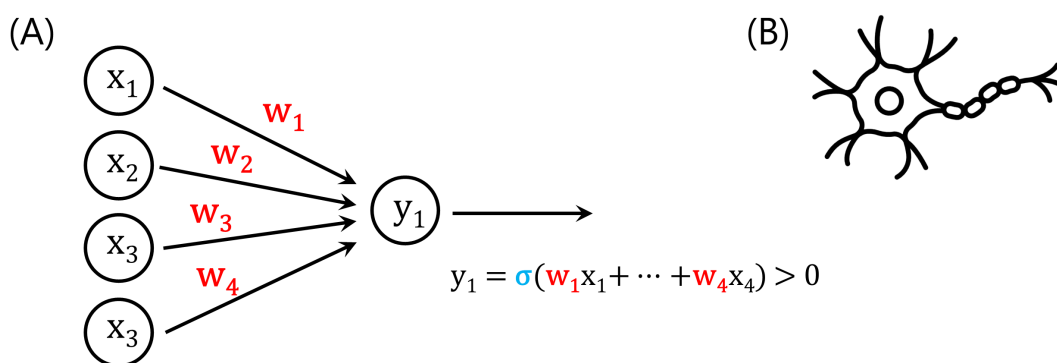
Figure 5.1: (A) Representation of *perceptron* network: the structural building block of deep learning. A central node y_1 receives input signaling from four nodes x_i. Each node x_i transmits on a channel modulated by the weight w_i. The node y_1 will transmit further (i.e., it will fire) if and only if enough input power, given by the sum of all modulated input connection, overshoot a null threshold. The output power is modulated by non-linearity σ that gives plasticity to both reactivity and processing of y_1. (B) The analogy to human brain neurons.

Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. A neuron (or node) receives signals then, processes them, and can signal neurons connected to

it. The *signal* at a connection is a real number (typically), and the output of each neuron is computed by deploying a non-linear function of the sum of its inputs. The connections' firing capability, or in other words, the strength of a connection, is modulated by a weight that adjusts as learning proceeds.

Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs and aggregate information on a different level of complexity. An example is given in Fig.5.2, where digitalized images represent handwritten numbers. Each pixel of the image encodes a value from 0 to 1 (i.e., pixel brightness), Fig.5.2(A). Pixels assemble into edges, and carefully selected combinations of edges will constitute shapes or patterns. Such patterns can be organized in images, like digits in this case.
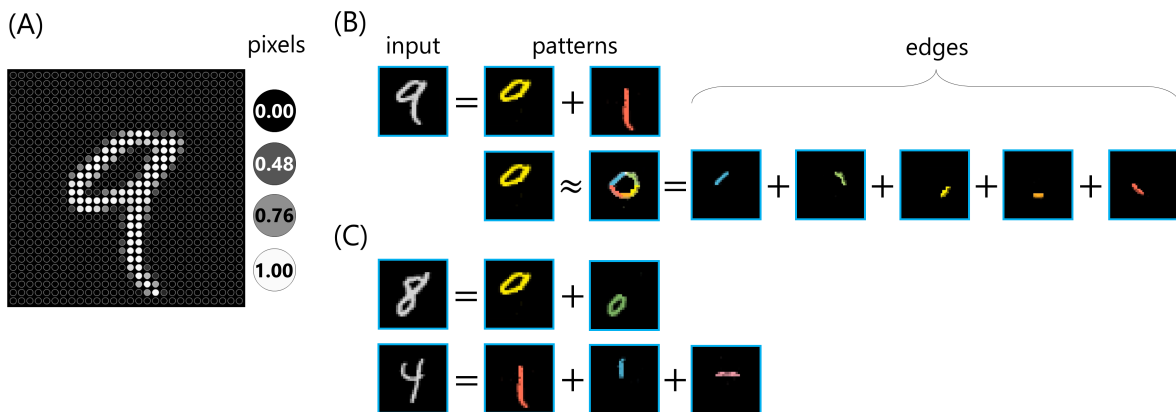
Figure 5.2: Interpretation of DL learnable features for images of handwritten digits. (A) A digitalized handwritten 9 is visualized as a set of pixels whose intensity reflects brightness. (B) The image can be decomposed into a superimposition of simpler images depicting high level patterns such as circles and lines. Such patterns can be further decomposed into simpler features like straight edges. (C) Different handwritten digit share common patterns, detectable by the same combination of edges.

Such logic directly translates into a deep architecture of fully connected nodes, Fig.5.3. Signals travel from left to right, from the first layer (the input layer) to the last layer (the output layer), traversing the layers multiple times (i.e., forward propagation). In image processing, lower layers typically identify low-level features, such as edges, while higher layers may identify the concepts relevant to a human, such as patterns and shapes.

The correct mocked state in Fig.5.3 assumed the network properly trained. Neural networks learn (or are trained) by processing examples, each of which contains a known *input* and *output* (i.e., the label or ground truth) forming probability-weighted associations between the two, which are stored within the data structure of the network itself. The training of a neural network from a given example is usually conducted by determining the difference between the processed output of the network (i.e., the prediction) and the target output. This difference is formally

known as the *cost function* (or error function) and can be propagated from the output layer back to the single weight of each connection in the network. The network then adjusts these weights according to a learning rule using this propagated error value. The overall philosophy of propagating error back through the network's connection is known as *back-propagation* whereas its implementation is known as *gradient descent algorithm*. Successive adjustments will cause the neural network to produce output that is increasingly similar to the target output. After a sufficient number of adjustments, the training can be terminated based upon convergence criteria driven by sheer *minimization loss*, *regularization*, *early-stopping* or *patience*.
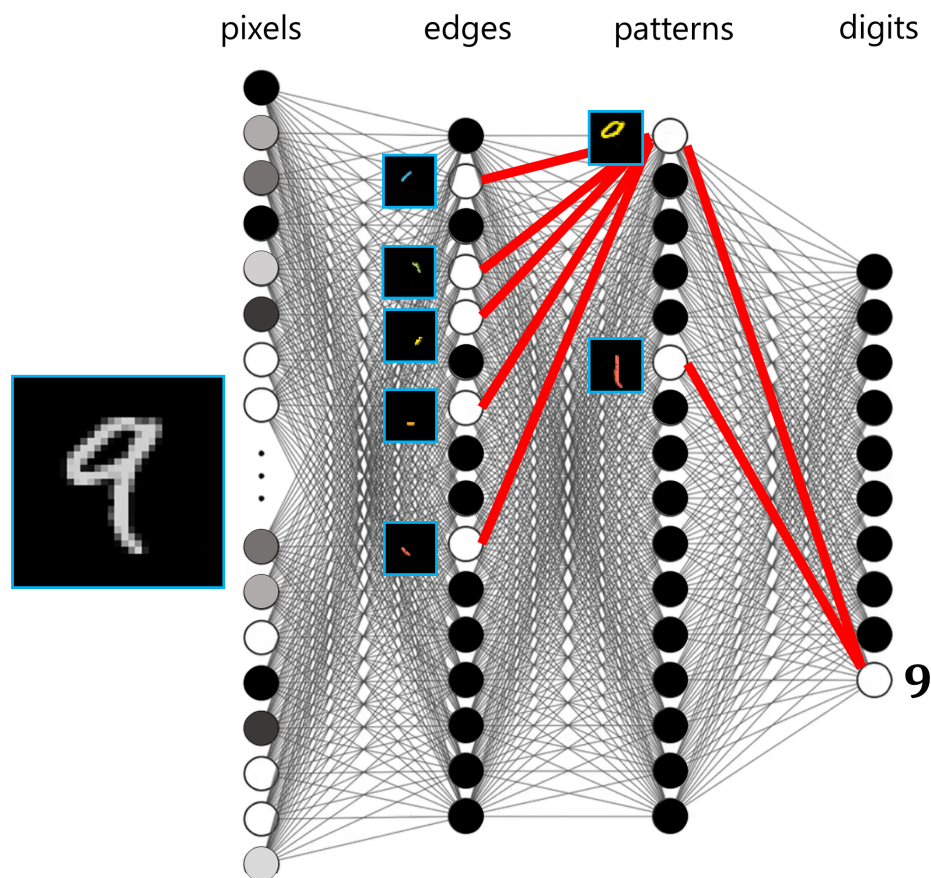
Figure 5.3: A deep fully-connected neural network is composed of four fully connected layers with a different number of nodes. The network mocks an activation state (in red) upon network feeding with the handwritten 9 as the input image. The matrix of pixels is re-arranged into a vector, and nodes in the 1st layer are activated with an activation level proportional to the correspondent pixel's brightness. On the 2nd layer, five nodes yielding edges that define a circular shape are activated. Their activation is propagated to the 3rd layer, where a circular shape is properly detected and activated. At the 3rd layer, a vertical pattern is as well activated from an independent pathway through the 1st and 2nd layers (here not mocked). The two nodes fire at last, with consequent activation for the 9th node in the output layer, which the user interprets as digit 9.

5.2.2.   Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are a class of widely used deep learning models designed to automatically and adaptively learn spatial hierarchies of features from input data. Their architectures cleverly and drastically reduce the number of parameters when compared to dense, fully connected counterparts.

A CNN architecture, depicted as an example in Fig.5.4, typically consists of an input layer, several hidden layers, and an output layer. The hidden layers include multiple *convolutional layers*, which use a small kernel (or filter) to scan over the input data and learn local patterns. They are typically followed by *pooling layers*, which down-sample the data to reduce its dimensionality and increase the network's invariance to small changes in the input. The output layer is usually a fully connected layer, which combines the features learned by the convolutional and pooling layers to make a final prediction.
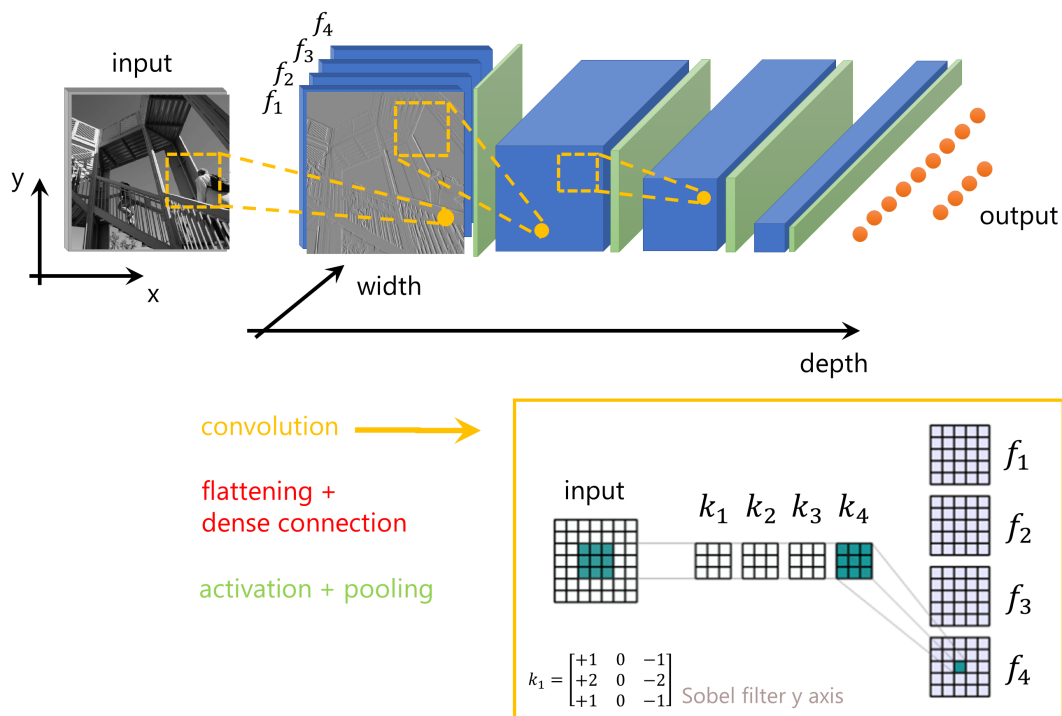
Figure 5.4: Convolutional Neural Network: architecture. A grayscale image is provided as input to a 4-layer CNN network. The 1st layer has 4 features f_i. Each feature f_i is calculated via convolution operation over the whole input image. The convolution operation is modulated by a filter (or kernel) k_i that explores the input image like a sliding window. If a *Sobel y-axis filter* is assumed as kernel k_1, the resulting feature f_1 will display vertical edges (i.e., rapid changes of pixel intensity) of the input image. Width: defines the number of features explored in each layer. Depth: defines the number of layers.

CNNs have been applied to various computer vision tasks such as image classification, object

detection, and semantic segmentation. They are also used in natural language processing tasks such as machine translation and speech recognition or for other applications such as drug discovery, protein folding, and audio processing.

Architectures

In recent years, the architecture of CNNs has been continuously evolving and improving, with new architectures being proposed that have achieved state-of-the-art performance on various benchmarks. Depending on the specific task, different CNN architectures can be used, including those for classification, regression, and processing, Fig.5.5.
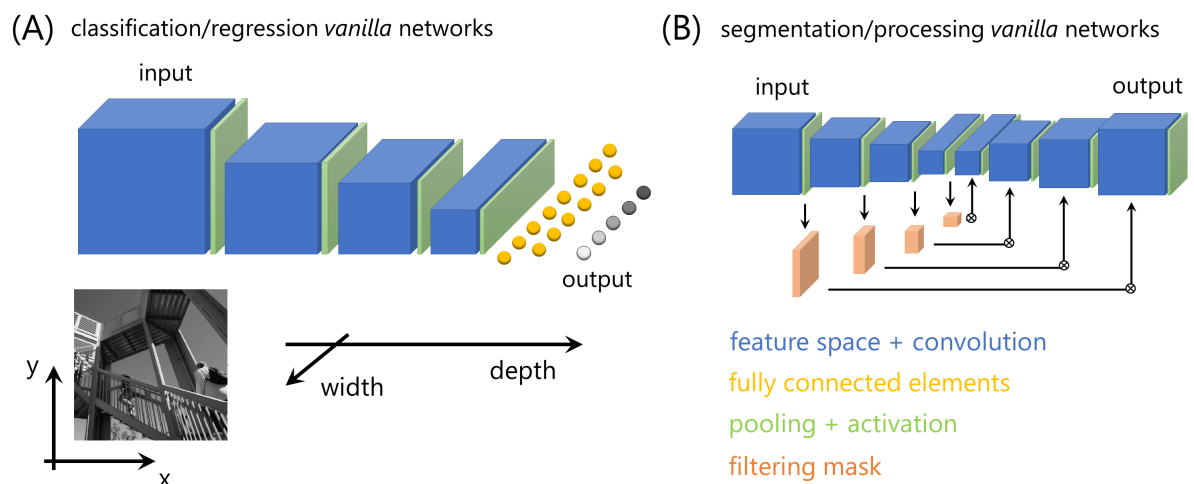
Figure 5.5: Task-oriented CNN architectures. (A) Classification or regression v*anilla* architecture: data are passed through a funnel-like structure where the output answer to a specific task and is fully unrelated to the input domain. (B) Segmentation or processing *vanilla* architecture: the encoder-decoder architecture aims to compress and extend the feature space to ensure input-output domain match.

Classification CNNs are used to predict a class or category for a given input. The output layer is usually a fully connected layer, which combines the features learned by the convolutional and pooling layers to make a final class prediction. One of the most popular classification CNNs architectures is *AlexNet* proposed in 2012 [50], which achieved state-of-the-art performance on the *ImageNet* dataset and set a new benchmark for image classification tasks.

Regression CNNs are used to predict a continuous value for a given input. These architectures typically consist of a similar structure as the classification CNNs, with the key difference being that the output layer is a fully connected layer where neurons produce a single continuous function as output. A popular regression CNN architecture is *C3D* proposed in 2015 [51], which is designed for video analysis and can be used for tasks such as action recognition and video captioning.

Processing CNNs are used to process the input data. These architectures typically consist of an *encoder-decoder* structure, where the encoder is used to extract features from the input (and encode them in a smaller subspace, usually known as *latent* space), and the decoder is used to produce the final processing, by decoding the encoded features and mapping them in the original input space. Among processing CNNs, specific fields have risen, such as **segmentation**, to extract spatial features from images or videos and segment them into different regions or segments, **denoising**, to alleviate nuisance component or remove artifacts, and **generative CNNs**, used for data synthesis, image-to-image or text-to-image translation [52]. One of the most popular processing CNN architectures is *U-Net* proposed in 2015 [53], which is designed for biomedical image segmentation and has been extended to various other applications such as object detection and scene segmentation, representing still, after almost 10 years, one of the main building blocks for current state-of-the-art DL solutions.

Architectures are not just task-oriented. The continuous evolution and improvement of CNN architectures are based as well on overcoming limitations or state-of-the-art performances and are based on a variety of benchmarks. Some of the most popular architectures that are worth mentioning for the purpose of this work follow.

Residual Networks (ResNets) are designed to address the problem of vanishing gradients, which can occur in deep networks, by introducing residual connections (also known as *skip* connections) between layers, Fig.5.6(A). These connections allow the network to learn the residual mapping between the input and output rather than the entire mapping, making possible the deployment of very deep architecture in the current practice [54].

Inception Networks (InceptionNet) are designed to address the problem of computational efficiency by introducing the idea of *inception modules* that use multiple filters of different sizes in parallel, rather than using one filter of a single size, Fig.5.6(B). This allows the network to learn a variety of features at different scales, which improves its performance and reduces the number of parameters that it needs to learn [55].

DenseNet is an architecture that uses dense connections between layers. Each layer is connected to all the previous layers, which allows the network to reuse features from previous layers and reduce the number of parameters it needs to learn. This also helps to mitigate the vanishing gradients problem. It allows the training of much deeper networks [56].

The most recent **EfficientNet** is designed to be both efficient and effective. EfficientNet uses a compound scaling method to scale the depth, width, and resolution of the network, which improves the performance of the network while keeping the number of parameters low. This architecture also uses a novel architecture search method to find the optimal architecture for a given dataset, which made it possible to achieve state-of-the-art performance on a wide range of benchmarks [57].
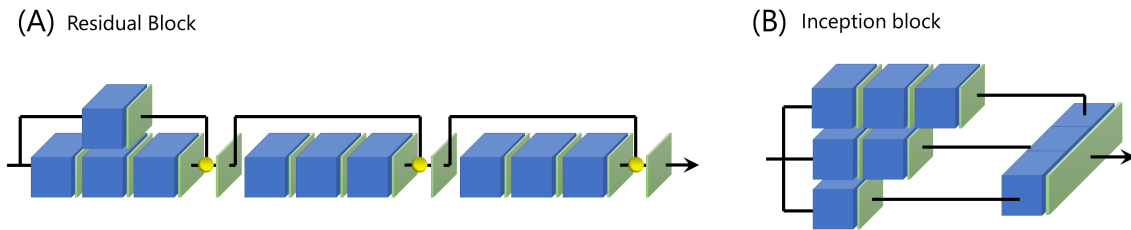
Figure 5.6: Visualization of the main building blocks for (A) ResNet and (B) InceptionNet. (A) Residual connections (also known as *skip* connections) are designed to skip three convolutions linking two separated feature spaces. (B) Different filters are applied throughout the three branches of a layer. Each filter runs a different level of feature extraction and down-sampling.

5.2.3. The issue of interpretability

Despite DL potential, which is well explained by sometimes vague or ill-defined ranking performances but still outperforms humans in many meaningful tasks, performances and applications have been questioned due to the lack of interpretability. A model is typically considered interpretable because it is small and basic enough to be completely comprehended. Ideally, the user should understand the learning process well enough to realize how it forms the decision limits from the training data and why the model has these rules. For ordinary users, DL models have been treated as black-box universal function approximators, where prediction explanations are no longer available as their traditional counterparts (e.g., machine learning tools such as *logistic regression* and *random forests*). Lack of interpretability hinders the wide application of DL models in critical domains like healthcare. In addition, due to bias in datasets or models, decisions made by DL algorithms are prone to be unfair, where an individual or a group is favored compared with the others owing to their inherent traits (e.g., in a modern society, pathological cases may be represented as a minority class in a dataset which models the overall population). As a result, more and more concerns about interpretability, fairness, and biases have been raised recently in the healthcare domain, where human lives are at stake. Comprehensive analyses of model interpretability, dataset bias, algorithmic fairness, and the interaction between interpretability, fairness, and robustness should come as a prerequisite to lift concerns to the development and usage of any DL-based algorithm [58, 59].

An example of a rich and explorable user interface that unifies many interpretability methods individually developed by the DL community is summarized in Fig.5.7. It shows *what* a network detects and explains *how* it develops its understanding while keeping the amount of information on a *human scale* [60]. Nevertheless, despite the overall grasp that it gives on the undergoing process, it remains very complex to process and fully explain. A lot of work is left ahead to build powerful and trustworthy interfaces for interpretability, especially because (1) the stakes can be high (as in safety and fairness) and (2) because ideas like training models with interpretability feedback put interpretability techniques in the middle of an adversarial setting [60].

INPUT IMAGE	ACTIVATIONS of neuron groups

NEURON GROUPS based on matrix factorization

color key				
feature visualization of each group				

EFFECT of neuron groups on output classes

Labrador retriever	4.340	-0.981	-2.386	1.155
beagle	1.295	0.136	-0.566	-0.135
tiger cat	-1.189	-2.305	0.254	0.209
lynx	-0.690	-0.519	-0.159	1.141
tennis ball	1.654	0.067	-1.189	1.238
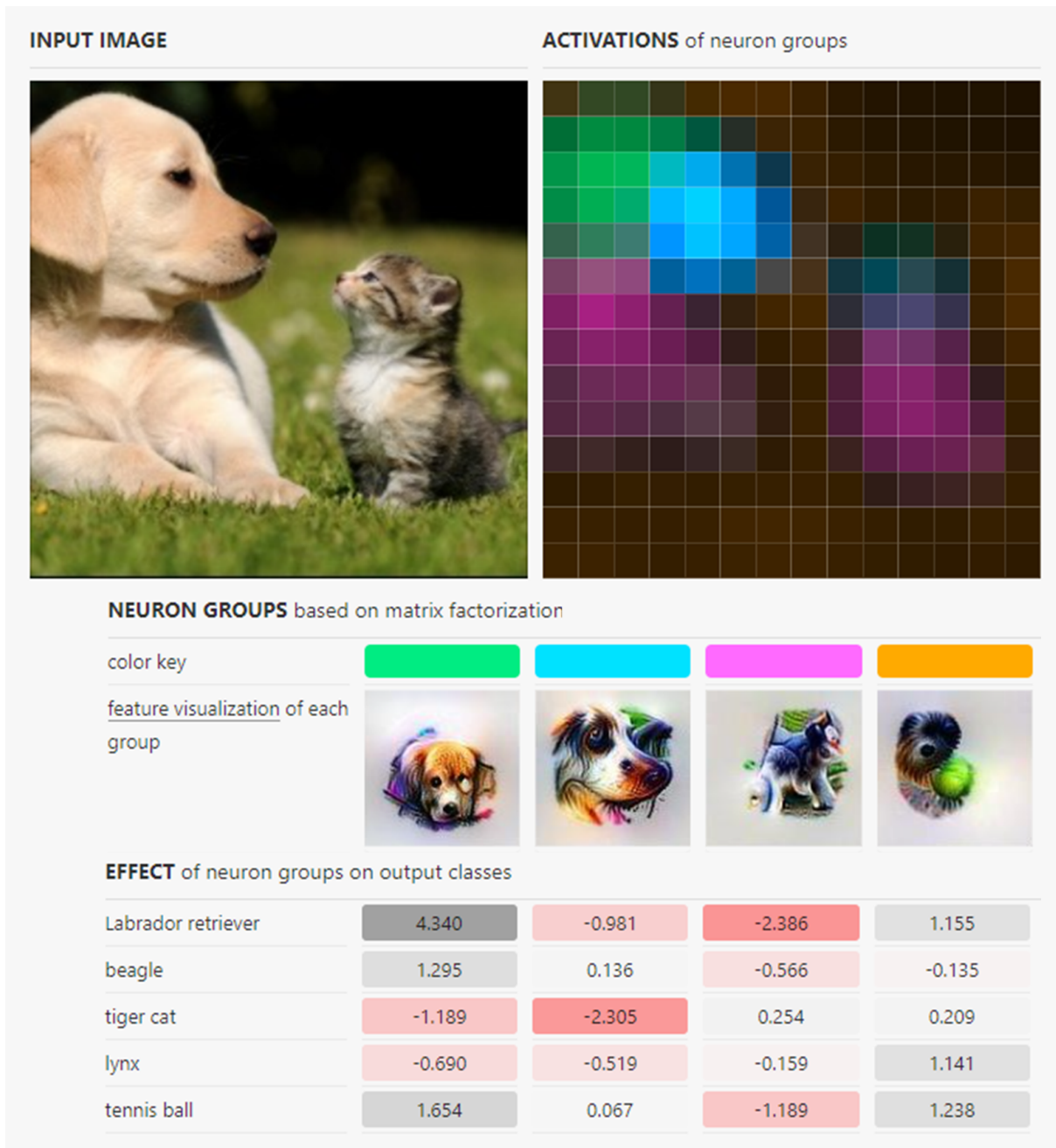
Figure 5.7: Prediction and interpretability of an InceptionNet-like architecture for classification of images. (Top) Input image and activation map. The activation map shows the image-domain contribution of the four most likely classes that were pointed out by the classification network [61]. (Middle) Neurons are grouped in four color-coded collections of pixels that partition the input image (i.e., orange is viewed as the background dark-ish orange in the activation map). The *feature visualization* of each group gives a visual understanding of the network activations upon feeding the network with the color-coded portion of the input image (i.e., a view of what the network sees). (Bottom) The effect in prediction for the four groups reflects the likelihood of pinpointing a certain class based on the pixel selection. Interesting the ambiguity among dog breeds while considering either head (green) or muzzle (blue). Figure from [60].

5.2.4.   Reliability and robustness

Despite the recent improvements in overall accuracy, deep learning systems still exhibit low levels of robustness. Robustness is defined as *the ability of a system to withstand (keep regular and anticipated function) in spite of exceptional, unforeseen events, stressful conditions such as component failures, and extreme conditions beyond the expected operating environment* [62]. From its definition it is straightforward to guess that detecting possible failures is critical for a successful clinical integration of these systems, where each data point corresponds to an individual patient [63]. Reliability, the extent to which the algorithm performances can be reproduced upon repetition under the same conditions, comes along when failure is to be monitored. A visualization of robustness and reliability issues is given in Fig.5.8.
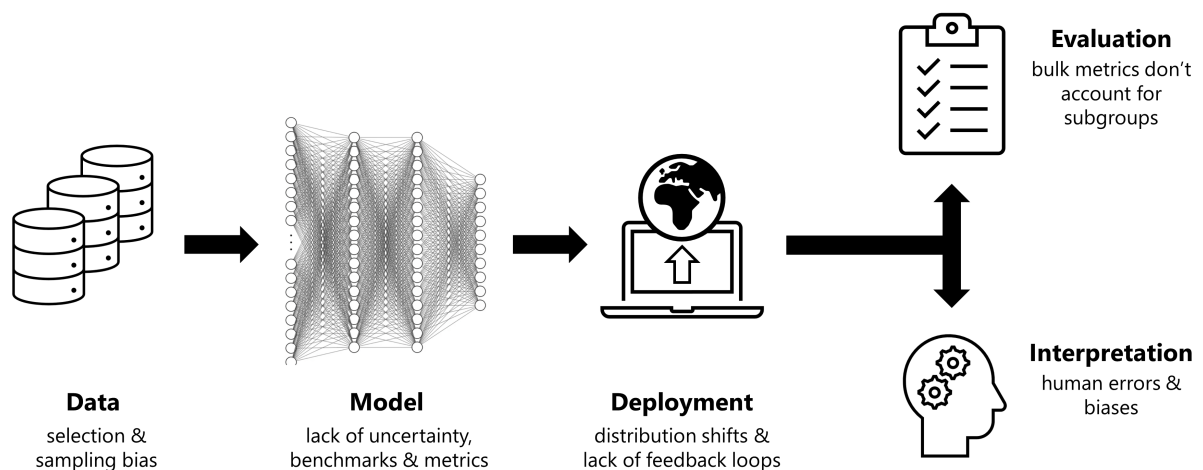
Figure 5.8: Visualization of reliability and robustness issues in a DL lifecycle. Questions of robustness come up at different level of a DL lifecycle and should be accounted for when developing distributed algorithm, especially when they aim to constitute black-box tools in the hands of average/naive users.

The current challenges for robust deep learning can be wrapped up in two categories: *bias handling* and *uncertainty estimation*. Bias handling refers to understanding, monitoring, and preventing model skewness by sensitive input features. A clamorous exemption in a real-world application was the race and gender bias detected by DL algorithms in facial recognition tasks, which came into play as the natural consequence of class imbalance in the training set, where the incidence of male and white people was higher than female or other ethnicities. The impact of training distribution has also been witnessed in the MR field, influencing the estimation of either tissue microstructure [64] or metabolite concentrations [65]. Approaches to mitigate class imbalance act either on the dataset, if possible by appropriately weighting the distribution of classes and features evenly, or on the network's feature handling, by *debiasing* the latent feature space to generate a fairer distribution where to learn from. Sometimes *debiasing* is also referred to as *model calibration* [66, 67, 68].

Uncertainty

Measures of uncertainty are a promising direction to improve failure detection since they measure the system's confidence [69]. This information, in turn, can be used to leverage the decision-making process for a user (e.g., how much to trust a given prediction) or to enable optimization of the deployed algorithm (e.g., adversarial training). Although many uncertainty estimation methods have been proposed for deep learning, (1) little is still known about their benefits and challenges for various medical applications [63], (2) general acceptance and agreement upon methodology is far from established [70] and (3) human-level readability and interpretability are yet open issues.

There is widespread agreement on classifying two main types of uncertainty: *aleatoric* and *epistemic* uncertainties [71]. The irreducible uncertainty in data that gives rise to uncertainty in predictions is named *aleatoric uncertainty* (also known as *data uncertainty*). This type of uncertainty is not a property of the model but rather an inherent property of the data distribution. Hence, it is irreducible. In contrast, *epistemic uncertainty* (also known as *knowledge uncertainty*) occurs due to inadequate knowledge. In other words, let's assume one can define models to answer different questions in model-based prediction. On the one hand, for data-rich problems, there may be massive data collection that is information poor (i.e., pointing to higher data-driven uncertainties). On the other hand, a given DL-based method may be found to be an inefficient estimator for characterizing the emergent task-oriented features of the data (i.e., pointing to model-driven higher uncertainties). A toy example of visualizing the effect of uncertainty is given in Fig.5.9.
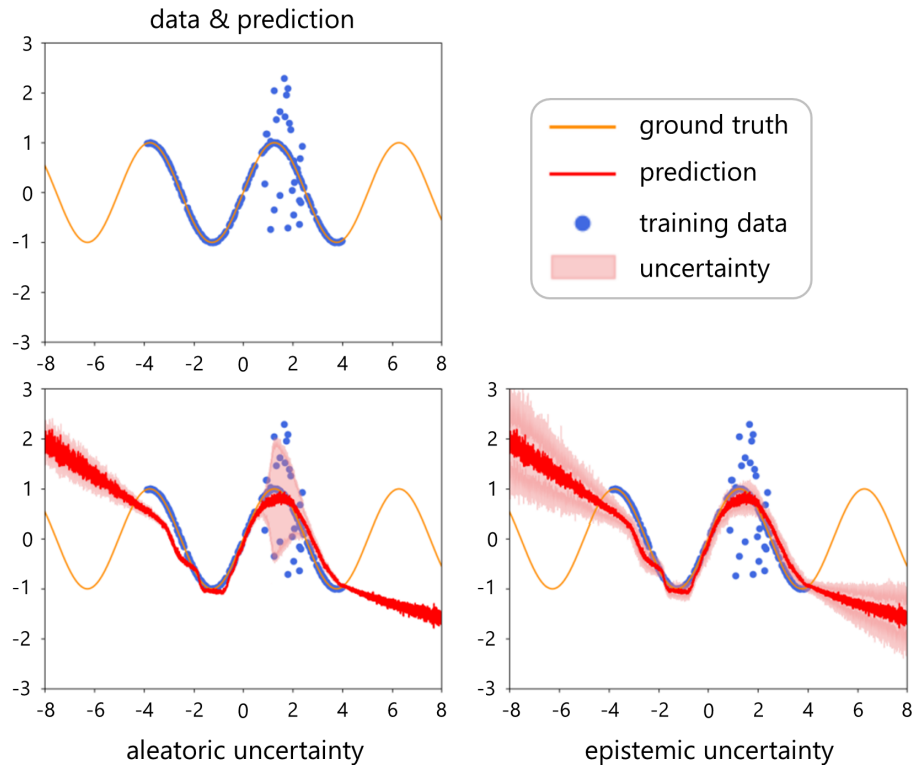
Figure 5.9: (Top) A toy example of a data set (training set, blue) is distributed in a 1D domain $D = [-4, +4]$ following a mostly sinusoidal pattern that matches ground truth. A DL model (not displayed) decently predicts the sinusoidal wave (red) within the training data domain D. Visualization of aleatoric (bottom-left) and epistemic (bottom-right) uncertainties. Despite good predictions, a measure of uncertainty relates to the interval $[1, 3]$ where training data suffered a bigger variance (aleatoric). A second measure of uncertainty relates to training distribution, displaying higher uncertainty when predictions are asked for data outside the training domain D (epistemic). Modified from [72].

To teach a model to recognize when it doesn't know the answer, different uncertainty estimation approaches have been introduced [70]. Among all, for this work, the reader will be introduced to *Monte Carlo dropout* and *Bootstrapping*: two computationally friendly strategies to model epistemic and aleatoric uncertainty, respectively. Monte Carlo dropout relies on NNs deploying dropout layers. During the training process, these layers randomly drop some units of the NN to prevent excessive co-tuning. Therefore, in the training phase, it works as a regularization term to solve overfitting problems. Dropout layers are typically switched off in the testing phase (i.e., all the units of the NN are exploited to predict the output). However, the idea of predicting multiple outputs from multiple identical networks where a portion of units is every time randomly switched off (i.e., Monte Carlo dropout) may suggest insights into the robustness of the network model itself (epistemic). Bootstrapping is, by definition, a statistical procedure that resamples a single dataset to create many simulated samples. This process allows for calculating standard errors, constructing confidence intervals, and performing hypothesis

testing. Thus, bootstrapping will allow the resampling of the testing set to evaluate measures of uncertainty, such as the expected estimated value and coefficient of variance for a subset of predictions, mirroring mainly the property of the dataset itself and therefore addressing the aleatoric component of uncertainties [69]. The two methodologies are visualized in Fig.5.10.
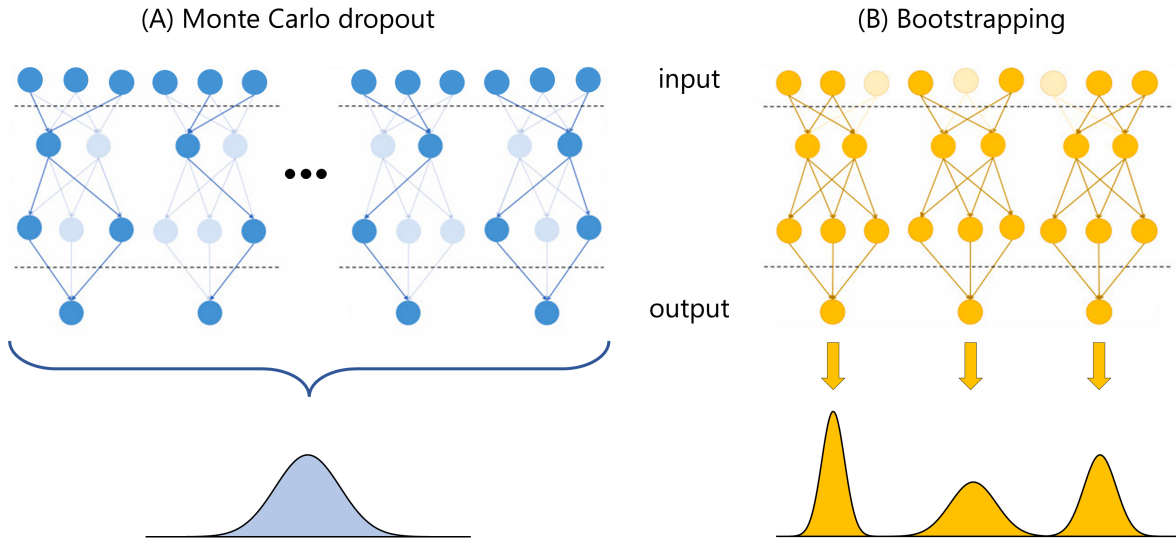
Figure 5.10: Visualization of (A) *Monte Carlo dropout* and (B) *input bootstrapping* as uncertainty estimators. Monte Carlo dropout estimates output predictions multiple times for the same input set across *different* networks. The ensemble of the networks' predictions yields a distribution which statistics give insights into epistemic uncertainty. Bootstrapping splits the input sets into subsets. Each subset's prediction gives a distribution of predictions suggesting insights into aleatoric uncertainties.

Need to be stressed that the mathematical definition of uncertainties for DL is currently under debate and far from general acceptance [69]. Across different options, the bridge between theoretical assumption and practical implementation often relies on trade-offs and approximations that may compromise or limit the usability of such tools according to either specific task-oriented networks or architectures. Despite flourishing literature addressing uncertainty estimation as a complementary tool for DL interpretability, a full-scale analysis of such models' robustness, reliability, and readability is still challenging [63].

5.3. Application of DL in *in-vivo* MRS

The application of DL in the *in-vivo* MRS field is dated back to 1997 when automatic tumor grading was first explored by deploying MRI and MRS data twinned to NNs. DL in *in-vivo* MRS remains a pretty small research niche, with few published papers compared to other fields (i.e., one paper per year from 1997 to 2017), such as computer vision or medical imaging analysis. Nevertheless, exponential growth in the reporting of DL tools deployed for *in-vivo* purposes came

into play in 2018, a year with four published papers. It ended with thirteen articles published in 2022[1]. To date, most of the approaches rely on supervised DL, with very few recent exceptions [74, 75].

Besides pathology classification, three common practical problems are appealing for DL methods and applications in *in-vivo* MRS:

- the lack of time for acquisition

- the urge for speed and model-free postprocessing

- real-time and model-free quantification

The lack of time fo acquisition is a well-known limitation for MRS, where prolonged acquisitions are needed to enforce sufficient spectral resolution and SNR. Such an issue extends naively to MRSI, where in the first place, a new spatial dimension is wanted, and its encoding shall be introduced. Secondly, intrinsic chemical shifts do not wish to get scrambled by frequency encoding gradient, prolonging the acquisition time for conventional scans even further. DL stands as a powerful alternative to other conventional methods providing efficient reconstruction techniques to repair information loss when any of these limitations is overtaken by either signal truncation [76], data undersampling [77, 78] or data generation [79]. DL applications for signal denoising and consequent enhancement of *apparent* SNR are also recently investigated as a trade-off between signal quality and acquisition time [80, 81, 82]. However, their acknowledgment by the community has yet to come (they are mostly either pre-prints or found in conference proceedings) compared to conventional possibilities [83, 84, 85, 86].

Computational time relates to speed when it comes to postprocessing and quantifying MRS data. For the time being, MRS suffers from weak clinical integration due to a lack of standardization of postprocessing and quantification routines, in the first place due to the intrinsic complexity of the technology that hampers a community consensus, but in the second place, directly due to the matter of haste that hinders real-time complete quantitative evaluation. For example, intuitively, the vast amount of data generated by advanced fast MRSI acquisition currently challenges any conventional postprocessing pipeline. Here DL has already shown its potential in time-and-data compression [87].

Likewise, the complexity of modeling and setting prior knowledge to assess a single MR spectrum quantitatively challenges any average user at the scanner. DL power as a representative technique capable of discovering essential features embedded in large datasets and therefore determining complex nonlinear mappings between inputs and outputs suggested its early stage

---

[1]Results analyzed on *Web of Science*: keyword (`TI=(deep learning OR neural network OR network OR artificial intelligence) AND AB=(MR Spectroscopy OR Magnetic Resonance Spectroscopy OR Magnetic Resonance spectroscopic)`) and filtering based on *in-vivo* MRS applications [73]

deployment in model-free contexts, where prior knowledge was absent or formal assumptions were complex, unsure, or untrue. As a matter of fact, *model-free* artifacts are common in the MRS field (e.g., spurious echoes, *some cases of* lipid contamination, motion). Despite the possibility of minimizing their impact in the acquisition phase, it can either be hard to eliminate their presence entirely, model it with appropriate prior knowledge or even discern their source of origin. Postprocessing by DL is currently investigated and does not simply stop to anomaly/artifact detection or spectral quality assessment [88, 75], but challenges the nature itself of MR signals by introducing postprocessing correction and reconstruction pipelines to eliminate various artifact sources restoring value to the acquired data [32, 89, 74].

5.3.1. Quantification of MR spectra

DL for quantification of MR spectra has boosted feature extraction to perform time-efficient and accurate quantification of MR spectra by peak picking routines. It has been reported that, on the one hand, these routines may lift, limit or change the paradigm to which prior knowledge is supplied to the algorithm while simultaneously reporting equivalence in quantitation performance compared to traditional model fitting-based quantification approaches [90, 91, 76, 92]. On the other hand, they introduce a substantial gain in speed (i.e., time is spent in training, and at run-time, the algorithm is much faster than a minimization problem in the least square fitting approach) and usability for the average user (i.e., black box, ready-to-use and user-friendly tools which do not require fine parameter tuning at run-time).

Nevertheless, questions arise concerning the robustness of DL algorithms. In fact, DL has shown excellent performance for classification or segmentation tasks but may suffer from inherent weaknesses in subsets of representative outlier samples. DL architectures for MRS quantitation have mostly been investigated for sample distributions of near-healthy spectral metabolite content. Hence, it can be suspected that high accuracy and precision are mainly found when DL is deployed for new entries of similar near-normal types. However, inaccurate estimates may result in tests with atypical datasets [64].

Estimates of uncertainty from DL predictions are crucial in the MRS context, where measurement repeatability is hindered by compliance with clinical time frame restrictions. It is fundamental to access a CRLB-comparable measure. Related work has barely explored the tip of this research field [93].

The complex nature of the data

DL, especially regarding convolutional neural networks, was initially driven by computer vision tasks, which dealt with data structures in 2D frames (i.e., images) in either one or three color channels (red, green, and blue - RGB). In MRS, we are familiar with complex domain 1D data, either FIDs in the time domain or spectra in the frequency domain. Given their definition as

temporal signals, they are prone and suited to explore time-domain processing/conversion to 2D frames such as spectrograms or scalograms. On the one hand, if DL is considered a feature-agnostic tool, where features should not be provided or engineered, one would question why such processing is added to the intrinsic nature of the data. On the other hand, such processing steps shift, enhance or suppress inherent features in the data (in fact, they behave as data scaling and normalization, which are routine procedures to prepare a dataset for DL purposes). A clear example of clever feature extraction by spectrograms was the time-frequency separability of ghost artifacts, otherwise overlapping in spectral relevant areas, as shown by [32]. Interestingly, the current literature explores various approaches to deal with MRS data, its interpretation of channels as well as the exploration of either 1D frames [91, 95, 92, 76, 96, 93, 90, 65, 97, 94] or 2D frames [98, 32, 65]. Insights and visualization of possible preprocessing pipelines to arrange MRS data for DL purposes are introduced in Fig.5.11.

(A) 1D-representations
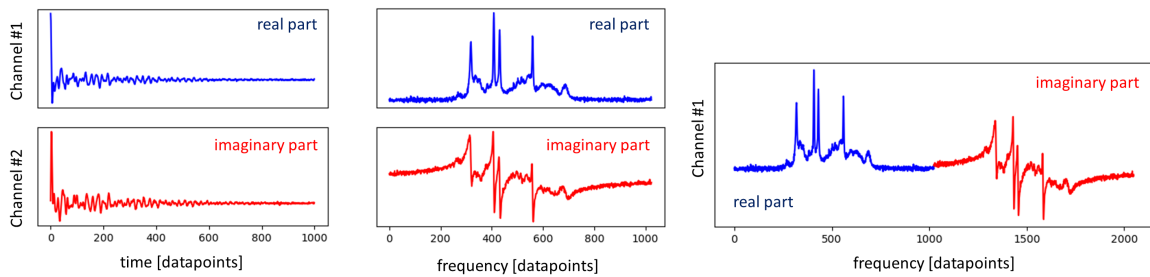
(B) 2D-representations

Spectrogram
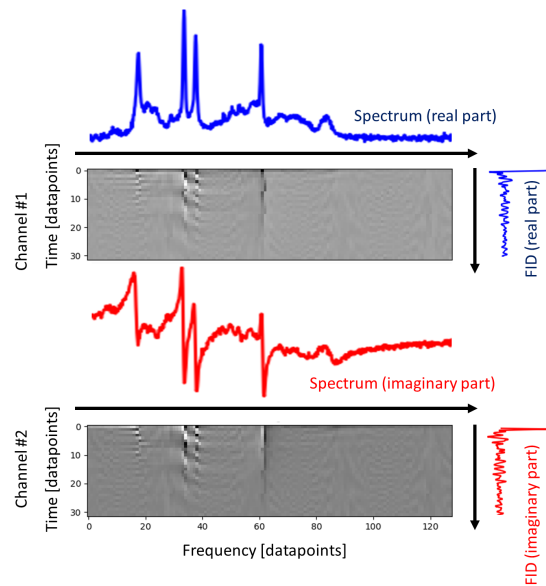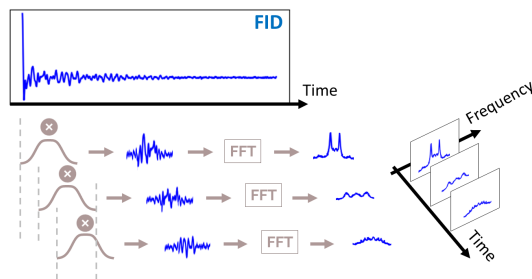
Scalogram
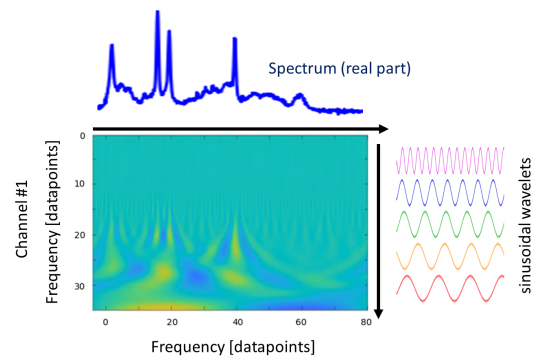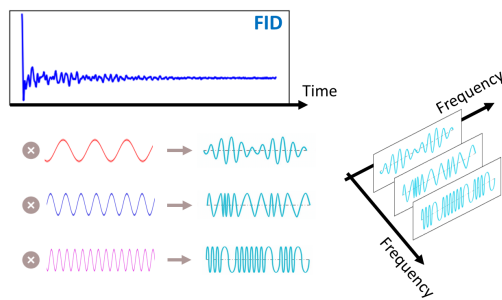
Figure 5.11: **A simulated pathological MRS signal represented for deep learning applications: (A) 1D frame and (B) 2D frame.** A 1D MRS signal can be represented in a 2-channel fashion (real and imaginary) in either time (A-left) or frequency (A-middle) domain or in a 1-channel fashion (A-right, a concatenated version of real and imaginary parts). 2D frames are shown as spectrograms (B-top) or scalograms (B-bottom). On the left side is reported the processing procedure, whereas on the right side is displayed either a 2-channel spectrogram (real and imaginary parts) or a 1-channel scalogram (only real part, inspired by [82]). Spectrograms model frequency variability over time and are calculated by Fourier transform (FFT) over a series of truncated sliding windows that chop the original FID in time-varying portions. Scalograms model the frequency content of the FID by wavelet modulation.

The philosophy of learning

The interpretability of MR spectra given the chemical structure of the metabolites of interest and the physics of the NMR experiment is shown in section 2.2. From there, any prior knowledge deployed in model fitting approaches naturally follows, and it sums up to (1) a series of model parameters and (2) the metabolite basis set, as discussed in section 2.6.2. Mimicking model fitting approaches, many different _philosophies_ of how to distill knowledge in DL algorithms arose over the past years.

A first distinction can be drawn by considering the network's output. The straightforward approach of DL for MRS quantification concerns **learning concentrations**. These approaches refer to supervised or unsupervised algorithms where the network's output consists of a set of numbers directly representing either the relative or absolute content of any metabolite in a given supplied input spectrum, Fig.5.12(A) [90, 96, 65, 97]. Contrarily, perhaps a more interpretable output is provided by **learning patterns**. These algorithms are typically metabolite-specific (i.e., a set of networks is developed where each network is tailored to quantify a specific target metabolite) and aim to learn the metabolite basis set pattern by means of its chemical fingerprint, Fig.5.12(B) [91, 92, 95, 65, 94, 97].
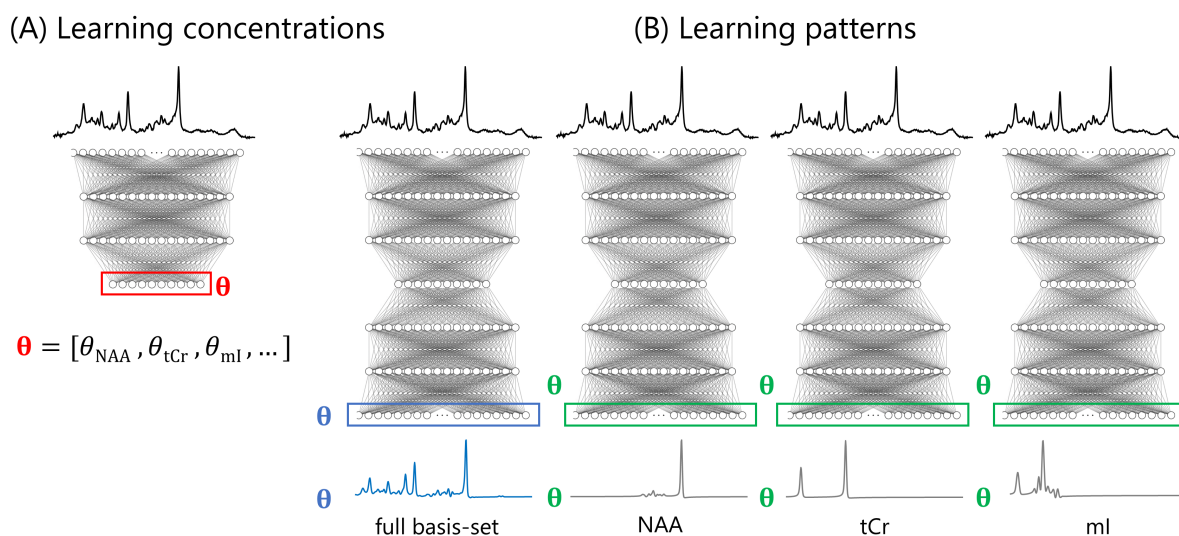
Figure 5.12: Learning strategies based on architecture's output. In all cases, a fully connected NN, either in the decoder (A) or encode-decoder (B) configuration, is assumed. The input data is assumed as a 1D channel real part only. (A) The output layer θ extracts the direct metabolite concentration θ_{met} for each node. (B) The output layer produces a filtered version of the input projected onto the identical domain (i.e., chemical shift). The filtering may aim to produce the full metabolic basis set (left, blue) or a selected target metabolite (right, green) where output integration yields the concentration.

A second distinction follows the concepts that drive the algorithms' design and is typically found independent on the networks' output. In contrast to traditional deep learning, where the

network's parameters are designed and optimized simply by referring to input and output relationships, **model-aware learning** enforces nodes or, more in general, portions of the network to learn specific parameters that typically straightly relate to the physics of the NMR experiment (e.g., phase shifts, frequency drifts, shim, etc.). These cases dive in-between machine learning and deep learning strategies, where feature tuning and engineering are embedded in deep architectures. Furthermore, feature engineering is usually found to help the overall interpretability of the deep model [92, 94]. Fig.5.13(A) reports an example where a latent space of an auto-encoder architecture is exploited to map specific spectral parameters.

Lastly, and most importantly, **risk-aware learning** converts existing models into risk-aware variants capable of efficiently identifying risk during training and deployment. It consists of unifying state-of-the-art algorithms for quantifying neural network risks ranging from underrepresentation bias, epistemic uncertainty, and aleatoric uncertainty to any learning philosophy or algorithm mentioned here [95, 93, 99]. An example is given in Fig.5.13(B), where output prediction is twinned to a set of scores that gives right away a grasp on the reliability of the results.
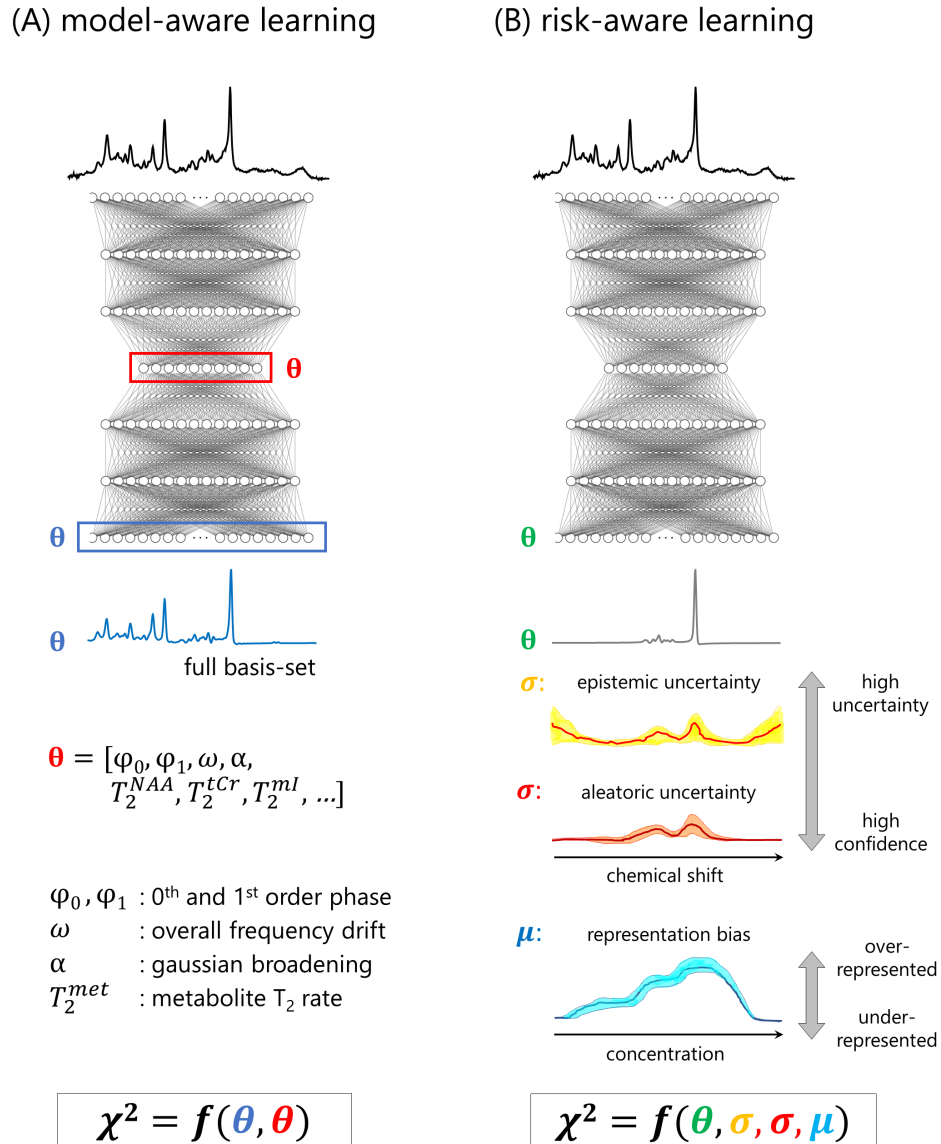
Figure 5.13: (A) Model-aware learning. A fully connected auto-encoder produces the full metabolite basis set from input provided as real domain spectra. The cost function (and its χ^2 minimization) includes specific modeling of the latent space (i.e., the bottleneck of the architecture) by which each node learns specific physical properties of the NMR experiment. (B) Risk-aware learning. A fully connected auto-encoder produces NAA basis set from real domain spectra. The cost function includes specific modeling of epistemic and aleatoric uncertainty. Moreover, it accounts for any representation bias given by a training set with favorable subsets of NAA concentrations (i.e., over-representation for average-healthy concentration levels and under-representation for pathological cases).

Results

6 | Published material

6.1. Quantification of MR spectra by DL: optimizations and pitfalls

References:

Rizzo R, Dziadosz M, Kyathanahally S P, Shamaei A, Kreis R.

Quantification of MR spectra by deep learning in an idealized setting: investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias.

Magn Reson Med. 2022;1-21. doi: 10.1002/mrm.29561

Rizzo R, Dziadosz M, Kyathanahally S P, Reyes M, Kreis R.

Reliability of quantification estimates in MR spectroscopy: CNNs vs. traditional model fitting.

Med Image Comput Comput Assist Interv - MICCAI 2022 Lect Notes Comput Sci. 2022;13438:715-724. doi: 10.1007/978-3-031-16452-1_68

Contribution to the studies

Software development to generate simulated spectra, simulation of datasets, training-testing of DL algorithms, statistical analysis, manuscripts drafting, part of the reviewing process.

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias

Rudy Rizzo[1,2,3,4] | **Martyna Dziadosz**[1,2,3,4] | **Sreenath P. Kyathanahally**[5] | **Amirmohammad Shamaei**[6,7] | **Roland Kreis**[1,2,4]

[1] MR Methodology, Department for Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland

[2] Department for Biomedical Research, University of Bern, Bern, Switzerland

[3] Translational Imaging Center (TIC), Swiss Institute for Translational and Entrepreneurial Medicine, Bern, Switzerland

[4] Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland

[5] Department of System Analysis, Integrated Assessment and Modelling, Data Science for Environmental Research Group, EAWAG, Dübendorf, Switzerland

[6] Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic, Brno, Czech Republic

[7] Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic

Correspondence
Roland Kreis, MR Methodology, Department for Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland.
Email: roland.kreis@insel.ch

Purpose: The aims of this work are (1) to explore deep learning (DL) architectures, spectroscopic input types, and learning designs toward optimal quantification in MR spectroscopy of simulated pathological spectra; and (2) to demonstrate accuracy and precision of DL predictions in view of inherent bias toward the training distribution.

Methods: Simulated 1D spectra and 2D spectrograms that mimic an extensive range of pathological in vivo conditions are used to train and test 24 different DL architectures. Active learning through altered training and testing data distributions is probed to optimize quantification performance. Ensembles of networks are explored to improve DL robustness and reduce the variance of estimates. A set of scores compares performances of DL predictions and traditional model fitting (MF).

Results: Ensembles of heterogeneous networks that combine 1D frequency-domain and 2D time-frequency domain spectrograms as input perform best. Dataset augmentation with active learning can improve performance, but gains are limited. MF is more accurate, although DL appears to be more precise at low SNR. However, this overall improved precision originates from a strong bias for cases with high uncertainty toward the dataset the network has been trained with, tending toward its average value.

Conclusion: MF mostly performs better compared to the faster DL approach. Potential intrinsic biases on training sets are dangerous in a clinical context that requires the algorithm to be unbiased to outliers (i.e., pathological data). Active learning and ensemble of networks are good strategies to improve prediction performances. However, data quality (sufficient SNR) has proven as a bottleneck for adequate unbiased performance—like in the case of MF.

KEYWORDS
active learning, bias, deep learning, ensemble of networks, model fitting, magnetic resonance spectroscopy, quantification

----

1 | INTRODUCTION

MR Spectroscopy (MRS) provides a noninvasive means for extracting biochemical profiles from in vivo tissues. Metabolites are encoded with different resonance frequency patterns, and their concentrations are directly proportional to the signal amplitude.[1,2] Metabolite quantification is traditionally based on model fitting (MF), where a parameterized model function is optimized to explain the data via a minimization algorithm. Metabolite parameters are usually estimated by a nonlinear least-squares fit (either in time or frequency domain) using a known basis set of the metabolite signals.[3] However, despite various proposed fitting methods,[3–7] robust, reliable, and accurate quantification of metabolite concentrations remains challenging.[8] The major problems influencing the quantitative outcome are: (1) overlapping spectral patterns of metabolites, (2) low SNR, and (3) unknown background signals and line shape (no exact prior knowledge). Therefore, the problem is ill-posed, and current methods address it with different regularizations and constraint strategies (e.g., parameter bounds, penalizations, choice of the algorithm), with discrepancies in the results from one method to another.[9]

Supervised deep learning (DL) utilizes neural networks to discover essential features embedded in large data sets and to determine complex nonlinear mappings between inputs and outputs.[10] Thus, DL does not require any prior knowledge or traditional assumptions. Given the success of the method in different areas,[10–14] DL has been introduced into MRS as an alternative to conventional methods.[15–22] Quantification of MRS datasets has been explored as follows: (1) DL algorithms identify datasets' features and either help reduce the parameter space dimension or set reliable starting conditions for the fit (i.e., combining knowledge on the physics with DL). It showed rapid spectral fitting of a whole-brain MRSI datasets.[23] (2) Convolutional neural networks (CNNs) have been deployed to investigate combinations of spectral input of edited human brain MRS, which showed improved accuracy of straight metabolite quantitation when compared to traditional MF techniques.[24] (3) Regression CNNs have been used to mine the real part of rat brain spectra to predict highly resolved metabolite basis set spectra with intensities proportional to the concentrations of the contributions,[17] with results comparable to traditional MF approaches and showing readiness for (pre)clinical applications.[22] (4) Targeting localized correlated spectroscopy (L-COSY) datasets, DL algorithms have reported faster data reconstruction and quantification compared to alternative acceleration techniques.[16]

Nevertheless, despite the reported equivalence in quantitation performance compared to traditional MF,[14,17,22,23] questions arise concerning the robustness of

DL algorithms. A robust use within a clinical MRS context requires the algorithm to be unbiased also for pathological spectra. In imaging, DL has shown excellent performance for classification or segmentation tasks but may suffer from inherent weaknesses in subsets of representative outlier samples.[11,25] DL architectures for MRS quantitation have mostly been investigated for sample distributions of near-healthy spectral metabolite content. Hence, it can be suspected that high accuracy and precision are mainly found when DL is deployed for new entries of similar near-normal types. However, inaccurate estimates may result for tests with atypical datasets.[26] Here, strongly variable metabolite concentrations that vary uniformly and independently over the entire plausible parameter space are used in the training set. This mimics the full range from healthy to strongly pathological spectra, that is, the full complexity of a clinical setup.

MRS signals are acquired in time domain but viewed in frequency domain. Traditional MF works in either of the two equivalent domains, and fit packages may allow the user to switch from one to the other for fitting and viewing. However, DL architectures for MRS quantification have mainly explored the frequency domain, mostly motivated by the reduced overlap between the constituting metabolite signals. Spectrograms[18] present an extension into a simultaneous time/frequency domain representation and offer a 2D signal support that matches the input format for the original usage of CNN algorithms in computer vision. This work introduces a dedicated high-resolution spectrogram calculation focusing on signal-rich areas in both domains to be used as input for different CNN architectures. They are compared to other inputs and networks, inspired by previous MRS publications. Specifically, 24 network designs are investigated with differing input–output dataset types with a combined focus on depth (i.e., number of layers) and width (i.e., number of nodes/kernels) of the networks. This focus was motivated by the fact that the exploitation of spectrograms in deep learning has shown top-notch performance for speech and audio processing when deploying architectures with few layers and large convolutional kernels.[27–29] Moreover, wide and shallow networks are more suitable to detect simple and small but fine-grained features. In addition, they are easier and faster to train.[30] Network linearity (i.e., activation function) and locality (i.e., kernel size) are also investigated.

Besides investigating multiple architectures and input formats, two established main strategies for improving the outcome of predictions are also explored: *active learning*[31] (data augmentation for critical types of spectra) and *ensemble learning*[32,33] (combination of outputs from multiple architectures).

Active learning can improve labeling efficiency,[31,34,35] where the learning algorithm can interactively select a

subset of examples that needs to be labeled. This is an iterative process where (1) the algorithm selects a subset of examples; (2) the subset is provided with labels; and (3) the learning method is updated with the new data.[36] *Uncertainty sampling*[37] is a specific strategy used in active learning that prioritizes selecting examples whose predictions are more uncertain (i.e., targeted data augmentation). Because these cases are usually close to the class separation boundaries, they contain most of the information needed to separate different classes.[38,39] In different applications, uncertainty sampling has been shown to improve the effectiveness of the labeling procedure significantly.[34,35,37,40]

DL algorithms are sensitive to the specifics of the training.[41] Hence, they usually find a different set of weights each time they are trained, producing different predictions.[10] A successful approach for reducing the variance is to train multiple networks instead of one and combine their predictions.[41] This is called *ensemble learning*, where the model generalization is maintained, but predictions improve compared to any of the single models.[33] From a range of different techniques,[42–44] here, *stacking of models* is implemented.[32]

To evaluate pros and cons of all these approaches, in silico ground truth (GT) knowledge is used (and hence no in vivo data was included in this evaluation) to assess performances via a dedicated set of metrics based on bias and SD. The CNN-predicted distributions of concentration are then compared to those from traditional MF. Furthermore, to emphasize the analysis at the core of the quantification task, the focus is placed on an idealized simulated setting with typical single-voxel spectra that have been preprocessed to eliminate phase as well as frequency drifts.[3] This assumption aims at (1) freeing the MF algorithm from problems with local χ^2 minima and (2) designing DL models optimized for the quantification task only.

2 | METHODS

2.1 | Simulations

This work is based on in silico simulations. A dataset of 22,500 entries was randomly split into 18,000 for training, 2000 for validation, and 2500 for testing. Larger dataset sizes are also explored, see section 2.4.

2.1.1 | MR spectra

Brain spectra were simulated using actual RF pulse shapes for 16 metabolites at 3 T using *Vespa*[45] for a semi-LASER[46]

protocol with TE = 35 ms, a sampling frequency of 4 kHz, and 4096 datapoints.

Further specifics of the simulations include: (1) Voigt line shapes, (2) metabolite concentration range, (3) addition of macromolecular background signal (MMBG), (4) noise generation, and (5) spectrum or spectrogram calculation.[47] Metabolite concentrations vary independently and uniformly between 0 and twice a normal reference concentration for healthy human brain.[1,48–50] Maximal concentrations in mM units—NAA 25.8, tCr (1:1 sum of creatine + phosphocreatine spectra): 18.5, mI (myo-inositol): 14.7, Glu (glutamate): 20, Glc (glucose): 2, NAAG (N-acetylaspartylglutamate): 2.8, Gln (glutamine): 5.8, GSH (glutathione): 2, sI (syllo-inositol): 0.6, Gly (glycine): 2, Asp (aspartate): 3.5, PE (phosphoethanolamine): 3.3, Tau (taurine): 2, Lac (lactate): 1, and GABA (γ-aminobutyric acid): 1.8. The concentration for tCho (1:1 sum of glycerophosphorylcholine + phosphorylcholine spectra) ranges from 0 to 5 mM to mimic tumor conditions.[51] A constant downscaled water reference (64.5 mM) is added at 0.5 ppm to ease quantitation. Metabolite T_2s in ms (and hence Lorentzian broadening) are fixed to reference values from literature—tCr (CH_2): 111, tCr (CH_3): 169, NAA (CH_3): 289, and all other protons: 185.[49,52,53,54] MMBG content, shim, and SNR mimicked in vivo acquisitions and varied independently and uniformly (time-domain water referenced SNR 5–40, Gaussian shim 2–5 Hz, MMBG amplitude ±33%). The MMBG pattern was simulated as a sum of overlapping Voigt lines as reported in Refs. 49 and 55 (Figure 1A).

2.1.2 | Spectrograms

A spectrogram is a complex 2D representation of a spectrum, where frequencies vary with time: Every image column represents the frequency content of a particular time portion of the FID. Time information is binned along every row of the image. It is calculated via application of a short-time Fourier transform,[18] where, depending on the size of the Fourier analysis window, different levels of frequency and time resolution can be achieved. A long window size modulated via zero-filling combined with a small overlap interval is chosen to increase frequency resolution and minimize the expense of time resolution (Figure 1B). Diagonal downsampling is designed to reduce the spectrogram size, keeping the original resolution grid at least as part of the time-frequency information on consecutive bins and reducing the spectrogram size (Figure 1C) to allow reasonable computation time for a CNN architecture (i.e., 128 frequency bins × 32 time bins) (Figure 1D).
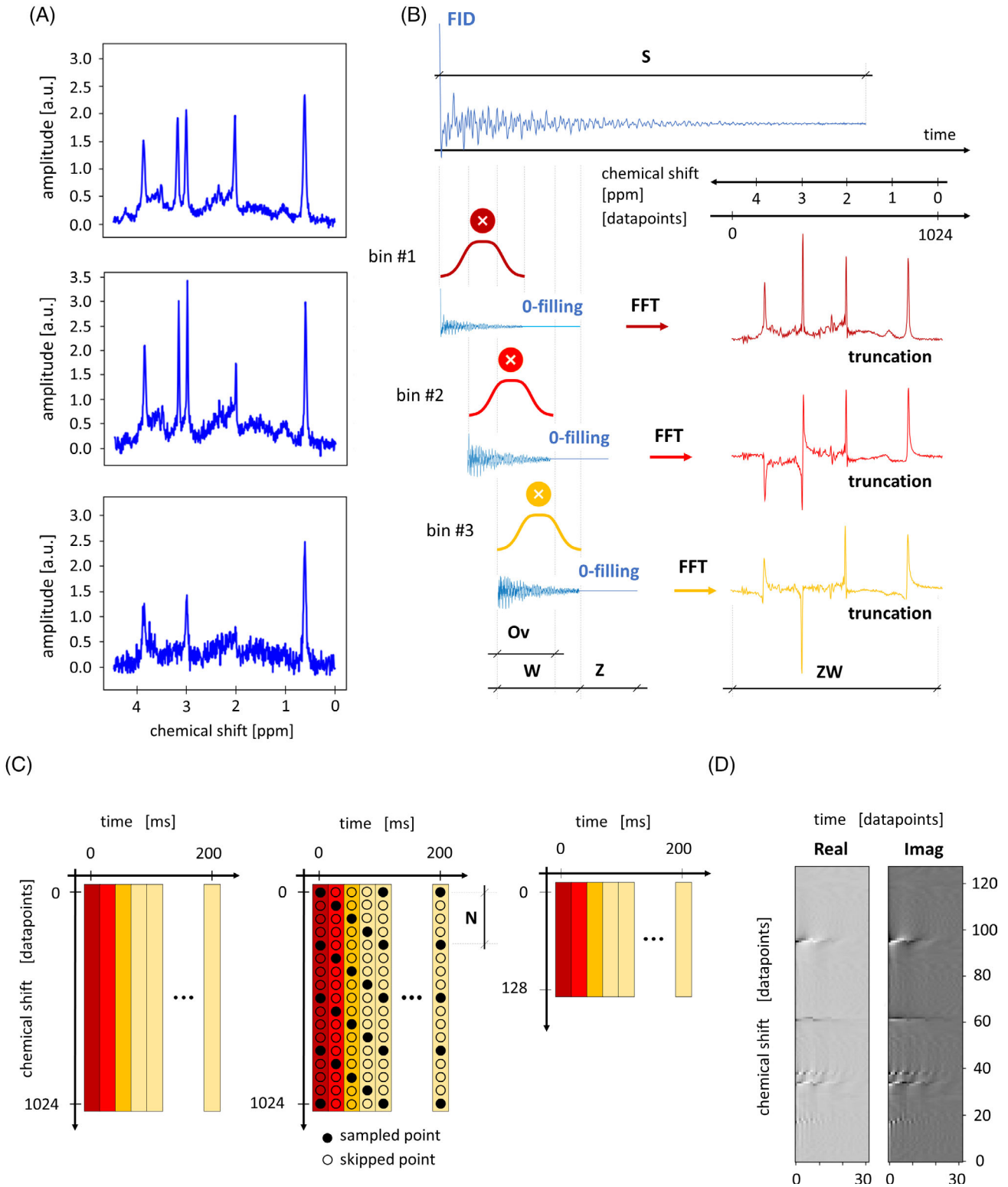
FIGURE 1 Illustration of input formats. (A) Samples of spectra, real part, view of the central 1024 points. (B) Spectrogram computation via short-time Fourier transform. Specifically, in datapoints units (corresponding to time and frequency resolution of 0.25 ms and 1 Hz, respectively): S = 4096, Z = 6000, W = 1024, Ov = 1000, ZW = 1024. Zero-filling is tuned to select the relevant part of the spectrum with W = 1024 datapoints. (C) (Left) Arrangement on a 2D frame of short-time Fourier transforms over time bins. Color code reference to windows in part (B). A truncation at 32 bins (200 ms) in time domain is used to limit the matrix space, given an almost complete T_2^* relaxation of the FID at that point. (C) (Middle) Diagonal undersampling reduces the vertical (frequency domain) matrix size. Size reduction is about a factor $N = 8$. (C) (Right) Undersampled spectrogram: 128 × 32 datapoints. (D) Example of constructed spectrogram matrix. FFT, fast Fourier transform; S, support of the signal; Ov, window overlap; W, Hamming window size; Z, zero filling; ZW, truncated support of zero-filled FFT.

2.2 | Design and training of CNN architectures

A total of 24 different CNN architectures combined with different spectroscopic input representations are compared for MRS metabolite quantification. Current state-of-the art networks have been taken as reference models and adapted to the purpose and datasets used.

Scripts were written in Python[56] using Keras library[57] on a Tensorflow[58] backend. Code ran on either of three graphic-processing units (GPUs; NVIDIA [Santa Clara, USA] Titan Xp, Titan RTX, or GeForce RTX 2080 Ti) or Google [Mountain View, USA] Colaboratory.[59] Samples of the design are reported in Figure 2. Overall network designs are given in Table S1; Figures S1, S2, S3, S4, S5; and Text S1.

2.2.1 | Architectures for straight numeric quantification of concentrations

A total of 22 architectures were fed with 1D (spectra) or 2D (spectrograms) input and mapped as output a vector of 17 normalized concentrations (i.e., in [0–1] interval) of 16 metabolites and the water reference, as listed in Table S1. Networks fed with 1D input exploit one channel with truncated spectra of 1024 datapoints from −0.5 to +6 ppm with concatenated real and imaginary parts (i.e., $2048 \times 1 \times 1$ datapoints, Figure 2A). Networks fed with 2D input can either be configured in two channels (real and imaginary components of the spectrogram, $32 \times 128 \times 2$ datapoints) or one channel (real and imaginary components concatenated, $64 \times 128 \times 1$ datapoints, Figure 2B).

Five networks receive 1D input: two deep convolutional neural networks (*DeepNet*),[60] two residual networks (*ResNet*)[61] and one inception network (*InceptionNet*).[62–64]

This work investigates deep and shallow architectures either exploiting large or small convolutional kernel sizes. A total of 10 networks receive two-channel spectrograms as input. Given the limited size of the input FOV, the architecture is limited to be shallow (i.e., pooling operations to downsampling features directly following a convolutional layer are limited). However, a deeper architecture with multiple convolutional operations with sparse pooling is also compared. A further comparison is performed regarding the optimal activation function, comparing batch normalization + rectified linear unit (ReLU) versus exponential linear unit (ELU).[65,66] Seven networks receive one-channel spectrograms as input. With this configuration, deeper architectures are explored: two *DeepNets*, four *ResNets*, and one *InceptionNet*.

Architectures are analyzed either in a preconfigured parameter state or in a parameter space that had been optimized via Bayesian hyperparameterization.[67] The optimization procedure is given in Text S1. In addition, to limit biases around zero for small concentrations,[68] all network designs are characterized by a final layer with linear activation, allowing the prediction of negative concentrations.

2.2.2 | Architectures for estimation of metabolite base spectra

1D input (real part only, 0–4.7 ppm, $1406 \times 1 \times 1$ datapoints after zero-filling of original FID) was used to input and output to/from the CNNs. U-Net architectures[69] analogous to those of Ref. 22 are implemented here to map the ideal high-resolved noiseless base spectrum of a target metabolite as output. CNNs are trained one by one for each metabolite such that each CNN filters out signals only from the designated target metabolite. A base U-Net design (Figure 2C) is optimized for individual metabolites as follows:

1. *UNet-1DR-hp* : A total of 17 different networks with the same base architecture but adapted weights for each metabolite;
2. *Unet-1DR-hp-met*: A total of 17 different networks with adapted Bayesian-optimized architecture and weights for each metabolite.

Configurations are reported in Figure S5. First, metabolite concentrations are evaluated by feeding an input spectrum to the 17 metabolite-specific CNNs. Integration of the predicted metabolite base spectrum is then referenced to the integrated water reference to produce concentrations for a fully automated quantification pipeline.[22]

2.2.3 | Training

Training and validation sets were randomly assigned for training the CNN on a maximum of 200 epochs with batch normalization of 50. The adaptive moment estimation algorithm (ADAM)[70] was used with dedicated starting learning rates for each network.[71,72] The loss function was the mean-squared error (MSE). Visualization of training and validation loss over epochs combined with implementing an early-stopping criterion monitoring minimization of validation loss with patience = 10 has been used for tuning the network parameter space.[57] Training time and test loss function are listed in Table S1.
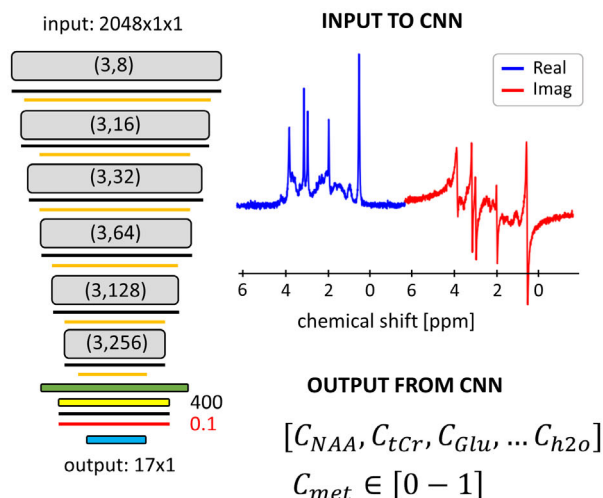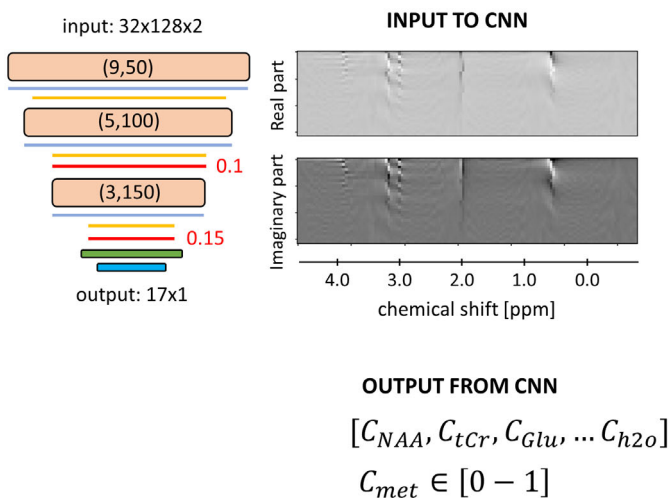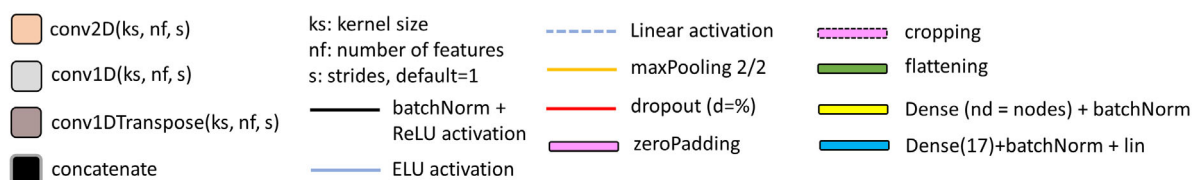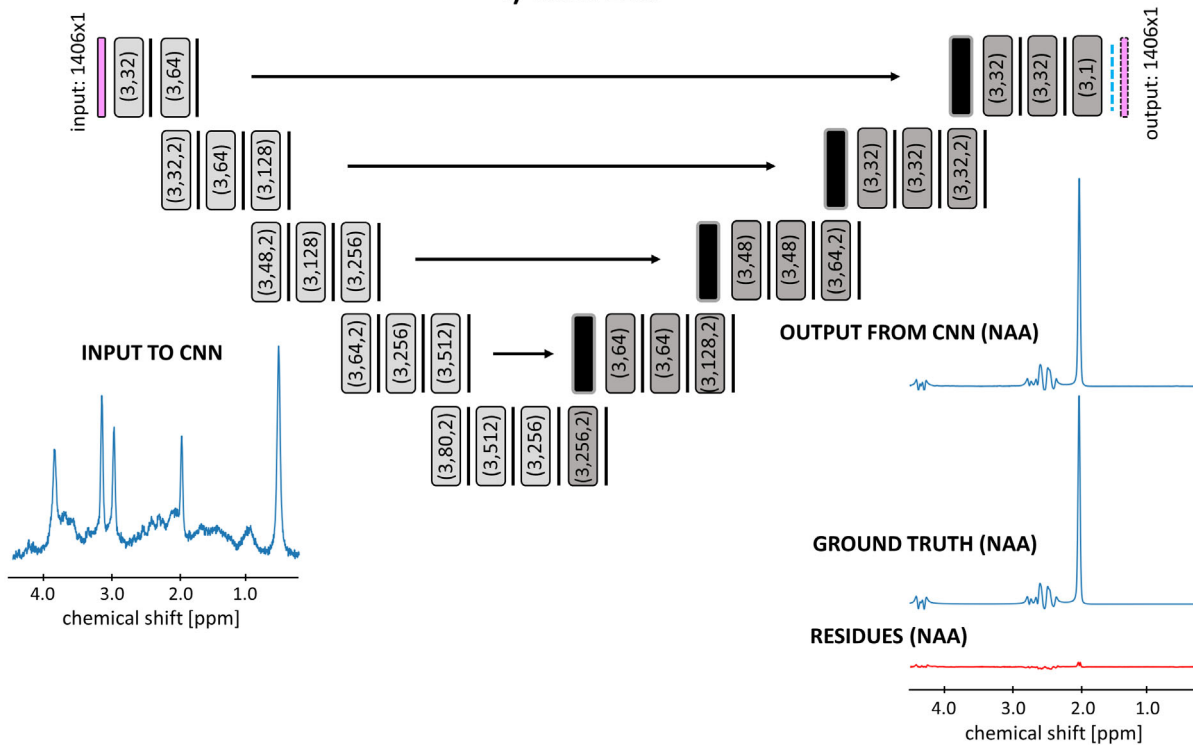
FIGURE 2 Examples of three CNN structures and schematic input–output relationships. (A) and (B) depict architectures for straight quantification, with metabolites relative concentrations as output. (C) depicts a U-Net architecture similar to what was proposed in Ref. 22 for NAA basis set prediction. Input details: (A) *Deep neural network* with 1D-spectral input from concatenated real and imaginary parts (-1D). (B) *Shallow neural network* with 2D-spectral input from two-channel spectrograms (-2D2c). (C) U-Net architecture fed with only the real part of a spectroscopic input (-1DR). CNN, convolutional neural network.

2.2.4 | Evaluation

Regression plots mapping GT concentrations versus CNN predicted concentrations from the whole test set are taken as indicators of the network's prediction performance. Four scores are defined:

- a (slope of the regression line): must be close to 1 for ideal mapping of concentrations over the whole range of simulated metabolite content;
- q (intercept of the regression line, mM): must be close to 0 to minimize prediction offsets/biases;
- R^2 (coefficient of determination): must be close to 1 to assess full model explanation of the variability of the data;
- σ (RMS error [RMSE] of prediction vs. GT, mM): as low as possible. However, expected to be comparable to Cramer Rao Lower Bounds (CRLBs) from MF.[73]

To easily compare different networks and input setups quantitatively in the Results section, these scores or combinations thereof have been used. The combinations are referred to as *concise scores*: $a \cdot R^2$ as measure of linearity, σ to compare with CRLBs. q was excluded because it is mostly negligible.

2.3 | Influence of inclusion of water reference peak

For the evaluation of the potential benefit of including a water reference peak, two slightly different *ShallowNet-2D2c-hp* networks are compared. *Network A* outputs 17 neurons (16 metabolites and water), whereas *network B* outputs 16 neurons only (no water output). Two adapted datasets are used for the investigation, one with (*dataset A*), and one without (*dataset B*) downscaled water reference at 0.5 ppm. Metabolite concentrations are calculated for both cases (assuming known water content in case A). Networks have been independently trained

five times to monitor network variability over multiple trainings.

2.4 | Active learning and dataset size

In this part, data augmentation techniques to smartly generate training sets are investigated. Subsets with 5000 new entries of the dataset where predictions scored worst are defined: specific subsets of spectrally weakly represented metabolites in either very low or very high concentrations and spectra with low SNR. New weighted datasets of 25,000 entries (20,000 training – 5000 validation set) or 40,000 entries (35,000 training – 5000 validation set) are generated (example in Figure 3, full description in Figure S6). Datasets with matching size and the testing set are kept unchanged from the previous simulation, thus with uniformly distributed concentrations and SNR. *ShallowNet-2D2c-hp* is selected as architecture and trained 10 times with a given augmented training set to minimize training variance.

Complementarily, given the network trained on a uniform span of concentrations, active learning is investigated in the testing phase on three different test sets where concentrations are clipped to a progressively smaller range of 20%–80%, 20%–80% with SNR >20, and 40%–60% concentration range relative to the training set.

2.5 | Ensemble of networks

In this section, ensembles of networks are implemented via *stacking of models*.[32] This consists of designing a DL architecture called *stacking model* (a multilayer perceptron (MLP) with two hidden layers is selected for this case) that will take as input the combination of a given number of independently pretrained models. The stacking model aims at weighting predictions from single models. It is trained using the same training and validation sets

FIGURE 3 Examples of dataset augmentation techniques representing sample distributions for two metabolites (NAA and GABA). (A) Dataset size increment with uniformly distributed concentrations. (B) Active learning weighted on higher occurrences of low and high concentrations for all metabolites. GABA, γ-aminobutyric acid.
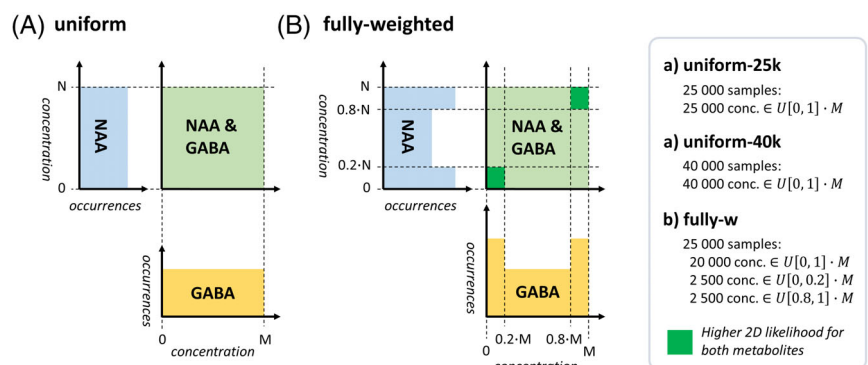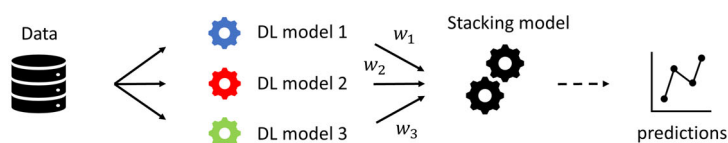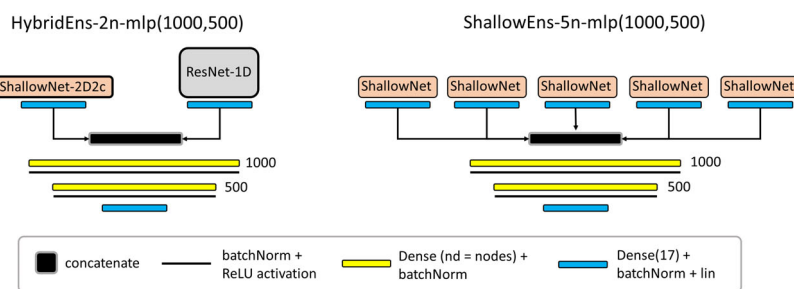
FIGURE 4 Illustration of ensemble learning. (A) Stacking model concept. (B) Examples of considered models: the stacking model consists of the two-layer MLPs (i.e., first layer with 1000 neurons, second layer with 500 neurons). *HybridEns*: an ensemble of two different networks (-2*n*). In this study, *ShallowNet-2D2c* and *ResNet-1D* are combined with two or 10 networks. *ShallowEns*: an ensemble of five different networks (-5*n*) of the same type, specifically *ShallowNet-2D2c*. HybridEns, hybrid ensemble; MLP, multi-layer perceptron; ResNet, residual network; ShallowNet, shallow network.

used to train single models while keeping the weights of the pretrained input models fixed. Three different ensembles are investigated: *ShallowEns-5n* groups five identical *ShallowNet-2D2c-hp* architectures, whereas *HybridEns* tests heterogeneous inputs grouping either two or 10 different networks (*ShallowNet-2D2c-hp* and *ResNet-1D-hp*) (Figure 4).

2.6 | Model fitting

Spectra are fitted using FiTAID[7] given its top performance in the ISMRM fitting challenge[9] and to be expected for the spectra as used in the current setup (in particular, without undefined spurious baseline). The model consists of a linear combination of the metabolite base spectra with Voigt lineshape, where the Lorentzian component was kept fixed at the known GT value. The areas of the metabolites are restricted in a range corresponding to $[-0.5 + 2.5\,\mu]$, where μ is the average concentration in the testing set distribution (i.e., the normal tissue content). These bounds mimic the effective boundaries of the DL algorithms. CRLBs are used as a precision measure[74] and are considered for three subgroups of the testing set (high [SNR > 28.4], medium [16.7 < SNR < 28.4], and low [SNR < 16.7] relative SNR, respectively).

3 | RESULTS

3.1 | S1Metabolite quantification referenced to the downscaled water peak

As illustrated for three different networks, Figure 5 shows that CNN predictions perform better if the spectra are

referenced to a downscaled water peak: Regression slope a and R^2 are closer to 1; σ is appreciably lower. Moreover, the spread of the scores is on average reduced, displaying improved stability over multiple trainings. Extended results are presented in Figures S7 and S8.

3.2 | Network design

Figure 6 reports CNN predictions versus GT values of a *ResNet-1D-hp* architecture for nine metabolites (see Figures S9 and S10 for extended results on 16 metabolites or different CNN architecture). Distributions of GT and predicted values are displayed for the test set (as for all results). Predictions relate very well to the GT for well-represented metabolites (top row). However, for metabolites with lower relative SNR, predicted distributions of concentrations tend to be less uniform and are biased toward average values of the GT distributions. Thus, concentrations at distribution boundaries are systematically mispredicted, particularly for low SNR. This is reflected in lower a and R^2 values and higher σ. Figures S11 and S12 include a comparison of multiple networks via bar graphs (which are ill-suited to express the systematic bias) and a plot of distributions of predictions.

The performance of all networks and fitting models for nine metabolites is reported in Figure 7 via a 2D plot of the *concise scores* $a \cdot R^2$ and σ (see Figure S13 for extended results on 16 metabolites). Top performance corresponds to the top-left corner where $a \cdot R^2$ approaches 1 and σ is low. Metabolites can roughly be divided into three groups:

1. *Well-represented* metabolites: NAA, tCho, tCr, mI, Glu with averaged DL scores $a \cdot R^2 > 0.80$ and $\sigma < 15\%$, as well as MF scores $a \cdot R^2 > 0.95$ and $\sigma < 10\%$;

FIGURE 5 Boxplot statistics of the prediction scores for four metabolites showing the effect of water referencing. Results reported for *ShallowNET-2D2c-hp*, *ResNet-1D-hp*, and *U-Net-hp* trained and tested on datasets with (red, black, or blue) and without (yellow, gray, or green) water reference (mean values plotted in orange). On average, water referencing yields better performance with higher coefficients a and R^2 as well as lower offset q and lower RMSE σ. RMSE, RMS error.
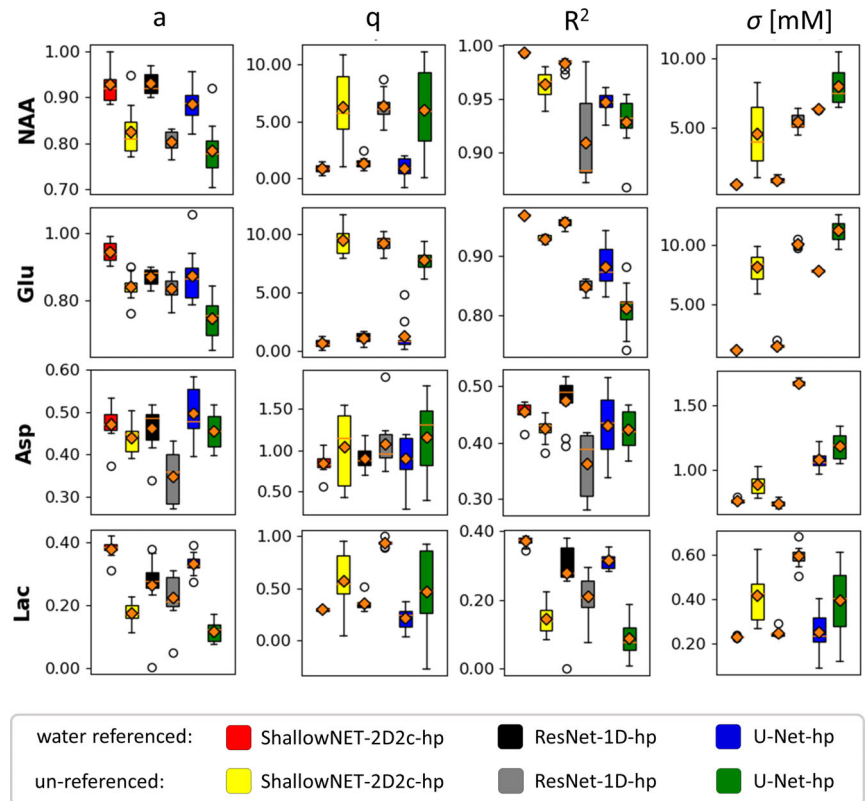
2. *Medium-represented* metabolites: Glc, NAAG, Gln, GSH with averaged DL scores $0.50 < a \cdot R^2 < 0.75$ and $20\% < \sigma < 35\%$, as well as MF scores $0.75 < a \cdot R^2 < 0.90$ and $15\% < \sigma < 35\%$;
3. *Weakly represented* metabolites: sI, Gly, Asp, PE, Tau, Lac, GABA with averaged DL scores $a \cdot R^2 < 0.40$ and average $\sigma > 35\%$, as well as MF scores $a \cdot R^2 < 0.65$ and $\sigma > 35\%$.

Overall, multiple DL networks perform similarly, but some general differences are noteworthy. Optimized spectrogram representation via two channels combined with a shallow architecture (i.e., dark blue squares) is found to be well suited for MRS quantification, showing mostly better performances than alternative deeper designs (i.e., light blue, pink, and gray squares), with one-channel designs (diamonds) or 1D spectra as signal representation (circles). Benefits are evident for medium and weakly represented metabolites. Performances of direct quantification and two-step quantification via base spectrum prediction followed by integration (stars) are similar. MF is found superior to DL for all medium- and weakly represented metabolites with significant average improvements for $a \cdot R^2$. However, σ tends to be higher for many cases. A more detailed presentation of performance is given in Figures S14 and Text S2.

Figure 8 displays plots of prediction errors (i.e., $\Delta = prediction - GT$) and their spread σ as a function of SNR and shim for tCho, NAAG, and sI. Prediction uncertainties increase with noise level approximately linearly with 1/SNR and reach a plateau for weakly represented metabolites when the spread represents essentially the whole training range. No dependence on shim is apparent for the investigated range.

3.3 | Dataset size, active learning, and ensembles of networks

Figure 9 reports on performance improvements by active learning in training phase and dataset sizing (part 9A) as well as by using an ensemble of networks (part 9B) for four metabolites as reflected by *concise scores*. Outcomes of emulated active learning approaches in limiting the testing sets are illustrated through regression plots for Gln in Figure 9C. Detailed comparisons for 16 metabolites are given in Figure S15, Table S2, Table S3, Figure S16, and Table S4.

3.3.1 | Dataset size

The performance showed moderate improvements for most metabolites when dataset size was increased from 25,000 to 40,000 samples (Figure S9).
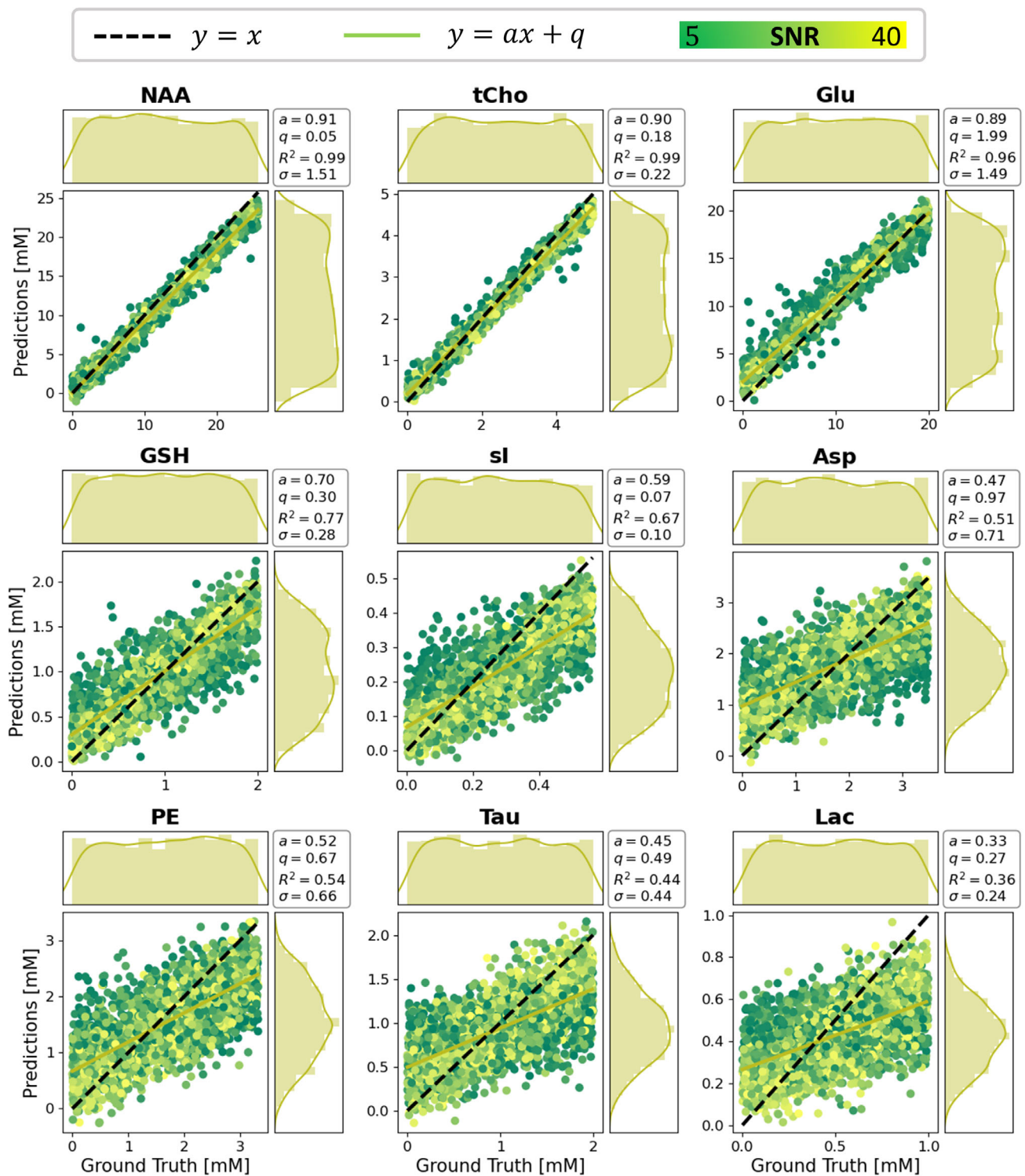
FIGURE 6 Maps and marginal distributions of predictions versus GT for a *ResNet_1D_hp* network. Results for nine metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE (σ) is reported as an overall measure of variability. A regression model ($y = ax + q$) is also provided to judge prediction quality. R^2 measures how well a linear model explains the overall data. Mispredictions can be monitored either by a decrease in a and R^2 or by visual biases in distributions of predictions (bell shape). The prediction bias toward the mean value of the training distribution is evident for medium- to weakly represented metabolites (e.g., sI, Asp, PE, Tau, Lac). On average, metabolites with lower SNR yield higher errors (q and σ in mM units). Further metabolite results are shown in Figure S11 and results for *ShallowNet-2D2c-hp* in Figure S15. GT, ground truth; Asp, aspartate; Lac, lactate; PE, phosphoethanolamine; sI, syllo-inositol; Tau, taurine.
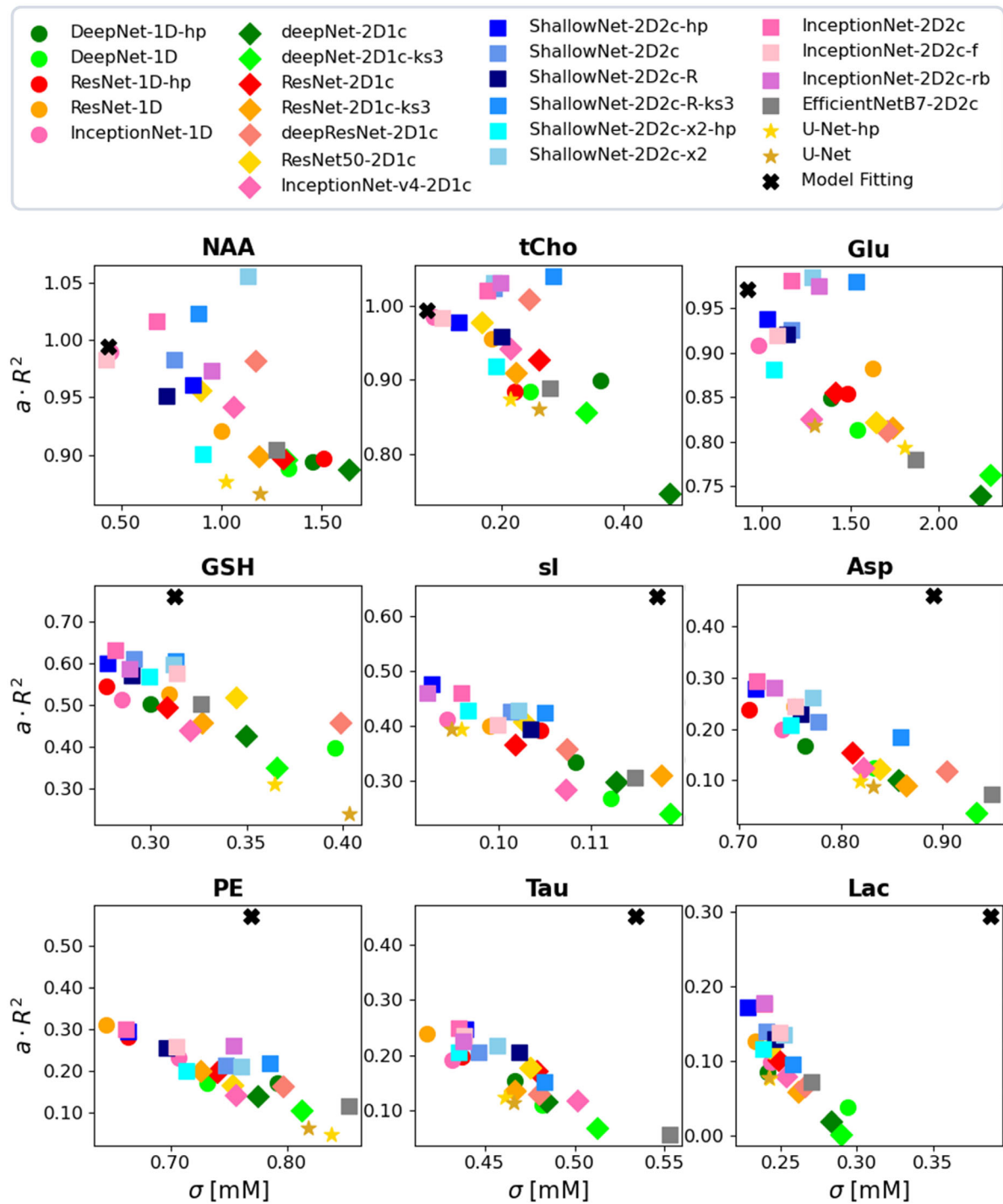
FIGURE 7 *Concise scores* presented to compare quantification quality for different networks and input setups (all with water reference). Network identification is chosen as follows: *NetworkType-InputType-properties*. 1c, 1 channel; 1D, spectra; 2D, spectrograms; f, factorized convolution; hp, Bayesian hyperparameterized architecture; ks3, convolutional kernel size = 3; R, exploiting ReLU activations; rb, downsampling via reduction blocks; x2, double convolution before MaxPooling.

3.3.2 | Active learning

Dataset augmentation to favor training with combinations of low or high concentrations of weakly represented metabolites (see Figure S6B–S6D) does not substantially improve performance (Figure 9A, Figure S15, Table S2). Mild improvements (<6% for a, q, R^2 and σ) are seen for GABA and sI, respectively, when exploiting metabolite-specifically augmented datasets (*GABA-w*, *sI-w*). Increased dataset size combined with data augmentation to favor high and low concentrations of different metabolites (*GSPT-w*) moderately improves performances
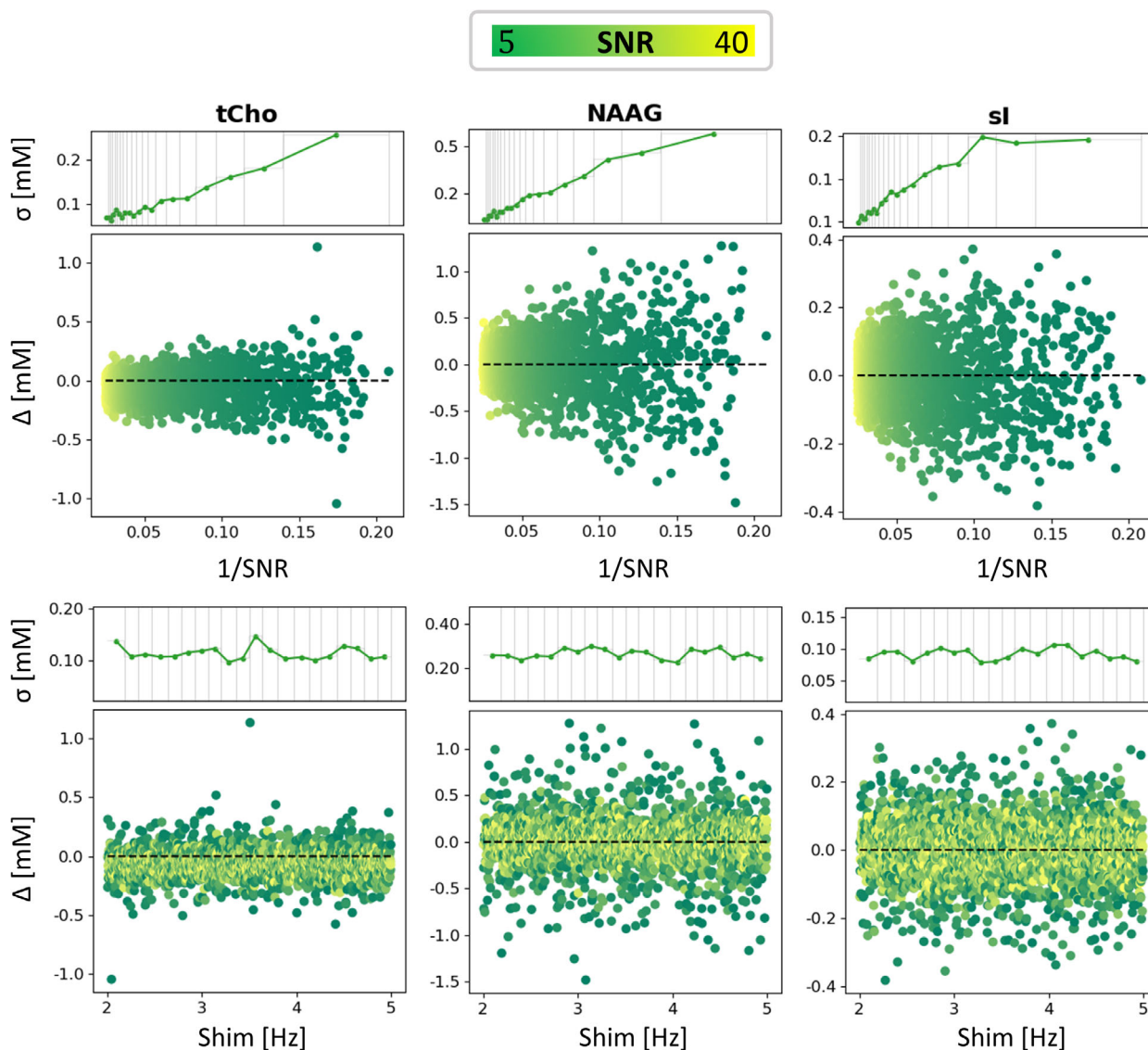
FIGURE 8 Illustration of the SNR and shim dependence of prediction quality. The CNN's prediction error Δ (*prediction – GT*) and the RMSE (σ) are plotted as a function of SNR (top row) and shim (bottom row) for four metabolites. Results reported for network type *ShallowNet-2D2c-hp* with water reference. RMSE is averaged over bins with an equal number of samples. Bins' width increases for low SNR values. Errors scale approximately linearly with *1/SNR* and are insensitive to different shim setups.

for the augmented metabolites (GABA, sI, PE, Tau). It also extends mild improvements to medium- to weakly represented metabolites that have not undergone data augmentation (e.g., Lac, Gly, Gln). A dataset that is strongly weighted toward extreme combinations of low or high concentration for all metabolites (*fully-w*) or a dataset weighted toward low SNR (*SNR-w*) deteriorated performances.

Clipping the test set to 20%–80% or 40%–60% of the concentration range in training renders improved performances (on average $a + 4.5\%$, $q - 10.2\%$, $\sigma -23.9\%$ and $a + 4\%$, $q - 37.5\%$, $\sigma -36.2\%$, respectively), which is even enhanced further when the testing set includes samples with higher SNR (on average $a + 15.4\%$, $q - 45.4\%$,

$\sigma -36.2\%$). Given the limited range on the y-axis, R^2 is less representative (Figure 9C, Table S3).

3.3.3 | Ensemble of networks

Ensembles of Bayesian-optimized networks show consistent and relevant $a \cdot R^2$ improvements for medium- to weakly represented metabolites without deteriorating performance for well-defined metabolites. A hybrid ensemble outperforms the ensemble of networks of the same type. The performance of the ensemble increases with the number of combined networks (Figures 9B, S16) (Table S4).
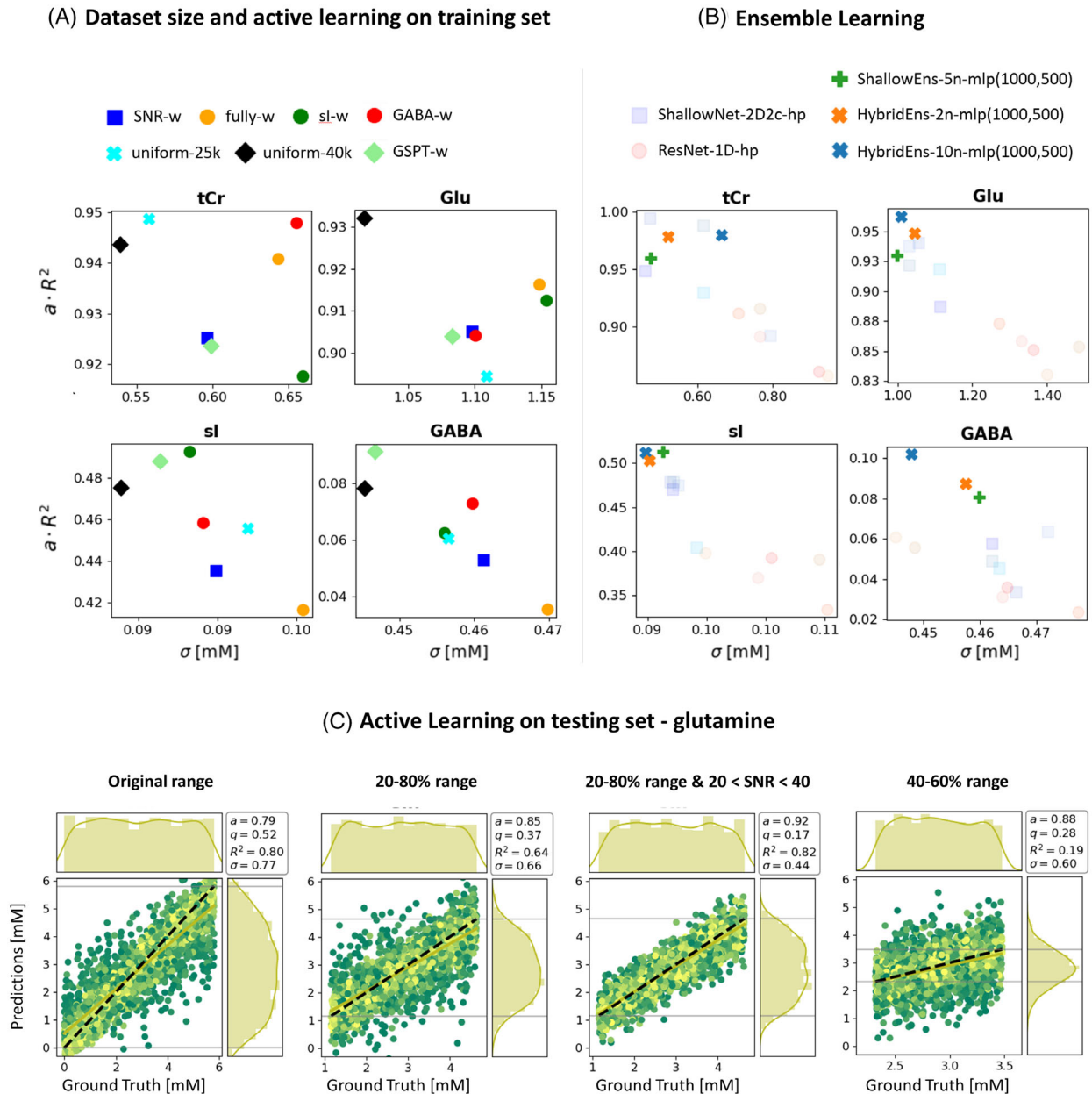
FIGURE 9 Outcome comparison for the influence of dataset size, active learning approaches, and ensemble of networks (all with water reference). *Concise scores* evaluated on the same testing set for tCr, Glu, sI, and GABA in different setups: (A) Dataset size and active learning on the training set (for abbreviations alluding to types of active learning extensions, see Results 3.3.2). (B) Ensemble of networks (for naming, see Figures 3 and 4). Ensemble models improve predictions for weakly to medium-represented metabolites without worsening the already good single-network performances for well-represented metabolites (higher $a \cdot R^2$ and lower σ). (C) Active learning on the testing set monitored via maps and marginal distribution of predictions versus GT for glutamine. Improvements for clipped concentration ranges can be monitored via scores. However, the 40%–60% interval shows a significant number of outliers. Prediction distributions are still far from being uniform. GABA, γ-aminobutyric acid; Glu, glutamate; sI, syllo-inositol; tCr, total creatine.

3.4 │ CNN predictions versus model fitting estimates

A general juxtaposition of CNN and MF performance is contained in Figure 7. In Figure 10, detailed results are presented for two metabolites in the form of regression plots for *ShallowNet-2D2c-hp* and MF with FiTAID. In addition, the estimated CRLBs from MF are displayed and then compared in subgroups of SNR with the variance found in MF estimates and CNN predictions.

Area-constrained MF shows biases at the parameter boundaries for weakly represented metabolites (e.g., Asp). However, traditional MF outperforms quantification via DL: regression lines show less bias (a and q), and the
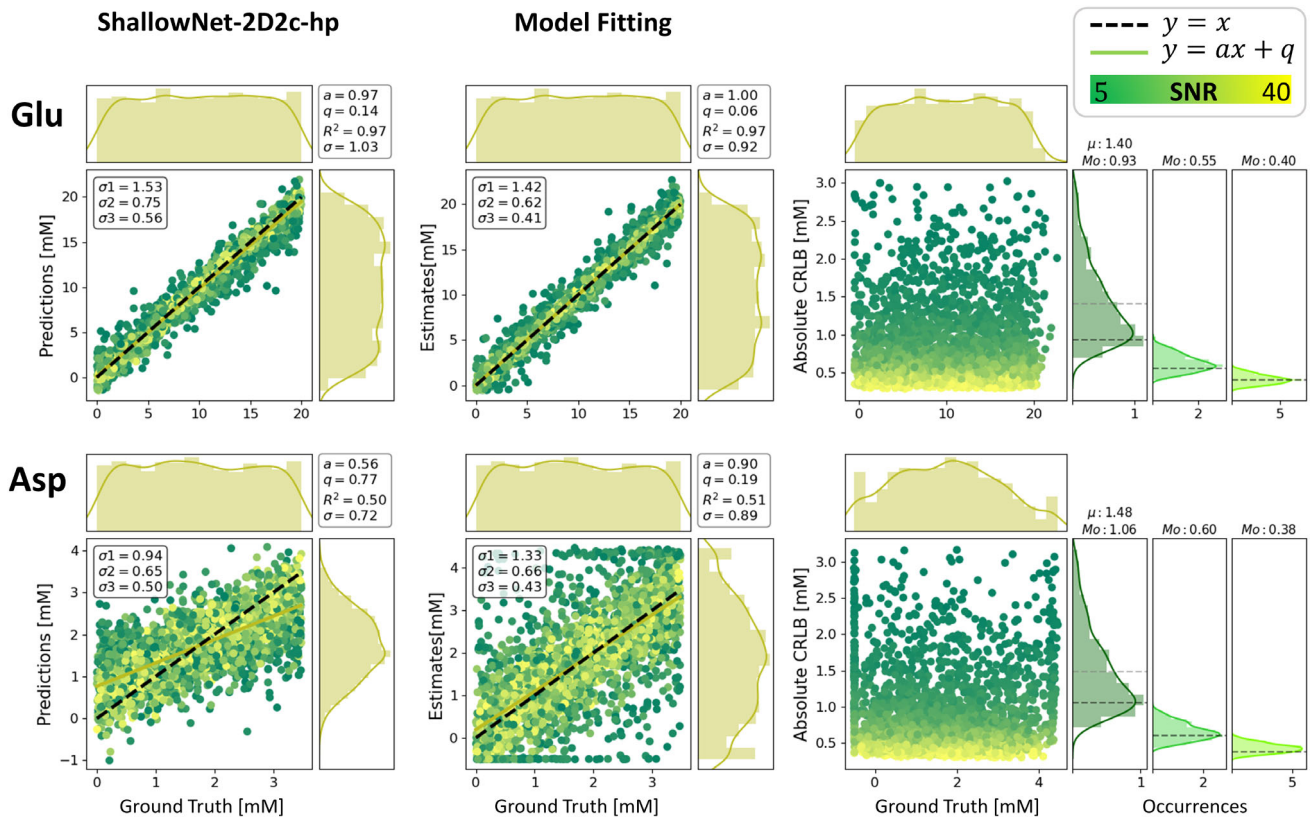
FIGURE 10 Comparison of performance for deep learning and model fitting reported for two illustrative metabolites. (Left) DL prediction versus GT mapped via *ShallowNet-2D2c-hp* with water reference. (Middle) Estimates versus GT for the MF approach. (Right) CRLBs evaluated on the fitted estimates. Histograms on the right group three subsets of an equal number of samples with different levels of SNR—group 1: SNR < 16.7, group 2: 16.7 < SNR < 28.4, and group 3: SNR > 28.4 displaying the distribution of estimated CRLBs. For group 1, given the skewness of distribution, mode (Mo) and mean (μ) values are reported. For comparison, RMSEs (σs) are reported as estimated for each SNR group for both DL and MF. DL's RMSEs (σs) underestimate CRLBs for low relative SNR metabolites.

distribution shape of estimates is closer to a uniform pattern within the GT range. RMSEs (σs) are higher in the case of MF for medium- to weakly represented metabolites (e.g., Asp) but lower for well-defined metabolites (e.g., Glu) (as formerly noted in Figure 7). Consequently, although σs of MF are bigger than the CRLBs estimated for their SNR reference group, σs of DL overestimate CRLBs for well-defined metabolites and underestimate CRLBs for weakly represented metabolites.

4 | DISCUSSION

Quantitation of brain metabolites using deep learning methods with spectroscopy data in 1D, 2D, and a combined input format was implemented in multiple network architectures. The main aim of the investigation was to compare the core performance of quantification in an idealized setting of simulated spectra. In fact, the analysis of the optimal performance of both, MF and DL, may otherwise be blurred by additional experimental inaccuracies or

artifacts from actual in vitro or in vivo spectra. Moreover, these nuisance contributors may be tackled in separate traditional or DL preprocessing steps that are beyond the current analysis. Many of the methods proved successful in providing absolute concentration values even when using a very large concentration range for the tested metabolites that goes way beyond the near-normal range that has often been used in the past. In addition, different forms of network input were tested, including a specifically tailored time-frequency domain representation and a downscaled water peak for easing of quantification. Whereas data augmentation by active learning schemes showed only modest improvements, ensembles of heterogeneous networks that combine both input representation domains improve the quantitation tasks substantially.

Results from DL predictions were compared to estimates from traditional MF, where it was found that MF is more accurate than DL at high and modest relative noise levels. MF yields higher variance at low SNR, with estimated concentrations artificially aggregated at the boundaries of the fitting parameter range. Predictions

obtained with DL algorithms delusively appear more precise (lower RMSE) in the low SNR regime, which may misguide nonexperts to believe that the DL predictions are reliable even at low SNR. However, these predicted concentrations are strongly biased by the dataset the network has been trained with. Hence, in case of high uncertainty (e.g., metabolites with low relative SNR or present in concentrations at the edge of the parameter/training space), the predicted concentration tends toward the most likely value: the average value from the training set.

4.1 | Forms of input to networks

Previously, 1D spectra have mostly been used as input for DL algorithms. Here, they have been compared and combined with 2D time-frequency domain spectrograms that had explicitly been designed to be of manageable size while retaining those areas of the high-resolved standard spectrogram that contain the most relevant information, that is, rich in detail in frequency domain to distinguish overlapping spectral features but also maintaining enough temporal structure to characterize T_2^* signal decay. This comes at the cost that the spectrogram creation cannot be reversed mathematically. However, this is irrelevant when serving as input to a DL network. It was found that this tailored time-frequency representation as input in combination with a shallow CNN architecture performs best and outperforms the use of traditional 1D frequency-domain input for straight quantification or for metabolite basis spectrum isolation with subsequent integration. Furthermore, DL quantitation performance improved upon the inclusion of a downscaled water peak for reference, likely solving scaling issues if no reference is provided.

4.2 | Active learning

Active learning has been explored by extending the training dataset with cases that appeared challenging to predict in the original setup. In particular, new training data with nonequal distribution of metabolite concentrations have been used with a predominance of single or multiple metabolites at low or high concentrations. None of these trials led to substantial improvements, although it might be helpful if specific metabolites are targeted primarily. Such data augmentation for all metabolites simultaneously even deteriorated the overall network performance. This can be understood given that augmentation at the border of the concentration range inherently leads to an underrepresentation of intermediate cases, which are

equally relevant for the overall performance. Extending the size of the training set even further in an unspecific manner appears to still yield modest improvements.[75] In addition, an unconventional way of active learning was probed by using unequal dataset ranges in training and testing by limiting testing on the central portion of the training range. This setup clearly ameliorated some of the issues at the edges of the testing range found in the typical setup. This approach was only implemented by reducing the test range rather than expanding the training range, which would yield better comparable outcome scores (e.g., R^2). However, expanding the training range to negative concentrations may be questionable.

While data augmentation with a bigger proportion of low SNR spectra leads to worse performance, the theoretical prediction limits for good SNR data are probed in the noiseless scenario in which training and testing are run with GT data. Example results for a *ShallowNet* architecture are reported in Figure S17 for NAA, GSH, and Lac. This, combined with the results discussed, suggests that the bottleneck that limits higher prediction performances is SNR, just like in traditional MF, regardless of the implementation of state-of-the-art networks, network optimization, or dataset augmentation. It thus reflects limitations in clinical applications where high enough SNR is just not available. According to this study, DL cannot do miracles unless one accepts the bias toward training conditions.[73]

4.3 | Ensemble of networks

An ensemble of networks has been implemented, and it shows improvements for quantifying metabolites. A combination of networks is less sensitive to the specifics of the training and helps reduce the variance in the predictions. Furthermore, ensembles of networks where multiple noise-sensitive predictions are weighted are more robust to noise. However, even the thus optimized networks underperform in comparison to MF. For MF, CRLBs clearly indicate limits for the confidence in the fit results. For DL, including the optimized ensemble of networks, such limits can only vaguely be deduced from the distributions of predicted values with the major danger of bias toward training data norms.[76,77] The CRLB would provide good guidance for the valid range of DL predictions as well—although of course they are not readily available without the model. New tools to estimate precision and replace CRLB in the case of DL[76,77] still have to prove their value in practice. The situation will be different again if the DL quantification is trained to include cleaning of spectra from artifacts (ghosts, baseline interference) where CRLBs are not available.

4.4 | Low SNR regime

Both MF and DL show lower reliability in quantifying metabolites in the low SNR regime. Clear-cut SNR limits for validity of concentration estimates are not available, neither for MF nor for DL, although SNR values are often indicated as measure of spectral quality. While CRLBs provide a widely used and easy-to-interpret reliability measure that includes the influence of SNR, a similar widely accepted concept does currently not extend to DL approaches.[77] Obviously, a SNR threshold for DL reliability would have to be metabolite-SNR specific, but already the definition of a meaningful metabolite-specific SNR would be cumbersome given that peak-splitting patterns and number of contributing protons as well as lineshape introduce ambiguity. On top, such a metabolite SNR would depend on the estimated metabolite content, whose reliability is at stake. Therefore, just like for MF, global or metabolite-specific SNR will not be informative enough. An uncertainty measure is needed that is based on the predictions and noise distribution but also integrating the uncertainty propagation of the DL model prediction[78,79] (like the inverse of the Fisher information matrix used in the CRLB definition[74]). Despite flourishing literature,[80,81] addressing uncertainty estimation as a complementary tool for DL interpretability, a full-scale analysis of the robustness and reliability of such models is still challenging.[82–84] First attempts to extend these concepts in DL for MRS quantification are just subject of recent investigations[76,77] but far from general acceptance.

4.5 | Limitations

The current investigation focused on probing multiple DL techniques and input forms for a full range of metabolite concentrations but a limited range of spectral quality. In particular, the shim remained in a broadly acceptable range, no phase or frequency jitter was considered, and no artifactual data was included. Such features could have been integrated in the current setup to arrive at a more realistic framework. However, the core of the findings (performance of the actual quantification step) is expected to remain in place. In addition, it is recommended to add separate preprocessing steps to prepare the data for the presented algorithms rather than to combine processing and quantification in a single process.[3] They could be realized in the form of dedicated DL networks, such as those proposed for phase and frequency drift corrections,[20,85,86] and stacked before the quantification model. This would also ensure the essential gain in speed expected from DL quantification models.

Direct comparison with previously proposed successful DL quantification implementations like Ref. 22 was not possible or meaningful for lack of open access network details and differences in the considered spectra.

Our particular implementation used to create spectrograms was optimized to maintain relevant resolution but downweights the initial part of the FID (initialization of Hamming window). CNN inputs may thus not be fully susceptible to changes in broad signals. Alternative recipes with, for example, prefilled filters or circular datasets, were not explored.

Furthermore, active learning has been explored for a single network type and could in principle be more beneficial for other networks or types of input than what has been found here.

5 | CONCLUSIONS

Quantification of MR spectra via diverse and optimized DL algorithms and using 1D and 2D input formats have been explored and have shown adequate performance as long as the metabolite-specific SNR is sufficient. However, as soon as SNR becomes critical, CNN predictions are strongly biased to the training dataset structure.

Traditional MF requires parameter tuning and algorithm convergence, making it more time consuming than DL-based estimates. On the other hand, we have shown that ideally (i.e., with simulated cases) and statistically (i.e., within a variable cohort of cases), it can achieve higher performances when compared to a faster DL approach. DL does not require feature selection by the user, but the potential intrinsic biases at training set boundaries act like soft constraints in traditional modeling,[9] leading estimated values to the average expected concentration range, which is dangerous in a clinical context that requires the algorithm to be unbiased to outliers (i.e., pathological data).

Active learning and ensemble of networks are attractive strategies to improve prediction performances. However, data quality (i.e., high SNR) has proven as bottleneck for adequate unbiased performance.

Engineering Research, University of Bern, Switzerland) for very helpful discussions.

DATA AVAILABILITY STATEMENT
The main part of the code will be available on GitHub (https://github.com/bellarude). In addition, simulated datasets will be available on MRSHub (https://mrshub.org/). For questions, please contact the authors.

ORCID
Rudy Rizzo https://orcid.org/0000-0003-4572-5120
Sreenath P. Kyathanahally https://orcid.org/0000-0002-7399-8487
Amirmohammad Shamaei https://orcid.org/0000-0001-8342-3284
Roland Kreis https://orcid.org/0000-0002-8618-6875

REFERENCES
1. De Graaf RA. *In Vivo NMR Spectroscopy: Principles and Techniques*. 3rd ed. John Wiley & Sons; 2019.
2. Kreis R, Boer V, Choi IY, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: background and experts' consensus recommendations. *NMR Biomed*. 2020;34:e4347. doi:10.1002/nbm.4347
3. Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed*. 2021;34:e4257. doi:10.1002/nbm.4257
4. Ratiney H, Sdika M, Coenradie Y, Cavassila S, van Ormondt D, Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR Biomed*. 2005;18:1-13. doi:10.1002/nbm.895
5. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn Reson Med*. 1993;30:672-679. doi:10.1002/mrm.1910300604
6. Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data. *Magn Reson Med*. 2011;65:1-12. doi:10.1002/mrm.22579
7. Chong DGQ, Kreis R, Bolliger CS, Boesch C, Slotboom J. Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a fitting tool for arrays of interrelated datasets. *MAGMA*. 2011;24:147-164. doi:10.1007/s10334-011-0246-y
8. Bhogal AA, Schür RR, Houtepen LC, et al. 1H–MRS processing parameters affect metabolite quantification: the urgent need for uniform and transparent standardization. *NMR Biomed*. 2017;30:e3804. doi:10.1002/nbm.3804
9. Marjańska M, Deelchand DK, Kreis R, et al. Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. *Magn Reson Med*. 2022;87:11-32. doi:10.1002/mrm.28942
10. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444. doi:10.1038/nature14539
11. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29:102-127. doi:10.1016/j.zemedi.2018.11.002
12. Lam F, Li Y, Peng X. Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models. *IEEE Trans Med Imaging*. 2020;39:545-555. doi:10.1109/TMI.2019.2930586
13. Klukowski P, Augoff M, ZieRba M, Drwal M, Gonczarek A, Walczak MJ. NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics*. 2018;34:2590-1597. doi:10.1093/bioinformatics/bty134
14. Hatami N, Sdika M, Ratiney H. Magnetic resonance spectroscopy quantification using deep learning. *Lect Notes Comput Sci*. 2018;11070:467-475. doi:10.1007/978-3-030-00928-1_53
15. Lee H, Lee HH, Kim H. Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy. *Magn Reson Med*. 2020;84:559-568. doi:10.1002/mrm.28164
16. Iqbal Z, Nguyen D, Thomas MA, Jiang S. Deep learning can accelerate and quantify simulated localized correlated spectroscopy. *Sci Rep*. 2021;11:8727. doi:10.1038/s41598-021-88158-y
17. Lee HH, Kim H. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med*. 2019;82:33-48. doi:10.1002/mrm.27727
18. Kyathanahally SP, Döring A, Kreis R. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. *Magn Reson Med*. 2018;80:851-863. doi:10.1002/mrm.27096
19. Gurbani SS, Schreibmann E, Maudsley AA, et al. A convolutional neural network to filter artifacts in spectroscopic MRI. *Magn Reson Med*. 2018;80:1765-1775. doi:10.1002/mrm.27166
20. Tapper S, Mikkelsen M, Dewey BE, et al. Frequency and phase correction of J-difference edited MR spectra using deep learning. *Magn Reson Med*. 2021;85:1755-1765. doi:10.1002/mrm.28525
21. Jang J, Lee HH, Park JA, Kim H. Unsupervised anomaly detection using generative adversarial networks in 1H-MRS of the brain. *J Magn Reson*. 2021;325:106936. doi:10.1016/j.jmr.2021.106936
22. Lee HH, Kim H. Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain. *Magn Reson Med*. 2020;84:1689-1706. doi:10.1002/MRM.28234
23. Gurbani SS, Sheriff S, Maudsley AA, Shim H, Cooper LAD. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. *Magn Reson Med*. 2019;81:3346-3357. doi:10.1002/mrm.27641
24. Chandler M, Jenkins C, Shermer SM, Langbein FC. MRSNet: metabolite quantification from edited magnetic resonance spectra with convolutional neural networks. 2019 arXiv:1909.03836v1 [eess.IV]. 10.48550/arXiv.1909.03836
25. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. doi:10.1016/J.MEDIA.2017.07.005
26. Gyori NG, Palombo M, Clark CA, Zhang H, Alexander DC. Training data distribution significantly impacts the estimation of tissue microstructure with machine learning. *Magn Reson Med*. 2022;87:932-947. doi:10.1002/MRM.29014
27. Espi M, Fujimoto M, Kinoshita K, Nakatani T. Exploiting spectro-temporal locality in deep learning based acoustic event detection. *J Audio Speech Music Proc*. 2015;2015:26. doi:10.1186/s13636-015-0069-2

28. Thomas S, Ganapathy S, Saon G, Soltau H. Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2014:2519–2523. 10.1109/ICASSP.2014.6854054

29. Alaskar H. Deep learning-based model architecture for time-frequency images analysis. *Int J Adv Comput Sci Appl*. 2018;9:486-494. doi:10.14569/IJACSA.2018.091268

30. Zagoruyko S, Komodakis N. Wide residual networks. In *ArXiv*; 2017:arXiv:1605.07146. 10.5244/C.30.87

31. Cohn DA, Ghahramani Z, Jordan MI. Active learning with statistical models. *J Artif Intell Res*. 1996;4:129-145. doi:10.1613/JAIR.295

32. Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell*. 1990;12:993-1001. doi:10.1109/34.58871

33. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60:84-90. doi:10.1145/3065386

34. Patra S, Bruzzone L. A cluster-assumption based batch mode active learning technique. *Pattern Recognit Lett*. 2012;33:1042-1048. doi:10.1016/J.PATREC.2012.01.015

35. Maiora J, Ayerdi B, Graña M. Random forest active learning for AAA thrombus segmentation in computed tomography angiography images. *Neurocomputing*. 2014;126:71-77. doi:10.1016/J.NEUCOM.2013.01.051

36. Kutsuna N, Higaki T, Matsunaga S, et al. Active learning framework with iterative clustering for bioimage classification. *Nat Commun*. 2012;3:1032. doi:10.1038/ncomms2030

37. Lewis DD, Gale WA. A sequential algorithm for training text classifiers. In the *17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1994: 3–12. 10.1007/978-1-4471-2099-5_1

38. Tuia D, Ratle F, Pacifici F, Kanevski MF, Emery WJ. Active learning methods for remote sensing image classification. *IEEE Trans Geosci Remote Sens*. 2009;48:2218-2232. doi:10.1109/TGRS.2008.2010404

39. Silva C, Ribeiro B. Margin-based active learning and background knowledge in text mining. In the *4th International Conference on Hybrid Intelligent Systems*, 2005: 8–13. 10.1109/ICHIS.2004.70

40. Pedrosa de Barros N, McKinley R, Wiest R, Slotboom J. Improving labeling efficiency in automatic quality control of MRSI data. *Magn Reson Med*. 2017;78:2399-2405. doi:10.1002/mrm.26618

41. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press; 2005.

42. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55:119-139. doi:10.1006/jcss.1997.1504

43. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In the *22nd International Conference on Knowledge Discovery and Data Mining*; 2016: 785–794. 10.1145/2939672

44. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. In the *31st International Conference on Neural Information Processing Systems*; 2017: 3149–3157. 10.5555/3294996

45. Soher BJ, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. In *Proceedings of the 19th Annual Meeting of ISMRM*, Montréal, Québec, Canada. 2011, 1410.

46. Oz G, Tkac I. Short-echo, single-shot, full-intensity proton magnetic resonance spectroscopy for neurochemical profiling at 4 T: validation in the cerebellum and brainstem. *Magn Reson Med*. 2011;65:901-910. doi:10.1002/mrm.22708

47. The Mathworks Inc. MATLAB (R2019a).MathWorks Inc 2019.

48. Marjańska M, McCarten JR, Hodges J, et al. Region-specific aging of the human brain as evidenced by neurochemical profiles measured noninvasively in the posterior cingulate cortex and the occipital lobe using 1H magnetic resonance spectroscopy at 7 T. *Neuroscience*. 2017;354:168-177. doi:10.1016/j.neuroscience.2017.04.035

49. Hoefemann M, Bolliger CS, Chong DGQ, van der Veen JW, Kreis R. Parameterization of metabolite and macromolecule contributions in interrelated MR spectra of human brain using multidimensional modeling. *NMR Biomed*. 2020;33:e4328. doi:10.1002/nbm.4328

50. Bottomley PA, Griffiths JR. *Handbook of Magnetic Resonance Spectroscopy in Vivo: MRS Theory, Practice and Applications*. 1st ed. Hoboken, NJ, Wiley & Sons; 2016.

51. Oz G, Alger JR, Barker PB, et al. Clinical proton MR spectroscopy in central nervous system disorders. *Radiology*. 2014;270:658-679.

52. Träber F, Block W, Lamerichs R, Gieseke J, Schild HH. 1H metabolite relaxation times at 3.0 Tesla: measurements of T1 and T2 values in normal brain and determination of regional differences in transverse relaxation. *J Magn Reson Imaging*. 2004;19:537-545. doi:10.1002/jmri.20053

53. An L, Li S, Shen J. Simultaneous determination of metabolite concentrations, T1 and T2 relaxation times. *Magn Reson Med*. 2017;78:2072-2081.

54. Zhang Y, Shen J. Simultaneous quantification of glutamate and glutamine by J-modulated spectroscopy at 3 Tesla. *Magn Reson Med*. 2016;76:725-732.

55. Cudalbu C, Behar KL, Bhattacharyya PK, et al. Contribution of macromolecules to brain 1H MR spectra: experts' consensus recommendations. *NMR Biomed*. 2021;34:e4393. doi:10.1002/nbm.4393

56. Van RG, Drake FL. *Python 3 Reference Manual*. CreateSpace; 2009.

57. Gulli A, Pal S. *Deep Learning with Keras*. Packt Publishing; 2017.

58. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. In the *12th USENIX Symposium on Operating Systems Design and Implementation*; 2016: 265–283. 10.5555/3026877.3026899

59. Bisong E. Google colaboratory. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, New York City; 2019. doi:10.1007/978-1-4842-4470-8_7

60. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In *arXiv*; 2015:1409.1556.

61. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In the *2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016: 770–778. 10.1109/CVPR.2016.90

62. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In the *2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016: 2818–2826. 10.1109/CVPR.2016.308

63. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In *2015 IEEE Computer Society Conference on Computer Vision*

and Pattern Recognition (CVPR); 2015: 1–9. 10.1109/CVPR.2015.7298594

64. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In the *31st AAAI Conference on Artificial Intelligence*; 2017: 4278–4284. 10.48550/arXiv.1602.07261

65. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *ArXiv*; 2015:arXiv:1502.03167.

66. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). In: *ArXiv*; 2016:arXiv:1511.07289.

67. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. In: *ArXiv*; 2012:arXiv:1206.2944.

68. Rizzo R, Kreis R. Accounting for bias in estimated metabolite concentrations from cohort studies as caused by limiting the fitting parameter space. In *Proceedings of the 2021 ISMRM & SMRT Annual Meeting and Exhibition, Virtual meeting*, May 15–20, 2021. p. 2011.

69. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci*. 2015;9351:234-241. doi:10.1007/978-3-319-24574-4_28

70. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In *ArXiv*; 2017:arXiv:1412.6980.

71. Bengio Y. Practical recommendations for gradient-based training of deep architectures. *Lect Notes Comput Sci*. 2012;7700:437-478.

72. Bengio Y, Goodfellow IJ, Courville A. Optimization for training deep models. *Deep Learning*. MIT Press; 2016.

73. Landheer K, Juchem C. Are Cramér-Rao lower bounds an accurate estimate for standard deviations in in vivo magnetic resonance spectroscopy? *NMR Biomed*. 2021;34:e4521. doi:10.1002/nbm.4521

74. Bolliger CS, Boesch C, Kreis R. On the use of Cramér-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments. *Neuroimage*. 2013;83:1031-1040. doi:10.1016/j.neuroimage.2013.07.062

75. Hong S, Shen J. Impact of training size on deep learning performance in in vivo 1H MRS. In *Proceedings of the 2021 ISMRM & SMRT Annual Meeting and Exhibition, Virtual meeting*, May 15–20, 2021, p. 2015.

76. Lee HH, Kim H. Bayesian deep learning-based 1 H-MRS of the brain: metabolite quantification with uncertainty estimation using Monte Carlo dropout. *Magn Reson Med*. 2022;88:38-52. doi:10.1002/MRM.29214

77. Rizzo R, Dziadosz M, Kyathanahally SP, Reyes M, Kreis R. Reliability of quantification estimates in MR spectroscopy: CNNs vs traditional model fitting. *Med Image Comput Comput Assist Interv–MICCAI 2022 Lect Notes Comput Sci*. 2022;13438:715-724. doi:10.1007/978-3-031-16452-1_68

78. Gal Y. 2016 Uncertainty in deep learning. https://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf

79. Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? In the *31st Conference on Neural Information Processing Systems (NIPS)*; 2017.

80. Sanchez T, Caramiaux B, Thiel P, Mackay WE. Deep learning uncertainty in machine teaching. In *27th Annual Conference on Intelligent User Interfaces (IUI)*, Vol. 1, 2022. 10.1145/3490099.3511117

81. Abdar M, Pourpanah F, Hussain S, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion*. 2021;76:243-297. doi:10.1016/j.inffus.2021.05.008

82. Jungo A, Reyes M. Assessing reliability and challenges of uncertainty estimations for medical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Proceedings, Part II*. Berlin, Heidelberg, Springer-Verlag. 10.1007/978-3-030-32245-8_6

83. Ennab M, McHeick H. Designing an interpretability-based model to explain the artificial intelligence algorithms in healthcare. *Diagnostics*. 2022;12:1557. doi:10.3390/DIAGNOSTICS12071557

84. Meng C, Trinh L, Xu N, Enouen J, Liu Y. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Sci Rep*. 2022;12:1-28. doi:10.1038/s41598-022-11012-2

85. Ma DJ, Le HAM, Ye Y, et al. MR spectroscopy frequency and phase correction using convolutional neural networks. *Magn Reson Med*. 2022;87:1700-1710. doi:10.1002/MRM.29103

86. Shamaei AM, Starcukova J, Pavlova I, Starcuk Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. *bioRxiv*. 2022. doi:10.1101/2022.06.28.497332

87. Lin A, Andronesi O, Bogner W, et al. Minimum reporting standards for in vivo magnetic resonance spectroscopy (MRSin-MRS): Experts' consensus recommendations. *NMR Biomed*. 2021;34:e4484. doi:10.1002/nbm.4484

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Table S1. List of probed networks for straight quantification of metabolites and some of their characteristics. The listed characteristics includes the complexity (defined as number of trainable parameters), test loss performance, and training time in sec/epoch. The network identifications were chosen as follows: *NetworkType-InputType-properties*. 1D: spectra, 2D: spectrograms, 1c: 1 channel, ks3: convolutional kernel size = 3, hp: Bayesian hyper-parameterized architecture, R: exploiting ReLU activations, x2: double convolution before MaxPooling, f: factorized convolution, rb: down-sampling via Reduction-Blocks

Figure S1. Schemes of Residual Network configurations with 1D (a) and 2D (b) inputs, as well as a deep residual network (c). The basic network structure is sketched on the left, the architectures of Residual, Identity, and Convolutional Blocks are reported on the right, while specifications are detailed in the tables in the middle, and symbols are explained near the bottom. The deeper Residual Network configuration has two convolutional layers at the beginning without pooling.

Figure S2. Schemes of Deep CNN configurations with 2D (a) and 1D (b) inputs, as well as an InceptionNet with

1D inputs (c). Network specifications are detailed in the tables, while the architectures of Reduction Blocks are reported on the bottom right. Symbols are explained near the bottom.

Figure S3. Schemes of InceptionNet configurations with 2D inputs on 2 channels. Networks (a) and (b) share the same configuration but (b) exploits convolutional factorization to speed-up training time. (c) Simple concatenation in architectures (a) and (b) are replaced by Reduction Blocks. The architectures of the Reduction Blocks are reported in Figure-S6. Symbols are explained on the right.

Figure S4. Schemes of (a) InceptionNet with 2D inputs and 1 channel, (b) EfficientNetB7, (c) ResNet50 and (d–f) Shallow Network configurations. Networks (a), (b) and (c) are modified from [1–3], respectively. (d) Implements ELU activations, (e) implements RELU activations, whereas (f) implements a deeper configuration with consecutive convolutional layers with sparse pooling. Network specifications are detailed in the tables. Symbols are explained near the bottom.

Figure S5. Scheme and detail of U-Net-1DR-hp configurations for metabolite basis-set prediction. Metabolite-specific network specifications are detailed in the tables. Symbols are explained at the bottom left.

Text S1. Details of Bayesian hyper-parameterization

Figure S6. Examples of dataset augmentation techniques representing sample distributions for two metabolites (NAA and GABA). (a) Dataset size increment with uniform distributed concentrations. (b) and (c) Active Learning weighted on higher occurrences of small and high concentrations for all metabolites in (b) and for selected metabolites in (c). (d) Active Learning weighted on more occurrences of low SNR entries whereas concentration distributions are kept uniform.

Figure S7. Comparison of prediction scores for *well-represented* and *medium-represented* metabolites for three CNN architectures with datasets with (red, black, or blue) and without (yellow, gray, or green) water reference. Mean values in orange. On average, water referencing yields higher coefficients a and R^2 and lower offset q and RMSE σ.

Figure S8. Comparison of prediction scores for *medium-represented* and *weakly-represented* metabolites for three CNN architectures with datasets with (red, black, or blue) and without (yellow, gray, or green) water reference. Mean values in orange. On average, water referencing yields higher coefficients a and R^2 and lower offset q and RMSE σ.

Figure S9. Maps and marginal distributions of predictions vs. GT for a *ResNet_1D_hp* network. Results for 16 metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE (σ) is reported as an overall measure of variability. A regression

model ($y = ax + q$) is also provided to judge prediction quality. R^2 measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in a and R^2 or by visual biases in distributions of predictions (bell-shape). The prediction bias towards the mean value of the training distribution is evident for medium- to weakly-represented metabolites (e.g., sI, Gly, Asp, PE, Tau, Lac, GABA). On average, metabolites with lower SNR yield higher errors. (q and σ in mM units.).

Figure S10. Maps and marginal distributions of predictions vs. GT for a *ShallowNet-2D2c-hp* network. Results for 16 metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE (σ) is reported as an overall measure of variability. A regression model ($y = ax + q$) is also provided to judge prediction quality. R^2 measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in a and R^2 or by visual biases in distributions of predictions (bell-shape). The prediction bias towards the mean value of the training distribution is evident for medium- to weakly-represented metabolites (e.g., sI, Gly, Asp, PE, Tau, Lac, GABA). On average, metabolites with lower SNR yield higher errors. (q and σ in mM units.)

Figure S11. Boxplots comparing the distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting estimate distributions (MF) and uniform Ground Truth (GT) distributions. Mis-prediction is evident for *medium-* to *weakly*-represented metabolites (e.g., sI, Asp, Tau, Lac) and can be monitored by different degrees of skewness of the boxplot. However, the bias to training distribution is not evident given the visual limitation of boxplots. For better visibility of this outcome, see Figure S14.

Figure S12. Comparison of distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting's estimate distributions (MF) and Ground Truth (GT) uniform distributions. Mis-prediction is evident for *medium-* to *weakly*-represented metabolites (e.g., sI, Asp, Tau, Lac) and can be monitored by visual biases (bell-shape) towardstoward the mean value of the training distribution (i.e., regression to the mean). Note: y-axes scale inhomogeneously between different networks. However, all distributions integrate to 1.

Figure S13. *Concise scores* presented to compare quantification quality for different networks and input setups for 16 metabolites. Results reported using the proposed artificial water signal reference. Network identification is chosen as follows: *NetworkType-InputType-properties*. Keywords: 1D: spectra, 2D: spectrograms, 1c: 1 channel, ks3: convolutional kernel size = 3, hp: Bayesian hyper-parameterized architecture, R: exploiting ReLU activations, x2: double convolution before MaxPooling, f:

factorized convolution, rb: down-sampling via Reduction-Blocks.

Figure S14. Comparison of performance scores from different networks for 16 metabolites. Model fitting is included in the comparison.

Text S2. Comparison of predictions from different CNNs.

Figure S15. Comparison of outcomes of Active Learning approaches using concise scores.

Figure S16. Quantification outcome as reflected by *concise scores* for differently trained single networks and three ensembles of networks (identical training set for 16 metabolites).

Figure S17. Maps and marginal distributions of predictions vs. GT obtained for three metabolites using *ShallowNet-2D2c-hp* as contrasted for a realistic and noiseless dataset.

Table S2. Results of Active Learning on training set: scores of 16 metabolites for every augmented training set.

Table S3. Results of emulated Active Learning on test set: scores of 16 metabolites for every concentration range considered.

Table S4. Outcome for ensemble learning: scores for 16 metabolites for average network or ensemble of network considered.

Table S5. MRSinMRS checklist.[87]

Supporting Information

Table S1: List of probed networks for straight quantification of metabolites and some of their characteristics.

Figure S1: Schemes of Residual Network configurations with 1D (a) and 2D (b) inputs, as well as a deep residual network (c).

Figure S2: Schemes of Deep CNN configurations with 2D (a) and 1D (b) inputs, as well as an Inception Network with 1D inputs (c).

Figure S3: Schemes of Inception Network configurations with 2D inputs on 2 channels.

Figure S4: Schemes of (a) Inception Network with 2D inputs and 1 channel, (b) EfficientNetB7, (c) ResNet50 and (d-f) Shallow Network configurations.

Figure S5: Scheme and detail of U-Net-1DR-hp configurations for metabolite basis-set prediction.

Text S1: Details of Bayesian hyper-parameterization.

Figure S6: Examples of dataset augmentation techniques representing sample distributions for two metabolites (NAA and GABA).

Figure S7: Comparison of prediction scores for *well-represented* and *medium-represented* metabolites for three CNN architectures with datasets with and without water reference.

Figure S8: Comparison of prediction scores for *medium-represented* and *weakly-represented* metabolites for three CNN architectures with datasets with and without water reference.

Figure S9: Maps and marginal distributions of predictions vs. GT for a *ResNet_1D_hp* network.

Figure S10: Maps and marginal distributions of predictions vs. GT for a *ShallowNet-2D2c-hp* network.

Figure S11: Boxplots comparing the distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting's estimate distributions (MF) and Ground Truth (GT) uniform distributions.

Figure S12: Comparison of distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting's estimate distributions (MF) and Ground Truth (GT) uniform distributions.

Figure S13: *Concise scores* presented to compare quantification quality for different networks and input setups for 16 metabolites.

Figure S14: Comparison of performance scores from different networks for 16 metabolites.

Text S2: Comparison of predictions from different CNNs.

Figure S15: Comparison of outcomes of Active Learning approaches using concise scores.

Table S2: Results of Active Learning on training set: scores of 16 metabolites for every augmented training set.

Table S3: Results of emulated Active Learning on test set: scores of 16 metabolites for every concentration range considered.

Figure S16: Quantification outcome as reflected by *concise scores* for differently trained single networks and three ensembles of networks (identical training set for 16 metabolites).

Table S4: Outcome for ensemble learning: scores for 16 metabolites for average network or ensemble of network considered.

Figure S17: Maps and marginal distributions of predictions vs. GT obtained for three metabolites using *ShallowNet-2D2c-hp* are contrasted for a realistic and noiseless dataset.

Table S5: MRSinMRS checklist.

	network	complexity	test-loss	training time	epochs
1	ResNet-1D	1 078 593	0.0240	30	92
2	ResNet-1D-hp	1 457 997	0.0242	32	66
3	ResNet-2D1c	1 090 641	0.0287	37	38
4	ResNet-2D1c-ks3	509 009	0.0306	25	50
5	deepResNet-2D1c	1 094 385	0.0316	37	33
6	deepNet-2D1c	874 513	0.0350	9	38
7	deepNet-2D1c-ks3	609 041	0.0369	5	44
8	DeepNet-1D	3 433 217	0.0344	7	24
9	DeepNet-1D-hp	7 269 197	0.0281	8	29
10	InceptionNet-1D	13 584 699	0.0238	165	77
11	InceptionNet-2D2c	6 934 586	0.0229	33	30
12	InceptionNet-2D2c-f	2 646 036	0.0249	23	43
13	InceptionNet-2D2c-rb	10 688 886	0.0246	61	31
14	InceptionNet-v4-2D1c	13 602 305	0.0366	93	134
15	EfficientNetB7-2D2c	63 960 515	0.0349	129	72
16	ResNet50-2D1c	23 570 057	0.0543	36	71
17	ShallowNet-2D2c	431 685	0.0258	5	34
18	ShallowNet-2D2c-hp	2 998 235	0.0228	55	40
19	ShallowNet-2D2c-R	956 885	0.0256	9	45
20	ShallowNet-2D2c-R-ks3	345 051	0.0276	5	29
21	ShallowNet-2D2c-x2	18 348 171	0.0260	16	31
22	ShallowNet-2D2c-x2-hp	1 086 951	0.0244	94	87

Table S1: List of probed networks for straight quantification of metabolites and some of their characteristics. The listed characteristics include the complexity (defined as number of trainable parameters), test loss performance, and training time in sec/epoch. The network identifications were chosen as follows: *NetworkType-InputType-properties.* 1D: spectra, 2D: spectrograms, 1c: 1 channel, ks3: convolutional kernel size = 3, hp: Bayesian hyper-parameterized architecture, R: exploiting ReLU activations, x2: double convolution before MaxPooling, f: factorized convolution, rb: down-sampling via Reduction-Blocks

a) ResNet-1D-xx

input: 2048x1x1

(3,nf1)

(3,nf2)

Residual Block (n1,m1,s11,s12,s13)

Residual Block (n2,m2,s21,s22,s23)

Residual Block (n3,m3,s31,s32,s33)

d   nd

	xx	
Param.	-	hp
nf1	8	15
nf2	16	40
n1,m1,s11,s12,s13	32,128,3	90,150,3
n2,m2,s21,s22,s23	64,256,3	150,230,3
n3,m3,s31,s32,s33	128,512,3	140,600,3
nd	400	110
d	0.3	0.25
lrate	2e-4	3.19e-4

b) ResNet-2D1c-xx

input: 64x128x1

*Residual Block (n1,m1,s11,s12,s13)

*Residual Block (n2,m2,s21,s22,s23)

d   nd

	xx	
Param.	-	ks3
n1,m1	32,128	32,128
n2,m2	64,256	64,256
s11,s12,s13	13,11,9	3,3,3
s21,s22,s23	7,5,3	3,3,3
nd	400	110
d	0.1	01

c) deepResNet-2D1c

input: 64x128x1

(3,8)

(3,16)

*Residual Block (32,128,13,11,9)

*Residual Block (64,256,7,5,3)

0.1   400

*Residual Block implementation with 2D convolution

Convolutional Block (n,m,s)

(1,n)

(s,n)   (1,m)

(1,m)

Identity block (n,m,s)

(1,n)

(s,n)

(1,m)

Residual Block (n,m,s1,s2,s3)

Convolutional Block (n,m,s1)

Identity block (n,m,s2)

Identity block (n,m,s3)

conv2D(ks, nf, s)

conv1D(ks, nf, s)

ks: kernel size
nf: number of features
s: strides, default=1

conv1DTranspose(ks, nf, s)

concatenate

- - - batchNorm
—— batchNorm + ReLU
—— ELU
- - - lin
—— maxPooling 2/2
- - - avgPooling 2/2
—— dropout %
—— zeroPadding
···· cropping
—— flattening
—— Dense + batchNorm
—— Dense(17)+batchNorm + lin

Figure S1: Schemes of Residual Network configurations with 1D (a) and 2D (b) inputs, as well as a deep residual network (c). The basic network structure is sketched on the left, the architectures of Residual, Identity, and Convolutional Blocks are reported on the right, while specifications are detailed in the tables in the middle, and symbols are explained near the bottom. The deeper Residual Network configuration has two convolutional layers at the beginning without pooling.

a) DeepNet-2D1c-xx

input: 64x128x1

	xx	
Param.	-	ks3
ks_1	13	3
ks_2	11	3
ks_3	9	3
ks_4	7	3
ks_5	5	3
ks_6	3	3

b) DeepNet-1D-xx

input: 2048x1

	xx	
Param.	-	hp
nf_1	8	15
nf_2	16	40
nf_3	32	90
nf_4	64	150
nf_5	128	160
nf_6	256	280
nd	400	780
d	0.1	0.3
lrate	2e-4	3.3e-3

c) InceptionNet-1D

Figure S2: Schemes of Deep CNN configurations with 2D (a) and 1D (b) inputs, as well as an Inception Network with 1D inputs (c). Network specifications are detailed in the tables, while the architectures of Reduction Blocks are reported on the bottom right. Symbols are explained near the bottom.

Figure S3: Schemes of Inception Network configurations with 2D inputs on 2 channels. Networks (a) and (b) share the same configuration but (b) exploits convolutional factorization to speed-up training time. (c) Simple concatenation in architectures (a) and (b) are replaced by Reduction Blocks. The architectures of the Reduction Blocks are reported in Figure-S6. Symbols are explained on the right.

a) InceptionNet-v4-2D1c

input: 64x128x1

([9,9],50)

Inception Block a
Reduction Block a
Inception Block b
Reduction Block b
Inception Block c

0.5

Inception and Reduction Blocks are inspired from [kentsommer]

b) EfficientNetB7-2D2c

input: 32x128x2

Keras model

Model imported from Keras library [keras]

c) ResNet50-2D1c

input: 64x128x1

ResNet50 model

Model adapted from [jjuinni]. Converted from classification to regression net.

d) ShallowNet-2D2c-xx

input: 32x128x2

(ks1,nf1)

(ks2,nf2)   d1

(ks3,nf3)   d2

Param.	xx	
	-	hp
ks1,nf1	9,50	7,300
ks2,nf2	5,100	9,100
ks3,nf3	3,150	5,150
d1	0.1	0.35
d2	0.15	0.35
lrate	2e-4	5.4e-3

e) ShallowNet-2D2c-R-xx

input: 32x128x2

(ks$_1$,50)

(ks$_2$,100)   0.1

(ks$_3$,150)   0.15

Param.	xx	
	-	ks3
ks1	11	3
ks2	9	3
ks3	5	3

f) ShallowNet-2D2c-x2-xx

input: 32x128x2

(ks1,nf1)

(ks1,nf1)

(ks2,nf2)

(ks2,nf2)   d1

(ks3,nf3)

(ks3,nf3)   d2

Param.	xx	
	-	hp
ks1,nf1	9,50	3,60
ks2,nf2	5,100	11,310
ks3,nf3	3,150	5,280
d1	0.1	0.2
d2	0.15	0.45
lrate	2e-4	1.01e-2

conv2D(ks, nf, s)

conv1D(ks, nf, s)
ks: kernel size
nf: number of features
s: strides, default=1

conv1DTranspose(ks, nf, s)

concatenate

- - batchNorm
— batchNorm + ReLU
— ELU
- - lin
— maxPooling 2/2
- - avgPooling 2/2
— dropout %
zeroPadding
cropping

— flattening
— Dense + batchNorm
— Dense(17)+batchNorm + lin

Figure S4: Schemes of (a) Inception Network with 2D inputs and 1 channel, (b) EfficientNetB7, (c) ResNet50 and (d-f) Shallow Network configurations. Networks (a), (b) and (c) are modified from [1], [2] and [3], respectively. (d) Implements ELU activations, (e) implements RELU activations, whereas (f) implements a deeper configuration with consecutive convolutional layers with sparse pooling. Network specifications are detailed in the tables. Symbols are explained near the bottom.

a) UNet-1DR-hp

input: 1406x1

(3,nf10)
(3,nf20,2) (3,nf11)
(3,nf21)
(3,nf30,2) (3,nf22)
(3,nf31)
(3,nf40,2) (3,nf32)
(3,nf41)
(3,nf50,2) (3,nf42)
(3,nf51)
(3,nf52)
(3,nf60,2)
(3,nf61)
(3,nf62)
(3,nf70,2)
(3,nf71)
(3,nf72)
(3,nf80,2)
(3,nf81)
(3,nf82)
(3,nf90,2)
(3,nf91)
(3,nf92)
(3,1)

output: 1406x1

Legend:

- conv2D(ks, nf, s)
- conv1D(ks, nf, s)
- conv1DTranspose(ks, nf, s)
- concatenate

ks: kernel size
nf: number of features
s: strides, default=1

- batchNorm + ReLU activation
- ELU activation

- Linear activation
- maxPooling 2/2
- dropout (d=%)
- zeroPadding

- cropping
- flattening
- Dense (nd = nodes) + batchNorm
- Dense(17) + batchNorm + lin

Layer	Standard	tCho	NAAG	NAA	Asp	tCr	GABA	Glc	Glu	Gln	GSH	Gly	Lac	mI	PE	sI	Tau	H_2O
10, 92	32, 32	30	10	40	20	30	20	30	20	20	30	20	10	50	20	40	30	50
11, 91	64, 32	60	20	80	40	60	40	60	40	40	60	40	20	100	40	80	60	100
20, 90	32, 32	30	20	50	20	30	40	30	30	40	40	10	10	30	40	20	30	10
21, 82	64, 32	60	40	70	90	50	50	40	40	60	90	30	30	20	80	70	80	20
22, 81	128, 32	120	80	140	180	100	100	80	80	120	180	60	60	40	160	140	160	40
30, 80	48, 64	60	70	100	80	30	70	70	50	70	60	90	70	100	40	80	50	100
31, 72	128, 48	140	140	160	100	130	150	90	80	180	40	100	100	50	150	50	60	200
32, 71	256, 48	280	280	320	200	260	300	180	160	360	80	200	200	100	300	100	120	400
40, 70	64, 128	70	80	80	170	160	110	160	130	170	130	130	130	70	170	90	100	40
41, 62	256, 64	370	190	210	120	380	210	190	290	220	240	130	360	90	80	280	370	400
42, 61	512, 64	740	380	420	240	760	420	380	580	440	480	260	720	180	160	560	740	800
50, 60	80, 256	250	250	320	320	130	210	260	250	100	380	90	240	200	390	280	160	80
51	512	240	560	600	650	590	580	330	190	570	250	260	170	800	600	310	590	160
52	256	480	1120	1200	1300	1180	1160	660	380	1140	500	520	340	1600	1200	620	1180	320
lrate e^{-3}	0.2	0.885	1.127	0.652	6.458	1.581	1.278	2.948	0.925	0.502	8.353	7.774	1.483	2.844	2.223	3.988	9.667	0.434

Figure S5: Scheme and detail of U-Net-1DR-hp configurations for metabolite basis-set prediction. Metabolite-specific network specifications are detailed in the tables. Symbols are explained at the bottom left.

Text S1: Details of Bayesian hyper-parameterization

The Bayesian hyper-parameterization library from Keras tuner [50] was used. A total amount of 50 model trials was checked, with one execution per trial. Training was run for 200 epochs with batch size of 50, monitoring validation loss with an early stopping criterion with patience equal to 10. Finally, the best-tuned model with minimized validation loss was selected. Parameter space exploration started from a hard-coded version of the investigated networks and followed as reported below for each architecture:

ShallowNet-2D1c-hp and **ShallowNet-2D1c-2x-hp:**
1st layer: from 20 to 400 units, step of 10; kernel size from 3 to 11, step 2.
2nd layer: from 40 to 400 units, step of 10; kernel size from 3 to 11, step 2.
3rd layer: from 80 to 400 units, step of 10; kernel size from 3 to 11, step 2.
1st dropout: from 0.05 to 0.5, step of 0.05.
2nd dropout: from 0.05 to 0.5, step of 0.05.
Learning rate: from 2e-6 to 2e-2.

DeepNet-1D-hp:
1st layer: from 5 to 35 units, step 10.
2nd layer: from 10 to 60 units, step 10.
3rd layer: from 20 to 100 units, step 10.
4th layer: from 40 to 150 units, step 10.
5th layer: from 70 to 200 units, step 10.
6th layer: from 150 to 300 units, step 10.
Dense layer: from 100 to 1 000 units, step 10
Droput: from 0.05 to 0.8, step of 0.05.
Learning rate: from 2e-6 to 2e-2.

ResNet-1D-hp:
1st layer: from 5 to 40 units, step 10.
2nd layer: from 10 to 60 units, step 10.
1st residual block input: from 20 to 100 units, step 10.
1st residual block output: from 70 to 200 units, step 10.
2nd residual block input: from 50 to 150 units, step 10.
2nd residual block output: from 170 to 330 units, step 10.
3rd residual block input: from 100 to 200 units, step 10.
3rd residual block output: from 400 to 600 units, step 10.
Dense layer: from 100 to 1 000 units, step 10
Droput: from 0.1 to 0.8, step of 0.05.
Learning rate: from 2e-6 to 2e-2.

UNet-1DR-hp:

Layer 10 and 92: from 10 to 50 units, step 10. Layer 11 and 91 have doubled units of Layer 10.
Layer 20 and 90: from 10 to 50 units, step 10.
Layer 21 and 82: from 20 to 100 units, step 10. Layer 22 and 81 have doubled units of Layer 21.
Layer 30 and 80: from 20 to 100 units, step 10.
Layer 31 and 72: from 40 to 200 units, step 10. Layer 32 and 71 have doubled units of Layer 31.
Layer 40 and 70: from 40 to 200 units, step 10.
Layer 41 and 62: from 80 to 400 units, step 10. Layer 42 and 61 have doubled units of Layer 41.
Layer 50 and 60: from 80 to 400 units, step 10.
Layer 51: from 160 to 800 units, step 10. Layer 52 has doubled units of Layer 51.
Learning rate: from 2e-6 to 2e-2.

Note: Since exploitation of spectrograms in deep learning for speech and audio processing has shown best performances deploying network architectures with large convolutional kernels [48] [55] [56], kernel size was optimized only for these cases, whereas kept fixed to size 3 for networks fed with 1D spectra.

Figure S6: Examples of dataset augmentation techniques representing sample distributions for two metabolites (NAA and GABA). **a)** Dataset size increment with uniform distributed concentrations. **b)** and **c)** Active Learning weighted on higher occurrences of small and high concentrations for all metabolites in (b) and for selected metabolites in (c). **d)** Active Learning weighted on more occurrences of low SNR entries whereas concentration distributions are kept uniform.

Figure S7: Comparison of prediction scores for *well-represented* and *medium-represented* metabolites for three CNN architectures with datasets with (red, black, or blue) and without (yellow, gray, or green) water reference. Mean values in orange. On average, water referencing yields higher coefficients a and R^2 and lower offset q and RMSE σ.

Figure S8: Comparison of prediction scores for *medium-represented* and *weakly-represented* metabolites for three CNN architectures with datasets with (red, black, or blue) and without (yellow, gray, or green) water reference. Mean values in orange. On average, water referencing yields higher coefficients a and R^2 and lower offset q and RMSE σ.

Figure S9: Maps and marginal distributions of predictions vs. GT for a *ResNet_1D_hp* network. Results for 16 metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE (σ) is reported as an overall measure of variability. A regression model ($y = ax + q$) is also provided to judge prediction quality. R^2 measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in a and R^2 or by visual biases in distributions of predictions (bell-shape). The prediction bias toward the mean value of the training distribution is evident for medium- to weakly-represented metabolites (e.g., sI, Gly, Asp, PE, Tau, Lac, GABA). On average, metabolites with lower SNR yield higher errors. (q and σ in mM units.)

Figure S10: Maps and marginal distributions of predictions vs. GT for a *ShallowNet-2D2c-hp* network. Results for 16 metabolites are arranged in approximate decreasing order of relative SNR from top left to bottom right. RMSE (σ) is reported as an overall measure of variability. A regression model ($y = ax + q$) is also provided to judge prediction quality. R^2 measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in a and R^2 or by visual biases in distributions of predictions (bell-shape). The prediction bias toward the mean value of the training distribution is evident for medium- to weakly-represented metabolites (e.g., sI, Gly, Asp, PE, Tau, Lac, GABA). On average, metabolites with lower SNR yield higher errors. (q and σ in mM units.)

Figure S11: Boxplots comparing the distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting's estimate distributions (MF) and uniform Ground Truth (GT) distributions. Mis-prediction is evident for *medium-* to *weakly*-represented metabolites (e.g., sI, Asp, Tau, Lac) and can be monitored by different degrees of skewness of the boxplot. However, the bias to training distribution is not evident given the visual limitation of boxplots. For better visibility of this outcome, see Figure S14.

Figure S12: Comparison of distributions of predictions for 8 metabolites via 7 different CNN architectures vs. Model Fitting's estimate distributions (MF) and Ground Truth (GT) uniform distributions. Mis-prediction is evident for *medium*- to *weakly*-represented metabolites (e.g., sI, Asp, Tau, Lac) and can be monitored by visual biases (bell-shape) toward the mean value of the training distribution (i.e., regression to the mean). Note: y-axes scale inhomogeneously between different networks. However, all distributions integrate to 1.

Figure S13: *Concise scores* presented to compare quantification quality for different networks and input setups for 16 metabolites. Results reported using the proposed artificial water signal reference. Network identification is chosen as follows: *NetworkType-InputType-properties*. Keywords: 1D: spectra, 2D: spectrograms, 1c: 1 channel, ks3: convolutional kernel size = 3, hp: Bayesian hyper-parameterized architecture, R: exploiting ReLU activations, x2: double convolution before MaxPooling, f: factorized convolution, rb: down-sampling via Reduction-Blocks

Figure S14: Comparison of performance scores from different networks for 16 metabolites. Model fitting is included in the comparison.

Text S2: Comparison of predictions from different CNNs

Figure-S16 reports the measured scores of all 16 metabolites for each investigated network. A comparison with Fig.7 follows:

1D-input architectures (circles):
- ResNets yield slightly better performances than DeepNets, especially for weekly represented metabolites.
- InceptionNet shows performances comparable to ResNet.

2D-input 1-channel architectures (diamonds):
- Spectrogram information as 1-Channel mined via DeepNets or ResNets show worse performances than 2-Channels representation or 1D-input networks.
- ResNet architectures outperform DeepNets.
- Processing spectrogram inputs exploiting convolution operations with large kernels outperform convolutions with small kernels.
- State-of-Art ResNet-50 outperforms all the other attempted ResNet architectures. However, deeper ResNet architectures (i.e., double convolution and sparse pooling) do not perform better.
- InceptionNet-v4 shows comparable performances to ResNet architectures but slightly worse than ResNet50.

2D-input 2-channels architectures (squares):
- Shallow or Inception architectures exploiting signal representation as 2-channels show overall top performances.
- ReLU activation on shallow networks highlights comparable performances to ELU activation. However, ELU activation converges faster, TableS1. Usage of smaller kernels, on the contrary, entails worse performances, as confirmed for spectrogram 1-Channel representation. Deeper architectures slightly underperform on weakly defined metabolites.
- Inception networks stress top performances comparable to shallow networks although involving longer computational time given their higher complexity, Table.S1. Their highly engineered structures are open to further optimizations. Factorization does speed up computational time without strongly afflicting performance.
- EfficientNet strongly underperforms compared to the top scorers.

U-Net (stars):
- Performances are comparable on average with 1D-input architectures designed for straight quantification. Although their performances are below top scorers, specifically with lower a and higher q and σ, whereas R^2 model representability is comparable. Higher biases (q) are explained by integrating the isolated basis set: likewise, concentrations are biased to 0+ space in modeling.
- Optimization of U-Net mildly reduces, on average, the estimated bias q and standard deviation σ compared to its hardcoded version.
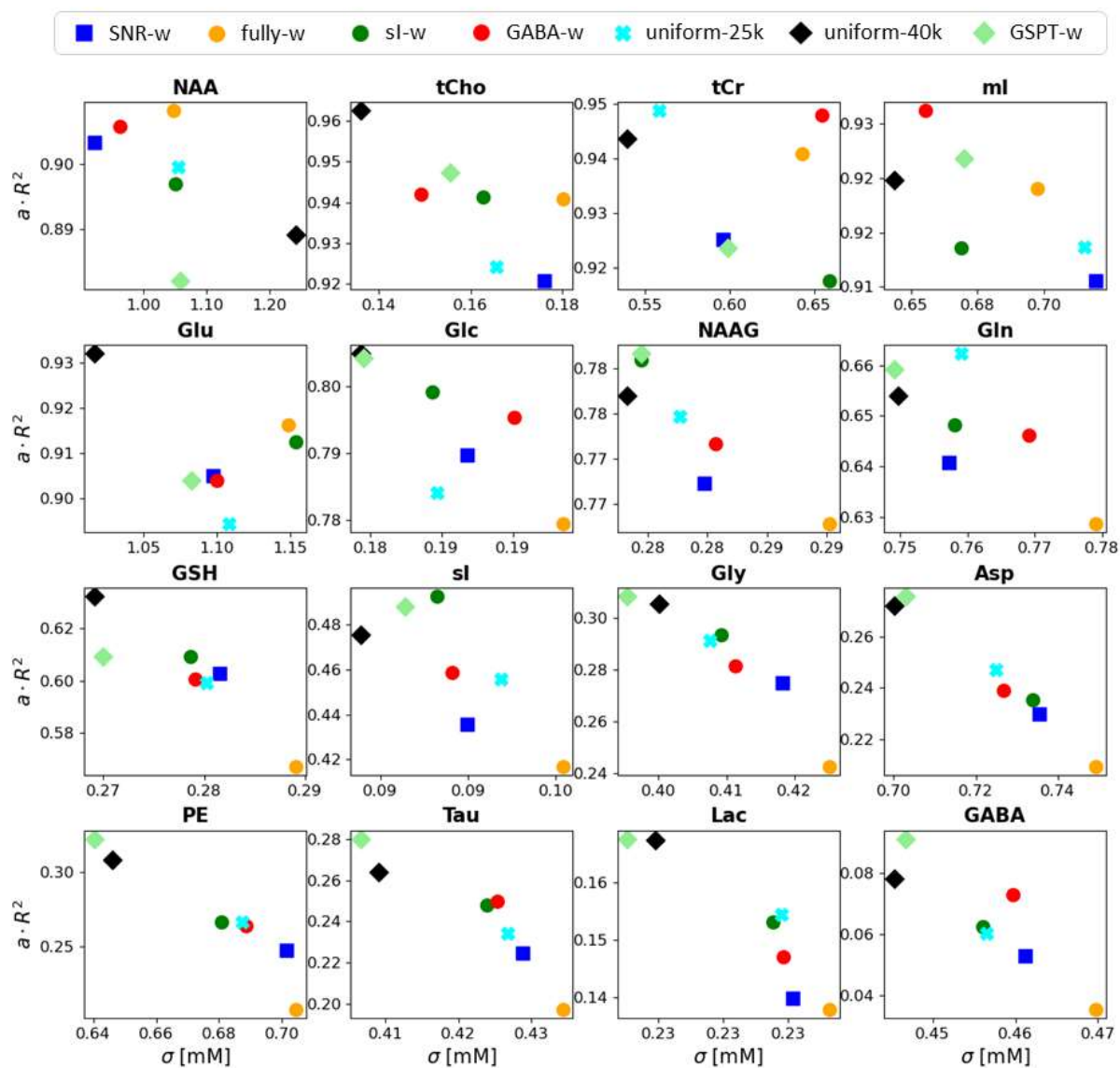
Figure S15: Comparison of outcomes of Active Learning approaches using concise scores. *Concise scores* on the same testing set of 16 metabolites for *ShallowNet-2D2c-hp* trained with different augmented datasets. Description of data augmentation can be found in Fig.3 of main text.

| met	param	training set augmentation type						
		uniform-25k	uniform-40k	SNR-w	fully-w	sl-w	GABA-w	GSPT-w
tCho	a	0.93	0.97	0.93	0.95	0.95	0.95	0.95
	q	0.13	0.05	0.17	0.09	0.08	0.11	0.05
	R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	σ[mM]	0.17	0.14	0.18	0.18	0.16	0.15	0.16
NAAG	a	0.88	0.87	0.87	0.87	0.88	0.87	0.88
	q	0.18	0.12	0.16	0.21	0.14	0.19	0.14
	R^2	0.88	0.89	0.88	0.88	0.89	0.88	0.89
	σ[mM]	0.28	0.27	0.28	0.29	0.27	0.28	0.27
NAA	a	0.91	0.89	0.91	0.91	0.90	0.91	0.89
	q	0.95	0.70	0.97	0.76	0.82	1.02	1.14
	R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	σ[mM]	1.06	1.24	0.92	1.05	1.05	0.96	1.06
Asp	a	0.50	0.52	0.48	0.45	0.49	0.49	0.52
	q	0.83	0.81	0.97	0.91	0.91	0.94	0.88
	R^2	0.49	0.52	0.48	0.47	0.48	0.49	0.53
	σ[mM]	0.72	0.70	0.74	0.75	0.73	0.73	0.70
tCr	a	0.95	0.95	0.93	0.95	0.92	0.95	0.93
	q	0.31	0.36	0.49	0.28	0.49	0.23	0.49
	R^2	0.99	1.00	0.99	0.99	0.99	0.99	1.00
	σ[mM]	0.56	0.54	0.60	0.64	0.66	0.65	0.60
GABA	a	0.27	0.29	0.24	0.19	0.27	0.31	0.33
	q	0.65	0.64	0.67	0.73	0.65	0.61	0.64
	R^2	0.23	0.27	0.22	0.18	0.23	0.24	0.28
	σ[mM]	0.46	0.45	0.46	0.47	0.46	0.46	0.45
Glc	a	0.87	0.88	0.87	0.87	0.88	0.88	0.88
	q	0.14	0.11	0.13	0.13	0.12	0.09	0.08
	R^2	0.90	0.91	0.90	0.90	0.90	0.90	0.91
	σ[mM]	0.18	0.18	0.19	0.19	0.18	0.19	0.18
Glu	a	0.92	0.96	0.93	0.95	0.94	0.93	0.93
	q	0.58	0.32	0.57	0.73	0.63	0.75	0.57
	R^2	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	σ[mM]	1.11	1.02	1.10	1.15	1.15	1.10	1.08
Gln	a	0.83	0.80	0.80	0.79	0.81	0.80	0.81
	q	0.49	0.53	0.56	0.66	0.61	0.55	0.59
	R^2	0.80	0.81	0.80	0.80	0.80	0.80	0.81
	σ[mM]	0.76	0.75	0.76	0.78	0.76	0.77	0.75
GSH	a	0.78	0.80	0.78	0.75	0.79	0.78	0.78
	q	0.21	0.18	0.23	0.26	0.22	0.23	0.23
	R^2	0.77	0.79	0.77	0.76	0.77	0.77	0.78
	σ[mM]	0.28	0.27	0.28	0.29	0.28	0.28	0.27
Gly	a	0.56	0.56	0.55	0.49	0.57	0.55	0.56
	q	0.45	0.43	0.47	0.53	0.43	0.45	0.45
	R^2	0.52	0.54	0.50	0.50	0.52	0.51	0.55
	σ[mM]	0.41	0.40	0.42	0.43	0.41	0.41	0.40
Lac	a	0.41	0.41	0.37	0.38	0.40	0.38	0.41
	q	0.28	0.30	0.33	0.31	0.31	0.33	0.31
	R^2	0.38	0.41	0.37	0.36	0.38	0.38	0.41
	σ[mM]	0.23	0.22	0.23	0.23	0.23	0.23	0.22
mI	a	0.93	0.94	0.93	0.94	0.93	0.95	0.94
	q	0.46	0.43	0.53	0.31	0.45	0.42	0.31
	R^2	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	σ[mM]	0.71	0.64	0.72	0.70	0.67	0.66	0.67
PE	a	0.53	0.55	0.50	0.44	0.52	0.52	0.57
	q	0.79	0.75	0.82	0.96	0.73	0.74	0.70
	R^2	0.51	0.56	0.50	0.47	0.51	0.51	0.56
	σ[mM]	0.69	0.65	0.70	0.70	0.68	0.69	0.64
sl	a	0.68	0.69	0.65	0.63	0.73	0.68	0.71
	q	0.09	0.09	0.10	0.09	0.08	0.09	0.08
	R^2	0.67	0.69	0.67	0.66	0.68	0.67	0.69
	σ[mM]	0.09	0.09	0.09	0.10	0.09	0.09	0.09
Tau	a	0.50	0.52	0.49	0.44	0.52	0.52	0.54
	q	0.50	0.47	0.52	0.58	0.47	0.49	0.46
	R^2	0.47	0.51	0.46	0.45	0.48	0.48	0.52
	σ[mM]	0.43	0.41	0.43	0.43	0.42	0.43	0.41

Table S2: Results of Active Learning on training set: scores of 16 metabolites for every augmented training set. The testing set is identical for every case.

met	param	Relative concentration range on test set				met	param	Relative concentration range on test set			
		[0-100]%	[20-80]%	[20-80]% & SNR in 20-40	[40-60]%			[0-100]%	[20-80]%	[20-80]% & SNR in 20-40	[40-60]%
tCho	a	0.98	0.99	1.00	0.98	Gln	a	0.79	0.85	0.92	0.88
	q	-0.02	-0.02	-0.03	0.02		q	0.52	0.37	0.17	0.28
	R^2	0.99	0.99	0.99	0.89		R^2	0.80	0.64	0.82	0.19
	σ [mM]	0.13	0.11	0.08	0.10		σ [mM]	0.77	0.66	0.44	0.60
NAAG	a	0.90	0.92	0.98	0.96	GSH	a	0.78	0.79	0.93	0.72
	q	0.14	0.11	0.03	0.05		q	0.17	0.17	0.03	0.24
	R^2	0.89	0.77	0.91	0.29		R^2	0.77	0.57	0.79	0.12
	σ [mM]	0.27	0.25	0.15	0.24		σ [mM]	0.28	0.25	0.17	0.23
NAA	a	0.97	0.96	0.97	0.96	Gly	a	0.59	0.62	0.76	0.56
	q	-0.17	-0.03	-0.07	0.10		q	0.37	0.35	0.17	0.39
	R^2	0.99	0.98	0.99	0.88		R^2	0.52	0.30	0.48	0.04
	σ [mM]	0.86	0.78	0.63	0.69		σ [mM]	0.41	0.36	0.29	0.33
Asp	a	0.56	0.64	0.76	0.66	Lac	a	0.45	0.47	0.57	0.41
	q	0.77	0.65	0.47	0.64		q	0.29	0.27	0.23	0.30
	R^2	0.50	0.30	0.49	0.05		R^2	0.39	0.21	0.31	0.02
	σ [mM]	0.72	0.62	0.50	0.57		σ [mM]	0.23	0.18	0.16	0.17
tCr	a	0.90	0.88	0.89	0.87	mI	a	0.97	0.99	0.99	0.98
	q	0.52	0.71	0.71	0.85		q	-0.25	-0.30	-0.26	-0.22
	R^2	0.99	0.98	0.99	0.89		R^2	0.98	0.96	0.98	0.73
	σ [mM]	0.79	0.65	0.56	0.53		σ [mM]	0.75	0.68	0.44	0.61
GABA	a	0.29	0.33	0.38	0.30	PE	a	0.56	0.59	0.67	0.63
	q	0.52	0.47	0.42	0.50		q	0.74	0.75	0.60	0.71
	R^2	0.22	0.12	0.20	0.01		R^2	0.53	0.30	0.51	0.05
	σ [mM]	0.47	0.37	0.33	0.32		σ [mM]	0.66	0.57	0.42	0.51
Glc	a	0.91	0.94	0.98	0.92	sI	a	0.71	0.74	0.89	0.76
	q	0.14	0.10	0.06	0.14		q	0.08	0.07	0.03	0.06
	R^2	0.90	0.81	0.91	0.35		R^2	0.67	0.44	0.66	0.09
	σ [mM]	0.19	0.17	0.12	0.16		σ [mM]	0.09	0.09	0.06	0.08
Glu	a	0.97	0.96	1.00	0.99	Tau	a	0.52	0.57	0.67	0.58
	q	0.14	0.20	-0.09	-0.03		q	0.36	0.29	0.18	0.29
	R^2	0.97	0.93	0.98	0.63		R^2	0.48	0.30	0.47	0.05
	σ [mM]	1.03	0.95	0.52	0.89		σ [mM]	0.44	0.36	0.31	0.32

Table S3: Results of emulated Active Learning on test set: scores of 16 metabolites for every concentration range considered. The training set is identical for every case.
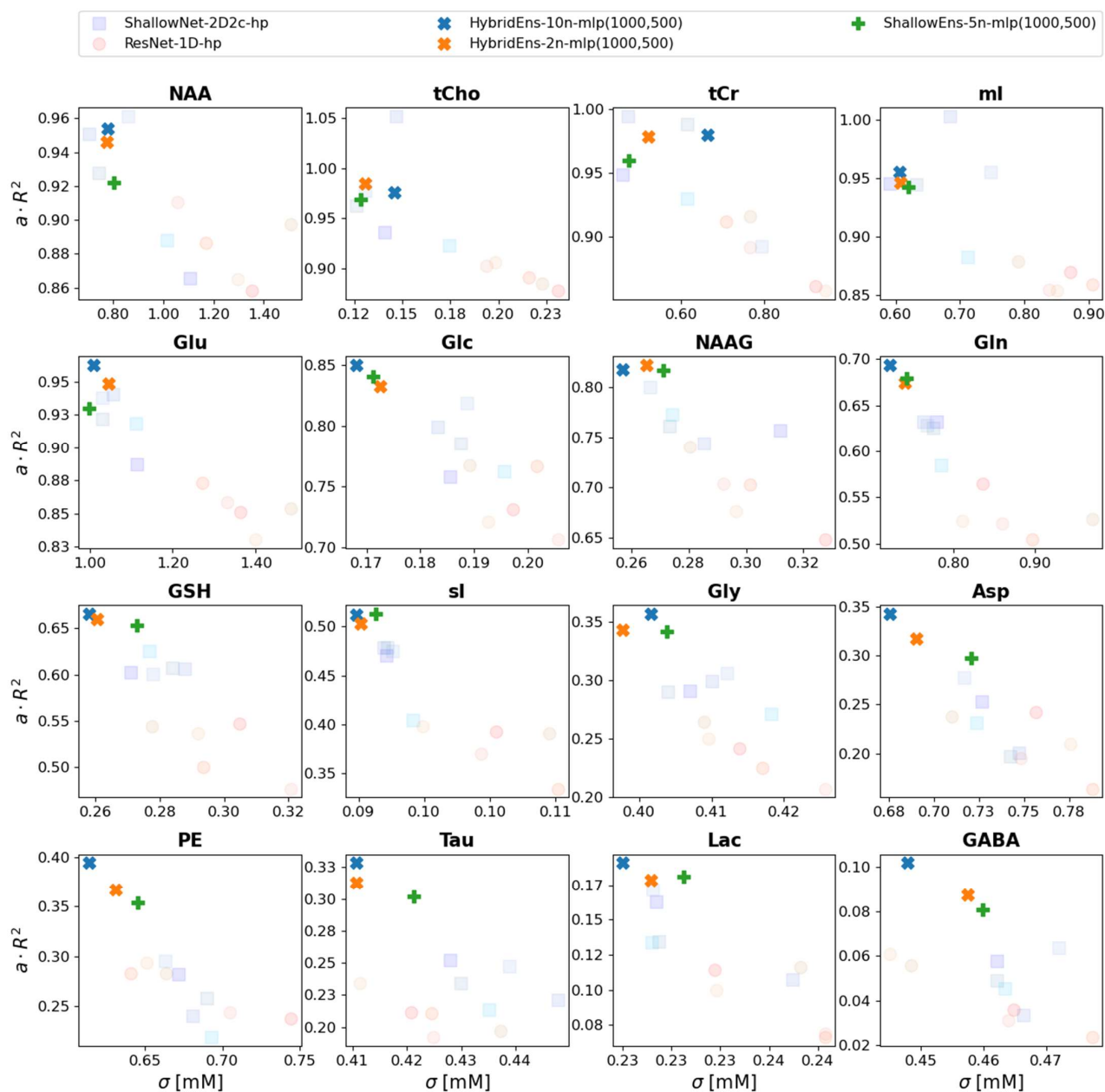
Figure S16: Quantification outcome as reflected by *concise scores* for differently trained single networks and three ensembles of networks (identical training set for 16 metabolites). Description of ensemble types to be found in Fig.4 of the main text.

| met | param | Network type and Ensemble type | | | | |
		ShallowNet-2D2c-hp-avg	ResNet-1D-avg	ShallowEns-5n	HybridEns-2n	HybridEns-10n
tCho	a	0.98	0.91	0.97	0.99	0.98
	q	0.03	0.22	0.01	0.03	-0.02
	R^2	0.99	0.99	0.99	0.99	0.99
	σ [mM]	0.14	0.21	0.13	0.13	0.15
NAAG	a	0.87	0.80	0.91	0.92	0.91
	q	0.14	0.31	0.04	0.12	0.14
	R^2	0.88	0.87	0.89	0.89	0.90
	σ [mM]	0.28	0.30	0.27	0.27	0.26
NAA	a	0.92	0.90	0.93	0.95	0.96
	q	0.65	0.91	0.76	0.75	0.70
	R^2	0.99	0.98	0.99	0.99	0.99
	σ [mM]	0.90	1.29	0.80	0.77	0.78
Asp	a	0.48	0.44	0.57	0.59	0.61
	q	0.87	0.98	0.89	0.63	0.62
	R^2	0.48	0.48	0.52	0.54	0.56
	σ [mM]	0.73	0.76	0.72	0.69	0.68
tCr	a	0.96	0.90	0.97	0.99	0.99
	q	0.28	0.80	0.16	-0.10	-0.35
	R^2	0.99	0.99	0.99	0.99	0.99
	σ [mM]	0.60	0.83	0.47	0.52	0.66
GABA	a	0.24	0.19	0.34	0.35	0.37
	q	0.67	0.71	0.56	0.55	0.54
	R^2	0.21	0.21	0.24	0.25	0.28
	σ [mM]	0.47	0.46	0.46	0.46	0.45
Glc	a	0.87	0.83	0.92	0.91	0.93
	q	0.13	0.19	0.10	0.10	0.07
	R^2	0.90	0.89	0.91	0.91	0.92
	σ [mM]	0.19	0.20	0.17	0.17	0.17
Glu	a	0.95	0.89	0.96	0.98	0.99
	q	0.39	1.39	0.66	-0.01	0.35
	R^2	0.97	0.96	0.97	0.97	0.97
	σ [mM]	1.07	1.37	1.00	1.05	1.01
Gln	a	0.78	0.69	0.84	0.83	0.85
	q	0.71	1.04	0.45	0.42	0.41
	R^2	0.80	0.77	0.81	0.81	0.82
	σ [mM]	0.77	0.88	0.74	0.74	0.72
GSH	a	0.79	0.69	0.83	0.83	0.83
	q	0.21	0.32	0.10	0.17	0.15
	R^2	0.77	0.75	0.79	0.80	0.80
	σ [mM]	0.28	0.30	0.27	0.26	0.26
Gly	a	0.56	0.47	0.63	0.63	0.65
	q	0.45	0.56	0.38	0.35	0.41
	R^2	0.52	0.50	0.54	0.55	0.55
	σ [mM]	0.41	0.42	0.40	0.40	0.40
Lac	a	0.38	0.28	0.48	0.46	0.47
	q	0.30	0.34	0.26	0.29	0.25
	R^2	0.37	0.33	0.38	0.39	0.40
	σ [mM]	0.23	0.24	0.23	0.23	0.23
mI	a	0.97	0.89	0.96	0.97	0.98
	q	0.20	0.74	0.09	0.17	0.20
	R^2	0.98	0.97	0.98	0.98	0.98
	σ [mM]	0.68	0.85	0.62	0.61	0.61
PE	a	0.51	0.49	0.63	0.63	0.66
	q	0.84	0.90	0.68	0.74	0.60
	R^2	0.50	0.54	0.56	0.59	0.60
	σ [mM]	0.68	0.68	0.65	0.63	0.61
sI	a	0.69	0.58	0.75	0.73	0.74
	q	0.09	0.12	0.06	0.08	0.07
	R^2	0.67	0.65	0.69	0.69	0.69
	σ [mM]	0.09	0.10	0.09	0.09	0.09
Tau	a	0.50	0.44	0.60	0.60	0.62
	q	0.45	0.55	0.46	0.34	0.45
	R^2	0.46	0.48	0.50	0.52	0.53
	σ [mM]	0.44	0.42	0.42	0.41	0.41

Table S4: Outcome for ensemble learning: scores for 16 metabolites for average network or ensemble of network considered.
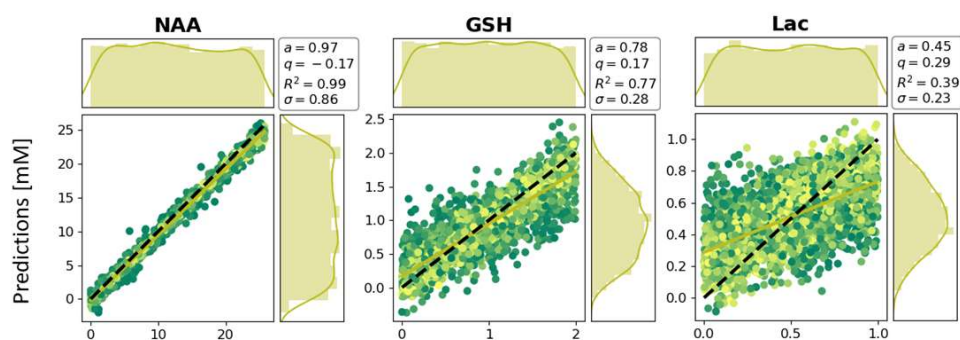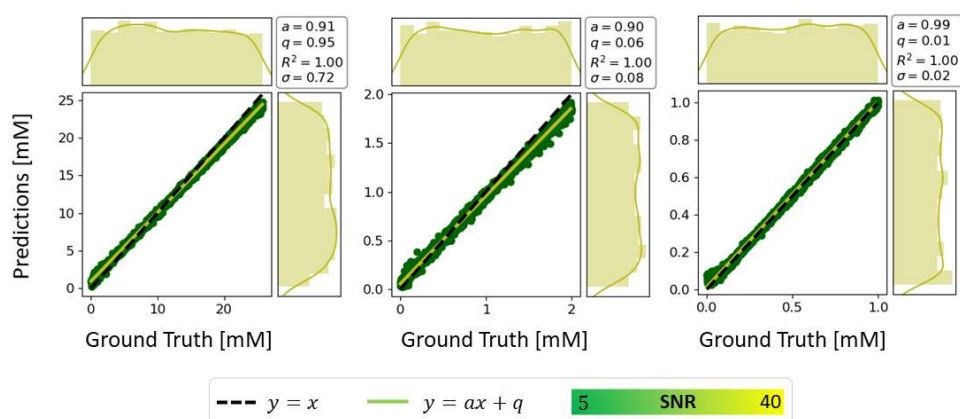
Figure S17: Maps and marginal distributions of predictions vs. GT obtained for three metabolites using *ShallowNet-2D2c-hp* as contrasted for a realistic and noiseless dataset. Predictions for the noiseless configuration in training and testing clearly show much higher accuracy and precision and sample the performance limits of the network. RMSE (σ) is reported as an overall measure of variability. A regression model ($y = ax + q$) is also provided to judge prediction quality. R^2 measures how well a linear model explains the overall data. Mis-predictions can be monitored either by a decrease in a and R^2 or by visual biases in distributions of predictions (bell-shape).

1. Hardware	
a. Field strength [T]	3 T
b. Manufacturer	N/A
c. Model (software version if available)	N/A
d. RF coils: nuclei (transmit/receive), number of channels, type, body part	N/A
e. Additional hardware	N/A
2. Acquisition	
a. Pulse sequence	Simulated Single-Voxel semi-LASER
b. Volume of Interest (VOI) locations	Brain mimicking various locations and conditions (including healthy to pathological cases) via variation in line broadening, SNR and metabolite concentrations.
c. Nominal VOI size [cm^3, mm^3]	N/A
d. Repetition Time (TR), Echo Time (TE) [ms, s]	TE 35ms TR assumed long enough to ensure full T1 recovery
e. Total number of Excitations or acquisitions per spectrum In time series for kinetic studies i. Number of Averaged spectra (NA) per time-point ii. Averaging method (e.g. block-wise or moving average) iii. Total number of spectra (acquired / in time-series)	N/A
f. Additional sequence parameters (spectral width in Hz, number of spectral points, frequency offsets) If STEAM:, Mixing Time (TM) If MRSI: 2D or 3D, FOV in all directions, matrix size, acceleration factors, sampling method	Spectral width: 4 kHz Number of spectral points: 4096 Frequency offset in simulation: -2.2ppm
g. Water Suppression Method	N/A
h. Shimming Method, reference peak, and thresholds for "acceptance of shim" chosen	N/A
i. Triggering or motion correction method (respiratory, peripheral, cardiac triggering, incl. device used and delays)	N/A
3. Data analysis methods and outputs	
a. Analysis software	Matlab: data pre-processing and analysis FitAID: fitting Python: statistics and plotting

b. Processing steps deviating from quoted reference or product	Spectra and spectrogram preparation (i.e., zero-filling, windowing, etc.)
c. Output measure (e.g., absolute concentration, institutional units, ratio)	Absolute concentration
d. Quantification references and assumptions, fitting model assumptions	Downscaled water referencing added to ease quantification by Deep Learning. No 0th or 1st order phase or frequency drifts were assumed.
4. Data Quality	
a. Reported variables (SNR, Linewidth (with reference peaks))	SNR 5-40 (time domain water referenced) Linewidth (Gaussian) 2-5Hz Simulated MMBG amplitude varying $\pm 33\%$
b. Data exclusion criteria	N/A
c. Quality measures of postprocessing Model fitting (e.g. CRLB, goodness of fit, SD of residual)	CRLBs from fitting and SD or RMSE reported for both fitting and deep learning algorithms. In depth analysis of performances via custom scores.
d. Sample Spectrum	Figure 1

Table S5: MRSinMRS checklist [4].

Supporting Information References

[1] kentsommer, "https://github.com/kentsommer/keras-inceptionV4/blob/master/inception_v4.py," Github. [Online]. [Accessed 7 2021].

[2] F Chollet, & others, *Keras,* GitHub. Available at: https://github.com/fchollet/keras, 2015.

[3] jjuinni, "https://jjuinni.medium.com/cnn-architecture-from-scratch-resnet50-with-keras-4414539521d1," 2021. [Online]. [Accessed 7 2021].

[4] Lin A, Andronesi O, Bogner W, et al. Minimum Reporting Standards for in vivo Magnetic Resonance Spectroscopy (MRSinMRS): Experts' consensus recommendations. *NMR Biomed*. 2021;34(5):e4484. doi:10.1002/nbm.4484

Reliability of Quantification Estimates in MR Spectroscopy: CNNs vs Traditional Model Fitting

Rudy Rizzo[1,2] , Martyna Dziadosz[1,2], Sreenath P. Kyathanahally[3] ,
Mauricio Reyes[4,5] , and Roland Kreis[1,2( )]

[1] MR Methodology, Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland
roland.kreis@insel.ch
[2] Translational Imaging Center, sitem-insel, Bern, Switzerland
[3] Data Science for Environmental Research Group, Department of System Analysis, Integrated Assessment and Modelling; EAWAG, Dübendorf, Switzerland
[4] Insel Data Science Center, Inselspital, Bern University Hospital, Bern, Switzerland
[5] ARTOG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

Abstract. Magnetic Resonance Spectroscopy (MRS) and Spectroscopic Imaging (MRSI) are non-invasive techniques to map tissue contents of many metabolites in situ in humans. Quantification is traditionally done via model fitting (MF), and Cramer Rao Lower Bounds (CRLBs) are used as a measure of fitting uncertainties. Signal-to-noise is limited due to clinical time constraints and MF can be very time-consuming in MRSI with thousands of spectra. Deep Learning (DL) has introduced the possibility to speed up quantitation while reportedly preserving accuracy and precision. However, questions arise about how to access quantification uncertainties in the case of DL. In this work, an optimal-performance DL architecture that uses spectrograms as input and maps absolute concentrations of metabolites referenced to water content as output was taken to investigate this in detail. Distributions of predictions and Monte-Carlo dropout were used to investigate data and model-related uncertainties, exploiting ground truth knowledge in a synthetic setup mimicking realistic brain spectra with metabolic composition that uniformly varies from healthy to pathological cases. Bias and CRLBs from MF are then compared to DL-related uncertainties. It is confirmed that DL is a dataset-biased technique where accuracy and precision of predictions scale with metabolite SNR but hint towards bias and increased uncertainty at the edges of the explored parameter space (i.e., for very high and very low concentrations), even at infinite SNR (noiseless training and testing). Moreover, training with uniform datasets or if augmented with critical cases showed to be insufficient to prevent biases. This is dangerous in a clinical context that requires the algorithm to be unbiased also for concentrations far from the norm, which may well be the focus of the investigation since these correspond to pathology, the target of the diagnostic investigation.

Keywords: Magnetic Resonance Spectroscopy · Convolutional neural networks · Model fitting · Quantification · Reliability · Uncertainties

1 Introduction

Magnetic Resonance Spectroscopy (MRS) and Spectroscopic Imaging (MRSI) are non-invasive methods for determining in-situ metabolic profile maps in humans or animals. A chemical-composition-specific response is evoked from localized tissue regions using an MRI scanner and allows the acquisition of a Voigt-damped time-domain signal, which results from a superposition of multiple metabolite signals. The resonance line patterns are metabolite-specific, reflecting the spin-systems, while their concentrations are proportional to the signal amplitude [1] (Fig. 1).
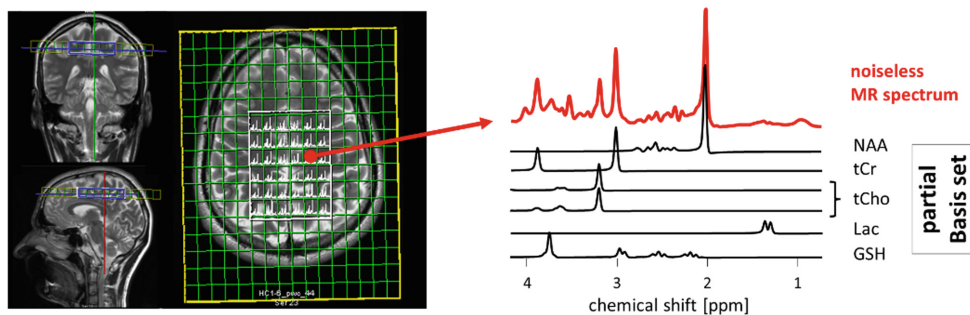
Fig. 1. MRSI acquisition with zoom-in to a sketch of a noiseless MR spectrum (real part) and relative spectral basis set outlined for five metabolites: N-Acetylaspartate (NAA), total-Creatine (tCr), total-Choline (tCho), Lactate (Lac), and Glutathione (GSH).

Quantification is traditionally based on parameter estimation with Model Fitting (MF), minimizing the difference between the data and a parameterized model function. Despite many fitting approaches [2–5], robust and accurate measurement of metabolite concentrations remains challenging [6, 7], mainly due to: (i) severely overlapping metabolite patterns, (ii) poor signal-to-noise ratio (SNR), and (iii) unknown background signals and peak lineshape (incomplete prior knowledge). As a result, the problem is ill-posed, and current techniques still hamper translation to clinical routine.

Supervised Deep Learning (DL) exploits neural networks to find key properties contained in large data sets and to generate complicated nonlinear mappings between inputs and outputs [8]. It thus requires no prior knowledge or formal assumptions. However, it is shown to be frequently biased towards the conditions prevalent in the datasets used in training [9]. DL in MRS quantification is increasingly explored [10–12] and has shown to speed up quantitation while reportedly preserving accurate estimates if compared to MF. Still, questions regarding the reliability of DL quantification have arisen.

Uncertainty measures provide information about how reliably or accurately a given algorithm performs a given task. This information in turn can be used to leverage the decision-making process for a user (e.g., how much to trust the estimated concentration of a metabolite) or to enable optimization of the acquisition technique or the algorithm employed to estimate results (e.g., focusing on areas of high uncertainty [13]). Given MRS restrictions to comply with clinical time frames, the repetition of multiple MRS measurements to determine repeatability is forbidding. Thus, estimates of uncertainty obtained from MRS model fitting of a single measurement are often taken as proxy:

the Cramer Rao Lower Bounds (CRLBs) [14] estimate uncertainties as function of the model (presumed to be true) and SNR; they represent the uncertainty limit for unbiased estimators. It is fundamental to access a CRLB-comparable uncertainty measure for MRS metabolite quantification by DL [15]. Neural network uncertainties originate from noise inherent in the data (aleatoric uncertainty) and uncertainty in the model parameters (epistemic uncertainty) [16, 17]. In the current work, an optimal-performance Convolutional Neural Network (CNN) architecture is designed to quantify metabolites, and metrics based on bias and spread of predicted distribution of concentrations are used to explain aleatoric uncertainties. Epistemic uncertainties are explored via Monte-Carlo dropout [18]. The reliability of MRS quantification is then compared between the two approaches. In-silico simulations guarantee knowledge of Ground Truth (GT).

2 Methods

2.1 Simulations

Spectral patterns were simulated for 16 metabolites recorded at 3T with a semi-LASER protocol [19, 20; TE = 35 ms, 4 kHz spectral width, 4096 points. To mimic pathological conditions, metabolite concentrations are varied independently and uniformly between 0 and twice a normal reference concentration for healthy human brain [1, 21–23]. A constant downscaled water reference (64.5 mM) is added at 0.5 ppm to ease quantitation. Macromolecular background (MMBG) signals and Gaussian broadening mimic in vivo conditions and were independently and uniformly varied (shim 2–5 Hz, MMBG amplitude ± 33%). Two datasets with 20'000 entries randomly split in training (80%), validation (10%), and testing sets (10%) are generated: one with independent, realistic white Gaussian noise realizations (time-domain water-referenced SNR 5–40), the other noiseless (Fig. 2).
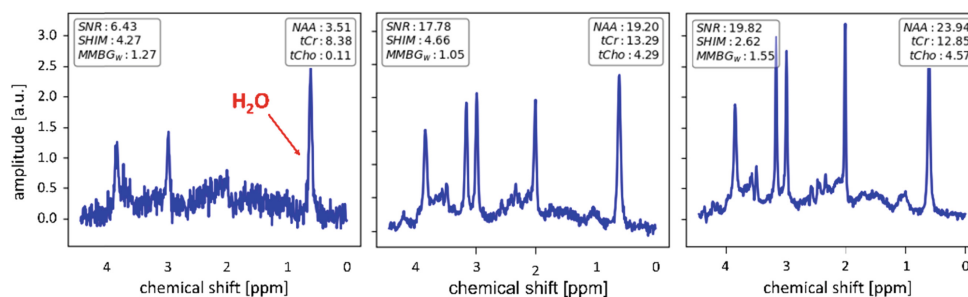
Fig. 2. Samples of realistic simulated spectra. SNR, shim, and MMBG intensity are indicated. Concentrations reported for 3 major metabolites in mM. Downscaled reference peak indicated.

High-frequency-resolved spectrograms [24] were used as input for the CNNs. A spectrogram is a complex, 2D time-frequency domain representation of a spectrum where each row reflects the frequency content of a specific time segment of the MRS signal. It is calculated using the Short-Temporal Fourier Transform (STFT), which allows for varying degrees of frequency and time resolution depending on the size of the Fourier

analysis window. A large window size established through zero-filling is paired with a tiny overlap interval to maximize frequency resolution while compromising temporal resolution (Fig. 3).

2.2 Quantification via Deep Learning

Two Bayesian hyper-parameterized [25] shallow CNNs [26] were trained and tested with the two datasets. They had emerged as optimal DL quantification methods from 24 tested scenarios with different architectures, input forms, and active learning data augmentation. Relative concentrations are provided as output but referencing to the water signal yields absolute concentrations (Fig. 3). Training and validation sets were randomly assigned to train the CNN on a maximum number of 100 epochs and with batch normalization of 50. The learning rate was modulated via the adaptive moment (ADAM) estimation algorithm [27]. Mean-squared error (MSE) served as loss function. Visualization of training and validation loss over epochs combined with the implementation of early-stopping criterion monitoring minimization of validation loss with patience ten was used as a reference for tuning the network parameter space.
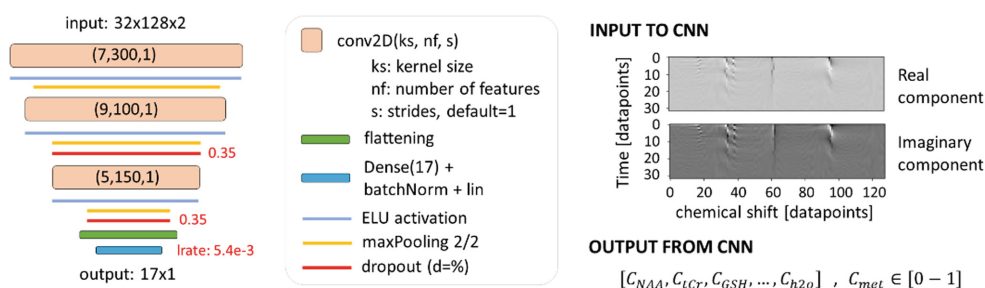
Fig. 3. Shallow CNN architecture, sample of input spectrogram, and output vector description.

Aleatoric uncertainties were evaluated via bias of the DL predictions from GT and spread of these predictions, both as estimated for 20 different bins that cover the whole GT concentration range of each metabolite (called Bin-Spread-Function in Fig. 4). Monte-Carlo dropout consisted of testing the trained model 100 times with activated dropout layers. Thus, the network structure slightly changed for each prediction (i.e., a different set of neurons was switched off) although preserving its weights. The 100 predictions yielded a distribution (called Point-Spread-Function in Fig. 4) for any sample in the test set. The bias and spread of these distributions were then calculated for every test sample, averaged for every GT value, and used as epistemic uncertainties. They highlight the susceptibility of predictions to model variation.

2.3 Quantification via Model Fitting

The test set spectra were fitted using fitAID [5]. The model consisted of a weighted sum of Voigt lines with fixed Lorentzian (GT value) and estimated Gaussian widths. Areas of the metabolites were restricted in $[-0.5 \cdot \mu, +2.5 \cdot \mu]$, where μ is the average

concentration in the testing and training set distribution for each metabolite, aiming to bound the fitting condition to known prior knowledge, mimicking the implicit restrictions of DL algorithms. Bias from GT and CRLB are used as uncertainty measures.

a) aleatoric uncertainties

$p(\hat{y}^*|\omega)$

$E[p(\hat{y}^*|\omega)]$

bin_1 | bin_2 | bin_3 | bin_4

$\hat{y}^*$

$E\left[y^*\big|_{bin}\right]$

$p(\hat{y}^*|\omega)$: Bin-Spread-Function of prediction $\hat{y}^*$ for a certain ground truth binned concentration $y^*|_{bin}$ given fixed parameterized model ω.

$$\Delta(y^*|\omega)\Big|_{bin} = E[p(\hat{y}^*|\omega)] - y^*\Big|_{bin}$$

$$\Sigma(y^*|\omega)\Big|_{bin} = \sqrt{E[\,(\hat{y}^* - E[p(\hat{y}^*|\omega)])^2\,]}\,\Big|_{bin}$$

b) epistemic uncertainties

$p(\hat{y}^*)$

$E[p(\hat{y}^*)]$

$\hat{y}^*$

y^*

$p(\hat{y}^*)$: Point-Spread-Function of prediction $\hat{y}^*$ for a certain ground truth concentration y^* under small variations of the model ω.

$$\Delta(y^*) = E[p(\hat{y}^*)] - y^*$$

$$\Sigma(y^*) = \sqrt{(E[\,(\hat{y}^* - E[p(\hat{y}^*)])^2\,]}$$

Fig. 4. Distribution of uncertainties: bias (Δ) represents a deviation from ground truth, and spread (Σ) represents variability of predictions around their center: the expected value (E[·]).
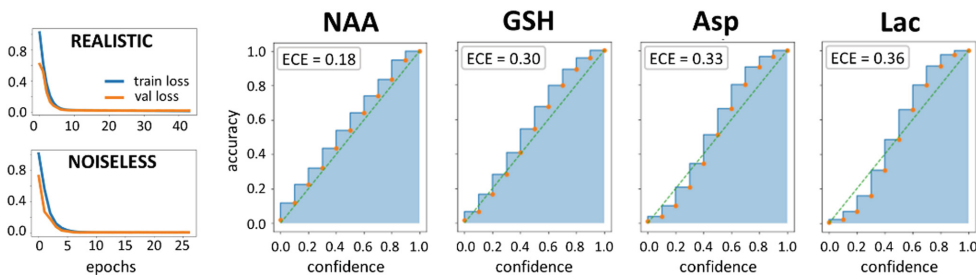
Fig. 5. Training and validation curves. Reliability diagrams and expected calibration error.

3   Results

Results are reported for four metabolites with progressively lower relative SNR: NAA, GSH, Asp, and Lac (c.f. Fig. 1). Figure 5 shows training and validation curves for both networks. Network calibration is investigated for regression, where the design is assumed to predict the Cumulative Distribution Function of relative metabolite concentrations [28, 29]. Reliability diagrams are reported for realistic simulations. Quantification of Lac and Asp is mildly overconfident for low concentrations and underconfident for high
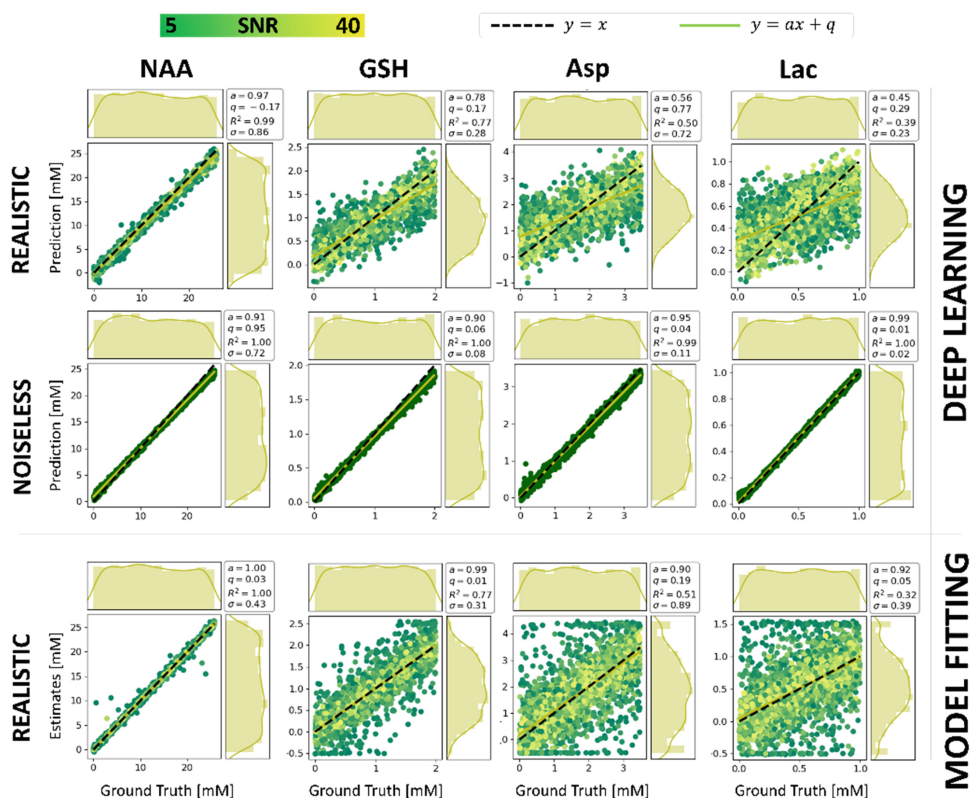
Fig. 6. Maps and marginal distribution of CNN predictions vs. GT for four metabolites and two datasets (realistic, noiseless) and model estimates vs. GT for the realistic case.

concentrations. However, the network can be considered well-calibrated for every target metabolite [13].

Figure 6 reports CNN predictions and MF estimates vs. GT values. A linear regression model is fit to estimate the quality of the prediction. Marginal distributions of GT and predicted values are displayed. Ideal predictions would display as a diagonal line ($y = ax + q$) with minimal spread (RMSE, σ). In line, distributions of predictions would mirror the uniform GT distributions. Considering the realistic case in DL and going from left to right, predicted distributions become less uniform and get more biased towards a mean expected value of the training range. This phenomenon is reflected in lower a and R^2 values and is emphasized for metabolites with low relative SNR (e.g., GSH, Asp, and Lac). Noiseless simulations show a significantly reduced bias. MF estimates show a better spread over the concentration range with $a \to 1$ and $q \to 0$ even for metabolites with low relative SNR. RMSE (σ) is lower in MF than DL for NAA but higher in the case of GSH, Asp, and Lac.

Figure 7 maps aleatoric and epistemic uncertainties as function of GT values for DL. Epistemic uncertainties indicate higher variability of predictions at the boundaries of the concentration ranges, which is paired with higher biases for aleatoric uncertainties. In the noiseless scenario, it is evident how the point-spread function is affected by a larger spread at the edges (i.e., U-shape). Training and testing with noise show the same trend if relative SNR is high enough. Model fitting appears unbiased (average bias as orange

line) except for a small effect at the parameter boundaries and biased larger outliers (blue line) for metabolites with low SNR. CRLB are concentration-independent (linear fit parameter). Moreover, average CRLBs are confirmed to represent a lower bound to standard deviation (σ) of the fit error.
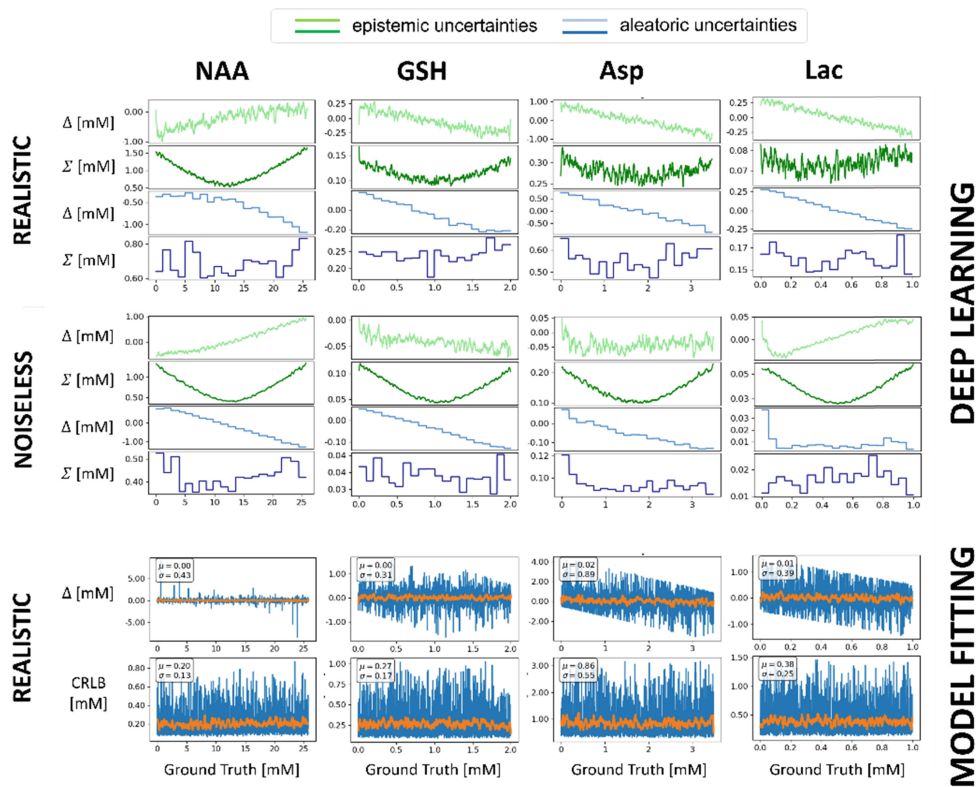
Fig. 7. CNNs' bias and spread of epistemic (green) and aleatoric (blue) uncertainties vs. GT values. MF bias and CRLBs vs. GT: single values (blue) and interpolated (orange). Estimated standard deviation (σ) of the bias can be compared to the estimated average (μ) CRLB. (Color figure online)

4 Discussion

Predicted concentrations should be unbiased, thus returning uniform distributions for uniform training and test data. However, our CNN predictions for real-world simulations tend towards the mean of the test data. Predictions at the boundaries of the testing range are folded back towards the mean value in case of strong uncertainty (i.e., low metabolite SNR), given the lack of knowledge outside the boundaries. Indeed, it is found that the prediction bias is influenced by the limited concentration ranges used in training: epistemic uncertainties indicate higher variability of predictions at the boundaries of the concentration ranges, which is paired with higher biases for aleatoric uncertainties. Though not exploring the exact same architecture, it is suspected that previous DL MRS

approaches may show similar deficiencies [10–12]. An ideal noiseless scenario also shows similar findings: uncertainty is lower in absolute values compared to the realistic case, but a similar dependency on the concentration range is found (U-shape), confirming that the DL prediction constitutes a biased estimator with uncertainties that depend on the placement of the test case in the training range. Training and testing with noise show the same trend if relative SNR is high enough. Metabolites in high concentrations suffer from comparable epistemic spread as those in low concentrations.

MF is confirmed to be unbiased on average. Individual estimates are mildly influenced by restrained concentration ranges (i.e., prior knowledge). CRLBs are confirmed to be a measure of variance that is independent of the estimated concentration.

The current study considers a limited synthetic dataset that does not cover the whole range of possible in-vivo sources of variability despite its aim to mimic realistic performances. Furthermore, a single CNN design tuned for metabolite's quantification is investigated, even if optimized via multiple iterations and with the best combination of input/output spectroscopic information (i.e., spectrograms and relative concentrations), it is not possible to draw general conclusion for MRS quantification via DL algorithms. Uncertainty investigation is limited to two uncertainty measures, which must be taken with their benefits, reliability, and limitations compared to other measures [13].

5   Conclusions

Four measures for aleatoric and epistemic uncertainties are provided, partly representing accuracy and precision of predictions. They scale with metabolite SNR but hint towards bias and increased uncertainty at the edges of the explored parameter space for (these) DL methods in many cases, even at infinite SNR.

Deep Learning does not require feature selection by the user, but the potential intrinsic biases at training set boundaries act like soft constraints in traditional modeling, leading estimated values to an apparently precise (low mean deviation) estimate reflecting an expectation value over the normal concentration range used in training. This is dangerous in a clinical context that requires the algorithm to be unbiased to outliers, which may well be the focus of the investigation corresponding to pathological data.

Further investigation to access more stable predictions is needed: (i) training with even larger concentration ranges, such that the region of interest is well inside the training range where uncertainties are limited, (ii) consider ensemble of networks to strengthen network performances for outliers or (iii) implementation of Batch Nuclear-norm Maximization to improve discriminability and diversity of the predictions [30].

6   Data Availability Statement

The simulated datasets and network architecture that support the findings of this study are available at https://github.com/bellarude.

References

1. de Graaf, R.A.: In Vivo NMR Spectroscopy: principles and techniques, 3rd ed., WILEY (2018)
2. Ratiney, H., et al.: Time-domain semi-parametric estimation based on a metabolite basis. NMR Biomed. **18**, 1–13 (2005)
3. Provencher, S.: Estimation of metabolite concentrations from localized in vivo. Magn. Reson. Med. **30**(6), 672–679 (1993)
4. Wilson, M., et al.: A constrained least-squares approach to the automated quantitation of in vivo 1h magnetic resonance spectroscopy data. Magn. Reson. Med. **65**(1), 1–12 (2011)
5. Chong, D.G.Q., et al.: Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets. Magn. Reson. Mater. Physics, Biol. Med. **24**, 147–164 (2011)
6. Bhogal, A.A., et al.: 1H-MRS processing parameters affect metabolite quantification: the urgent need for uniform and transparent standardization. NMR in Biomed. **30**, e3804 (2017)
7. Marjanska, M., et al.: Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. Magn. Reson. Med. **87**(1), 11–32 (2022)
8. Wick, C.: Deep Learn. Informatik-Spektrum **40**(1), 103–107 (2016). https://doi.org/10.1007/s00287-016-1013-2
9. Gyori, N.G., et al.: Training data distribution significantly impacts the estimation of tissue microstructure with machine learning. Magn. Reson. Med. **87**(2), 932–947 (2022)
10. Lee, H.H., et al.: Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain. Magn. Reson. Med. **84**(4), 1689–1706 (2020)
11. Gurbani, S.S., et al.: Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. Magn. Reson. Med. **81**, 3346–3357 (2018)
12. Hatami, N., Sdika, M., Ratiney, H.: Magnetic resonance spectroscopy quantification using deep learning. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Lecture Notes in Computer Science, vol. 11070, pp. 467–475. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_53
13. Jungo, A., et al.: Assessing Reliability and Challenges of Uncertainty Estimations for Medical Image Segmentation, arXiv:1907.03338v2. (2019)
14. Bolliger, C.S., et al.: On the use of Cramér-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments. Neuroimage **83**, 1031–1040 (2013)
15. Landheer, K., et al.: Are Cramer-Rao lower bounds an accurate estimate for standard deviations in in vivo magnetic resonance spectroscopy? NMR Biomed. **34**(7), e4521 (2021)
16. Gal, Y.: Uncertainty in Deep Learning, University of Cambridge (2016)
17. Kendall, A.: What uncertainties do we need in Bayesian deep learning for computer vision? arXiv:1703.04977v2. (2017)
18. Gal, Y. et al.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning, arXiv:1506.02142v6. (2016)
19. Soher, B.J., et al.: VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. Proc. Int. Soc. Magn. Reson. Med. **19**(19), 1410 (2011)
20. Oz, G., et al.: Short-echo, single-shot, full-intensity proton magnetic resonance spectroscopy for neurochemical profiling at 4 T: validation in the cerebellum and brainstem. Magn. Reson. Med. **65**(4), 901–910 (2011)
21. Marjańska, M., et al.: Region-specific aging of the human brain as evidenced by neurochemical profiles measured noninvasively in the posterior cingulate cortex and the occipital lobe using 1H magnetic resonance spectroscopy at 7 T. Neuroscience **354**, 168–177 (2017)

22. Hoefemann, M., et al.: Parameterization of metabolite and macromolecule contributions in interrelated MR spectra of human brain using multidimensional modeling. NMR Biomed. **33**(9), e4328 (2020)
23. Oz, G., et al.: Clinical proton MR spectroscopy in central nervous system disorders. Radiology **270**(3), 658–679 (2014)
24. Kyathanahally, S.P., et al.: Deep Learning approaches for detection and removal of ghosting artifacts in MR Spectroscopy. Magn. Reson. Med. **80**, 851–863 (2018)
25. Snoek, J.: Practical Bayesian otimization of Machine Learning Algoirthms. In: 25th International Conference on Neural Information Processing System, vol. 2, pp. 2951–2959 (2012)
26. Espi, M., et al.: Exploiting spectro-temporal locality in deep learning based acoustic event detection. J. Audio Speech Music Proc. **26** (2015)
27. Kingma, D.P., et al.: Adam: A method for stochastic optimization. Arxiv:1412.6980. (2014)
28. Niculescu-Mizil, A., et al.: Predicting good probabilities with supervised learning. In: 22nd ICML, pp.7–11 (2005)
29. Kuleshov, V., et al.: Accurate uncertainties for deep learning using calibrated regression. In: 35th ICML (2018)
30. Cui, S., et al.: Towards discriminability and diversity: batch Nuclear-norm Maximization under label insufficient situations. Arxiv:2003.12237v1. (2020)

7 | Preliminary published material

7.1. Simultaneous quantification of metabolite-specific concentrations and T_2 relaxation times

References[1]:

Rizzo R & Kreis R.

Multi-Echo Single-Shot Spectroscopy combined with simultaneous 2D model fitting for fast and accurate measurement of metabolite-specific concentrations and T_2 relaxation times

NMR Biomed, under review. Submitted on 24.02.2023.

Rizzo R, Stamatelatou A, Heerschap A, Scheenen T, Kreis R.

Simultaneous concentration and T_2 mapping of brain metabolites by multi-echo spectroscopic imaging

Proceedings of 32nd Annual Meeting of the International Society of Magnetic Resonance in Medicine - ISMRM, 2023: 03-08 June, Toronto, CA

Contribution to the studies

MR sequence development. Software development for postprocessing pipeline to translate scanner raw data to usable data. Data acquisition on phantom and volunteers, data postprocessing, and analysis. Model tuning and fitting. Simulations and in-silico analysis. Manuscripts drafting.

---

[1] The reference to the ISMRM abstract is a modified and extended version of the correspondent accepted work

Multi-Echo Single-Shot Spectroscopy combined with simultaneous 2D model fitting for fast and accurate measurement of metabolite-specific concentrations and T$_2$ relaxation times

Rudy Rizzo[1,2,3,4], Roland Kreis[1,2,3]

[1] *MR Methodology, Department for Diagnostic and Interventional Neuroradiology, University of Bern, Switzerland*

[2] *Department for Biomedical Research, University of Bern, Switzerland*

[3] *Translational Imaging Center (TIC), Swiss Institute for Translational and Entrepreneurial Medicine (sitem- insel), Bern, Switzerland*

[4] *Graduate School for Cellular and Biomedical Sciences, University of Bern, Switzerland*

Word count: 5402

Running head: Multi-Echo Single-Shot Spectroscopy combined with simultaneous 2D modeling

Corresponding author:
Prof. Dr. sc. nat. Roland Kreis,
MR Methodology, University Bern,
Freiburgstr. 3,
CH-3010 Bern, Switzerland
Tel: +41-31-632 8174
Email: roland.kreis@insel.ch

Abstract (262/300)

Purpose

To develop a novel single-voxel MR spectroscopy acquisition scheme to simultaneously determine metabolite-specific concentrations and transverse relaxation times within realistic clinical scan times.

Methods

Partly truncated multi-TE data are acquired as echo train in a single acquisition (Multi-Echo Single-Shot – MESS). A 2D multiparametric model fitting approach combines truncated, low-resolved short TE data with fully sampled, highly-resolved, longer TE data to yield concentration and T_2 estimates for major brain metabolites simultaneously. Cramer-Rao Lower Bounds (CRLBs) are used as a measure of performance. The novel scheme was compared with traditional Multi-Echo Multi-Shot (MEMS) methods. In-silico, in-vitro, and in-vivo experiments support the findings.

Results

MESS schemes requiring only 2:12 minutes for the acquisition of three echo times provide valid concentration and relaxation estimates for multiple metabolites and outperform traditional methods for simultaneous determinations of metabolite-specific T_2s and concentrations, with improvements ranging from 5-30% for T_2s and 10-50% for concentrations. However, substantial unsuppressed residual water signals may hamper the method's reproducibility, as observed in an initial experiment setup that prioritizes short TEs with severely truncated acquisition for the benefit of SNR. Nevertheless, CRLBs have been confirmed to be well-suited as design criteria, and within-session repeatability approaches CRLBs when residual water is removed in postprocessing by exploiting longer and less truncated data recordings.

Conclusions

Multi-Echo Single-Shot MRS combined with 2D model fitting promises comparable accuracy, increased precision, or inversely shorter experimental times when compared to traditional approaches. However, the optimal design must be investigated as a trade-off between SNR, the truncation factor, and TE batch selections, all influencing the robustness of estimations.

Keywords (8/8)

1. INTRODUCTION

Clinical magnetic resonance spectroscopy (MRS)[1] usually lacks subject-specific relaxation times needed for full quantification of spectra from single subjects.[2–4] In fact, to derive absolute concentrations of metabolites, two primary ingredients are needed: (i) reference data for absolute scaling (e.g., typically, the unsuppressed internal water signal is used as reference[5] together with segmentation results to define tissue composition) and (ii) relaxation correction factors for the metabolites of interest.[6,7] Furthermore, individually determined metabolite T_2 times may be of interest as such, potentially reflecting the cellular and sub-cellular microenvironment and thus providing insights into evolving pathological or physiological processes.[8]

Methods for a comprehensive evaluation of concentrations and relaxation times (and macromolecular background - MMBG) have been proposed that are based on combined evaluations of multiple different acquisitions (e.g., different TE, TR, or inversion times),[9–13] but given the additional scan time needed, are not in widespread use. Instead, clinical MRS relies on tabulated values that are listed for a few brain tissue types or cerebral locations, which do not account for inter-subject variability and are often lacking for specific pathologies. Such assumptions, therefore, introduce quantification errors and biases.[14]

Multi-echo data from a single acquisition – where a CPMG sequence may be deployed to prolong the range of high SNR to longer acquisition times – can be used as an alternative with the benefit of speed, though at the expense of spectral resolution. Furthermore, it introduces gaps in the acquired data where RF pulses and gradient crushers are applied,[15,16] which can partly be recovered by use of prior knowledge gained from a full resolution short TE spectrum.[16]

Similarly, multi-TE data is widely used in spectroscopic imaging. Like in RARE[17], multiple echo acquisitions can be used to cover k-space in different echo periods and speed the coverage of the full range.[18,19] Echo-planar spectroscopic imaging has proven similar speed gains with the acquisition of five echoes enabling T_2 quantification of three primary brain metabolites.[20] Alternatively, sparse multi-TE sampling with spatio-spectral encoding and subspace modeling has been suggested for mapping TE-dependent molecular changes within a clinical time frame.[21]

In this work, we propose a novel acquisition scheme of acquiring multi-TE data in a single acquisition to be used in a combined fitting process, where the half echo of the shortest TE is fitted with the full echoes recorded for later TEs, including the extended tail of the last echo that provides resolution information for the whole echo train.[22] The multi-TE combined fit enables the simultaneous estimation of concentrations and T_2s, deploying one single model equation. The primary goal of this study was to prove the feasibility of simultaneous quantification of metabolite concentrations and metabolite-specific T_2s by multi-echo single-shot (MESS) spectroscopy in a clinically compatible time frame and to evaluate its performance against the traditional acquisition approach with separately recorded multiple echo spectra.[23]

2. METHODS

2.1 Measurement setup

2.1.1 MESS acquisition scheme

A semi-LASER localization scheme[24,25] is complemented with a multi-echo single-shot design to increase sensitivity to T_2 relaxation, Fig. 1. Three subsequent spin-echo readouts are realized by adding two slice-selective Mao π pulses[26], which create the 2nd and 3rd recorded spin-echo after the 1st recorded spin-echo resulting from the semi-LASER adiabatic localization (as usual in semi-LASER, the first three spin-echoes cannot be used and are not recorded). The time between the additional RF pulses is used for signal acquisition characterized by a user-defined parameter Δ that (i) determines the echo timings and (ii) defines the truncation factor (i.e., the duration of the readouts for the 1st (TE1) and 2nd (TE2) recorded echo vs. the time needed for a complete echo sampling). Data for TE1 is acquired as FID (free induction decay), while TE2 and TE3 are acquired as partially sampled full echoes (partial-echo sampling - PES), where the last recording window extends to achieve an overall one second of total acquisition time (sum of all three readouts).[27]

The Mao pulses are optimized such that the slice thickness (ST) is large enough to ensure full refocusing of the adiabatically pre-selected volume of interest of the semi-LASER block (specifically, ST_{MAO} = 1.5x $ST_{semi-LASER}$). Moreover, a dedicated crusher scheme was optimized with phase cycling to minimize unwanted coherence pathways.[28] In addition, 16-step EXORCYCLE phase cycling is implemented to minimize spurious contributions of unwanted coherence pathways, though, of course, full phase-cycling considering all seven RF pulses is not possible.

Metabolite cycling (MC) is applied to simultaneously acquire water and metabolite data.[29]

2.1.2 Comparison to conventional acquisitions

Conventional acquisitions to assess the performance of the novel methodology were considered in different setups, Fig. 2.

- Multi-Echo Multi-Shot (MEMS) acquisitions consisted of separate fully sampled acquisitions of the three spin-echo experiments realized with multiple refocusing as with the MESS technique. This acquisition is evaluated with either (i) half-echo sampling (HES) or (ii) PES for TE2 and TE3.
- A single echo (SE) measurement is recorded at the shortest TE.
- A multi-TE measurement of standard semi-LASER (i.e., without additional refocusing pulses) at the corresponding TEs of MESS and MEMS.

2.1.3 In-silico data

MESS spectroscopy was investigated compared to short TE SE and MEMS schemes (HES and PES). Spectra of a metabolite mixture[2] specified in Table 1 were simulated at 3T with VeSPA[30] assuming 5 Hz Gaussian linewidth, 4 kHz spectral width, and 4096 data points. In addition, T_2s were set independently for the following molecules and molecule moieties: choline, split into Cho_{CH3} (trimethylammonium protons) and Cho_{REST}; creatine, split into Cr_{CH3}

(methyl singlet) and Cr_{CH2} (methylene singlet); glutamate (Glu); lactate (Lac); myo-inositol (mI) and N-acetylaspartate, split into NAA_{CH3} (acetyl group) and NAA_{REST} (aspartate group). A PRESS block implemented with ideal pulse rotations and crusher gradients replaced the semi-LASER block in Fig. 1. Three acquisition schedules with different Δ and TEs were investigated, Table 1. Ten spectra per schedule were simulated in a Monte-Carlo setup implementing random white Gaussian noise with fixed SNR (SNR=210, defined in time domain and referenced to the first point of the FID).

Metabolite		Concentration [mM]	T_2 [ms]
Creatine CH_2	Cr_{CH2}	10	129
Creatine CH_3	Cr_{CH3}		166
Choline Methylene	Cho_{REST}	3	129
Choline Trimethyl	Cho_{CH3}		166
Glutamate	Glu	12.5	124
Lactate	Lac	5	159
Myo-Inositol	mI	7.5	161
NAA Acetyl	NAA_{CH3}	12.5	343
NAA Aspartate	NAA_{REST}		310

Δ [ms]	TE1 [ms]	TE2 [ms]	TE3 [ms]
10	25	53	81
20	25	73	121
30	25	93	161

Table 1: Parameters for Monte-Carlo simulations. (top) Mixture of metabolites: concentrations and T_2 relaxation times. (bottom) Sequence timings: acquisition window (Δ) and echo times.

2.1.4 In-vitro measurements

MESS spectroscopy was investigated compared to MEMS schemes (PES) and traditional semi-LASER acquisitions recorded for different echo times. The multi-echo MESS and MEMS experiment consisted of two separate measurements with short and long TE batches:

- TE batch #1: TEs = 35/60.6/86.2 ms, with Δ = 8 ms (32 datapoints);
- TE batch #2: TEs = 140/213.6/287.2 ms, with Δ = 32 ms (128 datapoints);

Other parameters included TR: 2000 ms, spectral width: 4 kHz, 64 acquisitions, 2:12 min duration to acquire one TE in MEMS configuration or the whole 3 TE batch in MESS configuration. Spectra were acquired on a 3T MR scanner (Prisma, Siemens) with a 64-channel receive head coil for a "Braino" phantom (GE Medical Systems) containing an aqueous solution of the characteristic brain metabolites: mI (5.0 mM), NAA (12.5 mM), Glu (12.5 mM), Cr and creatinine (Cn, total of 10.0 mM), Lac (5.0 mM), and Cho (3.0 mM).

MESS, MEMS, and multi-TE approaches were investigated and compared for the three TEs batch #1 by itself (maximizing SNR for TE3) and for a six TEs setup, which combines TE batches #1 and #2 (balancing short TEs with high SNR and long TEs but higher resolution together with a better definition of the T_2 fit based on six data points rather than three).

2.1.5 In-vivo measurements

As detailed in Fig. 1, MESS spectroscopy was investigated compared to the MEMS scheme (PES). The multi-echo experiment consisted of two measurements with short and long TE batches with identical acquisition and investigation setup as in the in-vitro case. Ten healthy subjects (age 34 ± 12 years, 2 males, 8 females) were examined with the single voxel placed in the occipital cingulate cortex (OCC) with a size of $20x31x17mm^3$. Identical hardware and parameter setup was used as described above. In addition, five outer-volume suppression

bands were added for lipid and nasal cavity artifact suppression.[29] . The TE batch #2 was acquired only in five of the ten subjects. All scans were performed in accordance with the competent ethical review boards. Statistical evaluations were run with GraphPad Prism 9.4. Distributions of estimates were compared assuming *unpaired Welch t-tes*t with *Holm-Sidak* correction.
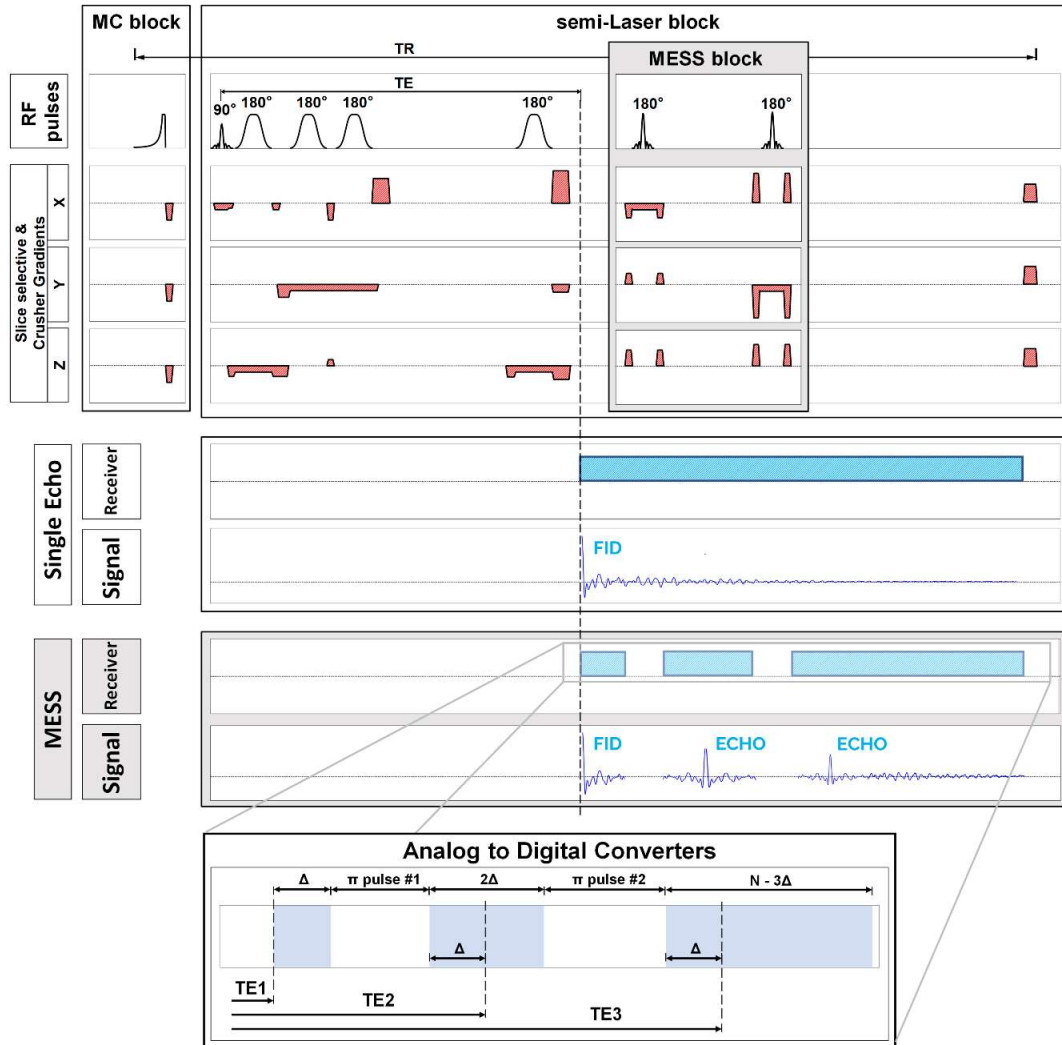
Figure 1: Acquisition scheme: standard semi-LASER localization (white background) and Multi-Echo Single-Shot (MESS) acquisition (gray background). (bottom) Zoom-in for dedicated echo timing: the parameter Δ (acquisition window) determines the overall sequence timing once TE1 is selected with semi-LASER localization. Maximized inter-pulse sampling is deployed: when TE1 is the shortest possible, MESS acquisitions are sampled as FID for TE1 (1st recorded echo) and as a partial echo for TE2 and TE3 (2nd and 3rd recorded echo, respectively).

Multi-Echo Single-Shot (MESS)
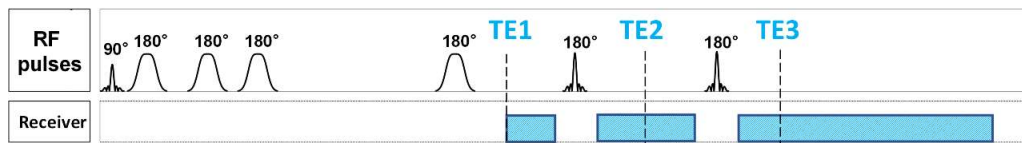
Multi-Echo Multi-Shot (MEMS): PES and HES
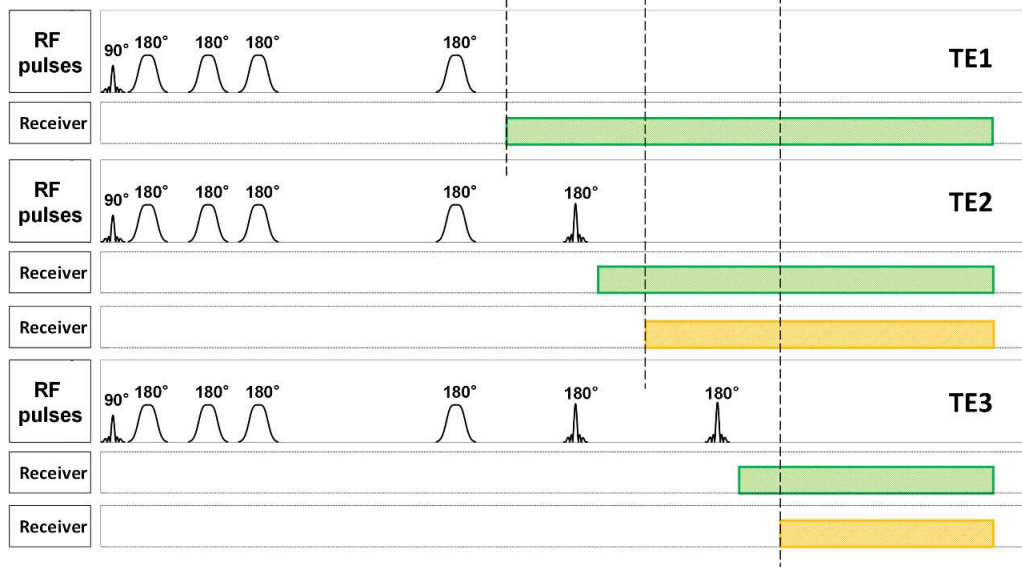
Figure 2: Schematic of the various acquisition schemes that are compared in the study. (Top, blue) Multi-Echo Single-Shot: three subsequent acquisitions at three different TEs with partly truncated sampling range, as reported in Fig.1. (Bottom) Multi-Echo Multi-Shot: three sequential fully sampled acquisitions at three different TEs. The spin-echo experiment is recorded as FID for TE1 and either as partial-echo (PES: Partial Echo Sampling, green) or FID (HES, Half Echo Sampling, orange). In-silico measurements exploit a PRESS module as the basis, instead of a semi-LASER module, and include a single short-TE PRESS experiment compared to the multi-echo acquisitions. In-vitro measurements compare MESS and MEMS with traditional multi-TE semi-LASER acquisitions (without Mao pulses for multiple refocusing), which are not displayed.

1.1.1 Repeatability

MESS acquisitions were repeated three times within the same session with the same 3-TE and 6-TE configurations for a cohort of five healthy subjects. For further exploratory measurements, see Texts S2 and S3.

1.2 Postprocessing

1.2.1 MESS

A dedicated postprocessing pipeline was developed in MATLAB and combined with the FID-A library[31] to treat raw measurement data (Siemens twix). Single coil channels from the three readouts were concatenated and optimally combined (weighting with the ratio of signal to noise squared).[32] Concatenating the three readouts implied using the same noise for all three readouts as determined from the signal-free end of the 3rd readout. Spectra

were automatically aligned and phase-corrected, then metabolite and water signals were separated, followed by eddy current-correction. Residual water removal using Hankel-Lanczos Singular Value Decomposition (HLSVD) filtering was possible only for the 3rd readout, where the number of data points is sufficient for the algorithm to untangle Lorentzian lines. However, TE1 and TE2 readouts were not amenable for HLSVD filtering, partly because difference signals from MC cannot be well decomposed into Lorentzian lines. 1st and 2nd readouts were zero-filled to match the number of data points acquired for the 3rd readout – needed for the input to the fit software, but only relevant for data displays, not for fitting or CRLB calculation, because the fit model base spectra had been treated identically (verified in simulations).

1.2.2 Traditional acquisitions

Processing was analogous to the MESS case, except that noise was estimated in each spectrum separately, and HLSVD filtering was performed for all TEs. Subsequent processing was done in JMRUI[33] in both cases.

1.3 Simultaneous 2D model fitting

Simultaneous 2D fitting was run in FitAID[34] with a time-domain model and χ^2 minimization in the frequency domain of the 1st dimension with simultaneous T_2 fit via mono-exponential damping in the 2nd dimension. Lineshape modeling was based on Voigt lines where the Lorentzian component is determined by T_2 times, and the Gaussian component represents any other peak broadening. Gaussian broadening is assumed to be identical for all metabolites. Further details on the fitting procedure are given in Text-S1. Estimated Cramer Rao Lower Bounds (CRLBs) were taken as the measure for achievable precision. The MESS was compared to the MEMS scheme for equivalent total experimental time. Therefore CRLBs of MEMS were adjusted by multiplication with $\sqrt{3}$, since a MEMS acquisition takes three times longer than MESS. Likewise, CRLBs for 6-TE experiments were adjusted by multiplication with $\sqrt{2}$.

1.3.1 In-silico experiments

One 2D model with three TEs was evaluated for MESS, MEMS/HES as well as MEMS/PES. Distinct T_2s were fitted for all metabolites.

1.3.2 In-vitro experiments

The basis set was simulated in VeSPA using the actual RF pulses but ignoring the slice selection gradients. T_2s were fitted and estimated independently for all metabolites considering three 2D models with three or six TEs each: (i) MESS, (ii) MEMS/PES, and (iii) semi-LASER multi-TE.

1.3.3 In-vivo experiments

The basis set was simulated for sixteen metabolites: Asp: aspartate, GABA: γ-aminobutyric acid, Glc: glucose, Gln: glutamine, Glu: glutamate, Gly: glycine, GSH: glutathione, Lac: lactate, mI: myo-inositol, NAA: N-acetylaspartate, NAAG: N-acetylaspartylglutamate, PE: phosphoethanolamine, sI: syllo-inositol, Tau: taurine, tCho: total choline (1:1 sum of glycerophosphorylcholine + phosphorylcholine), tCr: total creatine (1:1 sum of creatine + phosphocreatine), .

The macromolecular background (MMBG)[35] pattern was simulated as the sum of overlapping densely and equally spaced Voigt lines[36] and tailored to the acquired signal via fitting of a sub-cohort average of three subjects where the eleven groups modeling the prominent resonance peaks were fitted for single amplitude and Lorentzian broadening (i.e., Gaussian broadening kept fixed).

The fit model included metabolite-specific T_2s for five major moieties (Cr_{CH3}, Cr_{CH2}, Cho_{CH3}, NAA_{CH3} and mI).[37] A *global* T_2 was fitted for all other moieties. The residual water signal that could not be eliminated with HLSVD filtering was modeled as a singlet for TE1 and TE2 in the three TEs model and TE1, TE2, TE4, and TE5 in the six TEs model.

Two 2D models were evaluated with either three TEs (10 subjects) or six TEs (5 subjects). Absolute tissue concentrations were estimated referencing to water, including tissue-specific water relaxation and CSF partial volume correction. Metabolite concentrations were T_2 corrected within the FiTAID model, whereas T_1 corrections used values reported in the literature.[36]

Figure 3: Simulated spectra (real part) with Δ = 20 ms for three acquisition schemes. (A) MESS scheme (with TE1 and TE2 zero-filled for display purposes), (B) MEMS scheme sampled as HES, (C) MEMS scheme sampled as PES. The difference between HES with PES is evident in the 1st order phase feature introduced by sampling before the echo maximum in PES. MESS's truncated sampling for TE1 and TE2 is reflected in a low frequency resolution. However, MESS TE3 is comparable to PES TE3.

2. RESULTS

2.1 In-silico evaluations

Fig. 3 illustrates instances of the simulated spectra according to the three different methodologies (MESS, MEMS with HES, and MEMS with PES) for Δ=20 ms. As expected, the short TE spectra in MESS (Fig. 3A) show minimal resolution, with data acquisitions lasting only 20 ms (80 data points) and 40 ms (160 data points) for TE1 and TE2, respectively. In addition, MESS and PES spectra feature a linear phase shift according to the early start of data acquisition. However, all these frequency-domain issues are adequately represented in the time-domain model.

Fig. 4A shows fit uncertainties in the form of CRLBs, and standard deviations (SDs) over different noise realizations for the determined T_2s and concentrations as a function of the acquisition scheme averaged over different sequence timings (i.e., Δ settings according to

Table 1). A clear trend is evident: $CRLB_{MESS} < CRLB_{MEMS\ (PES)} < CRLB_{MESS\ (HES)}$, and this is confirmed by the SDs . However, a single short TE measurement (SE) yields the lowest CRLBs when only concentrations are investigated (i.e., prior knowledge on T_2 is kept fixed to assumed ground truth values).

Fig. 4B reports uncertainties for T_2s and concentrations as averaged over the metabolite set but varying as a function of measurement schemes and sequence timings. The precision for T_2 is best for a sequence timing with Δ=20 ms. The ranking for concentrations is less clearly influenced by the timings. Single short-TE measurements are confirmed to outperform MESS in terms of precision to estimate concentrations only.
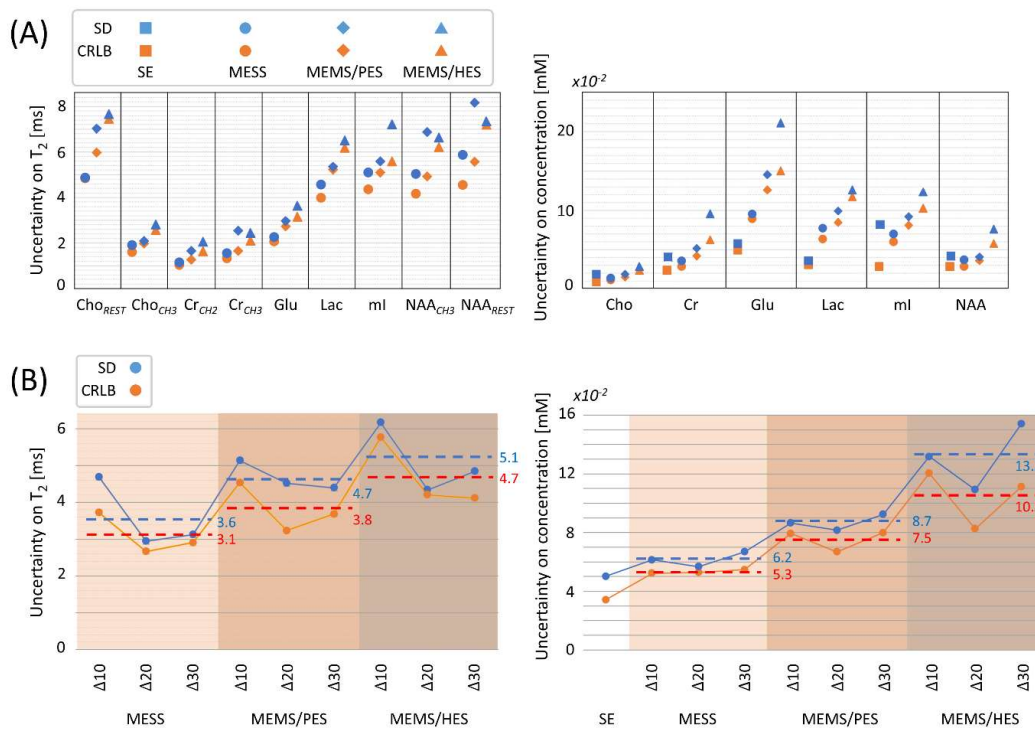
Figure 4: Uncertainties for in-silico experiment. (A) Standard deviation (SD) and CRLB for metabolite T_2s and concentrations as a function of the acquisition scheme averaged over different sequence timings (i.e., Δ settings in Table 1). (B) Standard deviation (SD) and CRLB for T_2s and concentrations as a function of measurement schemes and sequence timings (i.e., Δ settings in Table 1, Δ10 stands for Δ = 10 ms). The precision of T_2 determination is best for MESS, followed by PES and HES for all metabolites. The settings Δ20 and Δ30 are preferable for all methods. For concentrations, the ranking is similar, except that the choice of TE has limited influence. Single short-TE measurement (SE) outperforms MESS in terms of precision to estimate concentrations only.

2.2   In-vitro evaluations

Fig. 5 reports that estimated amplitudes and T_2s across the three acquisition techniques are not systematically distinct and confirms the in-silico expectation that CRLBs are best for MESS. MEMS/PES displays smaller CRLBs for amplitudes and T_2s compared to traditional semi-LASER mTE, which correspond in terms of precision essentially to MEMS/HES. The acquisition of six TEs shows the tendency to estimate slightly longer T_2s compared to models based on only three TEs (particularly evident for MEMS/PES and semi-LASER mTE for Lac

and NAA$_{CH3}$). The precision of T$_2$ estimation, as judged by CRLBs, is better for the 6-TE scheme for most metabolites (notably NAA, Glu, and Cn).
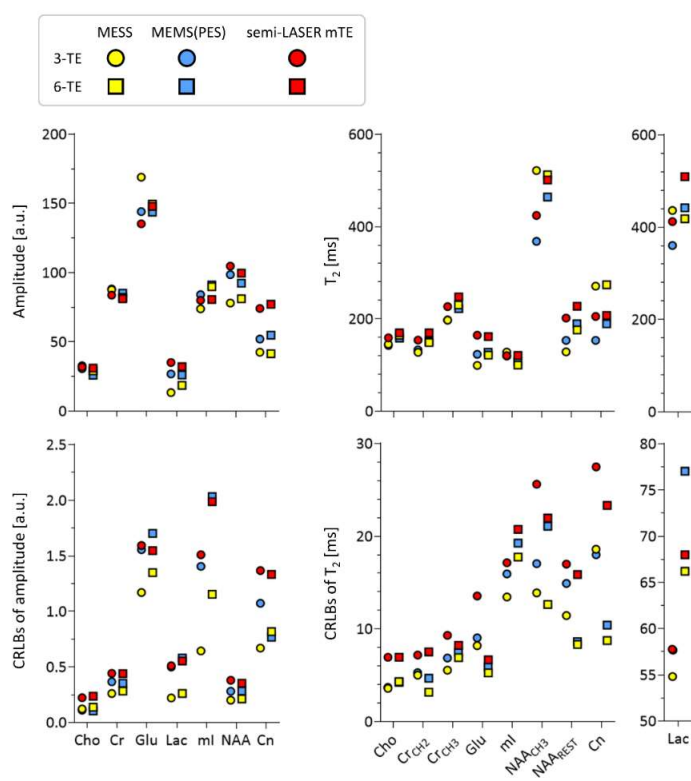
Figure 5: In-vitro estimated amplitudes - area under the curve - (top left) and T$_2$s (top right) together with corresponding CRLBs (bottom). Three acquisition techniques are reported: MESS (yellow), MEMS(PES) (blue), and multi-TE semi-LASER (red) considering protocols with 3-TE (circles) and 6-TE (squares). Estimated amplitudes and T$_2$s do not display systematic differences across methods. CRLBs are found superior for MESS. Partial echo sampling and multiple refocusing – MEMS - is found to report similar CRLBs for amplitudes and T$_2$s compared to traditional semi-LASER multi-TE.

2.3 In-vivo evaluations

Fig. 6 depicts the voxel selection in OCC and illustrates MESS and MEMS/PES acquired data, fits, and residues for a single subject. The "short" TE spectra in MESS (35, 60.6, 140, and 213.6 ms) show limited resolution, as expected. Zero-filling effects mostly of residual water from matching the duration of the 3rd readout (i.e., TEs 86.2 and 287.2 ms) are manifest. In addition, MESS spectra feature a linear phase shift and residual water components for TE1, TE2, TE4, and TE5 (see Methods 2.2.1).

Distributions of estimated concentrations (Fig. 7A, Table S1) and T$_2$s (Fig. 7B, Table S2) are reported together with CRLBs (Fig. 7) for the four acquisition schemes. Estimated concentrations and T$_2$s do not show systematic divergence across methods. Their CRLBs are smaller for MESS methods than corresponding MEMS acquisition, with significant improvements for concentrations with the MESS 3-TE compared to the MEMS 3-TE scheme (Fig. 7A). No difference in precision for concentrations is evident in terms of the choice of TE schedule. The precision of T$_2$ estimation as judged by CRLBs is better for the 6-TE scheme.
Fig. 8 illustrates the spectra and their variability within a cohort by displaying the cohort average ± the cohort SD range for the 5 subjects with the 6 TE scheme for MESS and MEMS. An apparently very large cohort variability for FIDs and spectra is entirely due to the large variability of the residual water signal.
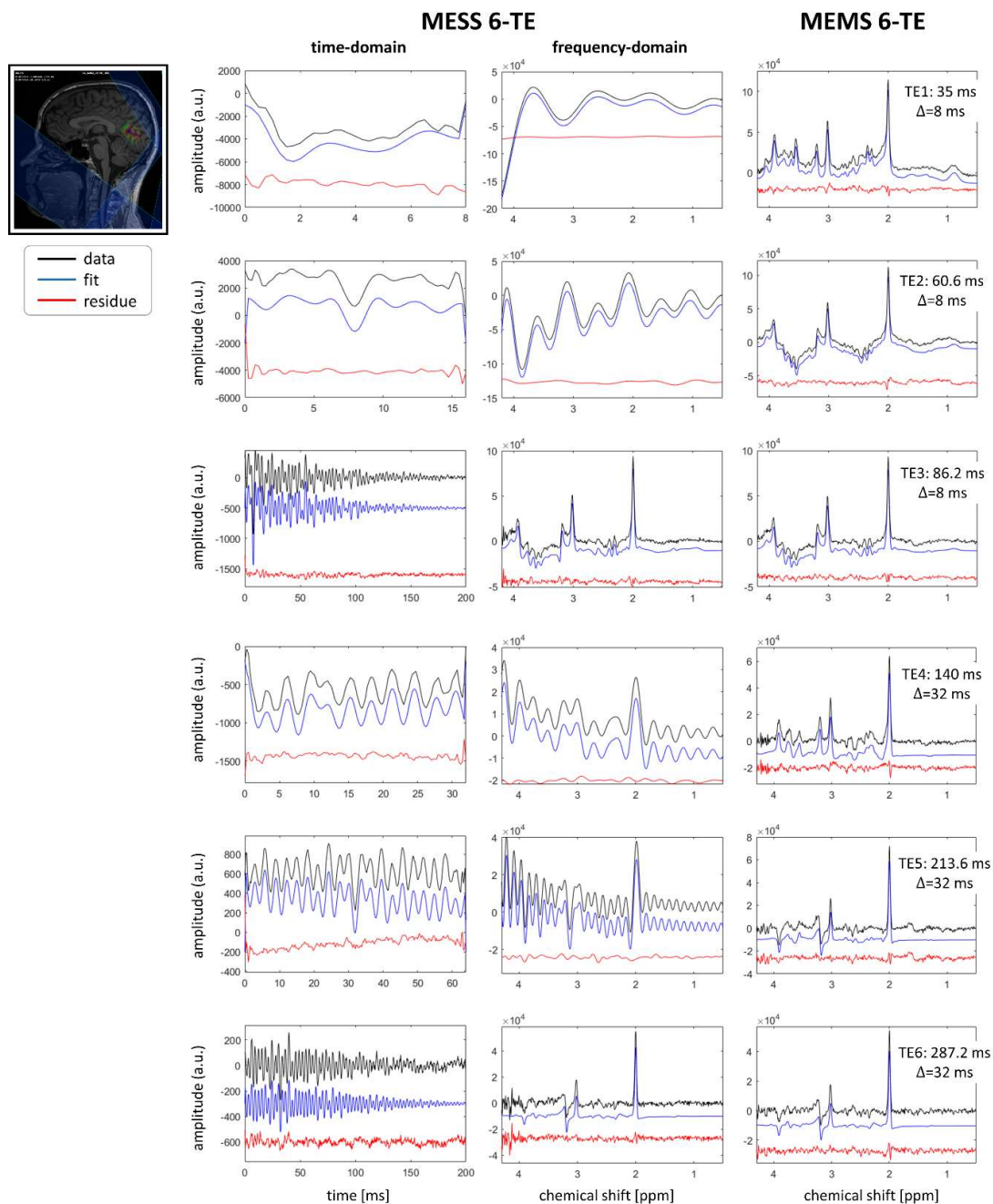
Figure 6 : Acquired data from MESS and MEMS 6-TE protocols in the occipital lobe from a single volunteer (black), fitted model (blue), and residues (red). For visual purposes, MESS TE1, TE2, TE3, and TE4 are zero-filled in the time domain. Residual water is visible by offsets in time-domain or partly cut-out resonance peaks around 4.2 ppm. Partial-echo sampling for 2nd and 3rd readouts displays a 1st order phase offset for MEMS acquisition at TE2 and TE3.
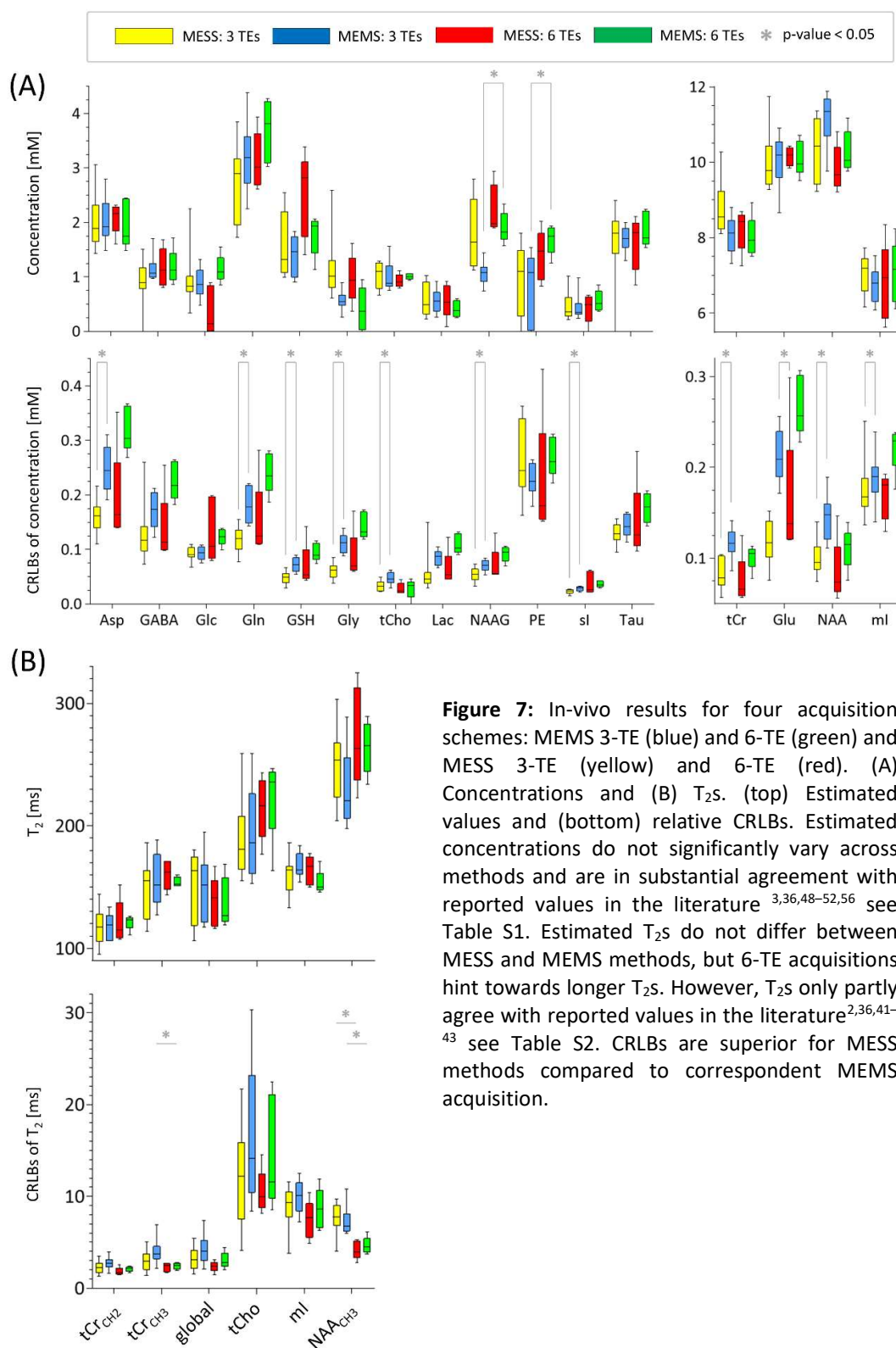
Figure 7: In-vivo results for four acquisition schemes: MEMS 3-TE (blue) and 6-TE (green) and MESS 3-TE (yellow) and 6-TE (red). (A) Concentrations and (B) T₂s. (top) Estimated values and (bottom) relative CRLBs. Estimated concentrations do not significantly vary across methods and are in substantial agreement with reported values in the literature [3,36,48–52,56] see Table S1. Estimated T₂s do not differ between MESS and MEMS methods, but 6-TE acquisitions hint towards longer T₂s. However, T₂s only partly agree with reported values in the literature [2,36,41–43] see Table S2. CRLBs are superior for MESS methods compared to correspondent MEMS acquisition.
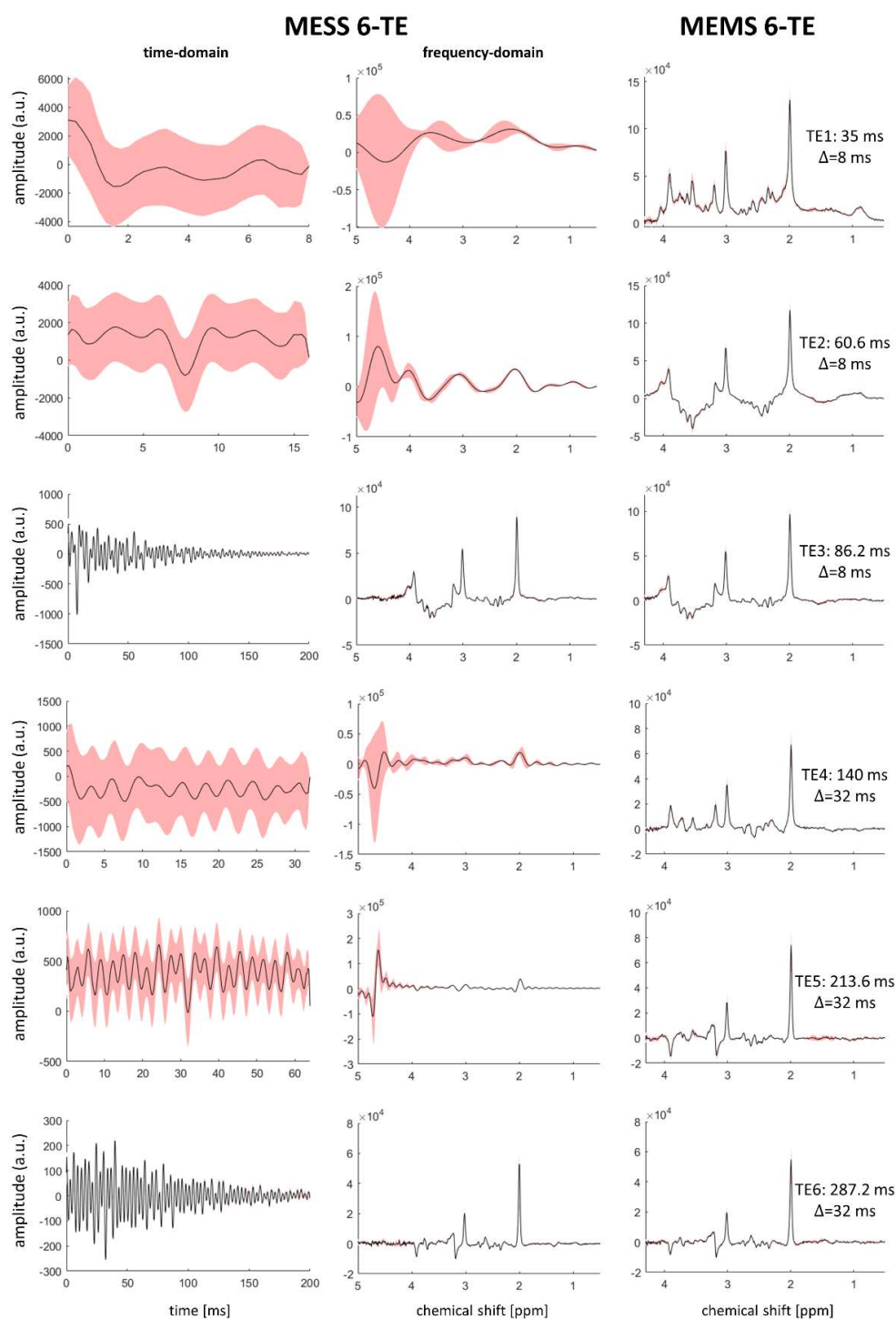
Figure 8: Measured in-vivo spectra (real part) for MESS and MEMS 6-TE scheme. (black) Average spectra from a cohort of five healthy volunteers, (red) signal variation (±1 SD). MESS acquisitions are displayed in time- (left) and frequency-domain (middle). MESS spectra for TE1, TE2, TE3, and TE4 are zero-filled for visual purposes. Truncation of readout and echo formation at 8 and 32 ms are visible (respectively for Δ=8ms in TE1 and TE2 and Δ=32ms in TE4 and TE5). The most apparent variability is driven by unsuppressed residual water. Partial-echo sampling for 2nd and 3rd readouts leads to a 1st order phase offset for both MEMS and MESS acquisitions

2.4   Repeatability

Subject-specific estimates and a measure of within-session repeatability (for MESS only) are reported for all metabolites across the cohort of five healthy subjects in Figures S1-S4. In Fig. 9, within-session repeatability of the MESS schemes (SD estimate from 3 spectra) is portrayed and contrasted with their minimal expected values derived from the CRLBs for both, concentrations and T_2s. This is plotted for each volunteer, as well as for the cohort medians. SDs are considerably larger than CRLBs for most subjects (average ratios of 5.9 and 8.0 for concentrations and 3.4 and 5.2 for T_2s in the 3-TE and 6-TE schemes, respectively).
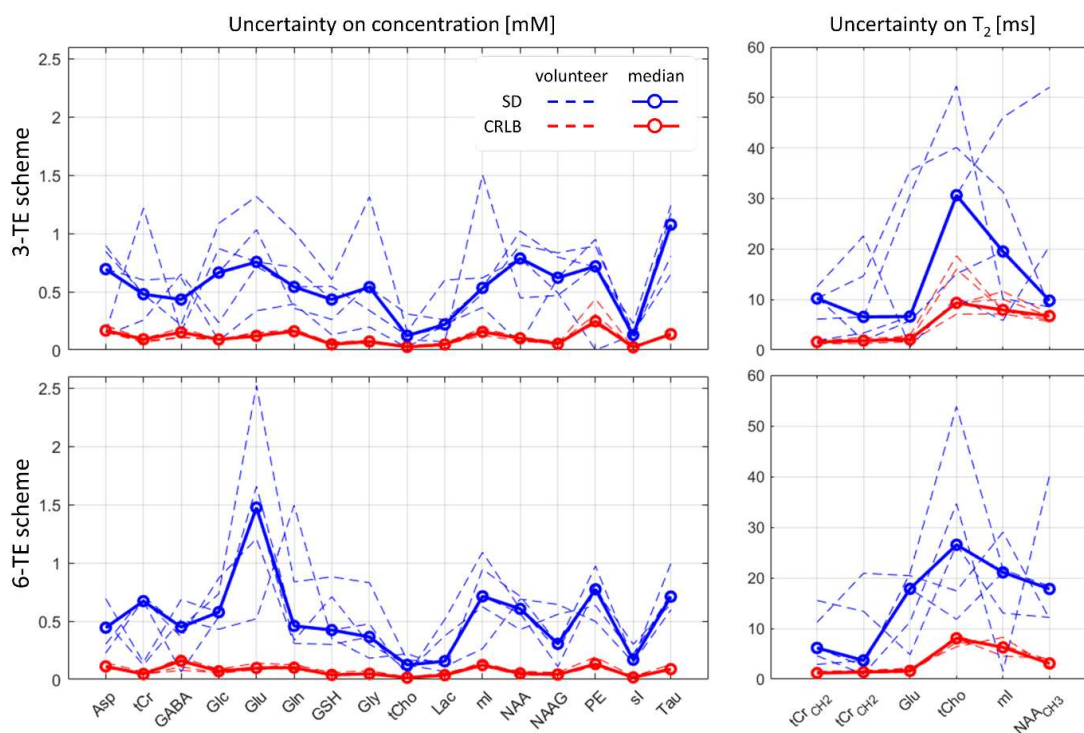
Figure 9: Repeatability of MESS: SD and CRLB estimates for concentrations and T_2s. SD is measured from estimates in three repetitions of MESS with 3-TE and 6-TE schemes in single volunteers and averaged over a cohort of five volunteers. Dashed lines: single volunteer data; bold lines: median values for the cohort. It is evident that SD overshoots the estimated CRLB consistently for all metabolites. The average ratio between the two values (5.9 and 8.0 for concentration and 3.4 and 5.2 for T_2s for the 3-TE and 6-TE schemes, respectively) is large. It is assumed driven by intra-subject inter-scan variability of residual water – acclaim that is substantiated in Figures S5, S6, S9.

3.   DISCUSSION

The proposed acquisition and processing schemes aim at evaluating the feasibility of simultaneously quantifying metabolite concentrations and metabolite-specific T_2s in a clinically feasible time frame. The study compared a novel and faster method (MESS) with its gold standard counterpart (MEMS), which targets the same multiple echo-train acquisitions but acquires the fully sampled spectra in separate acquisitions requiring a three times longer scan time. Furthermore, the proposed scheme was compared to single short TE

spectra (SE), the clinical standard to estimate concentrations alone, and to a multi-TE experiment that represents the clinical reference for estimating concentrations and T_2s.

It turned out that using an echo train and subsequent simultaneous modeling to arrive at concentrations and relaxation times makes such investigations possible in a clinically compatible time frame. Essentially, the relaxation information can almost be recorded for free at a time similar to the one previously used for measuring concentrations alone. CRLBs were used as a reference to judge the relative performance of the investigated sequences. They measure the minimum precision error and have previously proven to be a well-suited criterion for selecting optimal sampling strategies for many applications and fields.[38–40] Still, the CRLB-derived conclusions have been confirmed with Monte-Carlo evaluations in this work.

In-vivo, the performances of MESS compared to MEMS yielded CRLB improvements ranging from 5% to 20% for T_2s and from 10% to 20% for concentrations.

The MESS design implemented partial echo sampling to maximize the number of acquired data points with high SNR. The benefit of PES was investigated in-silico also for multi-shot approaches, where CRLB improvements ranged from 8% to 22% for T_2s and 16% to 37% for concentrations and MC-derived SDs confirmed these numbers (Fig. 4). These results are also in agreement with former studies on optimal MRS sampling strategies.[38] Moreover, evaluation of PES with multiple refocusing compared to traditional multi-TE spectra (sampled as half-echo FIDs) was investigated in-vitro (MEMS vs. semi-LASER mTE in Fig. 5). Average CRLB improvements around 17% for concentrations and 22% for T_2s were found when MEMS (PES) is used. Multiple vs. single refocusing would also be a difference between these two schemes, but this is unlikely to be a relevant factor for this overall performance gain. If anything, one would expect to find systematically different T_2s rather than improved precision with multiple refocusing, but this was not consistently seen in the in vitro study.

The efficacy of acquiring truncated multi-echo data to estimate metabolite-specific T_2s has also been demonstrated in a similar experiment at 7T.[16] There, a CPMG spectroscopic sequence was used, acquiring eight echoes, and was combined with a fully sampled short TE spectrum, acquired in a separate scan, that was used as prior to extrapolating the full resolution of the truncated CPMG spectra. The reconstruction was based on a short-TE data-driven singular value decomposition approach which integrated T_2^* relaxation. Its limitation lies in that a reliable estimation is limited to singlets. Our method does not aim to reconstruct the truncated spectra in the echo train. It exploits the resolution at the longest TE to regularize the fit over low-resolved spectra within a 2D frame in which priors are modeled for different TEs according to the known evolution based on simulated model spectra.[38] Therefore, it produces more reliable estimates for J-coupled spin systems than the previously proposed multi-echo approaches, which relied either on priors from only single short TE[16] or limited simulations for spectral singlets.[13] Furthermore, the single-shot echo-train approach used here and in Ref [16] promises to be more robust to patient motion than independent multi-shot acquisitions.[13] Of course, a separately acquired short TE spectrum with full-resolution could also be included in the simultaneous modeling of the current method, which would then combine our method with the approach from Ref [16], probably enhancing precision, but at the expense of additional scan time.

Multiparametric MRS in the form of MRS fingerprinting (MRSF) is an alternative to the currently proposed way of simultaneously extracting metabolite-specific concentrations and T_2s. It has shown the potential to improve precision and accuracy for T_2 by a factor of 2.5.[13] However, the complex modeling required to generate appropriate dictionary entries currently hampers the extension to J-coupled metabolites or the full range of detectable metabolites. Nevertheless, it appears promising to combine MESS with features of MRSF. In particular: (i) extending MESS with scans that are sensitive to T_1 and may thus also be more scan-time efficient or inversely (ii) enhancing the current MRSF acquisition schemes with MESS elements rather than separately recorded scans with differing TEs. Whether simultaneous traditional modeling or dictionary searches would be more successful in the evaluation of such extended schemes is open for investigation.

3.1   Quantification of T_2s

The proposed method leads to recording of subject-specific T_2 relaxation times from five prominent spectral peaks and one overall T_2 for all minor contributors to the OCC brain spectrum within clinical time frames.

The method has clear restrictions regarding which echo times can be recorded in a single shot and asks for compromises between resolution and increments of echo times. However, as shown for the 6-TE setup, multiple single-shot echo trains with different timings can be combined to adapt to specific needs for echo times to be covered.

The in-silico study did not reveal significant differences in precision for the explored TE batches but hinted that relatively large TE spacings (Δ=20 or Δ=30 ms) are favorable. However, this evaluation was limited to a few settings and the average result from the whole metabolite set was used as a measure. Furthermore, the schemes did not include very long TEs. According to a dedicated earlier study[38], such improvements can be achieved, mainly if focusing on specific target metabolites.

The real-world implementations based on semi-LASER rather than PRESS were restricted to longer minimal TE compared to the in-silico setup. To test basic feasibility and probe compromises with regard to SNR and resolution, both small and relatively large TE spacings were investigated (Δ = 8 or Δ = 32 ms). The multi-echo methods with 3 TEs considered a minimal TE spacing with TE3 shorter than 100 ms to maximize SNR in this last TE. MESS yielded higher precision than MEMS for this setup, but we also found that for T_2s, an optimal design combines these short with longer TEs, hence also small and larger TE spacing, confirming expectations from the in-silico evaluation and a former study.[38]

The differences in relaxation times found in vivo between metabolites or metabolite sub-moieties are broadly consistent with previous results,[2,36,41–43] with relaxation times for singlets decreasing in magnitude as follows: NAA_{CH3}> tCho > tCr_{CH3} > tCr_{CH2}.[16] However, the actual T_2 values found with the proposed multi-echo measurements are only partly consistent with the previous literature (Table S2) and overall yield somewhat shorter T_2 times (~ 20 to 40ms shorter).[43] In fact, due to the two pairs of adiabatic π pulses employed in the semi-LASER block combined with the added pair of Mao pulses, one would expect to find longer rather than shorter T_2 times here in comparison to traditional acquisitions, in particular those from PRESS. This is expected because of the quasi continuous refocusing, similar to the CPMG condition, and $T_{1\rho}$ rather than T_2 decay during the RF pulses.[44] Such a

finding could be caused by incomplete refocusing by the Mao pulses, but that is not verified in the in-vitro scans. Or it could be linked to a suboptimal choice of the TE batches. Indeed, it had been shown that a reliable and accurate T_2 estimation should include longer TEs (~1.3-1.5 times T_2;[45] e.g., for an assumed T_2^{NAA} = 350ms,a maximum TE = 490 ms is recommended) and larger TE spacing.[38] However, the currently mainly investigated design aimed to record the shortest TEs possible to maximize SNR in TE3 with best resolution. The effect of including longer TEs is confirmed by the in-vitro results, where estimates based on models with six TEs showed consistently longer T_2s. A benefit from longer TEs is clearly found in vivo based on the CRLBs for the T_2 estimates, in particular for NAA_{CH3}, with the longest T_2, where the improvement is statistically significant.

On the other hand, T_2 values from literature are scarce and inhomogeneous due to the intrinsic sensitivity of the measure to the chosen protocol [46,47] and thus difficult to use as gold standard. On top, when comparing to literature one should keep in mind that the values from this study resulted from 2:12 - 4:24 (MESS) or 6:36 - 13:12 (MEMS) minutes of acquisition time only, while literature values often originate from much longer scans. Moreover, it is paramount to stress that, for clinical use, precision (CRLBs) and variance within cohorts of equivalent subjects are more relevant than comparisons with literature values.

3.2   Quantification of Concentrations

For the purpose of estimating concentrations alone, a single SE experiment has provided the best performances in the simulated cases (Fig. 4). Of course, the main reason for this may be that in this case the fit model assumes fixed T_2 values, which simplifies the parameter space of the model and its corresponding partial derivatives in the Fisher matrix. Such an assumption translates to the traditional clinical approach when single SE experiments are used given that subject-specific T_2 values are unavailable.

The proposed MESS scheme clearly yields higher precision in identical scan time or promises shorter scan times for equal precisions as the standard MEMS approach when judgment is based on the CRLBs that resulted in silico, in-vitro and in-vivo. The increased precision is particularly striking for the in-vivo estimated concentrations of tCr, Glu, mI, and NAA. In spite of the benefit for T_2 estimation, the 6-TE scheme does not outperform the 3-TE scheme for concentration. The two yield similar performance.

Metabolite concentrations from this study include subject-specific T_2 corrections, which has not been performed often in the literature. Still, the obtained overall distribution of concentrations agrees well with the literature for occipital GM (see Table S1). [3,14,36,48–52].

When comparing the normal confidence range spanned by the mean and the cohort standard deviation, it is noted that the ranges are very similar or even narrower compared to literature, even though the scan times per spectrum are mostly much shorter than those in the quoted literature (Table S1) and are based on 3T data compared to some of the studies that used 7T spectra. This limited spread is surprising also because the simultaneous determination of T_2 increases the concentration uncertainty and some of the literature studies made use of outlier rejection based on relative CRLB or soft constraints stabilizing their estimation.

3.3 Repeatability

Based on the substantial difference between CRLB estimates from single in vivo measurements and SDs from within-session repeats (Fig. 9, and supplemental illustrations in Figures, S1-S4), CRLBs appear to be an ill-suited approximation for repeatability or precision of estimation for in-vivo MESS. And this might question the use of the CRLB as criterion of merit to judge between MESS and MEMS and the different acquisition schemes within MESS in spite of the wide use of CRLBs for such a purpose.[38–40] However, since the Monte-Carlo evaluations did not show such a divergence, the problem does not lie with the CRLBs per se, but rather with experimental issues for the investigated in-vivo implementations with minimal TEs. Considering the cohort average spectra in Fig. 8, it is highly likely that the large variability of the residual water signal in the spectra with limited resolution is the cause for unexpected scan-to-scan variability, a feature that was not built into the original in vitro setup. The residual water signal cannot readily be removed in postprocessing by HLSVD in those spectra (TE1, TE2, TE4, TE5) and it carries more complexity than what can be incorporated into the fit model (for MC, difference of large signals with effects of eddy currents and potential motion accentuating the difficulty for a proper lineshape model).

To prove that indeed the limited repeatability is an implementation issue rather than a basic flaw in the method, extended repeatability exams were performed in three subjects and two further tests were performed in hindsight. Figure S7(top) shows the variability from five repeated scans in one volunteer and displays clearly that intra-session fluctuations of the water signal plague in particular the TE1 and TE2 acquisitions with the lowest resolution. To test for the influence of the residual water signal on quantification results, a second Monte-Carlo study was performed where residual water fluctuations from in vivo scans were included in the setup as described in supporting information Text S2. The results from this in silico comparison of scans with and without residual water signals, as presented in Figures S5 and S6, clearly show that CRLBs are a valid correlate of repeatability unless residual water signals are not properly incorporated in the fit. Thus, to render the proposed MESS method robust and valid as a clinical tool, the influence of residual water has to be eliminated or substantially reduced. Four solutions are proposed and one of them has been initially tested: (i) improved elimination of the water signal in MC based on e.g. prospective motion correction or adapted inversion pulses; (ii) novel postprocessing techniques for optimal water signal removal for spectra with low spectral resolution, possibly based on deep learning for signal removal or for resolution restoration;[53] (iii) use of advanced WS techniques like VAPOR instead of MC[29]; (iv) alternative TE-batch designs with longer sampling windows where enough data points for TE1 and TE2 are acquired to enable HLSVD water removal. As a proof-of-concept, such a 3-TE MESS design with longer acquisition windows has been implemented and explored for the ratio between CRLB and repeatability in 5-fold repeated scans in three subjects, as detailed in Text S3. Figures S7(bottom) and S8-S9 document that this setup is only marginally plagued by residual water signals and would offer a robust implementation of MESS where the presumed reliability or acquisition speed derived from CRLB calculations is paralleled in measurement precision as determined by repeatability evaluations.

3.4 Limitations

The calculated CRLBs are only valid for the specifically addressed contexts (e.g., metabolite set, line width, optimal RF pulses, field strength) and in the case of complete knowledge of the model, where the handling of the baseline is critical as discussed in detail in the previous section. This limits the experiment design but allows fine-tuning of the acquisition strategy, given the exact experimental situation.[38]

The T_2 relaxation times of metabolites reported in this work are specific to the GM-rich occipital lobe region. The aim of the current study does not address T_2 differentiation between white and gray matter or different brain regions.[2,54]

The hypothesis of different T_2 for the tCr resonances at 3.0 ppm (or 3.9 ppm) to be modeled via a biexponential decay has not been investigated.[55] Similarly, extension to more individual metabolite-specific relaxation times should be addressed in further investigations, potentially using MESS schemes with more than three acquisition windows or multiple TE batches as used here.

4. CONCLUSIONS

A novel experimental scheme for simultaneous determination of metabolite and subject-specific T_2s and T_2-corrected concentrations for complex spectral patterns that combines short and long TE recordings from single acquisitions with 2D model fitting has been tested. The novel approach promises comparable accuracy, increased precision, or inversely shorter experiment times compared to traditional approaches. Therefore, it has proven to be a suitable alternative to conventional multi-echo multi-shot acquisitions.

Future work should explore optimal MESS design as a trade-off between TE spacing, TE batch and batch size, and data truncation/resolution to monitor the methods' robustness, stability, reliability, and error minimization. The experimental design may be optimized for target metabolites, with a potential focus on dynamic measurements of metabolite T_2 variations, particularly in time-resolved experiments where changes in metabolite T_2s over time may reflect physiological changes caused by activation, pathology or other modulating effects.[16]

Extensions to higher magnetic fields[37] that allow better separation of some of the metabolite patterns to explore individual relaxation times of more metabolites should be straight-forward.

List of Supporting Information

Text S1: Details of simultaneous model fitting for MESS.

Table S1: Estimated metabolite concentrations for GM from occipital cortex, obtained for four setups: MESS and MEMS, each for the 3-TE and 6-TE schemes.

Table S2: Estimated transverse metabolite relaxation times (T_2s) for GM from occipital cortex, obtained for four setups: MESS and MEMS, each for the 3-TE and 6-TE schemes.

Figure S1: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 1.

Figure S2: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 2.

Figure S3: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 3.

Figure S4: In-vivo estimated T_2s and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs.

Text S2: Influence of the residual water signal on the repeatability of MESS.

Text S3: Alternative TE-batch design for MESS, enabling residual water signal removal by HLSVD

Figure S5: Uncertainty of estimated concentrations from in-silico Monte-Carlo experiments with and without realistic residual water.

Figure S6: Uncertainty of estimated T_2s from in-silico Monte-Carlo experiments with and without realistic residual water

Figure S7: Cohort average datasets for MESS 3-TE schemes with different TE/Δ setups.

Figure S8: Exemplary MEMS datasets for the 3-TE schemes with the same TE/Δ setups as in Figure S7.

Figure S9: Repeatability of MESS upon different TE/Δ setups.

Table S3: MRSinMRS checklist.

Acknowledgments

References

1.   Kreis R, Boer V, Choi IY, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: Background and experts' consensus recommendations. *NMR Biomed*. 2020;34(5):e4347. doi:10.1002/nbm.4347

2.   Wyss PO, Bianchini C, Scheidegger M, et al. In vivo estimation of transverse relaxation time constant (T2) of 17 human brain metabolites at 3T. *Magn Reson Med*. 2018;80:452-461. doi:10.1002/mrm.27067

3.   Murali-Manohar S, Borbath T, Wright AM, Soher B, Mekle R, Henning A. T2 relaxation times of macromolecules and metabolites in the human brain at 9.4 T. *Magn Reson Med*. 2020;84(2):542-558. doi:10.1002/MRM.28174

4.   Wilson M, Andronesi O, Barker PB, et al. Methodological consensus on clinical proton MRS of the brain: Review and recommendations. *Magn Reson Med*. 2019;82(2):527-550. doi:10.1002/mrm.27742

5.   Gasparovic C, Song T, Devier D, et al. Use of tissue water as a concentration reference for proton spectroscopic imaging. *Magn Reson Med*. 2006;55(6):1219-1226. https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.20901

6.   Kreis R, Ernst T, Ross BD. Absolute quantitation of water and metabolites in the human brain. II. Metabolite concentrations. *J Magn Reson Ser B*. 1993;102:9-19.

7.   Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR Biomed*. 2021;34(5):e4257. doi:10.1002/nbm.4257

8.   Kulpanovich A, Tal A. What is the Optimal Schedule for Multiparametric MRS? A Magnetic Resonance Fingerprinting Perspective. *NMR Biomed*. 2021;34(5):e4196. doi:10.1002/nbm.4196

9.   Kreis R, Slotboom J, Hofmann L, Boesch C. Integrated data acquisition and processing to determine metabolite contents, relaxation times, and macromolecule baseline in single examinations of individual subjects. *Magn Reson Med*. 2005;54:761-768. doi:10.1002/mrm.20673

10. Just Kukurova I, Valkovič L, Bogner W, et al. Two-dimensional spectroscopic imaging with combined free induction decay and long-TE acquisition (FID echo spectroscopic imaging, FIDESI) for the detection of intramyocellular lipids in calf muscle at 7 T. *NMR Biomed*. 2014;27:980-987. doi:10.1002/nbm.3148

11. Bolliger CS, Boesch C, Kreis R. On the use of Cramér-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments. *Neuroimage*. 2013;83:1031-1040. doi:10.1016/j.neuroimage.2013.07.062

12. An L, Li S, Shen J. Simultaneous determination of metabolite concentrations, T1 and T2 relaxation times. *Magn Reson Med*. 2017;78:2072-2081. doi:10.1002/mrm.26612

13. Kulpanovich A, Tal A. The application of magnetic resonance fingerprinting to single voxel proton spectroscopy. *NMR Biomed*. 2018;31(11):e4001. doi:10.1002/nbm.4001

14. De Graaf RA. *In Vivo NMR Spectroscopy: Principles and Techniques*. 3rd ed. John Wiley & Sons; 2019.

15. Kiefer AP, Govindaraju V, Matson GB, Weiner MW, Maudsley AA. Multiple-echo proton spectroscopic imaging using time domain parametric spectral analysis. *Magn Reson Med*. 1998;39:528-538. doi:10.1002/mrm.1910390405

16. Ronen I, Ercan E, Webb A. Rapid multi-echo measurement of brain metabolite T2 values at 7T using a single-shot spectroscopic Carr-Purcell-Meiboom-Gill sequence and prior information. *NMR Biomed*. 2013;26(10):1291-1298. doi:10.1002/nbm.2951

17. Hennig J, Nauerth A, Friedburg H. RARE Imaging: A Fast Imaging Method for Clinical MR. *Magn Reson Med*. 1986;3:823-833. https://www.onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.1910030602

18. Duyn JH, Moonen CT. Fast proton spectroscopic imaging of human brain using multiple spin-echoes. *Magn Reson Med*. 1993;30(4):409-414. https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910300403

19. Dydak U, Pruessmann KP, Weiger M, Tsao J, Meier D, Boesiger P. Parallel spectroscopic imaging with spin-echo trains. *Magn Reson Med*. 2003;50(1):196-200. doi:10.1002/mrm.10495

20. Tsai S, Posse S, Lin Y, et al. Fast mapping of the T2 relaxation time of cerebral metabolites using proton echo-planar spectroscopic imaging (PEPSI). *Magn Reson Med*. 2007;57(5):859-865.

21. Wang Z, Li Y, Lam F. High-resolution, 3D multi-TE 1H MRSI using fast spatiospectral encoding and subspace imaging. *Magn Reson Med*. 2022;87(3):1103-1118. doi:10.1002/MRM.29015

22. Rizzo R, Kreis R. Potential benefits from Multi-Echo Single-Shot Spectroscopy with a combined fitting process. *Proc Intl Soc Mag Reson Med*. 2020;28:3668.

23. Rizzo R, Kreis R. Multiparametric Single-Shot Magnetic Resonance Spectroscopy for Fast Metabolite Specific Concentration and T2 Determination. *Proc Intl Soc Mag Reson Med*. 2022;30:311.

24. Oz G, Tkac I. Short-echo, single-shot, full-intensity proton magnetic resonance spectroscopy for neurochemical profiling at 4 T: Validation in the cerebellum and brainstem. *Magn Reson Med*. 2011;65:901-910. doi:10.1002/mrm.22708

25. Scheenen TW, Klomp DW, Wijnen JP, Heerschap A. Short echo time 1H-MRSI of the human brain at 3T with minimal chemical shift displacement errors using adiabatic refocusing pulses. *Magn Reson Med*. 2008;59(1):1-6. http://www.ncbi.nlm.nih.gov/pubmed/17969076

26. Mao J, Mareci TH, Andrew ER. Experimental study of optimal selective 180 radiofrequency pulses. *J Magn Reson*. 1988;79:1-10.

27. Lin A, Andronesi O, Bogner W, et al. Minimum Reporting Standards for in vivo Magnetic Resonance Spectroscopy (MRSinMRS): Experts' consensus recommendations. *NMR Biomed*. 2021;34(5):e4484. doi:10.1002/nbm.4484

28.   Landheer K, Juchem C. Dephasing optimization through coherence order pathway selection (DOTCOPS) for improved crusher schemes in MR spectroscopy. *Magn Reson Med*. 2019;81(4):2209-2222. doi:10.1002/mrm.27587

29.   Tkáč I, Deelchand D, Dreher W, et al. Water and lipid suppression techniques for advanced 1H MRS and MRSI of the human brain: Experts' consensus recommendations. *NMR Biomed*. 2021;34(5):e4459. doi:https://doi.org/10.1002/nbm.4459

30.   Soher BJ, Semanchuk P, Todd D, Steinberg J, Young K. VeSPA: Integrated applications for RF pulse design, spectral simulation and MRS data analysis. *Proc Intl Soc Mag Reson Med*. 2011;19:1410.

31.   Simpson R, Devenyi GA, Jezzard P, Hennessy TJ, Near J. Advanced processing and simulation of MRS data using the FID appliance (FID-A) - An open source, MATLAB-based toolkit. *Magn Reson Med*. 2017;77(1):23-33. doi:10.1002/mrm.26091

32.   Hall EL, Stephenson MC, Price D, Morris PG. Methodology for improved detection of low concentration metabolites in MRS: Optimised combination of signals from multi-element coil arrays. *Neuroimage*. 2014;86:35-42. doi:10.1016/j.neuroimage.2013.04.077

33.   Stefan D, Di Cesare F, Andrasescu A, et al. Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package. *Meas Sci Technol*. 2009;20:104035.

34.   Chong DGQ, Kreis R, Bolliger CS, Boesch C, Slotboom J. Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets. *Magn Reson Mater Physics, Biol Med*. 2011;24:147-164. doi:10.1007/s10334-011-0246-y

35.   Cudalbu C, Behar KL, Bhattacharyya PK, et al. Contribution of macromolecules to brain (1) H MR spectra: Experts' consensus recommendations. *NMR Biomed*. Published online 2020:e4393. doi:10.1002/nbm.4393

36.   Hoefemann M, Bolliger CS, Chong DGQ, van der Veen JW, Kreis R. Parameterization of metabolite and macromolecule contributions in interrelated MR spectra of human brain using multidimensional modeling. *NMR Biomed*. 2020;33(9):e4328. doi:10.1002/nbm.4328

37.   Terpstra M, Cheong I, Lyu T, et al. Test-retest reproducibility of neurochemical profiles with short-echo, single-voxel MR spectroscopy at 3T and 7T. *Magn Reson Med*. 2016;76(4):1083-1091. doi:10.1002/mrm.26022

38.   Bolliger CS, Boesch C, Kreis R. On the use of Cramer-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments. *Neuroimage*. 2013;83:1031-1040. http://www.sciencedirect.com/science/article/pii/S1053811913008318

39.   Scope Crafts E, Lu H, Ye H, Wald LL, Zhao B. An efficient approach to optimal experimental design for magnetic resonance fingerprinting with B-splines. *Magn*

Reson Med. 2022;88(1):239-253. doi:https://doi.org/10.1002/mrm.29212

40. Cercignani M, Alexander DC. Optimal acquisition schemes for in vivo quantitative magnetization transfer MRI. *Magn Reson Med*. 2006;56(4):803-810. doi:10.1002/mrm.21003

41. Träber F, Block W, Lamerichs R, Gieseke J, Schild HH. 1H Metabolite Relaxation Times at 3.0 Tesla: Measurements of T1 and T2 Values in Normal Brain and Determination of Regional Differences in Transverse Relaxation. *J Magn Reson Imaging*. 2004;19:537-545. doi:10.1002/jmri.20053

42. Mlynarik V, Gruber S, Moser E. Proton T (1) and T (2) relaxation times of human brain metabolites at 3 Tesla. *NMR Biomed*. 2001;14(5):325-331.

43. Deelchand DK, McCarten JR, Hemmy LS, Auerbach EJ, Eberly LE, Marjańska M. Changes in the intracellular microenvironment in the aging human brain. *Neurobiol Aging*. 2020;95:168-175. doi:https://doi.org/10.1016/j.neurobiolaging.2020.07.017

44. Lin M, Kumar A, Yang S. Two-dimensional J-resolved LASER and semi-LASER spectroscopy of human brain. *Magn Reson Med*. 2014;71:911-920. http://dx.doi.org/10.1002/mrm.24732

45. Beracha I, Seginer A, Tal A. Adaptive Magnetic Resonance. bioRxiv. doi:10.1101/2022.03.16.484410

46. Deelchand DK, Auerbach EJ, Kobayashi N, Marjańska M. Transverse relaxation time constants of the five major metabolites in human brain measured in vivo using LASER and PRESS at 3 T. *Magn Reson Med*. 2018;79(3):1260-1265. doi:10.1002/mrm.26826

47. Kreis R, Slotboom J, Hofmann L, Boesch C. Integrated data acquisition and processing to determine metabolite contents, relaxation times, and macromolecule baseline in single examinations of individual subjects. *Magn Reson Med*. 2005;54(4):761-768. doi:10.1002/mrm.20673

48. Marjańska M, McCarten JR, Hodges J, et al. Region-specific aging of the human brain as evidenced by neurochemical profiles measured noninvasively in the posterior cingulate cortex and the occipital lobe using 1H magnetic resonance spectroscopy at 7 T. *Neuroscience*. 2017;354:168-177. doi:10.1016/j.neuroscience.2017.04.035

49. Mekle R, Mlynarik V, Gambarota G, Hergt M, Krueger G, Gruetter R. MR spectroscopy of the human brain with enhanced signal intensity at ultrashort echo times on a clinical platform at 3T and 7T. *Magn Reson Med*. 2009;61(6):1279-1285. http://www.ncbi.nlm.nih.gov/pubmed/19319893

50. Deelchand DK, Van De Moortele PF, Adriany G, et al. In vivo 1H NMR spectroscopy of the human brain at 9.4T: Initial results. *J Magn Reson*. 2010;206:74-80. doi:doi: DOI: 10.1016/j.jmr.2010.06.006

51. Marjańska M, Auerbach EJ, Valabrègue R, Van de Moortele PF, Adriany G, Garwood M. Localized 1H NMR spectroscopy in different regions of human brain in vivo at 7 T: T2 relaxation times and concentrations of cerebral metabolites. *NMR Biomed*. 2012;25(2):332-339. doi:10.1002/NBM.1754

52. Terpstra M, Ugurbil K, Tkac I. Noninvasive quantification of human brain ascorbate concentration using1H NMR spectroscopy at 7 T. *NMR Biomed*. 2010;23(3):227-232. doi:10.1002/nbm.1423

53. Lee H, Lee HH, Kim H. Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy. *Magn Reson Med*. 2020;84(2):559-568. doi:10.1002/mrm.28164

54. Hasan KM, Walimuni IS, Kramer LA, Narayana PA. Human brain iron mapping using atlas-based T2 relaxometry. *Magn Reson Med*. 2012;67(3):731-739. doi:10.1002/MRM.23054

55. Ke Y, Cohen BM, Lowen S, Hirashima F, Nassar L, Renshaw PF. Biexponential transverse relaxation (T2) of the proton MRS creatine resonance in human brain. *Magn Reson Med*. 2002;47(2):232-238. doi:10.1002/mrm.10063

56. Mangia S, Tkáč I, Gruetter R, et al. Sensitivity of single-voxel 1H-MRS in investigating the metabolism of the activated human visual cortex at 7 T. *Magn Reson Imaging*. 2006;24(4):343-348. doi:10.1016/j.mri.2005.12.023

Supporting Information

Text S1: Details of simultaneous model fitting for MESS.

Table S1: Estimated metabolite concentrations for GM from occipital cortex, obtained for four setups: MESS and MEMS, each for the 3-TE and 6-TE schemes.

Table S2: Estimated transverse metabolite relaxation times (T_2s) for GM from occipital cortex, obtained for four setups: MESS and MEMS, each for the 3-TE and 6-TE schemes.

Figure S1: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 1.

Figure S2: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 2.

Figure S3: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 3.

Figure S4: In-vivo estimated T_2s and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs.

Text S2: Influence of the residual water signal on the repeatability of MESS.

Text S3: Alternative TE-batch design for MESS, enabling residual water signal removal by HLSVD

Figure S5: Uncertainty of estimated concentrations from in-silico Monte-Carlo experiments with and without realistic residual water.

Figure S6: Uncertainty of estimated T_2s from in-silico Monte-Carlo experiments with and without realistic residual water

Figure S7: Cohort average datasets for MESS 3-TE schemes with different TE/Δ setups.

Figure S8: Exemplary MEMS datasets for the 3-TE schemes with the same TE/Δ setups as in Figure S7.

Figure S9: Repeatability of MESS upon different TE/Δ setups.

Table S3: MRSinMRS checklist.

TEXT-S1: Details of simultaneous model fitting for MESS.

Basis sets: The MESS basis sets are equal to those for MEMS (simulations include the two Mao refocusing pulses in both cases) except that they are truncated for the 1st and 2nd readouts, respectively, at Δ ms and 2Δ ms, for both TE batches. The truncated basis sets are zero-filled to conform to the input format of FiTAID and to match the layout of the acquired data with equal number of datapoints for all interconnected datasets and zero-filling to accommodate that. In particular, not only the vespa-simulated basis sets, but also simple Voigt lines were prepared to match between model and experimental data

Fitting strategy: The starting conditions of any fit were derived from the MEMS fit of a cohort average from three of the ten datasets of the examined subjects. The MESS fit for a single subject is run with the following two-step fit strategy to ensure fitting stability.

1. Fit of the third TE acquisition only (or 3rd plus 6th readout for 6-TE acquisitions) with individual metabolite areas, overall 0th order phase, overall frequency offset, and overall Gaussian broadening fitted, while keeping T_2s fixed at starting values. (i.e. χ^2 minimization in frequency domain for complex data of 3rd dataset)
2. Fit of all responses with all parameters (i.e., incl. T_2s) free to fit. (i.e. χ^2 minimization in frequency domain for complex data of all datasets)

	MESS 3-TE	MEMS 3-TE	MESS 6-TE	MEMS 6-TE	[1]	[2]	[3]	[4]	[4]	[5]	[6]	[7]
Asp	2.02 ± 0.48	2.04 ± 0.38	2.09 ± 0.25	1.96 ± 0.4	1.74 ± 0.29	2.83 ± 0.32	4.84 ± 1.15	3.1 ± 0.3	2.9 ± 0.5	2.1 ± 0.5	2.9 ± 0.8	2.0 ± 0.4
tCr	8.76 ± 0.65	8.06 ± 0.47	8.22 ± 0.51	8.01 ± 0.49	9.33 ± 0.58	8.1 ± 0.96	10.10 ± 0.46	8.0 ± 0.3	8.0 ± 0.4	7.7 ± 0.4	8.7 ± 1.1	8.6 ± 1.2
GABA	0.92 ± 0.38	1.16 ± 0.24	1.17 ± 0.32	1.18 ± 0.29	0.96 ± 0.25	2.12 ± 0.42	1.87 ± 0.92	2.5 ± 0.4	1.3 ± 0.15	1.3 ± 0.4	1.5 ± 0.3	0.9 ± 0.1
Glc	0.94 ± 0.47	0.89 ± 0.25	0.36 ± 0.39	1.14 ± 0.23	—	1.77 ± 0.51	—	1.4 ± 0.3	—	—	—	—
Glu	10.05 ± 0.76	10.02 ± 0.72	10.15 ± 0.22	10.11 ± 0.41	9.47 ± 0.62	10.99 ± 0.55	10.90 ± 0.80	8.9 ± 0.9	9.9 ± 0.9	9.3 ± 0.9	9.6 ± 1.3	8.9 ± 0.3
Gln	2.73 ± 0.65	3.19 ± 0.6	3.13 ± 0.47	3.69 ± 0.51	2.95 ± 0.33	3.96 ± 0.46	7.61 ± 0.91	1.6 ± 0.4	2.2 ± 0.4	2.2 ± 0.2	1.5 ± 0.5	2.5 ± 0.2
GSH	1.55 ± 0.55	1.41 ± 0.35	2.5 ± 0.69	1.77 ± 0.34	1.00 ± 0.08	1.43 ± 0.13	1.72 ± 0.16	1.4 ± 0.11	1.3 ± 0.2	1.1 ± 0.3	1.1 ± 0.1	0.7 ± 0.1
Gly	1.15 ± 0.53	0.57 ± 0.17	0.97 ± 0.4	0.4 ± 0.36	—	—	1.11 ± 0.28	—	—	—	0.7 ± 0.1	—
tCho	1.03 ± 0.22	1.01 ± 0.27	0.93 ± 0.11	1.01 ± 0.05	1.05 ± 0.16	1.00 ± 0.18	1.03 ± 0.12	1.1 ± 0.13	1.1 ± 0.05	0.9 ± 0.2	0.9 ± 0.1	1.0 ± 0.3
Lac	0.57 ± 0.27	0.55 ± 0.2	0.56 ± 0.28	0.41 ± 0.14	—	—	—	—	0.7 ± 0.1	0.5 ± 0.1	0.8 ± 0.2	0.7 ± 0.1
mI	7.04 ± 0.48	6.75 ± 0.43	6.81 ± 0.93	7.07 ± 0.73	7.73 ± 0.62	6.11 ± 0.63	5.22 ± 0.45	5.3 ± 0.6	5.7 ± 0.5	5.3 ± 0.4	6.4 ± 0.8	—
NAA	10.39 ± 0.78	11.17 ± 0.61	9.84 ± 0.54	10.28 ± 0.5	12.24 ± 0.62	9.81 ± 0.40	12.61 ± 1.02	11.2 ± 0.8	11.8 ± 0.2	13.5 ± 1.6	12.6 ± 1.7	12.4 ± 0.7
NAAG	1.77 ± 0.59	1.07 ± 0.19	2.24 ± 0.4	1.91 ± 0.25	1.24 ± 0.13	2.09 ± 0.39	1.42 ± 0.19	1.0 ± 0.2	1.1 ± 0.4	1.1 ± 0.5	0.40 ± 0.2	1.2 ± 0.2
PE	0.94 ± 0.64	0.86 ± 0.58	1.39 ± 0.42	1.69 ± 0.24	1.24 ± 0.21	1.92 ± 0.51	2.28 ± 0.95	2.2 ± 0.18	2.5 ± 0.3	1.6 ± 0.4	2.3 ± 0.4	2.1 ± 0.2
sI	0.47 ± 0.28	0.44 ± 0.21	0.43 ± 0.24	0.55 ± 0.17	0.38 ± 0.21	0.33 ± 0.06	0.13 ± 0.06	0.4 ± 0.11	0.3 ± 0.12	0.3 ± 0.2	0.4 ± 0.1	0.3 ± 0.1
Tau	1.62 ± 0.64	1.7 ± 0.22	1.61 ± 0.44	1.87 ± 0.29	1.80 ± 0.25	1.70 ± 0.24	1.58 ± 0.37	1.4 ± 0.5	1.5 ± 0.3	1.3 ± 0.2	2.1 ± 0.3	1.7 ± 0.2
MMBG	modeled	modeled	modeled	modeled	measured	measured	measured	modeled	measured	measured	measured	measured
T2 correction	measured	measured	measured	measured	literature	measured	measured	no	no	no	measured	no
tissue fraction	yes	yes	yes	yes	yes	yes	yes	no	no	no	yes	80% content
reference	water	water	water	water	water	water	water	8mmol/kg tCr	8mmol/kg tCr	water	water	water
sequence	semi-LASER	semi-LASER	semi-LASER	semi-LASER	STEAM	PRESS	semi-LASER	SPECIAL	SPECIAL	STEAM	LASER	STEAM
field	3T	3T	3T	3T	7T	3T	7T	3T	7T	9.4T	7T	7T
VOI	OCC	OCC	OCC	OCC	OCC, average of young and old	Occipital-parietal Lobe	Occipital Lobe	Occipital Lobe	Occipital Lobe	Occipital Lobe	Occipital Lobe	Occipital Lobe

[1] Marjanska M. et al, Neuroscience 354:168-177 (2017)
[2] Hoefemann M. et al, NMR biomed 33e4328 (2020)
[3] Murali-Manohar S., et al, MRM 84(2):542-558 (2020)
[4] Mekle R. et al, MRM 61(6):1279-1285 (2009)
[5] Deelchand D.K. et al, Journal of MR 206(1):74-80 (2010)
[6] Marjanska M. et al, NMR Biomed 25(2):332-339 (2011)
[7] Terpstra M. et al, NMR biomed 23(3):227-232 (2010)

Table S1: Estimated metabolite concentrations for GM from occipital cortex, obtained for four setups: MESS and MEMS, each for the 3-TE and 6-TE schemes. Values are given in mM units. They are listed in comparison to values from the literature. Estimated values do not significantly vary between methods and compare well with the literature, including the cohort standard deviations, even though in the current study the acquisition times were minimal and data were obtained at 3T compared to 7T for some of the literature studies. Uncertainties converted from standard error to standard deviation for Ref [1] where needed. [1–7]

	MESS 3-TE	MEMS 3-TE	MESS 6-TE	MEMS 6-TE	[1]	[2]	[3]	[4]	[5]
tCr_{CH2}	118 ± 14	118 ± 10	121 ± 16	121 ± 5	121 ± 13	137 ± 4	112 ± 18	116 ± 9	135 ± 14
tCr_{CH3}	149 ± 21	156 ± 19	160 ± 11	154 ± 4	162 ± 16	202 ± 7	144 ± 17	152 ± 7	175 ± 16
tCho	190 ± 31	193 ± 36	214 ± 23	224 ± 31	222 ± 15	314 ± 21	222 ± 50	207 ± 16	319 ± 63
mI	159 ± 15	168 ± 10	164 ± 11	154 ± 9	—	229 ± 23	229 ± 15	—	179 ± 27
NAA_{CH3}	250 ± 30	231 ± 30	273 ± 36	264 ± 19	247 ± 13	349 ± 29	263 ± 43	247 ± 19	292 ± 34
global	149 ± 28	150 ± 25	138 ± 18	137 ± 18	—	205 ± 11 (Glu)	122 ± 32 (Glu)	—	173 ± 27 (Glu)
fitting	2D simultaneous	2D simultaneous	2D simultaneous	2D simultaneous	sequential	2D simultaneous	2D simultaneous	sequential	sequential
sequence	semi-LASER	semi-LASER	semi-LASER	semi-LASER	PRESS	PRESS	PRESS	STEAM	LASER
TEs [ms]	(3): 35/60.6/86.2		(6): 35/60.6/86.2/140/213.6/287.2		(5): 50/120/200/280/400 (24): 20-307.5, step 12.5 (100): 30-228, step 2 (5): 50/100/150/200/250 (6): 35/140/230/290/330/400				
VOI	OCC	OCC	OCC	OCC	OCC	OCC	OCC	OCC	OCC (average young & older)

[1] Träber F. et al. JMRI 19:537-545 (2004)
[2] Hoefemann M. et al, NMR biomed 33e4328 (2020)
[3] Wyss P.O. et al, MRM 80(2):452-461 (2018)
[4] Mlynárik V. et al, NMR Biomed 14:325-331 (2001)
[5] Deelchand D.K. et al, Neurobiology of Aging 95:168-175 (2020)

Table S2: Estimated transverse metabolite relaxation times (T_2s) for GM from occipital cortex, obtained for four setups: MESS and MEMS, each for the 3-TE and 6-TE schemes. The 6-TE schemes yield somewhat longer T_2s, especially for tCho and NAA_{CH3}. Values are given in ms. They are listed in comparison to values from literature. Uncertainties converted from standard error to standard deviation for Ref [5].[2,8–11]
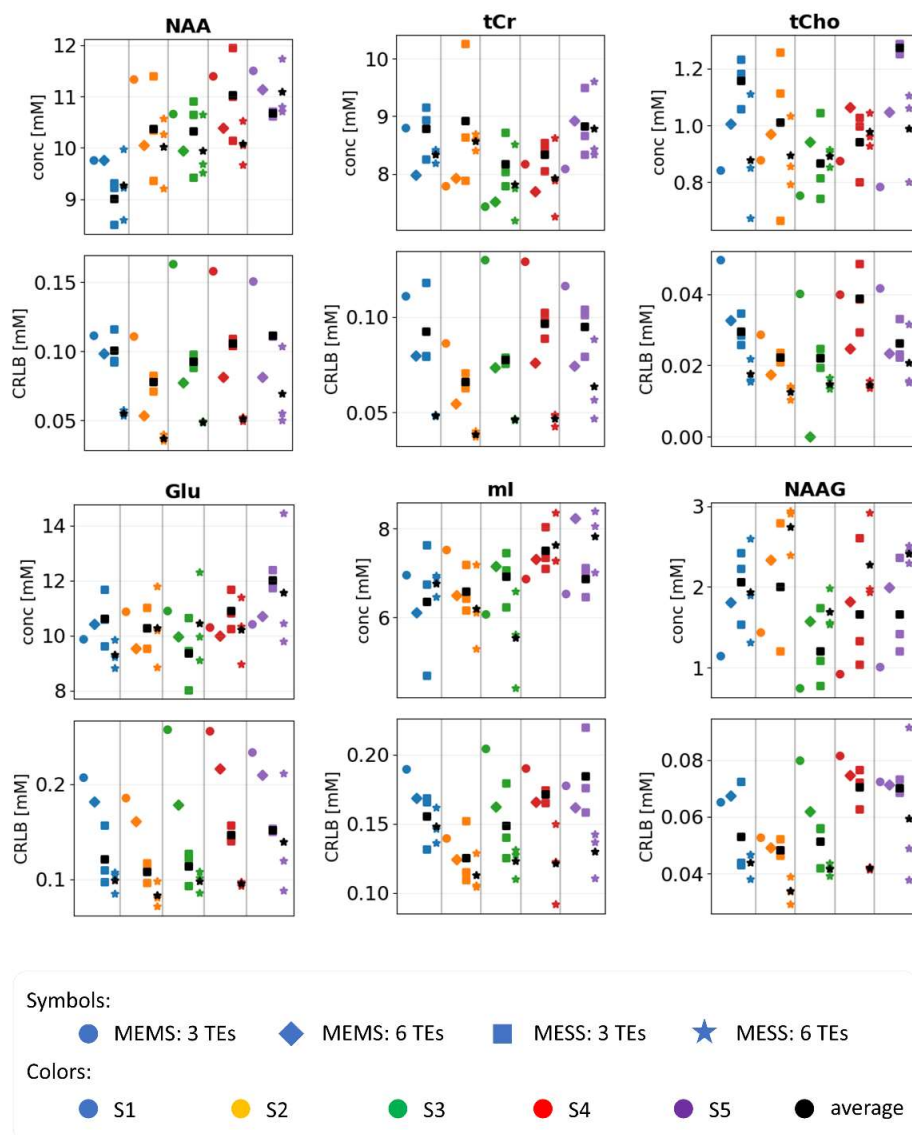
Figure S1: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 1. As proof of concept, acquisitions with MESS were repeated three times to match the scan time of MEMS. Estimated concentrations for the same subject do not significantly differ across the acquisition methods. CRLBs are confirmed to be superior for MESS. Variations across multiple acquisitions of MESS can be explained by variability of residual water signals. Six metabolites: NAA, tCr, tCho, Glu, mI, and NAAG. Different techniques are symbol-coded. Different subjects are color-coded.
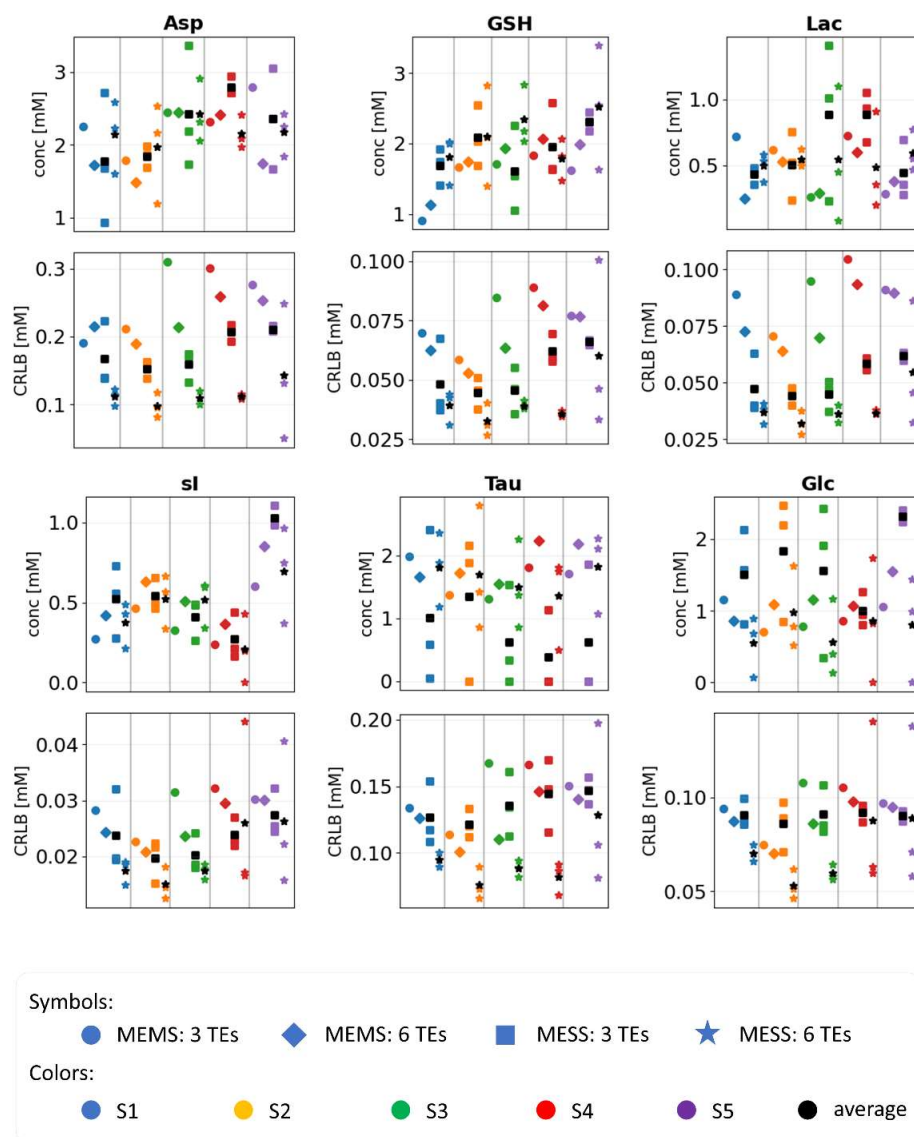
Figure S2: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 2. As proof of concept, acquisitions with MESS were repeated three times to match the scan time of MEMS. Estimated concentrations for the same subject do not significantly differ across the acquisition methods. CRLBs are confirmed to be superior for MESS. Variations across multiple acquisitions of MESS can be explained by variability of residual water signals. Six metabolites: Asp, GSH, Lac, sI, Tau, and Glc. Different techniques are symbol-coded. Different subjects are color-coded.
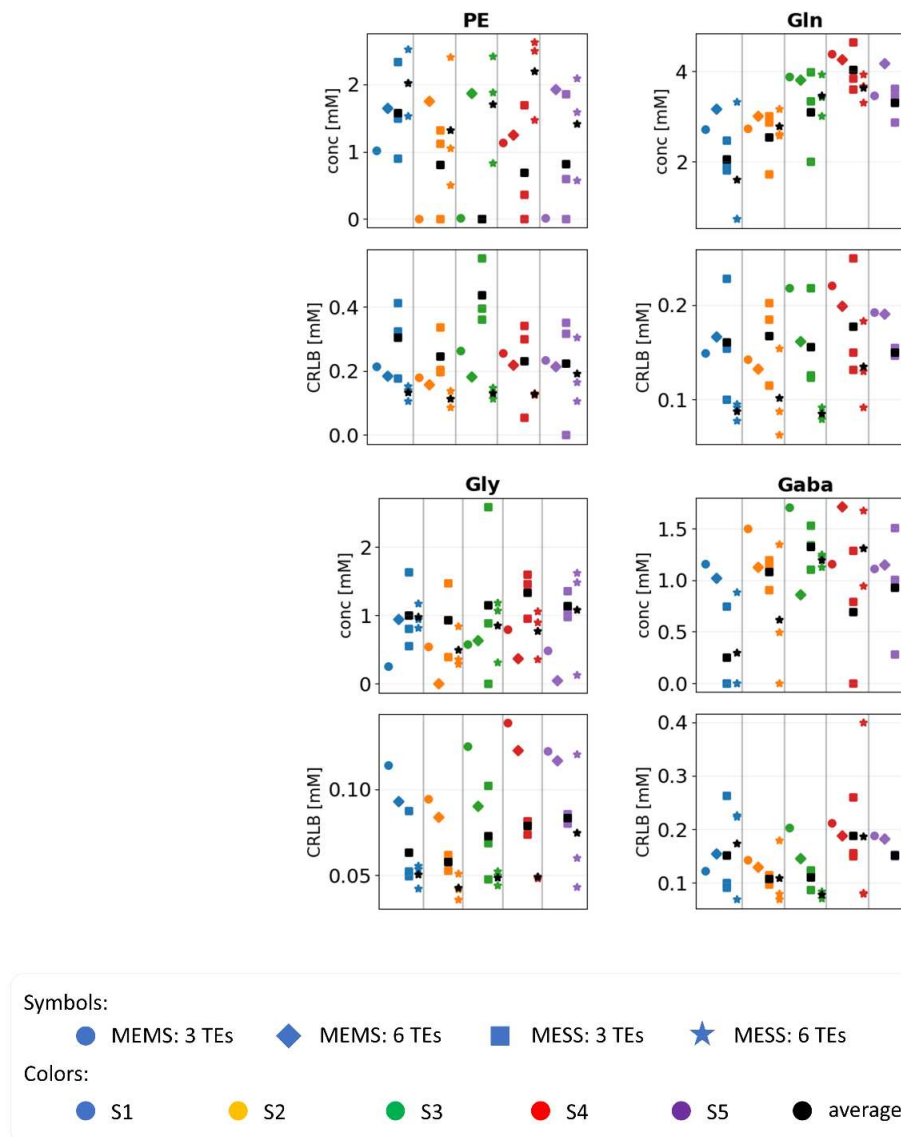
Figure S3: In-vivo estimated concentrations and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs: part 3. As proof of concept, acquisitions with MESS were repeated three times to match the scan time of MEMS. Estimated concentrations for the same subject do not significantly differ across the acquisition methods. CRLBs are confirmed to be superior for MESS. Variations across multiple acquisitions of MESS can be explained by variability of residual water signals. Four metabolites: PE, Gln, Gly, and GABA. Different techniques are symbol-coded. Different subjects are color-coded.
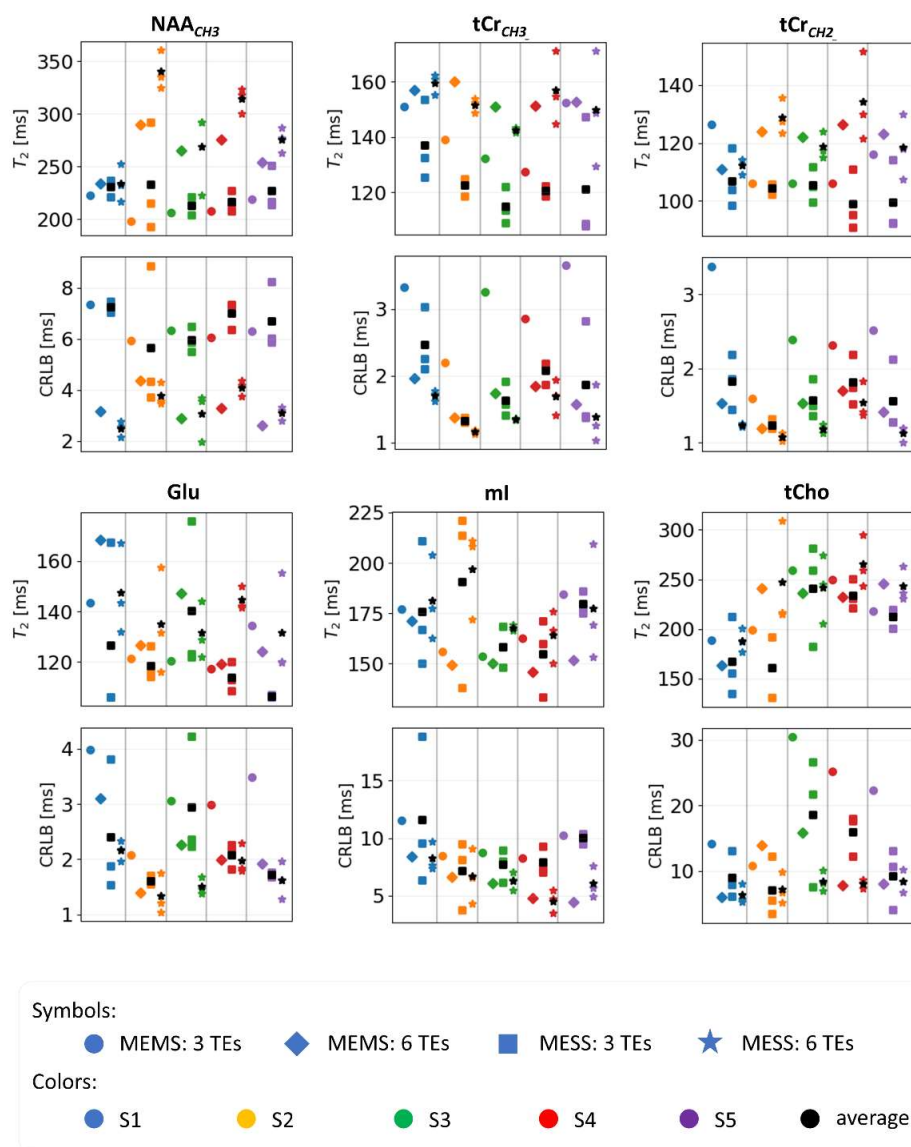
Figure S4: In-vivo estimated T_2s and CRLBs separately reported for five subjects and measured with MESS and MEMS methods with either three or six TEs. As proof of concept, acquisitions with MESS are repeated three times to match the scan time of MEMS. Estimated T_2s for the same subject do not significantly differ across the acquisition methods. CRLBs are confirmed to be superior for MESS. Variations across multiple acquisitions of MESS can be explained by increased sensitivity to motion and variability of residual water signals. Six moieties: NAA$_{CH3}$, tCr$_{CH3}$, tCr$_{CH2}$, Glu, mI, and tCho. Different techniques are symbol-coded. Different subjects are color-coded.

Text S2: Influence of the residual water signal on the repeatability of MESS

The effect of ill-modeled residual water that could not be removed in postprocessing by HLSVD filtering for MESS TE1 and TE2 was investigated with in-silico Monte-Carlo simulations for the MESS 3-TE protocol.

Methods: Simulations were run in the same setup as described in Methods 2.3.2 for a metabolite mixture in Table 1. 100-entries were generated with different noise realization but constant SNR, similarly as illustrated in Fig 3. The 100 entries generated for MESS were modified for TE1 and TE2 by the addition of a truncated residual water signal that had been isolated by HLSVD filtering from the measured MEMS datasets for the 10-volunteer in-vivo cohort. Thus, ten different residual water signals were distributed across ten Monte-Carlo entries. Bi-dimensional fitting was run as illustrated in Methods 2.3.3 when residual water was added to the data.
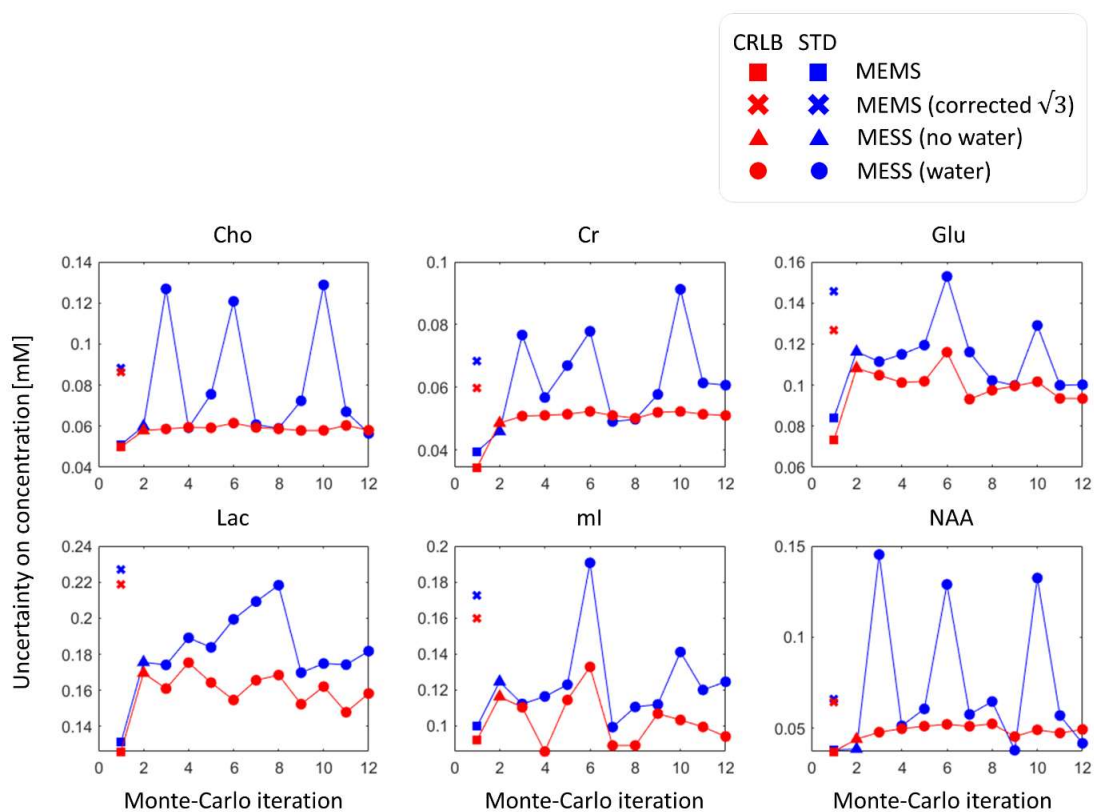
Results: Figs S5 and S6 display standard deviation (SD) and estimated CRLB for concentrations and T_2s, respectively. MEMS and water-free MESS show good convergence of SD to CRLBs, which supports the reliability of estimated CRLBs as a measure of precision in the absence of residual water. CRLBs systematically ill-represent measurement-SD when residual water is included in MESS schemes at TE1 and TE2. The effect varies across the ten realistic residual water lineshapes, given the different capabilities of the model (based on Voigt-line) to reproduce and fit the diverse residual water lineshapes.

Text S3: Alternative TE-batch design for MESS, enabling residual water signal removal by HLSVD

Alternative TE-batch designs for the MESS scheme are suggested to reduce the influence of residual water signal on quantitative results substantially. They feature longer minimal sampling windows, where enough data points are acquired for TE1 and TE2 such that HLSVD water removal is feasible.

Methods: A MESS 3-TE design with a longer acquisition window (Δ=56ms, 224 datapoints) and a TE batch of TE = 35, 156, and 278 ms was investigated and compared to the current MESS 3-TE TE/Δ setup. Measurement setup remains as described in Methods 2.1.1 and 2.1.5. Acquisitions were repeated 5-fold within the same session for both TE/Δ setups in a cohort of 3 healthy volunteers (age 28 ± 1 years, 1 male, 2 females).

Results: Time- and frequency-domain properties of both MESS 3-TE TE/Δ schemes are visualized in Figure S7. Residual water filtering can be deployed in postprocessing when a longer acquisition window is used, and significantly reduced variability is found for spectra of TE1 and TE2, Figure S7(bottom). A visualization of MEMS 3-TE is given in Figure S8 where fully resolved spectral features can be compared across different choices of TE batches. Figure S9 displays SD and CRLB on concentrations and T_2s for both MESS 3-TE TE/Δ setups. The divergence between CRLB and intra-session SD for Δ=8 ms (ratios of 4.4 and 2.6, respectively, for concentrations and T_2s) consistently reproduces the original data for the 10-volunteer cohort (Figure 9). However, a better water handling by HLSVD filtering at all TEs for Δ=56 ms yields closer agreement between CRLB and SD for both, concentrations and T_2s.

Figure S5: Uncertainty of estimated concentrations from in-silico Monte-Carlo experiments with and without realistic residual water with estimated SD in blue and estimated CRLB in red. MEMS/PES estimates with (cross) and without (square) $\sqrt{3}$-correction to equalize scantime. The MESS 3-TE scheme without residual water is plotted as triangle and with different residual water signals as circles. 10 Monte-Carlo entries (x-axis from 3 to 12) display the 10 different residual water lineshapes. Good convergence of SD to the estimated CRLB is reported for MEMS and water-free MESS. CRLBs systematically ill-represent measurement-SD when residual water is included in MESS schemes at TE1 and TE2.

Figure S6: Uncertainty on T_2s from in-silico Monte-Carlo experiments with and without realistic residual water with estimated SD in blue and estimated CRLB in red. MEMS/PES estimates with (cross) and without (square) $\sqrt{3}$-correction to equalize scantime. The MESS 3-TE scheme without residual water is plotted as triangle and with different residual water signals as circles. 10 Monte-Carlo entries (x-axis from 3 to 12) display the 10 different residual water lineshapes. Good convergence of SD to the estimated CRLB is reported for MEMS and water-free MESS. CRLBs systematically ill-represent measurement-SD when residual water is included in MESS schemes at TE1 and TE2.

Figure S7: Cohort average datasets for MESS 3-TE schemes with different TE/Δ setups. Short Δ=8ms (top, short TE batch) and long Δ=56ms (bottom, long TE batch) are reported in the time domain (real part, left) and frequency domain (absorption part, right). The measurements were repeated five times for each volunteer and setup: average spectra are portrayed in black, the variability (±1 SD) in red. Substantial variance throughout the fids and a large frequency range is visible for TE1 and TE2 in the upper panel with very short acquisition periods. (water signal effect spread out in frequency by zero-filling). For the long Δ scheme this effect is substantially reduced.
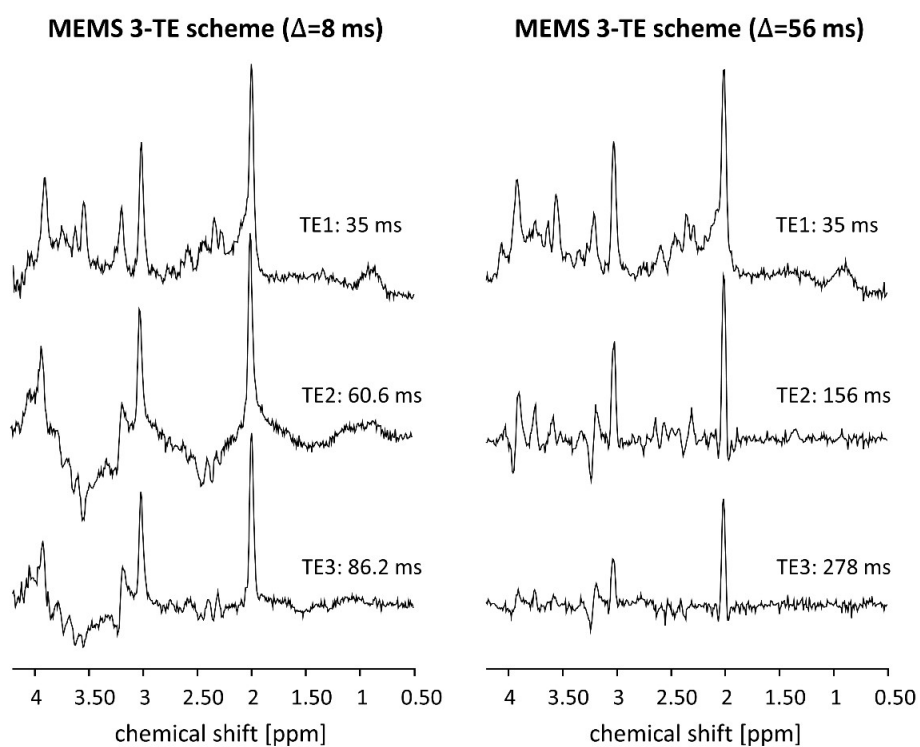
Figure S8: Exemplary MEMS datasets for the 3-TE schemes with the same TE/Δ setups as in Figure S7. Short Δ=8 ms (left, healthy subject #2) and long Δ=56 ms (right, healthy subject #2).
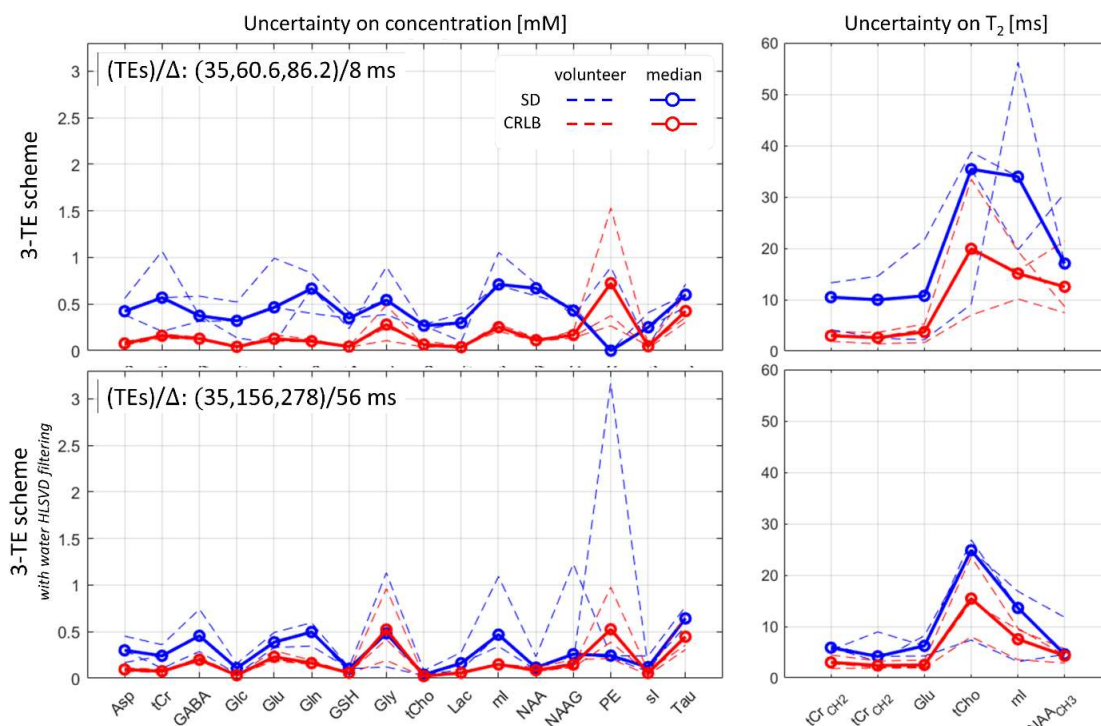
Figure S9: Repeatability of MESS upon different TE/Δ setups. (Top) short TE batch (Δ=8 ms), (bottom) long TE batch (Δ=56 ms). Within-session metabolite SDs (blue) and CRLB (red) are plotted for concentrations on the left and for T_2s on the right. The long TE batch (bottom) allows the removal of residual water via HLSVD filtering (Figure S5). Dashed lines: individual volunteer data, bold lines: median trace for the cohort. It is evident that CRLBs underestimate SDs consistently for all metabolites for Δ=8 ms (upper panel), as already shown in Figure 10. However, better water signal handling in the Δ=56 ms case (bottom panel) yields less divergence between CRLB and SD for both, concentrations and T_2s.

1. Hardware	
a. Field strength [T]	3 T
b. Manufacturer	Siemens (Erlangen, GE)
c. Model (software version if available)	Prisma (VE 11)
d. RF coils: nuclei (transmit/receive), number of channels, type, body part	Siemens receiver head coil, 64 channel
e. Additional hardware	N/A
2. Acquisition	
a. Pulse sequence	Modified semi-LASER sequence with addition of 2 Mao π pulses for multiple refocusing. Recordings as echo-train in single short (MESS) or repeated independent measures. A traditional semi-LASER sequence is used for comparison.
b. Volume of Interest (VOI) locations	Occipital Cyngulate Cortex (OCC)
c. Nominal VOI size [cm^3, mm^3]	20x31x17mm^3
d. Repetition Time (TR), Echo Time (TE) [ms, s]	TE batch #1: 35/60.6/86.2 ms TE batch #2: 140/213.6/287.2 ms TR = 2000 ms
e. Total number of excitations or acquisitions per spectrum In time series for kinetic studies i. Number of Averaged spectra (NA) per time-point ii. Averaging method (e.g. block-wise or moving average) iii. Total number of spectra (acquired / in time-series)	64 averages. Simple averaging after MC difference to untangle water and metabolite signal.
f. Additional sequence parameters (spectral width in Hz, number of spectral points, frequency offsets) If STEAM:, Mixing Time (TM) If MRSI: 2D or 3D, FOV in all directions, matrix size, acceleration factors, sampling method	Spectral width: 4 kHz, Number of spectral points: 4096. TE batch #1: 35/60.6/86.2 ms, with Δ = 8 ms sampled with 32 datapoints. TE batch #2: 140/213.6/287.2 ms, with Δ = 32 ms sampled with 128 datapoints.
g. Water Suppression Method	Metabolite Cycling (MC)
h. Shimming Method, reference peak, and thresholds for "acceptance of shim" chosen	Brain shimming routine. Acceptance of shim if < 5 Hz
i. Triggering or motion correction method (respiratory, peripheral, cardiac triggering, incl. device used and delays)	N/A
3. Data analysis methods and outputs	
a. Analysis software	Matlab: data pre-processing and analysis jMRUI: preprocessing

	FitAID: fitting
	Python: statistics and plotting
b. Processing steps deviating from quoted reference or product	N/A
c. Output measure (e.g., absolute concentration, institutional units, ratio)	Absolute concentration in millimolar and T_2s in millisecond
d. Quantification references and assumptions, fitting model assumptions	Reference: water signal simultaneously acquired via MC. Model Assumptions: simultaneous 2D fitting with time-domain modeling and frequency domain χ^2 minimization. 1st dimension: chemical shift 2nd dimension: simultaneous T_2 fit via mono-exponential signal decay. The fit along the 2nd dimension is run on a number of data points equal to the number of TEs (i.e., three or six). Model broadening assumed Voigt-line shapes: Lorentzian component models metabolic T_2 times, Gaussian component primarily represents non-refocusable line broadening and is assumed to be identical for all metabolites.
4. Data Quality	
a. Reported variables (SNR, Linewidth (with reference peaks))	Water linewidth (Gaussian) < 5Hz
b. Data exclusion criteria	N/A
c. Quality measures of postprocessing Model fitting (e.g. CRLB, goodness of fit, SD of residual)	Cramer Rao Lower Bounds (CRLBs) were taken as the measure for achievable precision. Goodness of the fit reported via residue display and monitored in fitting phase.
d. Sample Spectrum	Figure 3 and Figure 7

Table-S3: MRSinMRS checklist[12].

<u>Supporting Information References</u>

1. Marjańska M, McCarten JR, Hodges J, et al. Region-specific aging of the human brain as evidenced by neurochemical profiles measured noninvasively in the posterior cingulate cortex and the occipital lobe using 1H magnetic resonance spectroscopy at 7 T. *Neuroscience*. 2017;354:168-177. doi:10.1016/j.neuroscience.2017.04.035

2. Hoefemann M, Bolliger CS, Chong DGQ, van der Veen JW, Kreis R. Parameterization of metabolite and macromolecule contributions in interrelated MR spectra of human brain using multidimensional modeling. *NMR Biomed*. 2020;33(9):e4328. doi:10.1002/nbm.4328

3. Murali-Manohar S, Borbath T, Wright AM, Soher B, Mekle R, Henning A. T2 relaxation times of macromolecules and metabolites in the human brain at 9.4 T. *Magn Reson Med*. 2020;84(2):542-558. doi:10.1002/MRM.28174

4. Mekle R, Mlynarik V, Gambarota G, Hergt M, Krueger G, Gruetter R. MR spectroscopy of the human brain with enhanced signal intensity at ultrashort echo times on a clinical platform at 3T and 7T. *Magn Reson Med*. 2009;61(6):1279-1285. http://www.ncbi.nlm.nih.gov/pubmed/19319893

5. Deelchand DK, Moortele PF Van De, Adriany G, et al. In vivo 1H NMR spectroscopy of the human brain at 9.4 T: Initial results. *J Magn Reson*. 2010;206(1):74-80. doi:10.1016/j.jmr.2010.06.006

6. Marjańska M, Auerbach EJ, Valabrègue R, Van de Moortele PF, Adriany G, Garwood M. Localized 1H NMR spectroscopy in different regions of human brain in vivo at 7 T: T2 relaxation times and concentrations of cerebral metabolites. *NMR Biomed*. 2012;25(2):332-339. doi:10.1002/NBM.1754

7. Terpstra M, Ugurbil K, Tkac I. Noninvasive quantification of human brain ascorbate concentration using1H NMR spectroscopy at 7 T. *NMR Biomed*. 2010;23(3):227-232. doi:10.1002/nbm.1423

8. Wyss PO, Bianchini C, Scheidegger M, et al. In vivo estimation of transverse relaxation time constant (T2) of 17 human brain metabolites at 3T. *Magn Reson Med*. 2018;80:452-461. doi:10.1002/mrm.27067

9. Träber F, Block W, Lamerichs R, Gieseke J, Schild HH. 1H Metabolite Relaxation Times at 3.0 Tesla: Measurements of T1 and T2 Values in Normal Brain and Determination of Regional Differences in Transverse Relaxation. *J Magn Reson Imaging*. 2004;19:537-545. doi:10.1002/jmri.20053

10. Mlynarik V, Gruber S, Moser E. Proton T (1) and T (2) relaxation times of human brain metabolites at 3 Tesla. *NMR Biomed*. 2001;14(5):325-331.

11. Deelchand DK, McCarten JR, Hemmy LS, Auerbach EJ, Eberly LE, Marjańska M. Changes in the intracellular microenvironment in the aging human brain. *Neurobiol Aging*. 2020;95:168-175. doi:https://doi.org/10.1016/j.neurobiolaging.2020.07.017

12. Lin A, Andronesi O, Bogner W, et al. Minimum Reporting Standards for in vivo Magnetic Resonance Spectroscopy (MRSinMRS): Experts' consensus recommendations. *NMR Biomed*. 2021;34(5):e4484. doi:https://doi.org/10.1002/nbm.4484

Simultaneous concentration and T_2 mapping of brain metabolites by multi-echo spectroscopic imaging

Rudy Rizzo[1,2], Angeliki Stamatelatou[3], Arend Heerschap[3], Tom Scheenen[3], Roland Kreis[1,2].

1. Magnetic Resonance Methodology, Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland

2. Translational Imaging Center, sitem-insel, Bern, Switzerland

3. Department of Medical Imaging, Radboud Medical Center, Nijmegen, The Netherlands

Synopsis

A multi-parametric MR Spectroscopic Imaging (MRSI) experiment named Multi-Echo Single-Shot MRSI (MESS-MRSI) deploys partially sampled multi-echo trains from single readouts combined with simultaneous multi-parametric model fitting to produce 2D metabolite-specific T_2 and concentration maps on a 6x6 frame and with $1.34\,\mathrm{cm}^3$ resolution in only 7 minutes. It was tested in-vivo on a cohort of 5 subjects. Cramer-Rao Lower-Bounds (CRLBs) are used as the measure of performance. The novel scheme was compared with the (1) traditional Multi-Echo Multi-Shot (MEMS) method and (2) a truncated version of MEMS, which mimics the MESS acquisition (MESS-mocked). Results extended former findings for single voxel measurements with average improvements in CRLB ranging from 17-45% for concentrations and 12-23% for T_2s.

Summary of the main findings

Multi-Echo Single-Shot MRSI of the brain combined with 2D model fitting provides simultaneous concentration and T_2 maps in a 7-minute scan time, with average CRLB improvements between 12 and 45% compared to traditional Multi-Echo Multi-Shot MRSI.

1. Introduction

MR Spectroscopic Imaging (MRSI) aims to map spatial distributions of metabolite concentrations, which reflect tissues' biochemistry and provide insight into functionality and pathophysiology [100]. Metabolite relaxation rates, which mirror cellular and sub-cellular microenvironments, could hold additional valuable information but are rarely acquired within clinical scan times [101, 102, 103]. For example, the relaxation times of the neuronal marker *N-Acetylaspartate* (NAA) reflect the neuronal microenvironment and may operate as an independent marker of neurodegeneration or inflammation [39]. So far, there is clear evidence for age-dependence of metabolite relaxation times [40] but also altered values in pathologies such as multiple sclerosis [41], Alzheimer's disease [42], and cancer [43, 44].

Here, we extend a novel single-voxel acquisition scheme that acquires multi-TE data from single readouts twinned to a bi-dimensional fitting process [104] to produce metabolite-specific T_2 and concentration maps and the related CRLBs within clinical scan time.

2. Methods

A metabolite-cycled 2D-MRSI-sLASER scheme with weighted Cartesian k-space encoding was optimized to acquire three consecutive spin-echoes in one scan (multi-echo single-shot, MESS [104]). The 1st spin-echo was acquired at the shortest TE possible and was recorded as FID. The 2nd and 3rd spin-echoes were instead recorded as partially sampled full echoes, where the last recording window lasts to achieve an overall 1-second acquisition length (i.e., the sum of the 3 readouts). The echo train was generated by extending the sLASER block with two optimized slice-selective $Mao\ \pi$ pulses with 1.5-fold slice thickness. The acquisition setup explored a 16x16 FOV grid with 200x160 mm^2 resolution. The resolution of the MRSI-VOI was 80x60x15 mm^3. Sequence timings were set at $TR/(TEs)$ 1600/(35,156,278) ms. Spectra were recorded with spectral width (SW) of 4 kHz, and acquired with datapoints $total/(TEs)$ 4096/(224/448/3424) for a total of 4 weighted acquisitions in an overall 7-minute scan time. Measurements were performed on a 3T MR system (Siemens) with a 64-channel head coil.

Five healthy volunteers were examined with supraventricular VOI positioning. Next to MESS, we acquired (1) traditional multi-echo multi-shot MRSI sampling of three fully sampled spin-echoes (MEMS, 21 minutes scan time), and (2) a truncated version of the MEMS acquisition, which mimics the MESS setup (MESS-mocked). All scans were performed in accordance with the competent ethical review boards.

Simultaneous 2D fit ran in FitAID [16] with time-domain model and frequency domain χ^2-minimization. The half-echo of the shortest TE was fitted with the full-echo recorded for later TEs, including the extended tail of the last echo that provides resolution information for the whole echo train. FID and 2nd echo of MESS were zero-filled to match the duration of the 3rd echo for visual purposes.

The fitting model assumed Gaussian line-broadening with resulting Voigt-line shape where the Lorentzian component represents T_2 contribution. A 16-metabolite basis set was simulated in VeSPA [13] accounting for: Asp: aspartate, GABA: γ-aminobutyric acid, Glc: glucose, Gln: glutamine, GSH: glutathione, Gly: glycine, tCho: total choline (1:1 sum of glycerophosphorylcholine + phosphorylcholine), Lac: lactate, NAAG: N-acetylaspartylglutamate, PE: phosphoethanolamine, sI: syllo-inositol, Tau: taurine, tCr: total creatine (1:1 sum of creatine + phosphocreatine), Glu: glutamate, mI: myo-inositol, and NAA: N-acetylaspartate. Prior knowledge modeling of the macromolecular background (MMBG) pattern was simulated as a sum of overlapping densely, and equally spaced Voigt lines [23]. T_2s were fitted freely for five major metabolites (tCr, respectively for CH$_2$ and CH$_3$ moieties, singlets of tCho and NAA resonating

at about 3.2 ppm and 2.008 ppm, respectively, and mI) and MMBG, while T_2s were linked for all other metabolites (global T_2). Tissue concentrations were calculated referencing to the co-measured water signal, with T_1 corrections from the literature [23] and white (WM), gray (GM) matter, and CSF segmentation [105] to account for tissue-specific water relaxation, and tissue fraction corrections [3]. CRLBs were taken as the measure for achievable precision. To compare equivalent total experimental time, CRLBs of MEMS and MESS-mocked were corrected by $\sqrt{3}$.

The precision gain (CRLBs) of MESS was tested with statistical inference considering distributions of concentrations and T_2s on the cohort of 5 volunteers. Two subsets of either prevalent WM or GM voxels from the acquired 6x6 VOI grid were selected according to tissue segmentation (fractional volume of parenchymal water of WM or GM > 70%) and grouped across subjects to be used as the cohort population. Statistical evaluations were run on GraphPad Prism 9.4. Distributions of estimates and prediction errors (CRLBs) were compared across methods assuming *unpaired Welch t-test* with *Holm-Sidak* correction. No statistical inference regarding WM/GM differences was investigated.

3. Results & Discussion

Fig.7.1 illustrates the acquisition setup, MEMS and MESS data, fit, and residues for voxels in prevalently WM and GM. Short-TE MESS spectra show limited spectral resolution. TE_2 and TE_3 spectra are automatically phased to display positive absorption component for NAA and feature linear phase offsets (evident for tCho singlet at 3.185 ppm) due to partial-echo acquisitions. The overall quality of the fits is good. Residues are limited and follow a white-noise distribution for MEMS and MESS at TE_3. Signal truncation and zero-filling for TE_1 and TE_2 create visual ripples but acceptable residues.

Fig.7.2 and Fig.7.3 display estimated concentrations and T_2 values, respectively, for a subset of metabolites, reported spectrum-by-spectrum throughout the VOI. The evaluation considers fitting on a zero-filled k-space grid (i.e., 13x13 voxels) with cropped voxels at the edges (2 voxels per edge: 9x9 voxels) to minimize partial volume effects. Values and trends are comparable. The figure includes fit uncertainties (CRLBs). Precision for concentrations and T_2 estimates is equivalent or better for MESS (red). As expected, precision for MESS-mocked (yellow) scores the worst for most cases since it considers both a longer acquisition time (like MEMS) and signal truncation (i.e., reduced frequency resolution for TE_1 and TE_2). Results for a second healthy volunteer are reported in Supporting Figure 7.7 and 7.8.
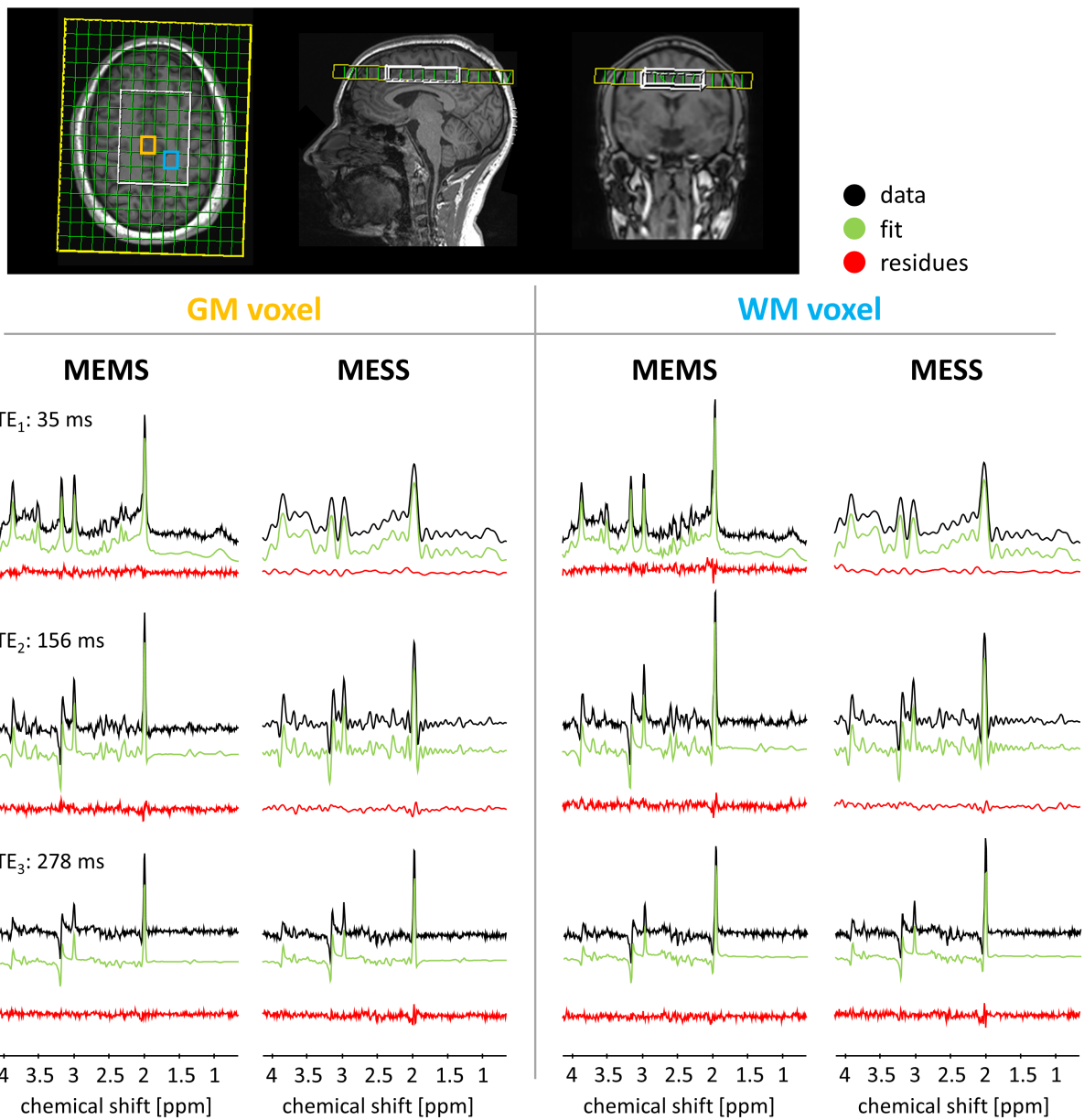
Figure 7.1: (Top) Display of MRSI FOV (green) and VOI (white) overlapped to a T_1w-MPRAGE anatomical reference from one acquisition. (Bottom) Acquired data (black), fitted model (green), and residues (red). MEMS and MESS acquisition are reported for two voxels, one in prevalent WM tissue (blue) and the other in prevalent GM tissue (orange).
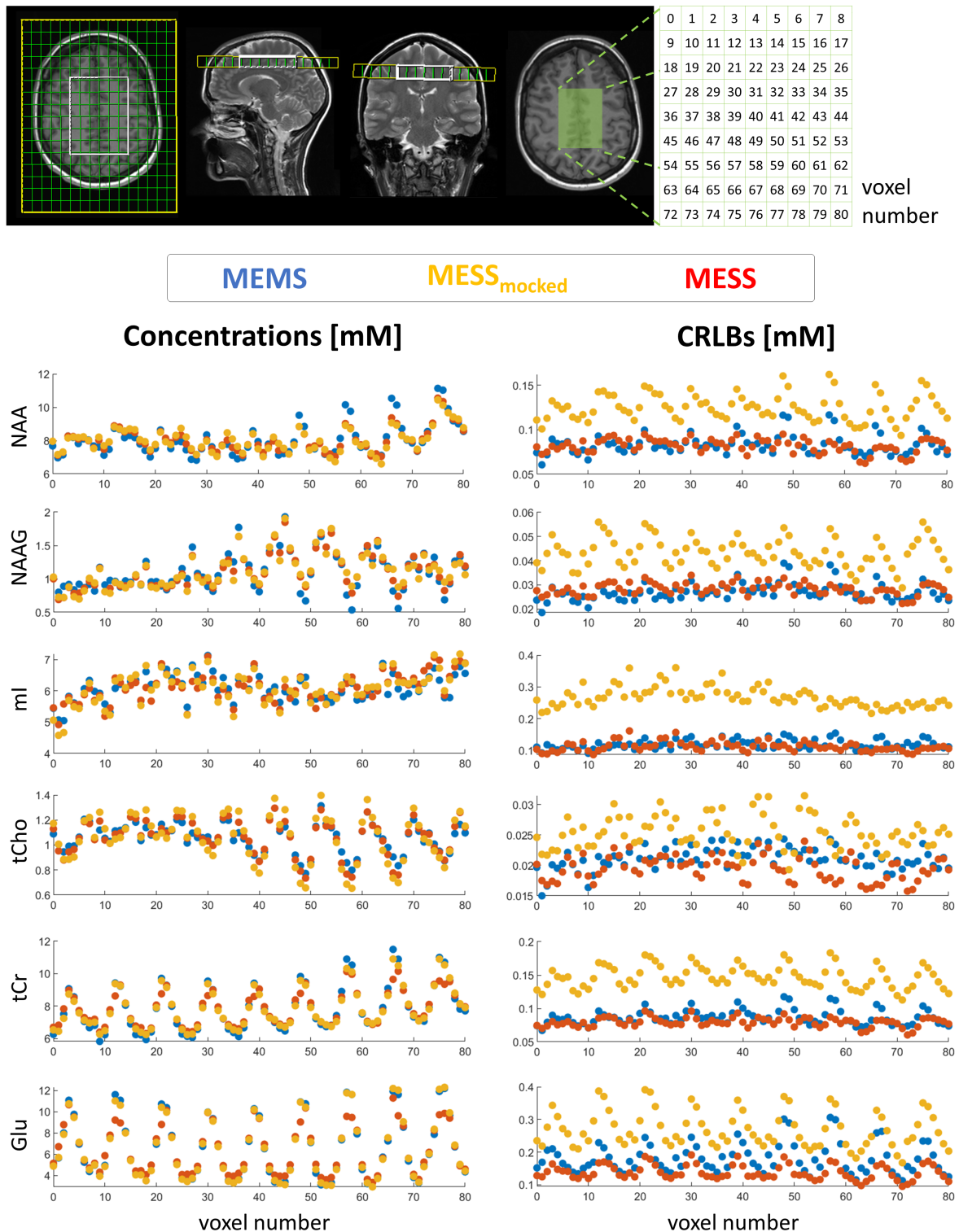
Figure 7.2: Estimates and uncertainties (CRLBs) for metabolite concentrations. (Top) T_1w and T_2w anatomical images display the MRSI FOV (green) with the voxel numbering referenced to the VOI (white). (Bottom-left) Estimated concentrations via MEMS (blue), MESS-mocked (orange), and MESS (red) overlap nicely, and their oscillation throughout the VOI reflects WM and GM variation. (Bottom-right) CRLBs report higher precision for MESS vs. MEMS.
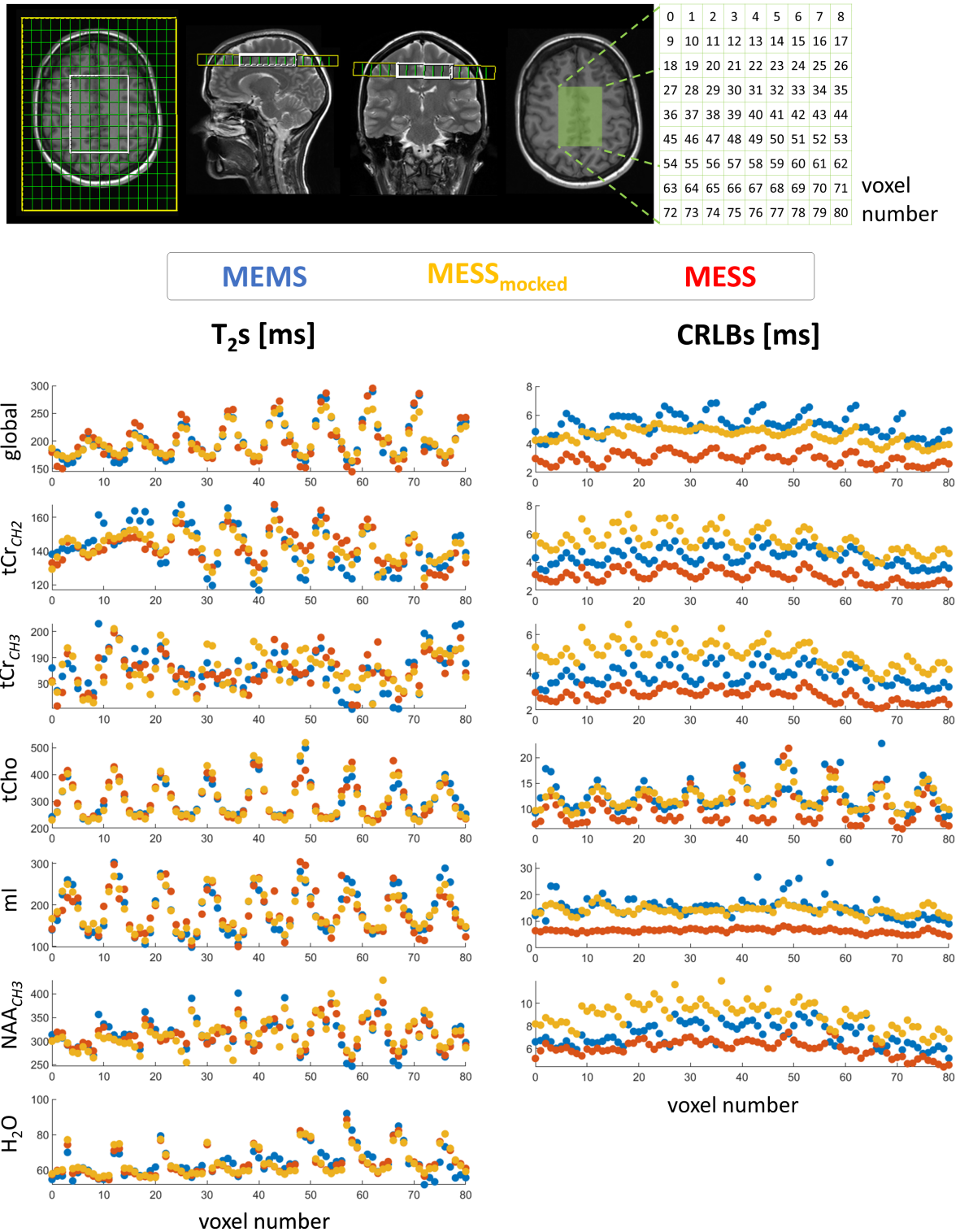
Figure 7.3: Estimates and uncertainties (CRLBs) for metabolite T_2s. (Top) T_1w and T_2w anatomical images display the MRSI FOV (green) with the voxel numbering referenced to the VOI (white). (Bottom-left) Estimated concentrations via MEMS (blue), MESS-mocked (orange), and MESS (red) overlap nicely, and their oscillation throughout the VOI reflects WM and GM variation. (Bottom-right) CRLBs report higher or comparable precision for MESS vs. MEMS.

A comparison of concentration and T_2 maps across two volunteers is reported in Fig.7.4 and Fig.7.5 for a subset of metabolites. MESS-mocked replicates MEMS adequately; thus, signal truncation is found suitable for concentration and T_2 mapping. MESS yields maps that overall agree with the 3-fold longer MEMS technique here considered the gold standard for comparison.

The distribution of concentrations and T_2s between GM and WM is reported in general agreement with other published studies which reported GM/WM differences [23, 1, 106, 107, 108, 109, 110]. Supporting Fig.7.9 displays the variation of concentration levels and T_2 rates across the three methodologies via boxplots representing the distributions for WM and GM cohorts. Neither regional differences nor statistical comparisons between levels of metabolite content or T_2s in GM vs. WM were considered. Supporting Tab.7.10 and 7.11 report cohort average and standard deviation compared to reported values in the literature for metabolite concentrations and T_2 rates, respectively.

Overall tCr levels are confirmed to be higher in GM than in WM since, biochemically, GM is expected to contain more neuronal cell bodies (i.e., WM features myelin layers) with more prevalence of mitochondria [111]. NAA presence is not limited to neuronal somata but is also found in synapto-soma subfractions that may account for the higher NAA levels found in GM [111]. Despite NAAG, like NAA, being considered a neuronal compound, the literature and the current study report higher concentration levels in WM given its abundance throughout the myelinated axons [107]. Glutamate is expected to be higher in GM since this neurochemical is highly concentrated within the neuronal body cell [106], as confirmed by our results. Further considerations are needed to account for other structural changes characteristic of WM/GM differences that go beyond cellular argumentation.

Relaxation rates for water align with the WM/GM contrast reported in the MRI literature where gold-standard techniques are used as estimators [112]. Longer T_2 rates for NAA are found for prevalent WM areas, whereas tCho and mI display longer T_2 rates in GM. These findings are supported by former studies [108, 110], confirming and highlighting that the local tissue composition and functional organization in the region of interest plays a major role in T_2 relaxation. Therefore, tissue composition correction with region-specific T_2 is advised.
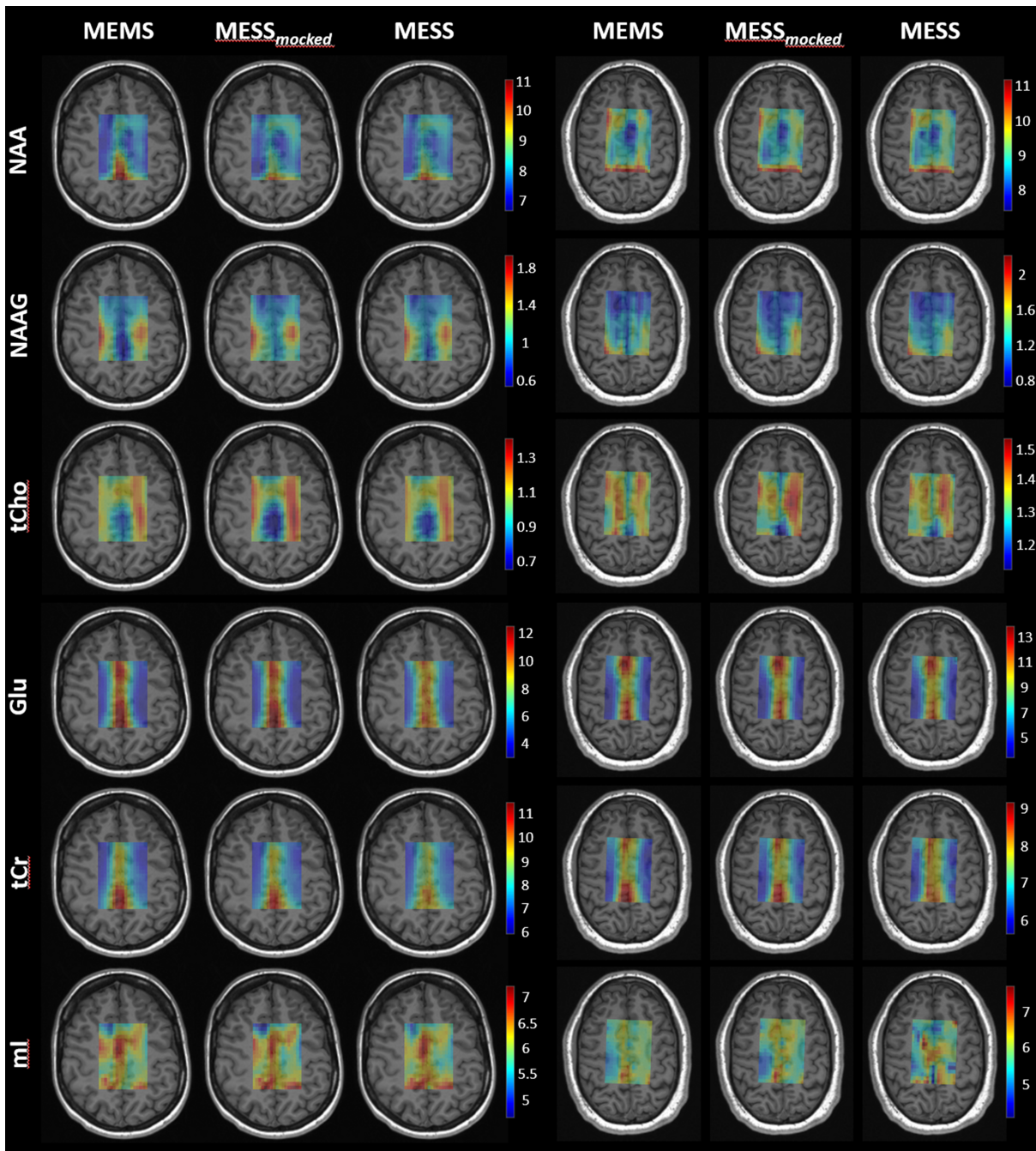
Figure 7.4: Concentration maps in milli-molar units [mM] for a subset of metabolites displayed for two subjects and the three methods (MEMS, MESS-mocked, and MESS). NAA: N-acetylaspartate, NAAG: N-acetylaspartylglutamate, tCho: total choline, Glu: glutamate, tCr: total creatine, and mI: myo-inositol. Maps are displayed with zero-filling in the spatial domain with cropped voxels at the edges (18x18 pixels).
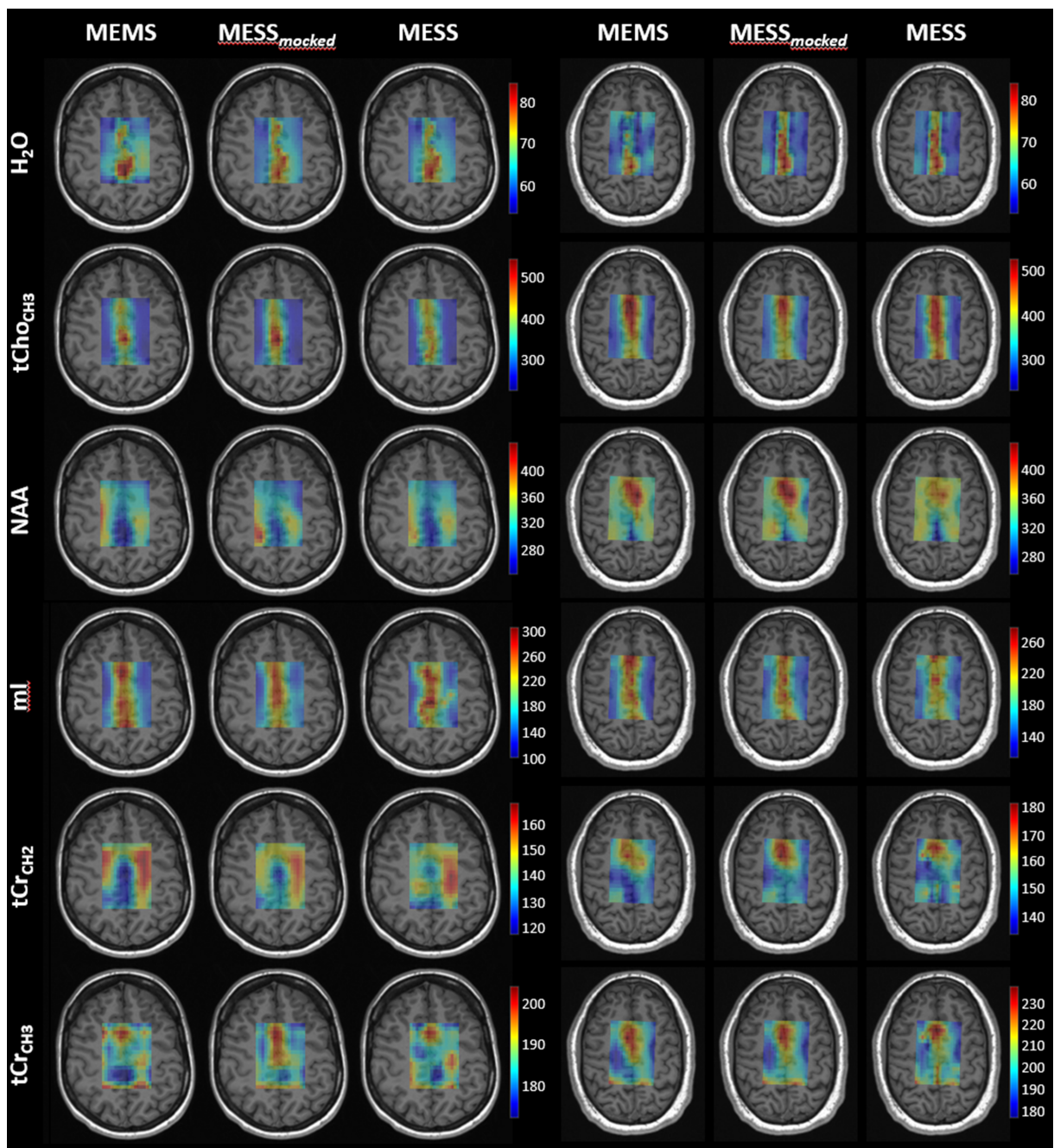
Figure 7.5: T_2 time maps in millisecond units [ms] of parenchymal water and a subset of metabolites, reported for two subjects and the three methodologies. H_2O: parenchymal water, $tCho_{CH3}$: total choline singlet at 3.2ppm, NAA: N-acetylaspartate singlet at 2ppm, mI: myo-Inositol, tCr_{CH2}: total creatine methylene resonance at 3.9ppm and tCr_{CH3}: tCr methyl resonance at 3ppm. Maps are displayed with zero-filling in spatial domain with cropped voxels at the edges (18x18 pixels).

Supporting Fig.7.9 reports the distribution of estimates considering WM/GM segmentation across the cohort of healthy subjects. Statistical deviations between methodologies are not systematically observed. Values are in line with estimates from former studies, which are included for overall comparison in Supporting Tab.7.10 and7.11, for concentrations and T_2s, respectively. Relevant WM/GM differences can be visually depicted for concentrations of Glu, NAA, GSH, Gln, Tau, and GABA, as well as for global T_2 and T_2s of tCho$_C$$H3$, mI and, as expected, water. However, no statistical analysis across WM/GM differences was investigated since out of the scope of the methodology investigation.

Fig.7.6 reports the cohort analysis of the methods' precision. MESS yields, on average, precision increments for concentrations ranging from 17% to 45% with significant differences across MESS vs. MEMS, especially evident for concentrations. An average improvement for T_2s between 12% to 23% is visible and compares to former single-voxel similar experiments [104]. As expected, MESS-mocked shows the lowest precision given signal truncation and 3-fold longer acquisition.

4. Conclusions

The novel MESS-MRSI approach yields metabolite-specific T_2 maps. It provides increased precision or inversely shorter experimental time (3-fold) compared to traditional approaches while achieving comparable accuracy of estimates, extending results from single-voxel experiments. This promises to be useful in functional or multi-parametric MRS, where concentrations provide insight into functionality and pathophysiology, and relaxation rates act as additional potential biomarkers of abnormality, mirroring information on cellular microenvironment.
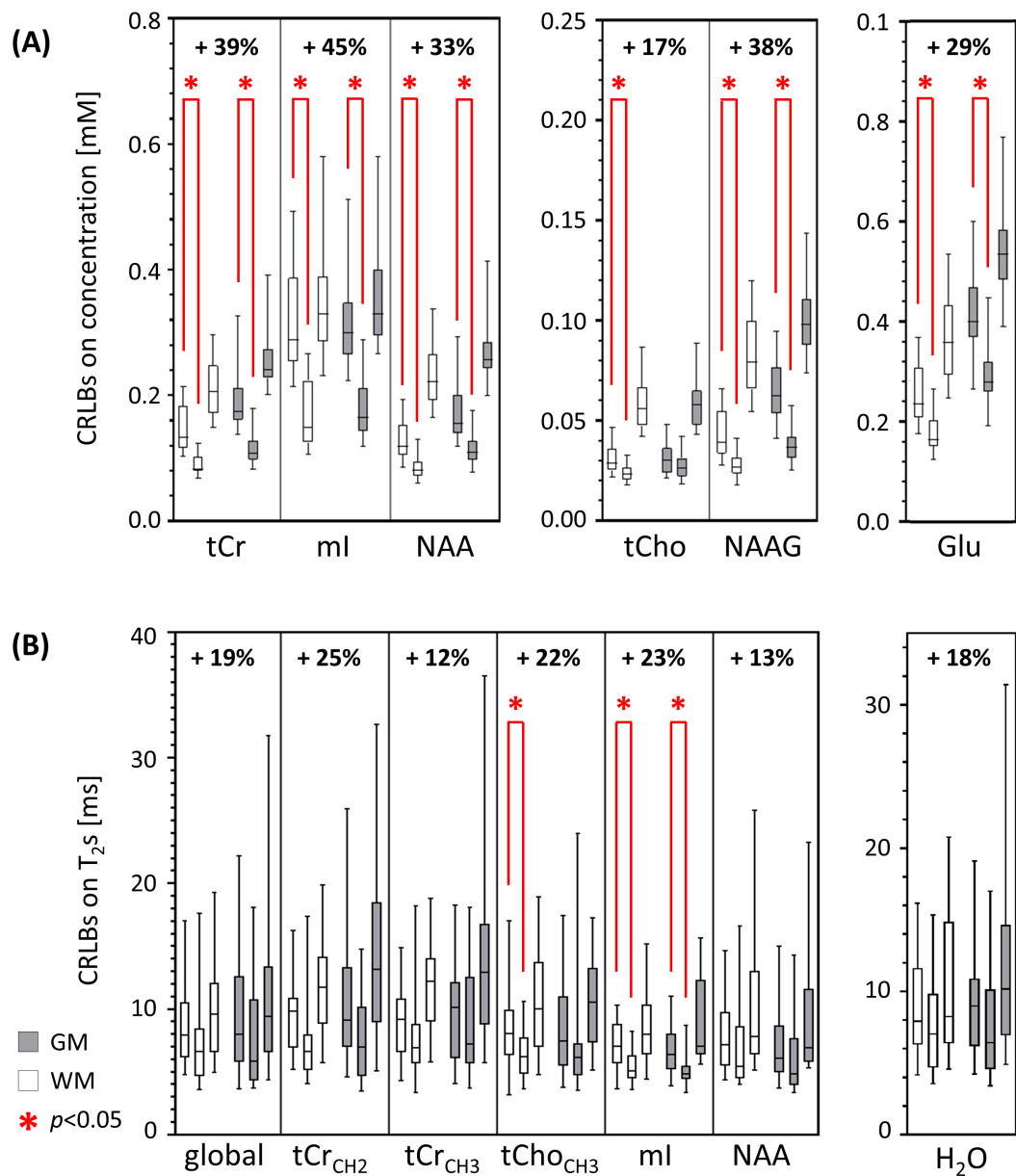
Figure 7.6: Comparison of methods' precision across the cohort of volunteers/voxels. Boxplots display the distributions for white (WM) and gray (GM) matter voxels for CRLBs of (A) concentrations and (B) T_2s. Results are reported for metabolite subsets, such as in Fig.7.4 and Fig.7.5. Methodologies are compared for each metabolite in three candles: MEMS, MESS, and MESS-mocked from left to right. The gain in precision of MESS vs. MEMS as averaged across the cohorts and tissue type is reported numerically on top. Statistical relevance is reported for MESS vs. MEMS comparison only (*, red).

5. Supporting Figures
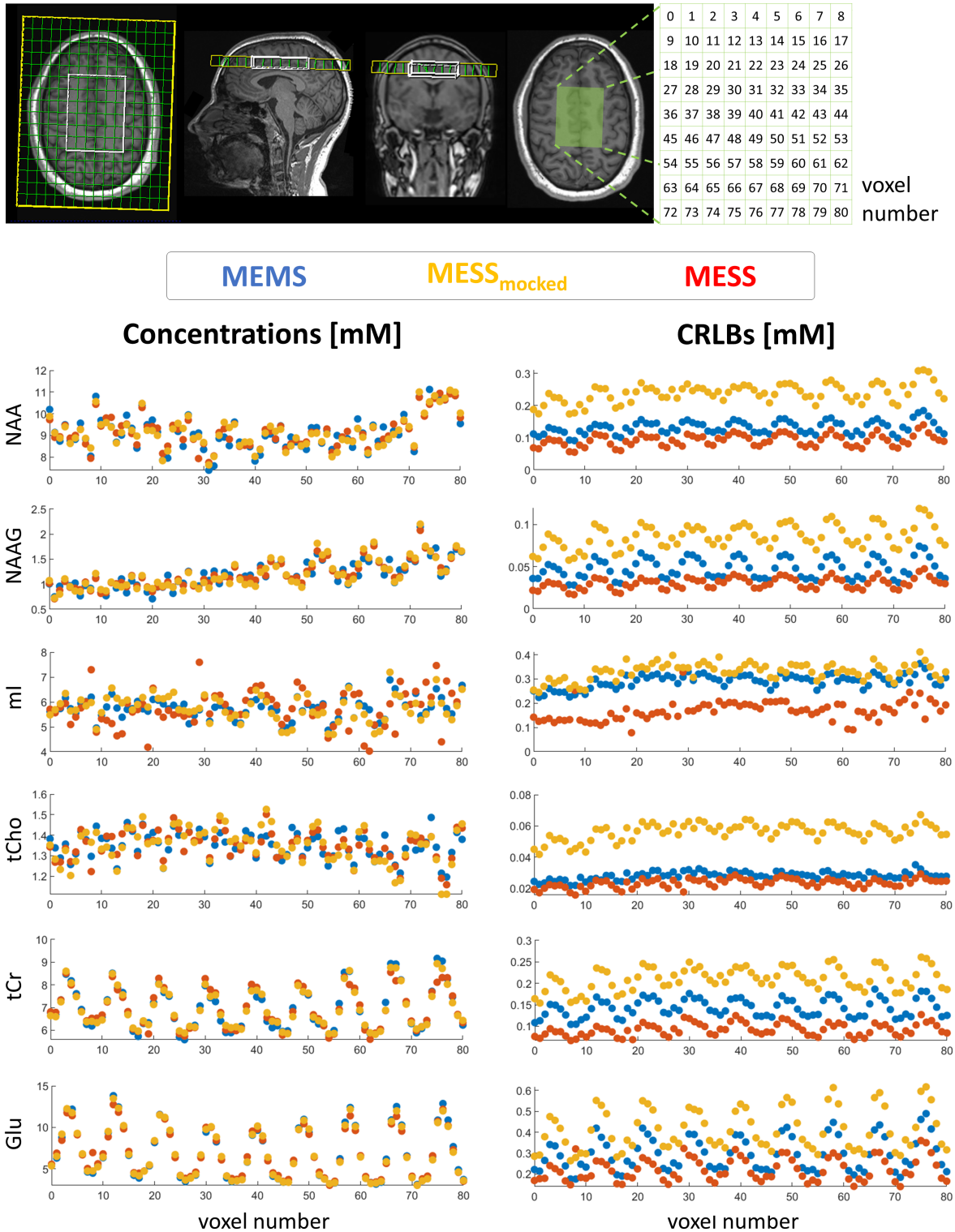
Supporting Figure 7.7: Estimates and uncertainties (CRLBs) for metabolite concentrations. (Top) $T_1 w$ anatomical images display the MRSI FOV (green) with the voxel numbering referenced to the VOI (white). (Bottom-left) Estimated concentrations via MEMS (blue), MESS-mocked (orange), and MESS (red) overlap nicely, and their oscillation throughout the VOI reflects WM and GM variation. (Bottom-right) CRLBs report higher or comparable precision for MESS vs. MEMS.
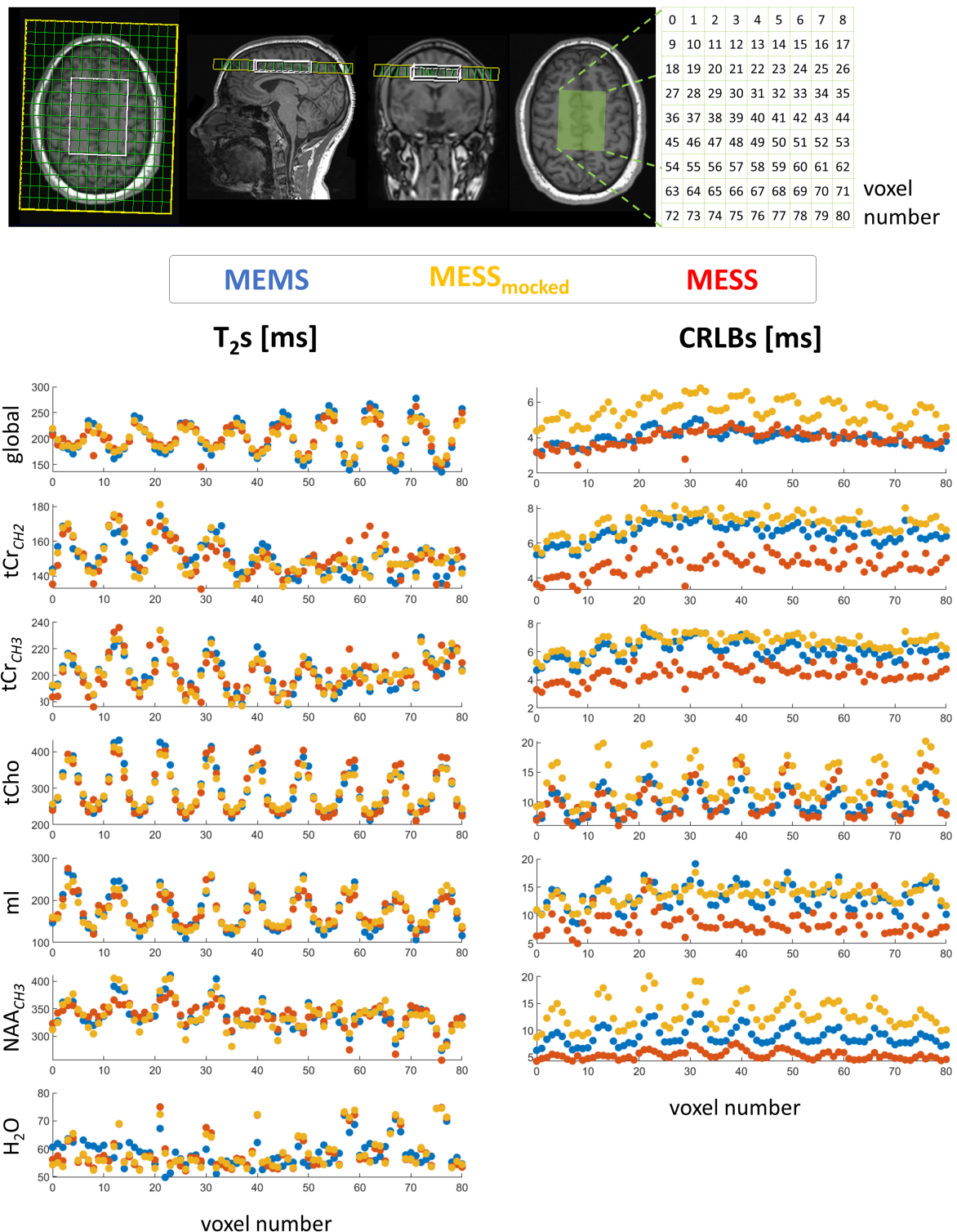
Supporting Figure 7.8: Estimates and uncertainties (CRLBs) for metabolite T_2s. (Top) T_1w anatomical images display the MRSI FOV (green) with the voxel numbering referenced to the VOI (white). (Bottom-left) Estimated concentrations via MEMS (blue), MESS-mocked (orange), and MESS (red) overlap nicely, and their oscillation throughout the VOI reflects WM and GM variation. (Bottom-right) CRLBs report higher or comparable precision for MESS vs. MEMS.
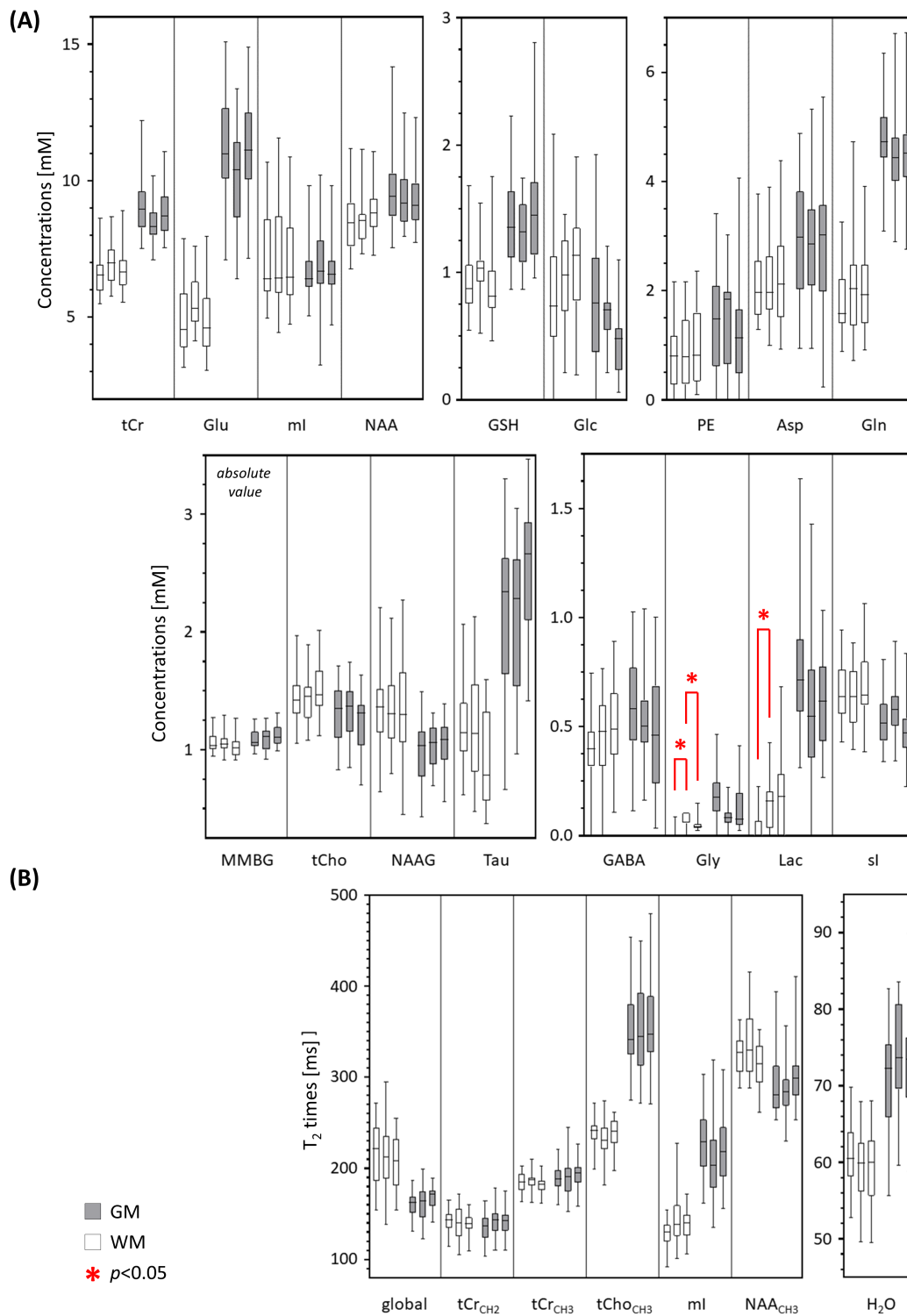
Supporting Figure 7.9: Distributions of estimated concentrations (A) and T_2 rates (B) grouped for WM and GM. Distributions do not show significant differences in spread or biases across methods and are well aligned with GM/WM values in the literature [23, 1, 106, 107, 108, 109, 110]. Statistical relevance is reported for MESS vs. MEMS comparison only (*, red).

	MEMS		MESS		MESS-mocked		[1]		[2]		[3]
	WM	**GM**	**WM**	**GM**	**WM**	**GM**	**WM**	**GM**	**WM**	**GM**	**GM**
tCr	6.62 ± 0.77	9.09 ± 1.08	7 ± 0.85	8.42 ± 0.77	6.79 ± 0.82	8.9 ± 0.96	6.2 ± 0.9	9.7 ± 0.9	5.16 ± 0.45	8.87 ± 0.56	8.89 ± 0.58
Glu	4.87 ± 1.3	11.32 ± 1.8	5.56 ± 0.97	10.08 ± 1.82	4.94 ± 1.36	11.25 ± 1.68	4.5 ± 0.6	9.4 ± 1.0	5.06 ± 0.57	10.37 ± 1.06	10.74 ± 0.62
mI	7.07 ± 1.69	6.74 ± 1.1	6.95 ± 1.76	6.88 ± 1.39	6.99 ± 1.77	6.81 ± 1.12	4.3 ± 0.5	7.7 ± 0.9			7.12 ± 0.58
NAA	8.42 ± 1.02	9.60 ± 1.32	8.53 ± 0.87	9.40 ± 1.12	8.81 ± 0.84	9.37 ± 1.08	8.5 ± 1.0	10.4 ± 0.8			10.82 ± 0.62
Asp	1.06 ± 0.09	2.96 ± 1.04	2.16 ± 0.72	2.97 ± 1.02	2.24 ± 0.92	2.86 ± 1.2					1.86 ± 0.29
GABA	0.38 ± 0.18	0.59 ± 0.23	0.44 ± 0.2	0.54 ± 0.22	0.49 ± 0.22	0.45 ± 0.26					0.78 ± 0.25
Glc	0.79 ± 0.46	0.74 ± 0.49	0.89 ± 0.37	0.68 ± 0.31	1.08 ± 0.41	0.43 ± 0.26					
Gln	1.78 ± 0.54	4.79 ± 0.74	2.06 ± 0.87	4.5 ± 0.85	2 ± 0.71	4.57 ± 0.77					3.18 ± 0.33
GSH	0.95 ± 0.3	1.41 ± 0.35	1.03 ± 0.19	1.31 ± 0.25	0.9 ± 0.31	1.52 ± 0.45					1.07 ± 0.8
Gly	0.01 ± 0.02	0.05 ± 0.02	0.07 ± 0.03	0.09 ± 0.05	0.18 ± 0.12	0.13 ± 0.1					
tCho	1.44 ± 0.21	1.3 ± 0.26	1.43 ± 0.2	1.33 ± 0.25	1.52 ± 0.22	1.23 ± 0.25	1.5 ± 0.1	1.8 ± 0.6			1.21 ± 0.12
Lac	0.04 ± 0.07	0.74 ± 0.27	0.14 ± 0.11	0.62 ± 0.33	0.19 ± 0.18	0.62 ± 0.22					
NAAG	1.36 ± 0.33	0.96 ± 0.27	1.33 ± 0.31	1.03 ± 0.18	1.34 ± 0.38	1.04 ± 0.23					1.17 ± 0.12
PE	0.8 ± 0.57	1.38 ± 0.99	0.9 ± 0.7	1.43 ± 0.87	0.94 ± 0.66	1.26 ± 1.08					2.12 ± 0.21
sI	0.67 ± 0.13	0.53 ± 0.12	0.63 ± 0.15	0.58 ± 0.12	0.7 ± 0.18	0.48 ± 0.13					0.26 ± 0.21
Tau	1.18 ± 0.33	2.16 ± 0.66	1.19 ± 0.43	2.11 ± 0.63	0.9 ± 0.39	2.52 ± 0.61					1.86 ± 0.25
MMBG	modeled						not reported		not reported		measured
T2 correction	measured						measured		measured		literature
T1 correction	literature						measured		literature		literature
reference	water, FitAID						water, LCModel		water		water
sequence	MRSI-sLASER, multi-echo						MRSI-PRESS, TE-averaged		SVS-PRESS, multi-echo		SVS-STEAM
field	3T						3T		3T		7T
VOI	supraventricular						supratentorial brain		medial PFC	LFC	PCC, young

	[4]		[5]				[6]				[7]	
	WM	**GM**	**WM**	**GM**	**WM**	**GM**	**WM**	**GM**	**WM**	**GM**	**WM**	**GM**
tCr	7.8 ± 1.1	8.1 ± 0.96	5.7 ± 0.6	6.5 ± 0.6	5.7 ± 0.5	6.4 ± 0.7					6.33 ± 0.14	8.04 ± 0.32
Glu	7.92 ± 0.43	10.99 ± 0.55	6.7 ± 1.8	8.2 ± 1.1	7.0 ± 2.6	8.5 ± 1.0						
mI	8.18 ± 1.00	6.11 ± 0.63	3.1 ± 0.6	4.3 ± 0.7	3.8 ± 0.9	4.3 ± 0.9					6.68 ± 0.44	7.26 ± 0.52
NAA	10.38 ± 1.00	9.81 ± 0.40	8.0 ± 1.0	8.2 ± 0.8	9.1 ± 0.9	7.7 ± 1.0	8.4 ± 0.9	8.4 ± 1.1	8.2 ± 1.0	8.9 ± 0.9	8.81 ± 0.18	9.24 ± 0.26
Asp	1.24 ± 0.55	2.83 ± 0.32		1.4 ± 0.8								
GABA	1.15 ± 0.41	2.12 ± 0.42		1.4 ± 0.8								
Glc	3.33 ± 0.60	1.77 ± 0.51		0.5 ± 0.6								
Gln	2.17 ± 0.52	3.96 ± 0.46	1.5 ± 1.3	3.8 ± 1.4	1.8 ± 1.6	4.4 ± 1.4						
GSH	1.30 ± 0.20	1.43 ± 0.13										
Gly	0.00 ± 0.00	0.47 ± 0.18										
tCho	1.24 ± 0.19	1.00 ± 0.18	1.68 ± 0.27	1.10 ± 0.14	1.78 ± 0.41	1.38 ± 0.17					1.58 ± 0.02	1.43 ± 0.08
Lac	0.53 ± 0.23	0.51 ± 0.20		0.6 ± 0.4								
NAAG	3.59 ± 0.65	2.09 ± 0.39	2.7 ± 1.2	0.5 ± 0.4	1.5 ± 0.9	0.7 ± 0.3	2.7 ± 1.2	0.6 ± 0.5	1.5 ± 0.9	0.7 ± 0.4		
PE	2.59 ± 0.66	1.92 ± 0.51										
sI	0.46 ± 0.10	0.33 ± 0.06		0.14 ± 0.09								
Tau	0.49 ± 0.77	1.70 ± 0.24		1.0 ± 0.4								
MMBG	measured		LCModel baseline		LCModel baseline		LCModel baseline		LCModel baseline		modeled	
T2 correction	measured		no		no		literature		literature		measured	
T1 correction	measured		no, long TR		no, long TR		no, long TR		no, long TR		measured	
reference	water		water, LCModel		water, LCModel		water, LCModel		water, LCModel		water	
sequence	SVS-PRESS		SVS-STEAM		SVS-STEAM		SVS-STEAM		SVS-STEAM		SVS-STEAM	
field	3T		2T		2T		2T		2T		3T	
VOI	PL	OPC	Parietal Lobe		Frontal Lobe		Parietal Lobe		Frontal Lobe		Occipito-Parietal Lobe	

[1]	Srinivasan R. et al, NeuroImage 2006, 30(4):1171-1178	**PFC**: prefrontal cortex	
[2]	Choi C. et al, MRM 2006, 56(5):971-977	**LFC**: left-frontal cortex	
[3]	Marjanska M. et al, Neuroscience 2017, 354:168-177	**PCC**: Parietal Central cortex	
[4]	Hoefemann M. et al, NMR biomed 2020; 33e4328	**OPC**: occipito-parietal cortex	
[5]	Pouwels P.J.W. et al, MRM 1998; 39:53-60	**PL**: parietal lobe	
[6]	Pouwels P.J.W. et al, NMR biomed 1998; 10:73-78		
[7]	Kreis R. et al, Journal of MR, Series B, 1993,102:9-19		

Supporting Table 7.10: Estimated concentrations grouped for WM and GM (average value ± standard deviation) compared to the literature.

	MEMS		MESS		MESS-mocked		[1]		[2]		[3]**	
	WM	GM	WM	GM	WM	GM	WM	GM	WM	GM	WM	GM
tCr-CH2	142 ± 13	137 ± 15	140 ± 17	142 ± 16	138 ± 12	142 ± 15	108 ± 6	110 ± 4	154 ± 6	156 ± 11	153 ± 46	154 ± 43
tCr-CH3	185 ± 10	189 ± 14	186 ± 10	190 ± 18	182 ± 9	194 ± 14						
mI	128 ± 15	229 ± 38	145 ± 28	209 ± 49	138 ± 16	220 ± 40						
NAA	325 ± 21	291 ± 27	338 ± 35	284 ± 27	315 ± 24	302 ± 33	206 ± 9	172 ± 15	291 ± 14	237 ± 15	255 ± 35	235 ± 30
tCho	239 ± 15	351 ± 43	232 ± 20	356 ± 53	239 ± 16	358 ± 47	136 ± 9	126 ± 4	216 ± 18	227 ± 25	179 ± 53	172 ± 49
global	215 ± 32	161 ± 13	215 ± 36	161 ± 19	205 ± 28	169 ± 12						
H2O	61 ± 4	71 ± 7	60 ± 5	71 ± 7	74 ± 7	73 ± 7						
MMBG	modeled						not reported		1st order polyomial baseline		machine learning separation	
fitting	2D simultaneous						TE-averaging fit		sequential		Multi-TE SPICE	
sequence	MRSI-sLASER, multi-echo						MRSI-PRESS, TE-averaged		Multi-TE MRSI – EPSI readout		Multi-TE MRSI – EPSI readout	
TEs [ms]	(3):35,156,278						(64): 35-195, step 2.5		(5):50-100-160-220-300		(3):30,80,130	
VOI	supraventricular						supratentorial brain		supraventricular		whole brain coverage	

	[4]		[5]		[6]		[7]		[8]	
	WM	GM	WM	GM	WM	GM	WM	GM	WM	GM
tCr-CH2	145 ± 11	137 ± 4	129 ± 13	134 ± 15	127 ± 13	128 ± 11	169 ± 15	164 ± 12	152 ± 7	116 ± 9
tCr-CH3	220 ± 13	202 ± 7	166 ± 11	148 ± 22	178 ± 9	161 ± 10			156 ± 20	141 ± 16
mI	149 ± 18	229 ± 23	161 ± 37	244 ± 61						
NAA	400 ± 0.5	349 ± 29	343 ± 34	253 ± 64	301 ± 18	254 ± 15			295 ± 29	247 ± 19
tCho	269 ± 18	314 ± 21	218 ± 22	274 ± 61	222 ± 17	265 ± 29			187 ± 20	207 ± 16
global	250 ± 4	205 ± 11	193 ± 29*	170 ± 35*			198 ± 22 (Glu)	201 ± 18 (Glu)		
H2O										
MMBG	measured		not reported		not reported		not reported		no baseline correction	
fitting	2D simultaneous		2D simultaneous		sequential		sequential		sequential	
sequence	SVS-PRESS		SVS-PRESS		SVS-PRESS		SVS-PRESS, multi-echo		SVS-STEAM	
TEs [ms]	(24): 20-307.5, step 12.5		(100): 30-228, step 2		(5): 50,120,200,280,400		(6):128,164,214,262,326,380		(5):50,100,150,200,250	
VOI	PL	OPC	PVWM	pACC	Occipital	Cingulate gyrus	LFC	medial PFC	Occipital area	

[1]	Srinivasan R. et al, NeuroImage 2006, 30(4):1171-1178	**PFC**: prefrontal cortex
[2]	Tsai S-Y. et al, MRM 2007, 57(5):859-865	**LFC**: left-frontal cortex
[3]	Wang Z. et al, MRM 2022, 87(3):1093-1637	**OPC**: occipito-parietal cortex
[4]	Hoefemann M. et al, NMR biomed 2020; 33e4328	**PVWM**: periventricular WM area
[5]	Wyss P.O. et al, MRM 2018, 80:452-461	**pACC**: pregenual anterior cingulate cortex
[6]	Träber F. et al, Journal of MRI 2004, 19(5):537-545	**PL**: parietal lobe
[7]	Choi C. et al, MRM 2006, 56(5):971-977	
[8]	Mlynarik V. et al. NMR Biomed 2001, 14:325-331	

*average value from: Cho 3.6ppm, NAA 2.47ppm, Gln, Glu and NAAG
** values reported for 1 healthy subject

Supporting Table 7.11: Estimated T_2 rates grouped for WM and GM (average value ± standard deviation) compared to the literature.

7.2. Diffusion weighted single-voxel MR Spectroscopy of the prostate

References[2]:

Stamatelatou* A, **Rizzo* R**, Simsek K, Van Asten S, Heerschap A, Scheenen T, Kreis R.

Diffusion-weighted MR spectroscopy of the prostate

Proceedings of 32nd Annual Meeting of the International Society of Magnetic Resonance in Medicine - ISMRM, 2023: 03-08 June, Toronto, CA

Rizzo* R, Stamatelatou* A, Simsek K, Van Asten S, Kreis R, Heerschap A, Scheenen T.

Initial exploration of the potential of diffusion-weighted MRS for the evaluation of prostate pathology

Proceedings of 32nd Annual Meeting of the International Society of Magnetic Resonance in Medicine - ISMRM, 2023: 03-08 June, Toronto, CA

Contribution to the studies

Data acquisition and postprocessing. Model tuning and fitting. Data analysis. Manuscript drafting. *Contribution equally shared with Stamatelatou A.*

---

[2]The following references are a modified and extended version of the correspondent ISMRM accepted abstracts.
*Shared first authorship

Diffusion-weighted MR spectroscopy of the prostate

Angeliki Stamatelatou[1], **Rudy Rizzo**[2,3], Kadir Simsek[4], Sjaak Van Asten[1], Arend Heerschap[1], Tom Scheenen[1], Roland Kreis[2,3].

1. Department of Medical Imaging, Radboud Medical Center, Nijmegen, The Netherlands

2. Magnetic Resonance Methodology, Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland

3. Translational Imaging Center, sitem-insel, Bern, Switzerland

4. Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom

Synopsis

Diffusion-weighted Magnetic Resonance Spectroscopy (DW-MRS) is ideally suited to explore complex microstructures with metabolites selectively distributed in different subspaces. So far, this technique has only been applied to the brain and muscles. In this work, we explored DW-MRS for the first time in the prostate, an organ with potentially more motion problems. Thus, dedicated acquisition and post-processing techniques were used, including the measurement of water next to that of metabolites for corrections. ADC values of citrate (Cit), total-choline (tCho), total-creatine (tCr), myo-Inositol (mI), and spermine (Spe) were estimated and evaluated according to the compartmental structure of the prostate, indicating hindered metabolite diffusion in the luminal space.

Summary of the main findings

In this work, we applied single-voxel DW-MRS on the prostate in nine volunteers. The ADC values of the metabolites Cit, tCho, Spe, tCr, and mI were evaluated and related to their respective compartmentation in prostate tissue suggesting new insights for microenvironments.

1. Introduction

Prostate tissue has a complex microstructure composed of three major components: epithelial cells, stromal cells, and large extracellular (luminal) spaces. MRI has been used to elucidate these spaces in prostate tissue by relaxometry and diffusion-weighted MRI [113, 114]. Diffusion-weighted MR Spectroscopy (DW-MRS) [45] is ideally suited to explore complex microstructure in vivo with metabolites selectively distributed in different subspaces. So far, this technique has been applied in the brain and muscle only, which are relatively stationary organs. In this work, we explore DW-MRS for the first time in the prostate, an abdominal organ prone to motion artifacts. We tackled motion by concurrently acquiring a water signal for corrections

and dealt with limited SNR by the simultaneous bidimensional fitting of spectral shape and signal attenuation at multiple b-values (i.e., leveraging high b-values, where diffusion weighting severely worsens the already intrinsically low SNR of MRS, by simultaneously including low b-values with better SNR).

2. Methods

Single-voxel DW-MRS was performed with a non–water-suppressed STEAM sequence with metabolite-cycling to measure metabolite and water signals concurrently [45]. Acquisition parameters were: TE/TM/TR 33/35/2500 ms. Interleaved b-values were acquired with 16 averages repeated four times while visually monitoring frequency drifts and subject motion via the spectroscopic interactive window at the scanner console. Severely motion-corrupted batches were discarded. To maximize the achievable b-value, identical diffusion gradient amplitude was applied in all three directions ($G = G_x = G_y = G_z$), yielding an effective gradient amplitude $G_{EFF} = \sqrt{3}G$. *In-vivo* acquired b-values: 124, 359, 776, 1353, 1988, 2516 s/mm^2. Due to low SNR, higher b-values were not acquired whereas *in-vitro* b-values 3106 and 3759 s/mm^2 were included. The *in-vitro* setup consisted of an 8 cm^3 voxel placed in a prostatic fluid mimic phantom containing citrate (Cit), spermine (Spe), and myo-Inositol (mI) in the presence of ions and 15 g/l BSA [115]. We examined the prostate of nine healthy volunteers (mean age 56 years, range 31-66 years) with a mean VOI of ∼13.9 cm^3, covering ∼60% central gland and ∼40% peripheral zone tissue. Measurements were performed on 3T MR systems (Siemens, Germany) with an external phased-array coil for signal reception.

In post-processing, a motion-compensation scheme for inter-acquisition distortions was applied, using the co-acquired water signal as the reference to correct phase and frequency fluctuations, eddy-current distortions, and to compensate for non-linear motion, restoring potential signal loss in individual acquisitions [45]. The processed MR spectra were fitted for signals of Cit, tCho (sum of glycerophosphorylcholine, phosphorylcholine, and choline), Spe, tCr (1:1 sum of creatine and phosphocreatine), and mI, and separately for water. Resonances of glutamate (Glu), syllo-Inositol (sI) and taurine (Tau) were visible only for setups with good shim and at low b-values; see Supporting Fig.7.15. Therefore, despite their respective basis sets being included in the model, as expected and indicated by ex-vivo metabolic fingerprints [116], neither their estimated concentration nor ADCs were evaluated and included in the findings. A Voigt-line modulated baseline with three components is added to regularize the fit given the lack of prior knowledge in the lipid area between $[0 - 2.5]$ ppm and the $[3.5 - 4]$ ppm area.

For the simultaneous 2D spectrum-ADC fitting, FiTAID [16] was used, employing metabolite spectra simulated in VeSPA[13] assuming ideal RF pulse shapes and modeling Voigt-lineshapes. The ADCs of the metabolites were modeled by a mono-exponential decay function, while water was represented bi-exponentially. A cohort-average spectrum with high SNR was also constructed to verify the model. In-vivo ADC values in Tab.7.1 were calculated as weighted cohort

averages using the inverse of the CRLBs as weights. Absolute quantification for concentrations was performed referencing to the total water signal [117], with corrections for T_1 and T_2 relaxation [118].

3. Results and Discussion

The ADC values derived from the decaying signals of the compounds in the phantom (Fig.7.12) show that those for Cit and Spe are lower than that of mI (Tab.7.1), which would agree with their lower T_2 values relative to mI measured in studies of similar phantoms, indicating (transient) complexation of these compounds [115].

Compound $x10^{-4}$ mm^2/s	ADC $\pm$ CRLB *in-vitro* (20°)	ADC *in-vitro* (37°)	ADC $\pm$ STD *in-vivo* cohort	ADC $\pm$ CRLB *in-vivo* average
tCho	-	-	1.47 $\pm$ 0.45	1.64 $\pm$ 0.13
Cit	4.28 $\pm$ 0.02	6.4	2.86 $\pm$ 0.51	2.48 $\pm$ 0.07
tCr	-	-	1.19 $\pm$ 0.56	1.34 $\pm$ 0.09
mI	5.98 $\pm$ 0.12	9.0	1.52 $\pm$ 0.22	1.41 $\pm$ 0.09
Spe	4.01 $\pm$ 0.02	6.0	1.59 $\pm$ 0.54	1.80 $\pm$ 0.06
H_2O	19.15 $\pm$ 0.01	29.0	(i) 4.80 $\pm$ 1.13	(i) 5.51 $\pm$ 0.00
			(e) 22.89 $\pm$ 5.25	(e) 30.00 $\pm$ 0.00

Table 7.1: Estimated ADC values for metabolites and water [$x10^{-4}$ mm^2/s]. In-vitro ADC values are listed as obtained from in vitro measurements in a prostatic fluid mimic solution at 20° and approximately temperature-corrected for 37°. In-vivo ADC values are reported as a weighted *cohort* average from the individual subject results and the estimated value from the cohort *average* spectra. Two ADCs are reported for water (H_2O) given bi-exponential fit assumption: (i) intracellular compartment, (e) extracellular compartment.

The MR spectra of healthy volunteers obtained at increasing b-values show good to moderate quality with average water FWHM[3] = 5.96 $\pm$ 3.28 Hz and SNR[4] = 2.72 $\pm$ 0.95 x10^3. Fig.7.13 compares the acquisition setup, spectral, and fit quality for two healthy volunteers. Fig.7.14 display spectra and fit quality on cohort average spectra. Excellent spectroscopic fits of the metabolites were achieved, as demonstrated by the residuals in Fig.7.13 and 7.14. Overall results integrating concentrations and ADCs for minor (and originally excluded) metabolites are reported in Supporting Fig.7.16. The *in-vivo* results of the metabolites ADCs showed considerable variance within the cohort. However, this was expected due to various reasons: different sizes and types of prostate tissues included and limited by one selected VOI, the broad age range

---

[3]FWHM: Full Width Half Maximum as estimated by water Gaussian broadening.
[4]SNR: time-domain water SNR

of the subjects included in the study, motion, SNR differences between the subjects and between b-values, and different conditions with induced variance in complexation mechanisms [115].
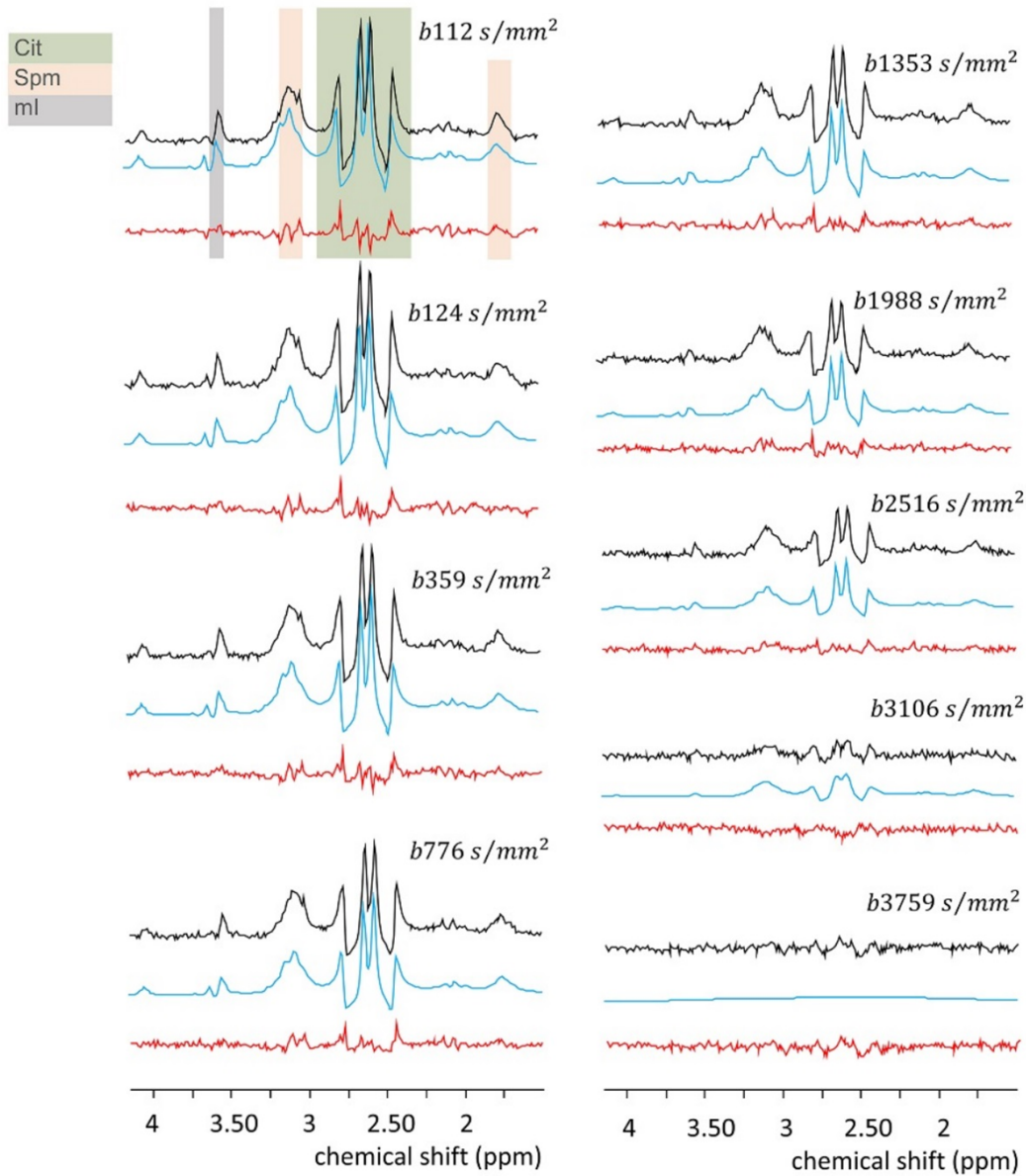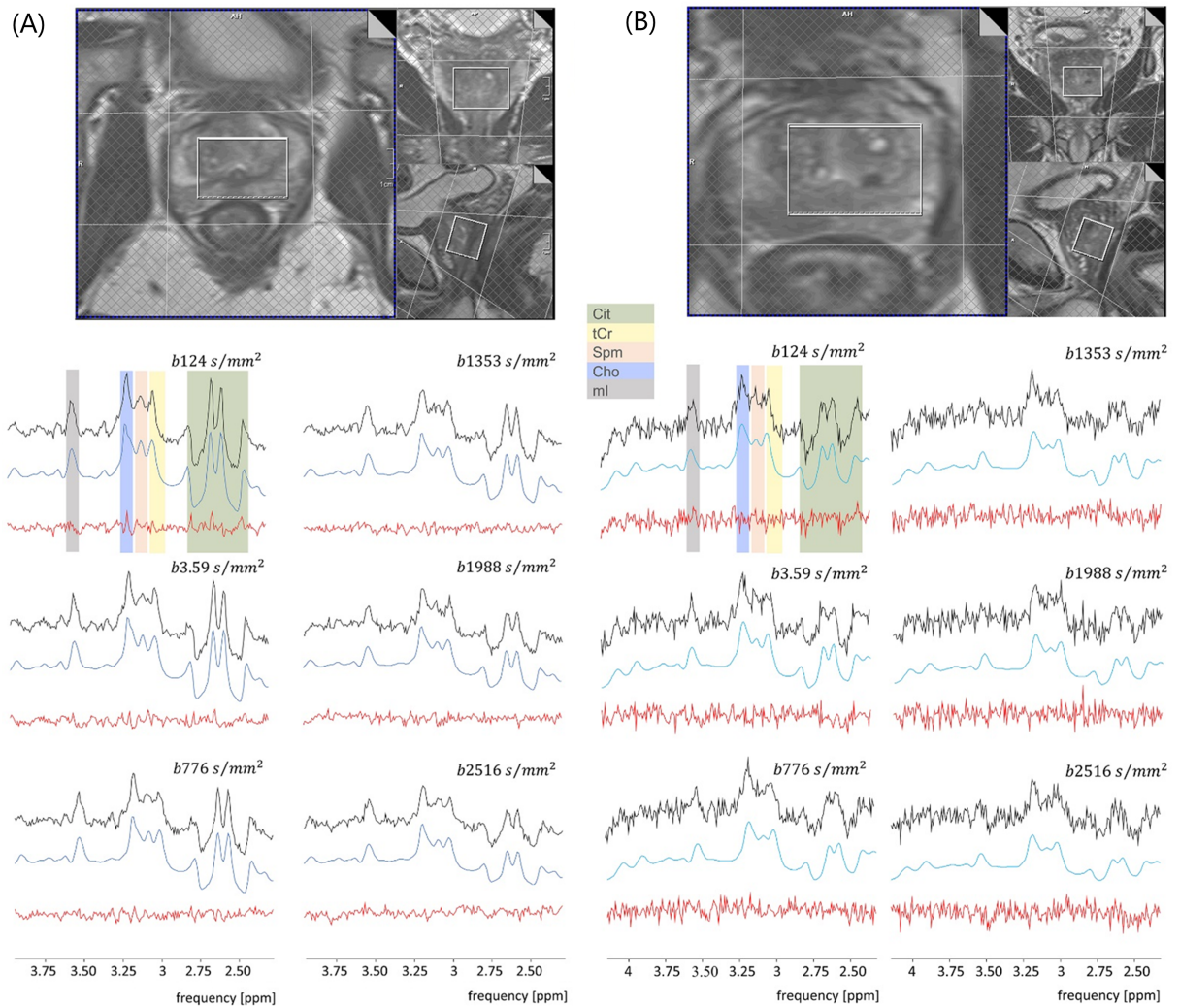

Figure 7.12: *In-vitro* single-voxel DW-MRS. Acquired and post-processed MR spectra for eight b-values (black), fit (blue), and residues (red). Residues display little but negligible signal contribution due to prior knowledge-related challenges on the spin system definition of both Spe and Cit. For details see section 3.2.1.

Figure 7.13: *In-vivo* single-voxel DW-MRS of good (A; water FWHM: 3.2Hz, water time-domain SNR: 4.5×10^3) and moderately good (B; water FWHM: 4.8Hz, water time-domain SNR: 2.5×10^3) quality. (Top) T_2w images showing the voxel positioning. (Bottom) acquired and corrected MR spectra for six b-values (black), fit (blue), and residues (red). Cit diffuses the fastest, while the other metabolites diffuse slower with Spe and mI unexpectedly slow (similarly to tCho and tCr) even though *prevalently* present in the luminal space.

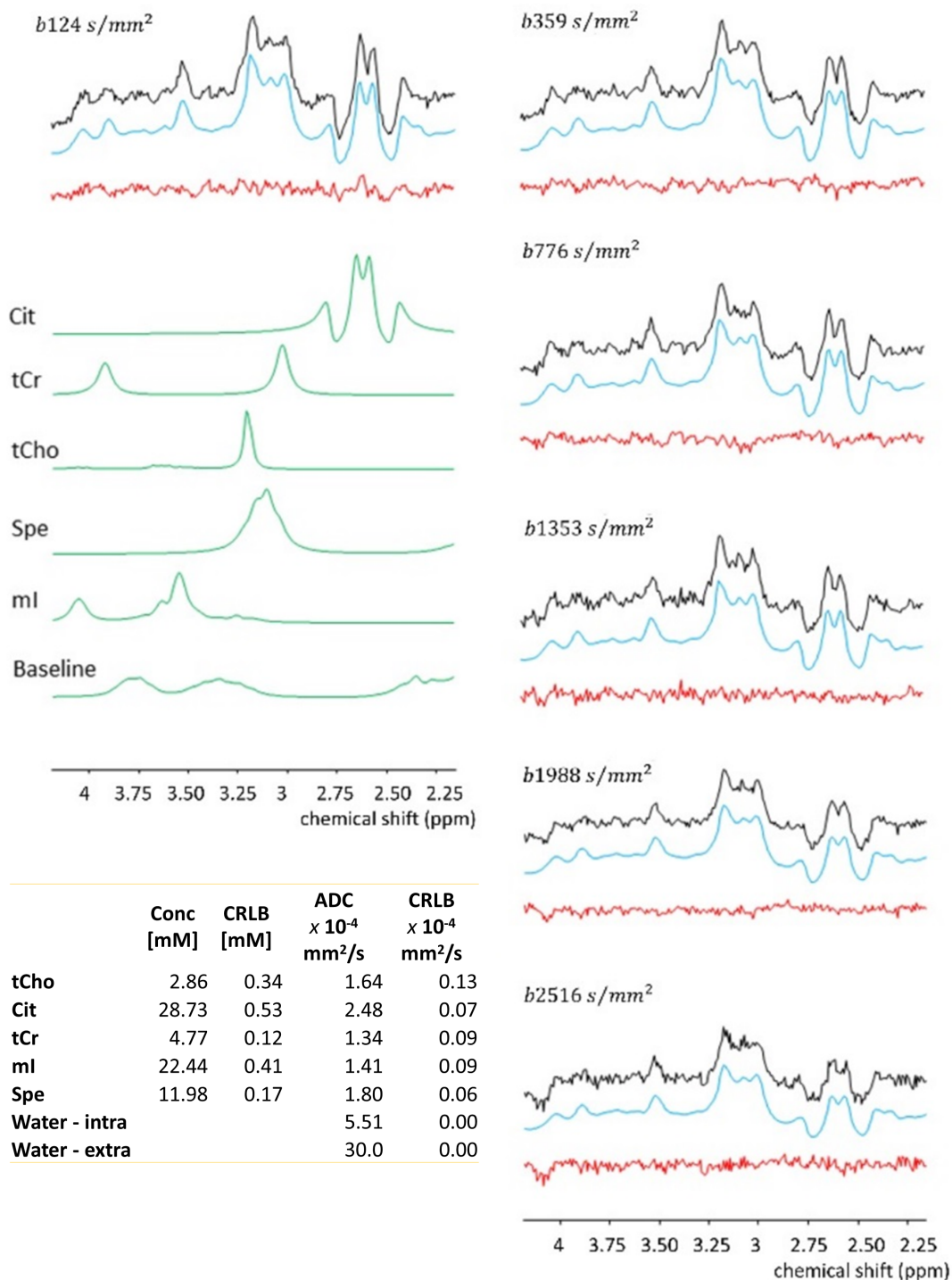|	Conc [mM]	CRLB [mM]	ADC x 10⁻⁴ mm²/s	CRLB x 10⁻⁴ mm²/s
tCho	2.86	0.34	1.64	0.13
Cit	28.73	0.53	2.48	0.07
tCr	4.77	0.12	1.34	0.09
mI	22.44	0.41	1.41	0.09
Spe	11.98	0.17	1.80	0.06
Water - intra			5.51	0.00
Water - extra			30.0	0.00

Figure 7.14: *In-vivo* cohort averaged spectra from nine volunteers (black), cohort average fit of metabolite signals (blue), and residues (red). The metabolite patterns are indicated as fitted for $b124$ s/mm² (green). Tissue concentrations and ADCs with their estimates' precision (Cramer Rao Lower Bound, CRLB) are reported for water and tCho, Cit, tCr, mI, and Spe.

Since no literature is available regarding the ADC values of the metabolites of interest for the prostate tissue, the results were interpreted according to the existing knowledge of the cellular structure in prostatic tissues. The ADC values derived from these experiments reveal that tCho and tCr have low weighted-average values (Tab.7.1), in agreement with their intracellular origin. In contrast, the higher values for Cit and Spe would agree to confirm that these compounds dominantly originate from the luminal space [119, 120] with fewer diffusion restrictions. However, the ADC of Cit is substantially higher than that of Spe (p<0.05), and inspection of the cohort-averaged spectra (Fig.7.14) confirms the relatively slow Spe signal decay visually. Furthermore, as shown by MRS studies of prostate tissue and prostatic fluid [119], there is a strong correlation between Cit and Spe levels suggesting transient complex-formation of these compounds as was demonstrated in-vitro by T_2 measurements [115]. Moreover, this and other studies indicated that this complex and its components transiently bind to proteins within the luminal space. As the luminal concentration of Cit is much higher than that of Spe and assuming a 1:1 complex, it follows that the fraction of free compound is much more significant for Cit than Spe, which could explain the higher ADC for Cit compared to Spe because of its larger free fraction in solution.

As the average luminal diameter is about 200 µM [36], one would expect nearly unrestricted diffusion for Cit, Spe, and mI with much higher ADCs for such small molecules than what we measured in this study (around 6-9 rather than 2-3 x10^{-4} mm^2/s), which further supports that protein-binding, as described above, occurs in luminal space. Nevertheless, this hypothesis needs further investigation through in-vitro setups with phantom solutions enriched with various prostate-specific proteins.

It needs to be noted that the estimated ADC value of mI is aligned with the one of Spe in the cohort distribution. However, its value reported by a more trustworthy fit of the cohort average is much lower than expected for a free small uncharged molecule that is commonly assumed prevalent in luminal space (Tab.7.1). The latter value seems aligned with the ADC value of tCr (i.e., a principal intracellular metabolite), yielding less obvious deductions. NMR spectra of prostatic fluid are dominated by signals of Cit, Spe, and mI [119, 120] of which the concentrations roughly agree with their prostate in-vivo concentrations taking the luminal fraction into account [113, 34]. However, some mI may still be located intracellularly.

Finally, we obtained two ADC values for water from a bi-exponential fit of its signal decay, Tab.7.1. The higher ADC component would represent water in the luminal space, and the other that of cellular water. In our findings, we report an average intra-cellular volume fraction of 62 ± 8%, which is aligned with some of the *in-vivo* MRI available literature based on T_2-defined water compartments [121, 122]. This, in turn, could be used to calculate compartment-specific concentrations.

4. Conclusions

We successfully demonstrate that DW-MRS of the prostate is feasible and provides a window into the microenvironment of prostatic compartments. The preliminary results indicate restricted or hindered diffusion of Cit and Spe in the luminal space despite its average large size, indicating complexation with macromolecules in this compartment.

The intra-extra cellular concentration ratio of metabolites with a contribution in both luminal and intracellular space, like mI, Cit, and Spe, could be investigated by diffusion-weighted bi-exponential modeling like it was done here for water. However, further experimental design challenges need to be tackled, such as acquiring more and higher b-values needed to ensure reliable bi-exponential fitting.

Given the substantial heterogeneity of prostate tissue, fast DW-MRSI technologies would be valuable approaches for exploring selective tissue types with the potential of monitoring the luminal space size variation across the prostate.

5. Supporting Figures
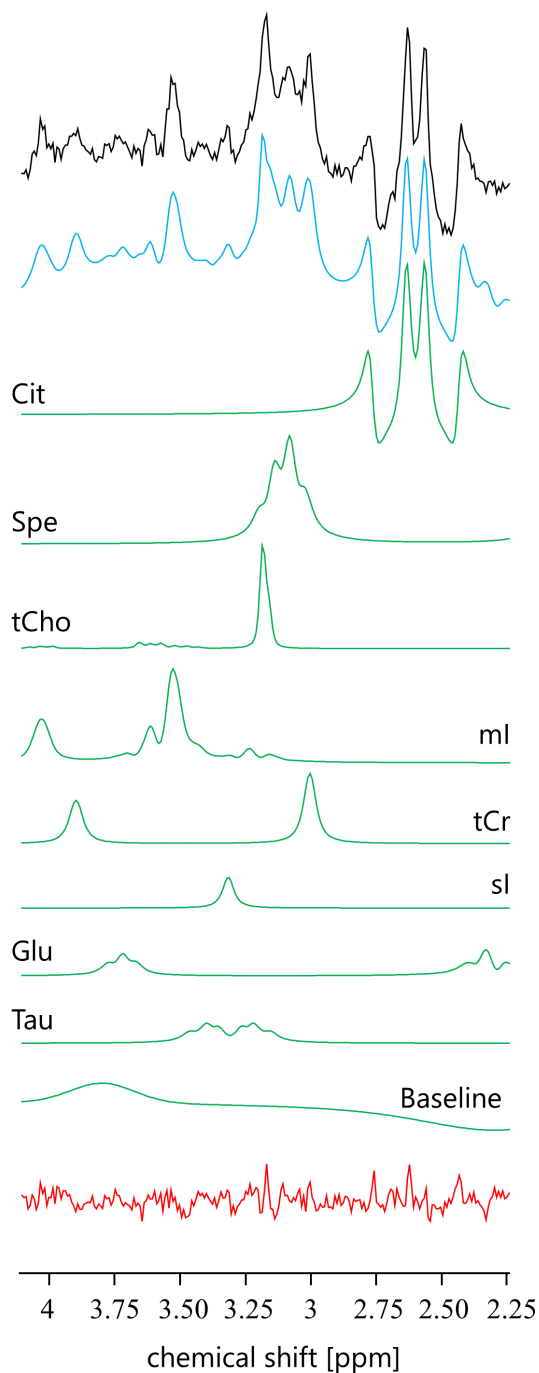
Supporting Figure 7.15: Acquired and post-processed good quality spectrum at b-value 124 s/mm^2 (black, same as Fig.7.13.A). Overall fitted metabolite signal (blue) and residues (red). Single metabolite fitted basis set (green). Weakly defined metabolites (sI, Glu, and Tau) can be easily pinpointed.
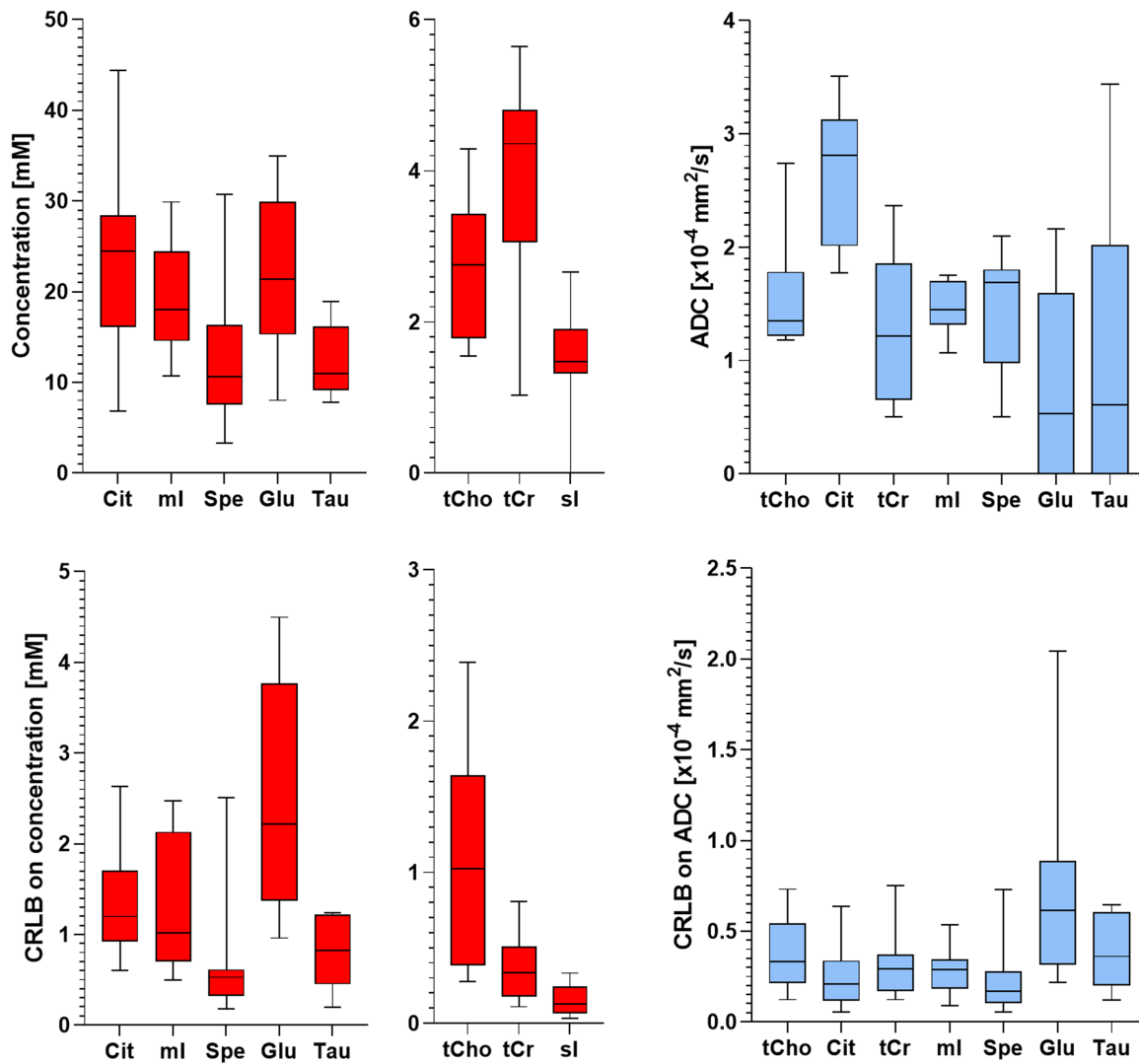
Supporting Figure 7.16: Boxplots reporting *in-vivo* estimated value (top) and CRLB (bottom) of concentrations (left, red) and ADCs (blue, right). ADCs of sI are assumed equal to the ADCs of mI, given their molecular similarity. ADCs for Glu ad Tau are distributed with a large spread.

Initial exploration of the potential of diffusion-weighted MRS for the evaluation of prostate pathology

Rudy Rizzo[1,2], Angeliki Stamatelatou[3], Kadir Simsek[4], Sjaak Van Asten[3], Roland Kreis[1,2], Arend Heerschap[3], Tom Scheenen[3]

1. Magnetic Resonance Methodology, Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland

2. Translational Imaging Center, sitem-insel, Bern, Switzerland

3. Department of Medical Imaging, Radboud Medical Center, Nijmegen, The Netherlands

4. Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom

Synopsis

Prostate pathologies like cancer and prostatitis cause alterations in tissue microstructure that are reflected in signal variations on diffusion-weighted MRI or MR relaxometry. Diffusion-weighted MR Spectroscopy (DW-MRS) allows specific characterization of tissue microstructure by quantifying both concentration and diffusion properties of MR-observable metabolites. In this work, an optimized DW-MRS protocol tailored to overcome the inherent challenges of prostate measurements is deployed. Initial results on potential alterations of metabolite concentrations and diffusivities (ADCs) are reported for prostate cancer and prostatitis. The preliminary findings are tentatively explained by microstructure alterations of the prostate tissue.

Summary of the main findings

A first preliminary exploration of potential alterations of concentrations and diffusion coefficients (ADCs) of water and metabolites in prostate cancer and prostatitis is performed using diffusion-weighted MR Spectroscopy (DW-MRS).

1. Introduction

Diffusion-weighted Magnetic Resonance Spectroscopy (DW-MRS) [123] allows the characterization of tissue microstructure by quantifying the diffusion properties of MR-observable metabolites [124, 125]. Prostate pathologies like cancer and prostatitis cause tissue microstructure alterations reflected in diffusion-weighted MRI or MR relaxometry. While MRS detects abnormal tissue content of metabolites, DW-MRS may pinpoint metabolic tissue content alterations and changes in the tissue microenvironment. In the prostate, DW-MRS is challenging, not only because of the organ size and location close to the bladder and rectum but also due to the sensitivity of DW-MRS to any motion, including peristaltic motion. The latter section (Section

7.2, abstract 1) outlined the optimization of DW-MRS for this organ in healthy subjects. Here, we explored apparent diffusion coefficients (ADCs) of water and metabolites in prostate cancer (PCa) and prostatitis (CPPS).

2. Method

Ten subjects with elevated PSA levels (≥ 9 ng/mL, mean age 72 years), suspicious for PCa, underwent a clinical multi-parametric MRI protocol followed by biopsy. Single-voxel DW-MRS was added as a 15-minute prolongation of the examination. In addition, MRS voxel positioning was performed to maximize the targeted tumor tissue content by inspection of T_2w and ADC images. We report on two patients eventually diagnosed with PCa and two with CPPS. Nine healthy subjects served as controls. Measurements were performed on a 3T MR system (Siemens) with external phased-array coils.

A metabolite-cycled DW-MRS STEAM sequence allowed simultaneous measurement of metabolite and water signals [31, 45]. The acquisition parameters were: TE/TM/TR 33/35/2500 ms; b-values: 124, 776, 1988 s/mm^2. Post-processing featured artifact correction and motion compensation using the co-acquired water signal as inherent reference as described for brain [45]. A 2D-fit with simultaneous χ^2-minimization in the spectral domain and the mono-exponential diffusion–decay dimension was run in FitAID [16] for both water and metabolites. Metabolites of interest include citrate (Cit), total-choline (tCho), spermine (Spe), myo-inositol (mI), and total-creatine (tCr). Absolute quantification was performed referencing to the total water signal [117], with corrections for T_1 and T_2 relaxation [118]. Details of the specifically optimized DW-MRS methodology and results from healthy subjects are reported in Section 7.2 (abstract 1), where 6 b-values were recorded, leading to better precision and relatively lower ADC values, and the potential to distinguish luminal and cellular space. Here, we investigate a minimal setup with 3 b-values respecting clinical time constraints. The fractional volume of tumor tissue was qualitatively estimated with SpectrIm [126] using ADC maps.

3. Results and Discussion

Fig.7.17(top) presents results for PCa. Good fit results can be monitored from white Gaussian-distributed residues. Given the hardly visible peaks for Cit and tCr, tissue concentrations and ADC values are only reported for tCho and water. CPPS patients are portrayed in Fig.7.17(bottom) with fit results for all metabolites (CPPS #1: acute prostatitis, CPPS #2: chronic inflammation). Fig.7.18 depicts average spectra and results from the healthy cohort. The metabolite tissue contents are in line with those published previously [118, 127] with Cit levels typical for the transition zone, in agreement that this area covered most of the VOI. Separated metabolite patterns are included to demonstrate spectral overlap.
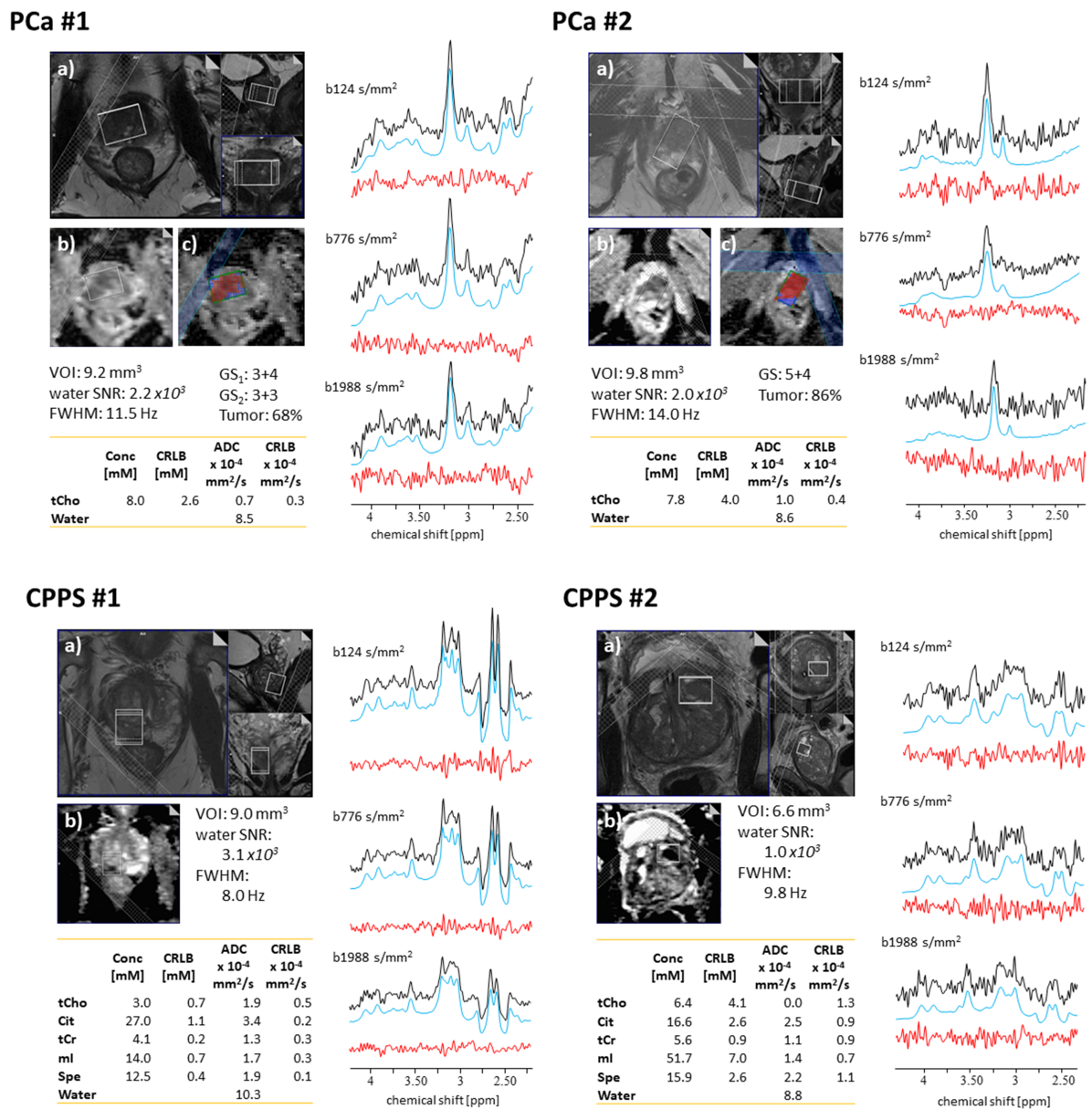
Figure 7.17: (Top) PCa and (bottom) prostatitis patients. (a) T_2w, (b) ADC axial images with (c) voxel segmentation. Measured spectra (black), fitted model (blue), and residues (red) for the 3 b-values. Assessment of spectral quality via water-referenced time domain SNR and Full-Width Half-Maximum (FWHM) averaged over b-values. Volume of Interest (VOI), Gleason Score (GS), and tumor fraction from the segmentation map are also reported. Tissue concentrations, ADCs, and estimates' precision (Cramer Rao Lower Bound, CRLB) are reported for water, tCho, Cit, tCr, mI, and Spe.

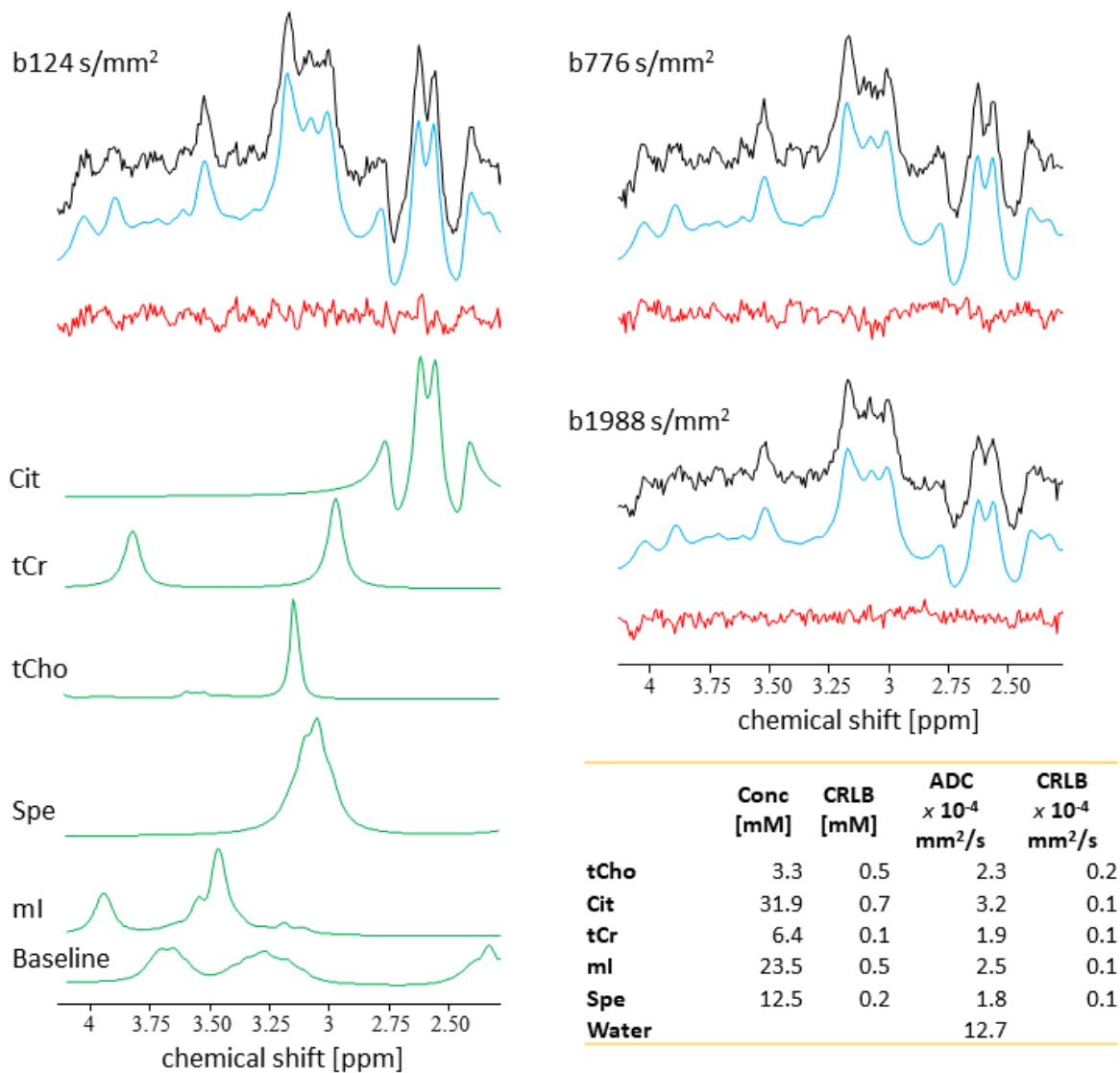|	Conc [mM]	CRLB [mM]	ADC x 10⁻⁴ mm²/s	CRLB x 10⁻⁴ mm²/s
tCho	3.3	0.5	2.3	0.2
Cit	31.9	0.7	3.2	0.1
tCr	6.4	0.1	1.9	0.1
mI	23.5	0.5	2.5	0.1
Spe	12.5	0.2	1.8	0.1
Water			12.7	

Figure 7.18: Healthy subjects. Averaged spectra obtained from nine volunteers, measured spectra (black), fitted model (blue), and residues (red). In green, the fitted metabolite patterns for the metabolites of interest at the lowest b-value. Tissue concentrations, ADCs, and estimates' precision (Cramer Rao Lower Bound, CRLB) are reported for water, tCho, Cit, tCr, mI, and Spe.

Fig.7.19 compares tissue concentrations and ADC values between the average healthy cohort and the pathologic cases. For metabolite ADCs, there are no values from the literature to compare to. They are instead critically reviewed with respect to plausibility given the knowledge on prostate cellular microstructure:

- PCa vs. healthy cohort:
  - The ADC for tCho is found to be substantially lower in both patients. Given that tCho resides intracellularly [34] and cells are denser in tumorous tissue, the reduction in ADC may be well explained given a diffusion time ensuring metabolite diffusion to be restricted by cell size. Surprisingly, there is hardly any and even inconsistent literature on this also from other organs [128, 129].
  - As expected, [117, 34] tCho concentration is higher in tumor tissue.

- CPPS vs. healthy cohort:
  - Based on only two cases, one could speculate that the ADCs of mI and tCr are lowered in CPPS.
  - Similarly, for metabolite contents, no certain trend is seen. Potentially, tCr and Cit may turn out to range consistently lower. Reduced Cit levels would reflect a smaller LS, expected in such pathology [130].
  - The chronic inflammation case (CPPS #2) shows comparable tCho and lower mI content, whereas the acute prostatitis case (CPPS #1) shows higher tCho and mI content.

Water ADCs are aligned with literature values [131, 132, 133]. Both PCa and CPPS report lower water ADCs compared to healthy tissue, as expected for reduced luminal space. ADCs are lower in PCa than CPPS with values within reported distributions [131].

4. Conclusions

Initial results for prostate cancer and prostatitis show the potential of DW-MRS for the characterization of microstructural changes in prostate pathology, even in a clinical time frame. Longer dedicated exams with more and higher b-values would no doubt provide more precise information, which may also include quantifying loss of luminal space from water data [133, 113]. The latter would also allow partial volume correction of metabolite concentrations, which might lead to enhanced specificity of MRS findings. Statistical investigations on a more significant patient cohort are obviously required to consolidate the initial findings – particularly for reduced ADCs of tCho in PCa.
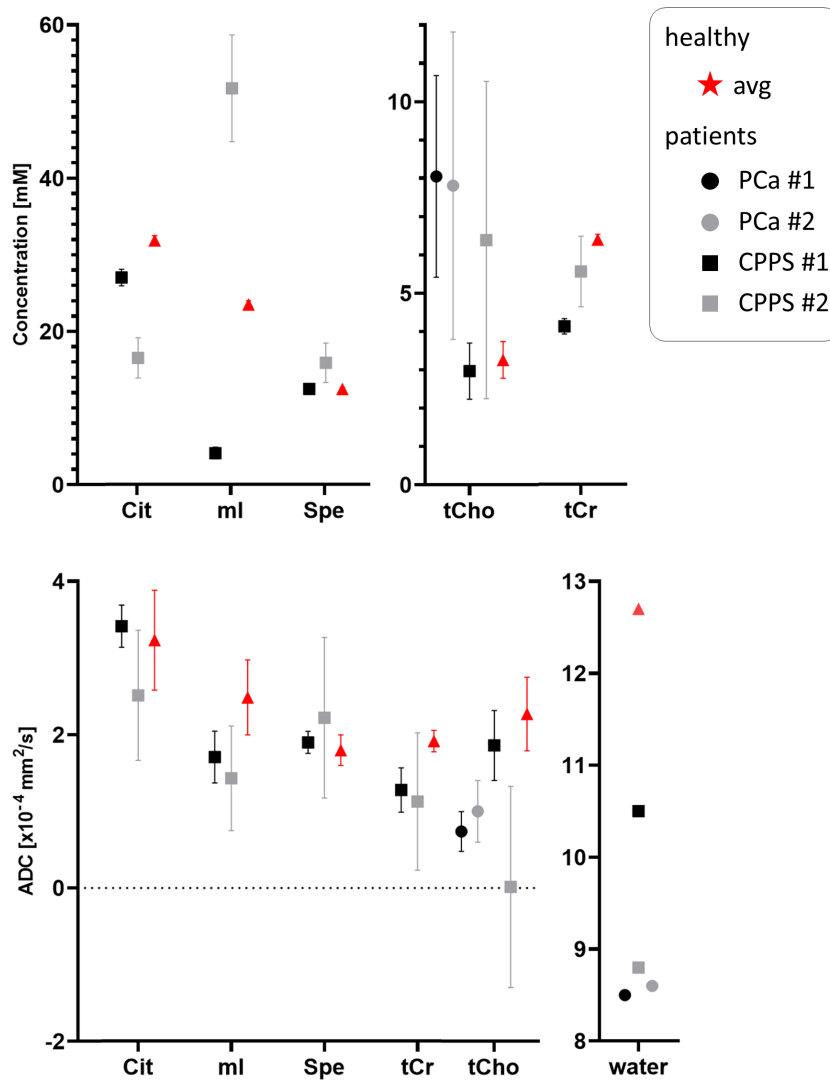
Figure 7.19: Comparison of findings for metabolite concentrations and ADCs for metabolite and water for the initial four patients with PCa (circles) and prostatitis (squares) related to the norm from 9 healthy subjects (red triangle). CRLBs are reported as error bars.

Conclusions & Outlook

The current work explored the development and the deployment of new methodologies for acquiring MR Spectroscopic data with the idea of enhancing the sensitivity of the NMR signal to the chemistry and the physical properties of human tissues to either yield a patient-specific more complete viewpoint or to untangle open questions of human biology. Secondly, a new window on applying deep learning methods for MRS quantification is explored, with care on providing prediction uncertainty and investigating potential biases inherited from the training phase.

Simultaneous quantification of metabolite-specific concentrations and T_2 relaxation times

Multiparametric MR spectroscopy has been deployed at first for the purpose of speed, aiming at a complete patient-specific description of the metabolic profile at a given location, in regards not just to metabolite concentration levels but introducing information on metabolite-specific T_2 rates as well. A faster acquisition protocol twinned to bi-dimensional data fitting has been proposed and tested in both single-voxel and MRSI experiments. It features compatible clinical time duration of the measurement and comparable accuracy without systematic biases. Furthermore, gains in precision by the measure of CRLBs go beyond the mere compression of the experiment duration.

The method comes with limitations such as (1) T_2 rates that are produced for a handful of metabolites and (2) the hindered readability of the spectral profile for short TEs. Furthermore, the handling of residual water guides the optimization of the experiment schedule towards longer TE batches with reduced SNR for the longest and better-resolved acquired readout. These effects impact the modeling of such a *compact* dataset for any application to a new volume of interest, which must refer to a small initial batch of multi-echo multi-shot complete acquisitions where each spectrum in the TE batch is available in full resolution and can be used for the model tune-up as well as debugging and monitoring phases of the fitting algorithm. Only once the model is set up, it can be translated to the multi-echo single-shot compact dataset and can be reliably run over any study cohort.

To further optimize the reconstruction algorithm, it is of interest the development and optimization of weighted summation strategies to treat metabolite-cycled (MC) responses beyond plain

averaging. Such concepts have already been tailored to the DW-MRS motion compensation algorithm with consequent promising results [45]. The rationale behind such an algorithm would aim to maximize the level of water suppression by MC, account for water lineshape (i.e., partial echoes instead of FID), and deploy concepts of local and global water suppression efficiency (i.e., locality of water suppression looks at the amplitude ratio in the range 4.6-4.7 ppm [134] whereas a global water suppression index would explore the power of the signal throughout chemical shift range, including water tails contaminating spectral ranges of interest).

Furthermore, future directions for the investigation of the methodology account for its deployment for MRSI acquisition of the prostate. A semi-LASER multi-echo acquisition protocol has been tailored to acquire prostate metabolic profiles. It deploys GOIA adiabatic modulation and water suppression with MEGA editing. Data has already been acquired for three healthy volunteers and a prostate-mimic phantom. The postprocessing pipeline and the simulation of the basis set are ready, while the model tuning phase will soon begin with a first approach to multi-echo multi-shot acquisitions.

A further exciting evolution of the methodology could be integrating T_1 and diffusion-weighting encoding in a single 15-min long single-voxel multiparametric MRS experiment. Sensitivity to T_1 could be integrated into the multi-echo experiment by embedding an inversion recovery pulse before the semi-LASER block. Diffusion-weighting will require the implementation of diffusion-encoding gradients within the semi-LASER block. Such a multiparametric schedule may be twinned to a 4D model where concentrations, T_1, T_2, and ADC are fitted together simultaneously. This method would offer an alternative approach to MR spectroscopy fingerprinting (MRSF) [135], where the modeling part would be more accessible and could provide better interpretability. However, integration of MRSF with multi-echo single-shot acquisition may also be of interest.

Diffusion weighted single-voxel MR Spectroscopy of the prostate

Diffusion-weighted Magnetic Resonance Spectroscopy (DW-MRS) featuring a dedicated acquisition setup and postprocessing technique was deployed to investigate the prostate, an organ intrinsically more prone to motion issues, for the first time. The prostate complex microstructure was investigated and disclosed according to (1) the selective distribution of metabolites in different subspaces and (2) the complex interaction between metabolites and proteins in the luminal space.

The current results are hindered by the single voxel approach, which considers a relatively large volume of interest to benefit SNR but does not account for the substantial prostate tissue heterogeneity, which would specify the different cellular distribution of shape, size, and morphology. Furthermore, on the one hand, the current disclosure of the complex bonding between proteins

and compounds, like spermine and citrate, does not find answers regarding the nature of the proteins interacting in such chemical bonds and the intensity to which these interactions take place. Therefore, further analysis of in vitro solutions doped with different protein substrates and various metabolite concentration levels is needed. On the other hand, the clinical potential of the method stressed by the differences between healthy cases and patients was supported by a very limited cohort. Therefore, the study will need the integration of a more significant cohort of patients.

The twinning of metabolic apparent diffusion coefficient to metabolite-specific T_2 rates can offer a complementary view of the tissue microenvironment. Therefore, the integration of diffusion-weighted MRS to multi-echo measurement by extending the multi-echo technology with the embedding of diffusion encoding gradients or by the meaning of integrating the current dataset with a standard multi-echo multi-shot measurement could offer a better understanding of the whole biological picture.

At last, overcoming the limitation of single-voxel DW-MRS towards a DW-MRSI experiment finds the perfect candidate organ for investigation in the prostate, where its tissue heterogeneity would find a more suitable description and assist even further its potential clinical value.

Quantification of MR spectra by DL: optimizations and pitfalls

Various approaches utilizing (1) different preprocessing techniques to highlight spectroscopic features and (2) different convolutional neural network architectures have been investigated to quantify metabolic profiles. Performances hint toward systematic biases of predictions linked to the training set's structure. Strategies to overcome such limitations are found to mitigate the issue. Nevertheless, care for the measures used to evaluate DL performances is pointed out, recommending the evaluation of a broader and complementary set of scores and suggesting an initial attempt to describe predictions' uncertainty.

Further challenges toward better model interpretability may account for integrating model-aware designs, where DL architectures are embedded with MRS prior knowledge inspired by MR physics in traditional modeling [70]. Moreover, upon the uprising of new trends in the fast-moving DL community, newer State of the Art models such as *diffusion models* [136] or *transformers* [137] won't take long before their adaptation meet the MRS domain.

The robustness and reliability of DL models should further be enhanced within clinical routines by the development of unified frameworks that not simply predict a target output but are aware of and quantify the risks related to performing such a task. In other words, the modern pervasiveness of large-scale deep neural networks is solely and unfortunately currently driven by the fulfillment of extraordinary performance on complex problems. Therefore, it is plagued by

their sudden, unexpected, and often catastrophic failures, particularly in challenging scenarios [138]. The community effort of developing DL tools should, therefore, not limit itself to performance but directly plug into existing training pipelines insight into bias (dataset distributions and imbalance) as well as aleatoric (data) and epistemic (model) uncertainty [70].

Nevertheless, open questions remain about the contribution of explainability to trust in AI [139].

Bibliography

[1]   R A De Graaf. *In vivo NMR spectroscopy: Principles and techniques*. 3rd ed. John Wiley & Sons, 2019.

[2]   U Fano. "Description of States in Quantum Mechanics by Density Matrix and Operator Techniques". In: *Rev. Mod. Phys.* 29 (1957), pp. 74–93.

[3]   J Near et al. "Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations". In: *NMR in Biomedicine* 34.5 (2021), e4257.

[4]   I Tkáč et al. "Water and lipid suppression techniques for advanced 1H MRS and MRSI of the human brain: Experts' consensus recommendations". In: *NMR in Biomedicine* 34.5 (2021), e4459.

[5]   I K Steinseifer et al. "Improved volume selective 1H MR spectroscopic imaging of the prostate with gradient offset independent adiabaticity pulses at 3 tesla". In: *Magnetic Resonance in Medicine* 74.4 (Oct. 2015), pp. 915–924.

[6]   M Mescher et al. "Simultaneous in vivo spectral editing and water suppression". In: *NMR in Biomedicine* 11 (1998), pp. 266–272.

[7]   M Wilson et al. "Methodological consensus on clinical proton MRS of the brain: Review and recommendations". In: *Magnetic Resonance in Medicine* 82.2 (2019), pp. 527–550.

[8]   R Kreis. "Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts". In: *NMR in Biomedicine* 17.6 (2004), pp. 361–381.

[9]   U Klose. "In vivo proton spectroscopy in presence of eddy currents". In: *Magnetic Resonance in Medicine* 14.1 (1990), pp. 26–30.

[10] O C Andronesi et al. "Motion correction methods for MRS: experts' consensus recommendations". In: *NMR in Biomedicine* 34.5 (2021), e4364.

[11] E Cabanes et al. "Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain". In: *Journal of Magnetic Resonance* 150.2 (2001), pp. 116–125.

[12] D Stefan et al. "Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package". In: *Measurement Science & Technology* 20 (2009), p. 104035.

[13] B J Soher et al. "VeSPA: Integrated applications for RF pulse design, spectral simulation and MRS data analysis". In: *19th Annual Meeting of ISMRM*. 2011, p. 1410.

[14] A A Maudsley et al. "Numerical simulation of PRESS localized MR spectroscopy". In: *J Magn Reson* 173.1 (2005), pp. 54–63.

[15] S W Provencher. "Automatic quantitation of localized in vivo (1)H spectra with LCModel". In: *NMR in Biomedicine* 14.4 (2001), pp. 260–264.

[16] D G Q Chong et al. "Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 24 (2011), pp. 147–164.

[17] K Landheer and Christoph Juchem. "Are Cramér-Rao lower bounds an accurate estimate for standard deviations in in vivo magnetic resonance spectroscopy?" In: *NMR in Biomedicine* 34.7 (2021), e4521.

[18] S Cavassila et al. "Cramer-Rao bounds: an evaluation tool for quantitation". In: *NMR in Biomedicine* 14.4 (2001), pp. 278–283.

[19] C S Bolliger, C Boesch, and R Kreis. "On the use of Cramér-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments". In: *NeuroImage* 83 (2013), pp. 1031–1040.

[20] A Van den Bos. "Parameter Estimation". In: *Handbook of Measurement Science*. Ed. by P H Sydenham. Vol. 1. New York: Wiley, 1982. Chap. 8, pp. 331–377.

[21] R de Beer and D van Ormondt. "Analysis of NMR Data Using Time Domain Fitting Procedures". In: *In-Vivo Magnetic Resonance Spectroscopy I: Probeheads and Radiofrequency Pulses Spectrum Analysis*. Ed. by M. Rudin. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 201–248.

[22] D Van Ormondt et al. "Time-domain methods for quantifying MR spectra". In: *eMagRes* 4.3 (2015), pp. 651–661.

[23] M Hoefemann et al. "Parameterization of metabolite and macromolecule contributions in interrelated MR spectra of human brain using multidimensional modeling". In: *NMR Biomed* 33.9 (2020), e4328.

[24] A Tal. "The future is 2D: spectral-temporal fitting of dynamic MRS data provides exponential gains in precision over conventional approaches". In: *Magnetic Resonance in Medicine* Early View (2022).

[25] A Horská, P B Barker, and D Phil. "Imaging of Brain Tumors: MR Spectroscopy and Metabolic Imaging". In: *Neuroimaging Clin N Am.* 20.3 (2010), pp. 293–310.

[26] G Oz et al. "Clinical proton MR spectroscopy in central nervous system disorders". In: *Radiology* 270.3 (2014), pp. 658–679.

[27] H Zhu and P B Barker. "MR Spectroscopy and Spectroscopic Imaging of the Brain". In: *Methods Mol Biol.* 711 (2011), pp. 203–226.

[28] V Sawlani et al. "Multiparametric MRI: practical approach and pictorial review of a useful tool in the evaluation of brain tumours and tumour-like lesions". In: *Insights into Imaging* 11.1 (Dec. 2020), pp. 1–19.

[29] F A Howe et al. "Metabolic profiles of human brain tumors using quantitative in vivo 1H magnetic resonance spectroscopy". In: *Magnetic Resonance in Medicine* 49.2 (2003), pp. 223–232.

[30] C Cudalbu et al. "Contribution of macromolecules to brain 1H MR spectra: Experts' consensus recommendations". In: *NMR in Biomedicine* 34.5 (2021), e4393.

[31] K Şimşek et al. "Macromolecular background signal and non-Gaussian metabolite diffusion determined in human brain using ultra-high diffusion weighting". In: *Magnetic Resonance in Medicine* 88.5 (Nov. 2022), pp. 1962–1977.

[32] S P Kyathanahally, A Döring, and R Kreis. "Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy". In: *Magnetic Resonance in Medicine* 80.3 (2018), pp. 851–863.

[33] J H Yacoub and Aytekin O. "MR Imaging of Prostate Zonal Anatomy". In: *Radiologic clinics of North America* 56 (2018), pp. 197–209.

[34] A Stamatelatou, T W J Scheenen, and A Heerschap. "Developments in proton MR spectroscopic imaging of prostate cancer". In: *Magma (New York, N.y.)* 35.4 (Aug. 2022), p. 645.

[35] A Saha, M Hosseinzadeh, and H Huisman. "End-to-end prostate cancer detection in bpMRI via 3D CNNs: Effects of attention mechanisms, clinical priori and decoupled false positive reduction". In: *Medical Image Analysis* 73 (2021), p. 102155.

[36] J R Packer and Norman J Maitland. "The molecular and cellular origin of human prostate cancer". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863.6, Part A (2016), pp. 1238–1260.

[37] C Juchem et al. "B0 shimming for in vivo magnetic resonance spectroscopy: Experts' consensus recommendations". In: *NMR in Biomedicine* 34.5 (2021), e4350.

[38] M Terpstra et al. "Test-retest reproducibility of neurochemical profiles with short-echo, single-voxel MR spectroscopy at 3T and 7T". In: *Magnetic Resonance in Medicine* 76.4 (2016), pp. 1083–1091.

[39] A Kulpanovich and A Tal. "What is the Optimal Schedule for Multiparametric MRS? A Magnetic Resonance Fingerprinting Perspective". In: *NMR Biomed* 34.5 (2021), e4196.

[40]   I I Kirov et al. "Age Dependence of Regional Proton Metabolites T2 Relaxation Times in the Human Brain at 3 Tesla". In: *Magn Reson Med* 60.4 (2008), pp. 790–795.

[41]   I I Kirov et al. "Brain Metabolite Proton T2 Mapping at 3.0 T in Relapsing-Remitting Multiple Sclerosis". In: *Radiology* 254.3 (2010), pp. 858–866.

[42]   M Dumoulin et al. "Increased brain metabolite T2 relaxation times in patients with Alzheimer's disease". In: *International Society for Magnetic Resonance in Medicine*. 2005, p. 1179.

[43]   T Isobe et al. "Quantification of cerebral metabolites in glioma patients with proton MR spectroscopy using T2 relaxation time correction". In: *Magnetic Resonance Imaging* 20.4 (2002), pp. 343–349.

[44]   Y Li et al. "Comparison of T 1 and T 2 metabolite relaxation times in glioma and normal brain at 3 T". In: *J Magn Reson Imaging* 28.2 (2008), pp. 342–350.

[45]   A Döring et al. "Diffusion-weighted magnetic resonance spectroscopy boosted by simultaneously acquired water reference signals". In: *Magnetic Resonance in Medicine* 80.6 (2018), pp. 2326–2338.

[46]   3blue1brown. *3blue1brown*. 2023. URL: https://www.3blue1brown.com/ (visited on 01/31/2023).

[47]   MIT Deep Learning team. *MIT Deep Learning 6.S191*. 2023. URL: http://introtodeeplearning.com/ (visited on 01/31/2023).

[48]   "Tools such as ChatGPT threaten transparent science; here are our ground rules for their use". In: *Nature* 613 (2023), p. 612.

[49]   H H Thorp. "ChatGPT is fun, but not an author". In: *Science* 379.6630 (2023), p. 313.

[50]   A Krizhevsky, I Sutskever, and G E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F Pereira et al. Vol. 25. Curran Associates, Inc., 2012.

[51]   A Krizhevsky, I Sutskever, and G E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F Pereira et al. Vol. 25. Curran Associates, Inc., 2012.

[52]   I J Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems* 27 (2014).

[53]   O Ronneberger, P Fischer, and T Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *Lecture Notes in Computer Science* 9351 (2015), pp. 234–241.

[54] K He et al. "Deep residual learning for image recognition". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778. arXiv: 1512.03385.

[55] C Szegedy et al. "Going deeper with convolutions". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2015, pp. 1–9.

[56] G Huang et al. "Densely connected convolutional networks". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-January (Nov. 2017), pp. 2261–2269. arXiv: 1608.06993.

[57] M Tan and Q V Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning* (2019).

[58] C Meng et al. "Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset". In: *Scientific Reports* 12.1 (May 2022), pp. 1–28. arXiv: 2102.06761.

[59] M Ennab and H McHeick. "Designing an Interpretability-Based Model to Explain the Artificial Intelligence Algorithms in Healthcare". In: *Diagnostics* 12.7 (July 2022), p. 1557.

[60] C Olah et al. "The Building Blocks of Interpretability". In: *Distill* (2018).

[61] C Olah, A Mordvintsev, and L Schubert. "Feature Visualization". In: *Distill* (2017).

[62] Body of Knowledge - For Humanity. *Accuracy, validity, reliability, robustness and resilience (AVR3)*. 2022. URL: https://forhumanity.center/bok/accuracy-validity-reliability-robustness-and-resilience-avr3/#_ftnref1 (visited on 02/01/2023).

[63] A Jungo and M Reyes. "Assessing Reliability and Challenges of Uncertainty Estimations for Medical Image Segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11765 LNCS. Springer Science and Business Media Deutschland GmbH, 2019, pp. 48–56. arXiv: 1907.03338.

[64] N G Gyori et al. "Training data distribution significantly impacts the estimation of tissue microstructure with machine learning". In: *Magnetic Resonance in Medicine* 87.2 (Feb. 2022), pp. 932–947.

[65] R Rizzo et al. "Quantification of MR spectra by deep learning in an idealized setting: Investigation of forms of input, network architectures, optimization by ensembles of networks, and training bias". In: *Magnetic Resonance in Medicine* (2022).

[66] C Guo et al. "On Calibration of Modern Neural Networks". In: *34th International Conference on Machine Learning (PLMR)*. 2017, 70:1321–1330. arXiv: 1706.04599v2.

[67] V Kuleshov, N Fenner, and S Ermon. "Accurate Uncertainties for Deep Learning Using Calibrated Regression". In: *Proceedings of the 35th International Conference on Machine Learning* 80 (2018).

[68] A Niculescu-Mizil and R Caruana. "Predicting Good Probabilities With Supervised Learning". In: *Proceedings of the 22nd International Conference on Machine Learning*. 2005.

[69] M Abdar et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges". In: *Information Fusion* 76 (Dec. 2021), pp. 243–297. arXiv: 2011.06225.

[70] A F Psaros et al. "Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons". In: *Journal of Computational Physics* 477 (Mar. 2023).

[71] A Kendall and Y Gal. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In: *31st Conference on Neural Information Processing Systems (NIPS)*. Long Beach, CA, USA, 2017.

[72] H Xu et al. "Digging into Uncertainty in Self-supervised Multi-view Stereo". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2021, pp. 6058–6067.

[73] Clarivate. 2022. URL: https://www.webofscience.com/ (visited on 12/31/2022).

[74] A Shamaei et al. "Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals". In: *Magnetic Resonance in Medicine* 89.3 (Mar. 2023), pp. 1221–1236.

[75] J Jang et al. "Unsupervised anomaly detection using generative adversarial networks in 1H-MRS of the brain". In: *Journal of Magnetic Resonance* 325 (2021), p. 106936.

[76] H Lee, H H Lee, and H Kim. "Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy". In: *Magnetic Resonance in Medicine* 84.2 (2020), pp. 559–568.

[77] D F Hansen. "Using Deep Neural Networks to Reconstruct Non-uniformly Sampled NMR Spectra". In: *Journal of Biomolecular NMR* 73.10-11 (Nov. 2019), pp. 577–585.

[78] K L Chan, T Ziegs, and A Henning. "Improved signal-to-noise performance of MultiNet GRAPPA 1H FID MRSI reconstruction with semi-synthetic calibration data". In: *Magnetic Resonance in Medicine* 88.4 (Oct. 2022), pp. 1500–1515.

[79] Z Iqbal et al. "Super-Resolution 1H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning". In: *Frontiers in Oncology* 9.October (2019), pp. 1–13. arXiv: 1802.07909.

[80] D Chen et al. "Magnetic Resonance Spectroscopy Deep Learning Denoising Using Few In Vivo Data". In: *arXiv*. Jan. 2021. arXiv: 2101.11442v4.

[81] M Dziadosz et al. "Denoising MR spectra by deep learning: miracle or mirage?" In: *31st International Society of Magnetic Resonance in Medicine (ISMRM)*. London, 2022.

[82] M Bastiaan Schilder. "Noise reduction and artefact removal in MR spectroscopy using wavelet analysis and deep learning". PhD thesis. 2023.

[83] A Abdoli, R Stoyanova, and A A Maudsley. "Denoising of MR Spectroscopic Imaging Data Using Statistical Selection of Principal Components". In: *Magma (New York, N.Y.)* 29.6 (Dec. 2016), p. 811.

[84] B C Rowland et al. "A comparison of denoising methods in dynamic MRS using pseudo-synthetic data". In: *medRxiv* (Feb. 2021), p. 2021.02.23.21252282.

[85] W T Clarke and M Chiew. "Uncertainty in denoising of MRSI using low-rank methods". In: *Magnetic Resonance in Medicine* 87.2 (Feb. 2022), pp. 574–588.

[86] J Mosso et al. "MP-PCA denoising for diffusion MRS data: promises and pitfalls". In: *NeuroImage* 263 (Nov. 2022), p. 119634.

[87] S Motyka et al. "k-Space-based coil combination via geometric deep learning for reconstruction of non-Cartesian MRSI data". In: *Magnetic Resonance in Medicine* 86.5 (Nov. 2021), pp. 2353–2367.

[88] S S Gurbani et al. "A convolutional neural network to filter artifacts in spectroscopic MRI". In: *Magnetic Resonance in Medicine* 80.5 (2018), pp. 1765–1775.

[89] S Tapper et al. "Frequency and phase correction of J-difference edited MR spectra using deep learning". In: *Magnetic Resonance in Medicine* 85.4 (2021), pp. 1755–1765.

[90] N Hatami, M Sdika, and H Ratiney. "Magnetic resonance spectroscopy quantification using deep learning". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11070 LNCS (2018), pp. 467–475. arXiv: 1806.07237.

[91] H H Lee and H Kim. "Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain". In: *Magnetic Resonance in Medicine* 82.1 (2019), pp. 33–48.

[92] S S Gurbani et al. "Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting". In: *MRM* September 2018 (2019), pp. 3346–3357.

[93] H H Lee and H Kim. "Bayesian deep learning-based 1 H-MRS of the brain: Metabolite quantification with uncertainty estimation using Monte Carlo dropout". In: *Magnetic resonance in medicine* 88.1 (July 2022), pp. 38–52.

[94]   A Shamaei, J Starcukova, and Z Starcuk. "Physics-informed Deep Learning Approach to Quantification of Human Brain Metabolites from Magnetic Resonance Spectroscopy Data". In: *bioRxiv* (Oct. 2022), p. 2022.10.13.512064.

[95]   H H Lee and H Kim. "Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain". In: *Magnetic Resonance in Medicine* 84.4 (Oct. 2020), pp. 1689–1706.

[96]   M Chandler et al. "MRSNet: Metabolite Quantification from Edited Magnetic Resonance Spectra With Convolutional Neural Networks". In: *arXiv*. 2019, arXiv.1909.03836. arXiv: 1909.03836.

[97]   J Songeon et al. "In vivo magnetic resonance 31P-Spectral Analysis With Neural Networks: 31P-SPAWNN". In: *Magnetic Resonance in Medicine* 89.1 (Jan. 2023), pp. 40–53.

[98]   Z Iqbal et al. "Deep learning can accelerate and quantify simulated localized correlated spectroscopy". In: *Scientific Reports* 11.1 (Apr. 2021), pp. 1–13.

[99]   R Rizzo et al. "Reliability of Quantification Estimates in MR Spectroscopy: CNNs vs Traditional Model Fitting". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science* 13438 (2022), pp. 715–724.

[100] A A Maudsley et al. "Advanced magnetic resonance spectroscopic neuroimaging: Experts' consensus recommendations". In: *NMR in Biomedicine* 34.5 (2021), pp. 1–22.

[101] R Kreis et al. "Integrated data acquisition and processing to determine metabolite contents, relaxation times, and macromolecule baseline in single examinations of individual subjects". In: *Magnetic Resonance in Medicine* 54.4 (2005), pp. 761–768.

[102] J I Kukurova et al. "Two-dimensional spectroscopic imaging with combined free induction decay and long-TE acquisition (FID echo spectroscopic imaging, FIDESI) for the detection of intramyocellular lipids in calf muscle at 7 T". In: *NMR in Biomedicine* 27 (2014), pp. 980–987.

[103] L An, S Li, and J Shen. "Simultaneous determination of metabolite concentrations, T1 and T2 relaxation times". In: *Magnetic Resonance in Medicine* 78.6 (2017), pp. 2072–2081.

[104] R Rizzo and R Kreis. "Multi-Parametric Single-Shot Magnetic Resonance Spectroscopy for Fast Metabolite Specific Concentration and T2 Determination". In: *International Society for Magnetic Resonance in Medicine*. 2022, p. 311.

[105] W T Clarke, C J Stagg, and S Jbabdi. "FSL-MRS: An end-to-end spectroscopy analysis package." eng. In: *Magnetic resonance in medicine* 85.6 (June 2021), pp. 2950–2964.

[106] R Srinivasan et al. "TE-Averaged two-dimensional proton spectroscopic imaging of glutamate at 3 T". In: *NeuroImage* 30.4 (2006), pp. 1171–1178.

[107] P J W Pouwels and J Frahm. "Differential distribution of NAA and NAAG in human brain as determined by quantitative localized proton MRS". In: *NMR in Biomedicine* 10 (1997), pp. 73–78.

[108] P O Wyss et al. "In vivo estimation of transverse relaxation time constant (T2) of 17 human brain metabolites at 3T". In: *Magnetic Resonance in Medicine* 80 (2018), pp. 452–461.

[109] F Träber et al. "1H Metabolite Relaxation Times at 3.0 Tesla: Measurements of T1 and T2 Values in Normal Brain and Determination of Regional Differences in Transverse Relaxation". In: *Journal of Magnetic Resonance Imaging* 19 (2004), pp. 537–545.

[110] R Kreis, T Ernst, and B D Ross. "Absolute quantitation of water and metabolites in the human brain. II. Metabolite concentrations". In: *Journal Magnetic Resonance Series B* 102 (1993), pp. 9–19.

[111] Y Wang and S J Li. "Differentiation of metabolic concentrations between gray matter and white matter of human brain by in vivo 1H magnetic resonance spectroscopy". In: *Magnetic Resonance in Medicine* 39.1 (1998), pp. 28–33.

[112] J Z Bojorquez et al. "What are normal relaxation times of tissues at 3 T?" In: *Magnetic Resonance Imaging* 35 (Jan. 2017), pp. 69–80.

[113] G Lemberskiy et al. "Characterization of prostate microstructure using water diffusion and NMR relaxation". In: *Frontiers in Physics* 6.SEP (Sept. 2018), p. 91.

[114] A Chatterjee et al. "Validation of Prostate Tissue Composition by Using Hybrid Multidimensional MRI: Correlation with Histologic Findings". In: *Radiology* 302.2 (Feb. 2022), pp. 368–377.

[115] M Jupin, F H A van Heijster, and A Heerschap. "Metabolite interactions in prostatic fluid mimics assessed by 1H NMR". In: *Magma (New York, N.Y.)* 35.4 (Aug. 2022), pp. 683–694.

[116] P R Braadland et al. "Ex vivo metabolic fingerprinting identifies biomarkers predictive of prostate cancer recurrence following radical prostatectomy". In: *British Journal of Cancer* 117.11 (2017), pp. 1656–1664.

[117] N Tayari, A J Wright, and A Heerschap. "Absolute choline tissue concentration mapping for prostate cancer localization and characterization using 3D 1 H MRSI

without water-signal suppression". In: *Magnetic resonance in medicine* 87.2 (Feb. 2022), pp. 561–573.

[118]   N Tayari et al. "In vivo MR spectroscopic imaging of the prostate, from application to interpretation". In: *Analytical Biochemistry* 529 (July 2017), pp. 158–170.

[119]   M J Lynch and J K Nicholson. "Proton MRS of human prostatic fluid: correlations between citrate, spermine, and myo-inositol levels and changes with disease." eng. In: *The Prostate* 30.4 (Mar. 1997), pp. 248–255.

[120]   N J Serkova et al. "The metabolites citrate, myo-inositol, and spermine are potential age-independent markers of prostate cancer in human expressed prostatic secretions." eng. In: *The Prostate* 68.6 (May 2008), pp. 620–628.

[121]   S Sabouri et al. "MR measurement of luminal water in prostate gland: Quantitative correlation between MRI and histology". In: *Journal of Magnetic Resonance Imaging* 46.3 (2017), pp. 861–869.

[122]   T H Storås et al. "Prostate magnetic resonance imaging: multiexponential T2 decay in prostate tissue." eng. In: *Journal of magnetic resonance imaging : JMRI* 28.5 (Nov. 2008), pp. 1166–1172.

[123]   I Ronen and J Valette. "Diffusion-weighted magnetic resonance spectroscopy". In: *eMagRes* 4.4 (2015), pp. 733–750.

[124]   D Le Bihan. "Molecular diffusion, tissue microdynamics and microstructure". In: *NMR Biomed* 8.7-8 (1995), pp. 375–386.

[125]   M Palombo et al. "Insights into brain microstructure from in vivo DW-MRS". In: *Neuroimage* 182 (2018), pp. 97–116.

[126]   N Pedrosa de Barros et al. "Automatic quality control in clinical 1H MRSI of brain cancer". In: *NMR in Biomedicine* 29.5 (2016), pp. 563–575.

[127]   J Weis et al. "Quantification of metabolite concentrations in benign and malignant prostate tissues using 3D proton MR spectroscopic imaging". In: *Journal of magnetic resonance imaging : JMRI* 45.4 (Apr. 2017), pp. 1232–1240.

[128]   M Harada et al. "Diffusion-weighted in vivo localized proton MR spectroscopy of human cerebral ischemia and tumor". In: *NMR in Biomedicine* 15.1 (2002), pp. 69–74.

[129]   J M Hakumäki et al. "Quantitative 1H nuclear magnetic resonance diffusion spectroscopy of BT4C rat glioma during thymidine kinase-mediated gene therapy in vivo: identification of apoptotic response." In: *Cancer Res* 58.17 (1998), pp. 3791–9.

[130]   A Shukla-Dave et al. "Chronic Prostatitis: MR Imaging and 1H MR Spectroscopic Imaging Findings—Initial Observations1". In: *Radiology* 231.3 (June 2004), pp. 717–724.

[131] K N A Nagel et al. "Differentiation of Prostatitis and Prostate Cancer by Using Diffusion-weighted MR Imaging and MR-guided Biopsy at 3 T". In: *Radiology* 267.1 (Apr. 2013), pp. 164–172.

[132] X Liu et al. "Biexponential apparent diffusion coefficients values in the prostate: Comparison among normal tissue, prostate cancer, benign prostatic hyperplasia and prostatitis". In: *Korean Journal of Radiology* 14.2 (Mar. 2013), pp. 222–232.

[133] H Shinmoto et al. "Biexponential apparent diffusion coefficients in prostate cancer". In: *Magnetic Resonance Imaging* 27.3 (Apr. 2009), pp. 355–359.

[134] R Kreis et al. "Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: Background and experts' consensus recommendations". In: *NMR in biomedicine* 34.5 (2020), e4347.

[135] A Kulpanovich and A Tal. "The application of magnetic resonance fingerprinting to single voxel proton spectroscopy". In: *NMR in Biomedicine* 31.11 (2018), e4001.

[136] A. Vaswani et al. "Diffusion Models: A Comprehensive Survey of Methods and Applications". In: (2022). arXiv: 2209.00796.

[137] L. Yang et al. "Attention Is All You Need". In: (2017). arXiv: 1706.03762.

[138] Anonymus Authors. *Capsa: a unified framework for quantifying risk in deep neural networks*. 2023. URL: https://openreview.net/pdf?id=_BSowr-_ED (visited on 01/31/2023).

[139] A Ferrario and M Loi. "How Explainability Contributes to Trust in AI". In: *SSRN Electronic Journal* (Jan. 2022).

List of Publications

as of April 2023

Journal publications

- **Rizzo R**, Dziadosz M, Kyathanahally S P, Reyes M, Kreis R. "Reliability of quantification estimates in MR Spectroscopy: CNNs vs. traditional model fitting". *Lecture Notes in Computer Science (MICCAI)*, 2022, **13438**:715-724

- **Rizzo R**, Dziadosz M, Kyathanahally S P, Shamaei A, Kreis R. "Quantification of MR spectra by deep learning in an idealized setting: investigation of forms of input, network architectures, optimization by ensembles of networks and training bias". *Magnetic Resonance in Medicine*, 2022:1-21, `doi:10.1002/mrm.29561`

- Dziadosz M, **Rizzo R**, Kyathanahally S P, Kreis R. "Denosing single MR spectra by deep learning: miracle or mirage?" *Magnetic Resonance in Medicine*, 2023, **under revision**.

- **Rizzo R**, Kreis R. "Multi-echo single-shot spectroscopy combined with simultaneous 2D model fitting for fast and accurate measurement of metabolite-specific concentrations and T_2 relaxation times". *NMR in Biomedicine*, 2023, **in review**.

Conference proceedings

- **Rizzo R**, Kreis R. "Potential benefit from multi-echo single-shot spectroscopy with a combined fitting process". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2020, abstract #**2904**.

- **Rizzo R**, Kreis R. "Accounting for bias in estimated metabolite concentrations from cohort studies as caused by limiting the fitting parameter space". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2021, abstract #**2011**.

- **Rizzo R**, Dziadosz M, Kyathanahally S P, Kreis R. "Uncertainties in metabolite quantitation: CNN vs. traditional fitting". *European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Magn. Reson. Mater. Phy.*, 2021, **34**(1):S177.

- Dziadosz M, **Rizzo R**, Kyathanahally S P, Kreis R. "Denoising 1H MR spectra in a time-frequency representation by deep learning". *European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Magn. Reson. Mater. Phy.*, 2021, **34**(1):S175.

- **Rizzo R**, Dziadosz M, Kyathanahally S P, Kreis R. "Uncertainties and bias in quantification by deep learning in MR spectroscopy". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2022, abstract #**2617**.

- Dziadosz M, **Rizzo R**, Kyathanahally S P, Kreis R. "Denoising MR spectra by deep learning: miracle or mirage?". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2022, abstract #**2541**.

- **Rizzo R**, Kreis R. "Multiparametric single-shot MR spectroscopy for fast metabolite-specific concentrations and T_2 determination". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2022, abstract #**311**.

- **Rizzo R**, Dziadosz M, Kyathanahally S P, Kreis R. "Uncertainties for quantification of metabolites in Magnetic Resonance Spectroscopy: deep learning vs. model fitting". *Life Sciences, Switzerland*, Zurich, Switzerland, 2022.

- **Rizzo R**, Stamatelatou A, Heerschap A, Scheenen T, Kreis R. "Toward simultaneous concentrations and T_2 mapping of brain metabolites by multi-echo spectroscopic imaging". *MRS workshop*, 2022, abstract #**S3A2**.

- **Rizzo R**, Dziadosz M, Kyathanahally S P, Reyes M, Kreis R. "Reliability of quantification estimates in MR Spectroscopy: CNNs vs. traditional model fitting". *XXIV Swiss NMR Symposium*, Bern, Switzerland, 2022.

- **Rizzo R**, Stamatelatou A, Heerschap A, Scheenen T, Kreis R. "Simultaneous concentration and T_2 mapping of brain metabolites by multi-echo spectroscopic imaging". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2023, accepted abstract submission ID #**6723**.

- Stamatelatou A, **Rizzo R**, Simsek K, Van Asten J, Heerschap A, Scheenen T, Kreis R. "Diffusion-weighted MR spectroscopy of the prostate". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2023, accepted abstract submission ID #**7160**.

- **Rizzo R**, Stamatelatou A, Simsek K, Van Asten J, Kreis R, Heerschap A, Scheenen T. "Initial exploration of the potential of diffusion-weighted MRS for the evaluation of prostate pathology". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2023, accepted abstract submission ID #**7797**.

- Shamaei A, **Rizzo R**. "Physics-informed deep learning approach to quantifying MR spectroscopy data with simultaneous uncertainty estimation". *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2023, accepted abstract submission ID #**7208**.

- Heershap A, van Heijster F, Stamatelatou A, Simsek K, Van Asten J, **Rizzo R**, Kreis R, Scheenen T. "The association of citrate and spermine in prostatic fluid in phantoms and in humans". *European Molecular Imaging Meeting (EMIM)*, 2023, abstract #**703**.

Acknowledgements

What a journey it has been. While figuring out the puzzle of my Ph.D. I had the opportunity and fortune to grow very much on a personal level, too.

I wish to thank, first, my supervisor, Roland, who has proven to be more than a scientific leader and knowledgeable master, but an inspiring personal mentor. From ecology, sustainability, and politics to critical thinking, scientific growth, and horizons exploration, many are the teachings I will never forget. Thank you for such an excellent opportunity to join the lab here in Bern.

The second thank must go to my partner Cecilia who has supported and sustained me along the whole way. So many were the viewpoints that talking to you has helped me see different perspectives, professionally and personally, that I couldn't see myself.

Professionally, I'd like to express my thanks to Mauricio for many thinking out-of-the-box discussions. Thanks for having supported and guided us with precious tips in our DL brainstorming sessions. Tom and Arend, I am grateful for the opportunity of such collaborations, and thank you very much for having hosted me up in Nijmegen. I loved our continuing productive confrontation on multiple aspects of the field: from MR physics through biology and down to chemistry.

I wish to thank my colleagues. Among all, Kadir was more of a friend than anything else. Thank you for the extreme patience you showed daily: inspiring. A sincere thought goes to Karin too, who, as I like to call her, is the real *emotional glue* of the MRM group. If without PIs, there won't be any scientific MRM group, I dare to think that without Karin, there won't be any emotional and so well personally connected MRM group either.

I am incredibly grateful and lucky for having been involved in such a personally well-linked European network. I must really acknowledge Elina for her kindness, curiosity, and friendship. How could I ever forget that fantastic spontaneous chicken soup you brought me after getting horribly sick just a day after I stepped into the Netherlands? I am happy to have met Jessie, too, a precious, sincere, and inspiring friend.

A thought goes to my friends at home and Fabbri. Always there when I need them. The circle of acknowledgments closes with my family, and among all, mum and dad. Never ever less supportive, motivating, and blind believers I could hope for. Thank you for being my first supporters in life.

Declaration of Originality

Last name, first name: Rudy Rizzo

Matriculation number: 19-116-789

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations.
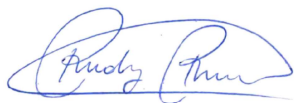
All data, tables, figures and text citations which have been reproduced from any other source, including the internet, have been explicitly acknowledged as such.

I am aware that in case of non-compliance, the Senate is entitled to withdraw the doctorate degree awarded to me on the basis of the present thesis, in accordance with the "Statut der Universität Bern (Universitätsstatut; UniSt)", Art. 69, of 7 June 2011.

Place, date

 Bern, 21.02.2023

Signature