NeuroImage 263 (2022) 119634

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

MP-PCA denoising for diffusion MRS data: promises and pitfalls

Jessie Mosso [1,2,3,*], Dunja Simicic [1,2,3], Kadir Şimşek [4,5,6], Roland Kreis [4,5], Cristina Cudalbu [1,2,#], Ileana O. Jelescu [7,#]

[1] *CIBM Center for Biomedical Imaging, Switzerland*
[2] *Animal Imaging and Technology, EPFL, Lausanne, Switzerland*
[3] *LIFMET, EPFL, Lausanne, Switzerland*
[4] *Magnetic Resonance Methodology, Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland*
[5] *Translational Imaging Center (TIC), Swiss Institute for Translational and Entrepreneurial Medicine, Bern, Switzerland*
[6] *Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[7] *Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland*

ARTICLE INFO

Keywords:
Marchenko-Pastur
PCA
diffusion-weighted MRS
denoising
brain

ABSTRACT

Diffusion-weighted (DW) magnetic resonance spectroscopy (MRS) suffers from a lower signal to noise ratio (SNR) compared to conventional MRS owing to the addition of diffusion attenuation. This technique can therefore strongly benefit from noise reduction strategies. In the present work, Marchenko-Pastur principal component analysis (MP-PCA) denoising is tested on Monte Carlo simulations and on *in vivo* DW-MRS data acquired at 9.4 T in rat brain and at 3 T in human brain. We provide a descriptive study of the effects observed following different MP-PCA denoising strategies (denoising the entire matrix versus using a sliding window), in terms of apparent SNR, rank selection, noise correlation within and across b-values and quantification of metabolite concentrations and fitted diffusion coefficients. MP-PCA denoising yielded an increased apparent SNR, a more accurate B_0 drift correction between shots, and similar estimates of metabolite concentrations and diffusivities compared to the raw data. No spectral residuals on individual shots were observed but correlations in the noise level across shells were introduced, an effect which was mitigated using a sliding window, but which should be carefully considered.

1. Introduction

Magnetic resonance spectroscopy (MRS) is a powerful technique that provides unique information about brain metabolite concentrations *in vivo*. Combined with diffusion weighting (DW), information on metabolites' diffusivities which are expected to reflect properties of the tissue microstructure can be extracted (Pfeuffer et al., 2000, Nicolay et al., 2001, de Graaf et al., 2001, Palombo et al., 2018, Ligneul et al., 2019, Najac et al., 2016, Ronen et al., 2014, Genovese et al., 2021). These properties include cell geometry, characteristic sizes of compartments, cytosol viscosity and molecular crowding. Unlike water, metabolites are naturally compartmentalized and probe the intracellular space almost exclusively. Some metabolites are even considered to be largely specific to glial cells, such as glutamine (Gln) or myo-inositol (mIns), some to neurons, such as N-acetylaspartate (NAA) or glutamate (Glu) (Palombo et al., 2018, Ligneul et al., 2019, Brand et al., 1993, Urenjak et al., 1993, Harris et al., 2015), while others are found in all cell types, such as creatine in all its forms (Rackayova et al., 2017) (total creatine: tCr). This intrinsic compartment specificity makes DW-MRS an extremely powerful tool to probe brain microstructure, in combination or in contrast to water diffusion MRI.

However, MRS is an inherently low signal-to-noise (SNR) technique due to the much lower concentration of metabolites relative to water, resulting in the need for substantial spectral averaging. For DW-MRS, even more extended averaging is needed to compensate for diffusion attenuation, and acquisition times become prohibitively long to parse multiple diffusion weightings (b-values), directions or diffusion times. DW-MRS data is typically acquired in single-voxel fashion. When fine spatial localization is required to study small structures, low SNR cannot be compensated by large voxel volumes. In this case, post-processing methods aiming to minimize the noise variance and its impact on the quantification of MRS signals are needed.

Several denoising schemes have been proposed, but remarkably none of them has been fully adopted by the MRS community (Ebel et al., 2006, Brender et al., 2019, Ahmed, 2005, Goryawala et al., 2020, Abdoli et al., 2016, Nguyen et al., 2013, Clarke and Chiew, 2022, Klauser et al., 2018, Simicic et al., 2021, Knoll et al., 2011, Pijnappel et al., 1992, Rowland et al., 2021, Doyle et al., 1994). Some of these denoising techniques, typically based on singular value decomposition (SVD) or another sparse

---

[*] Corresponding author.
E-mail address: jessie.mosso@epfl.ch (J. Mosso).
[#] equal contribution

https://doi.org/10.1016/j.neuroimage.2022.119634.
Received 5 May 2022; Received in revised form 7 September 2022; Accepted 14 September 2022

representation such as Fourier space or wavelets (Pijnappel et al., 1992, Rowland et al., 2021), have been implemented for spectroscopic imaging data (MRSI) (Goryawala et al., 2020, Abdoli et al., 2016, Nguyen et al., 2013, Clarke and Chiew, 2022, Klauser et al., 2018, Simicic et al., 2021), and mainly in clinical applications. These methods rely on linear predictability, partial separability of spatial-temporal modes, or both, of such data (Abdoli et al., 2016, Nguyen et al., 2013, Clarke and Chiew, 2022, Klauser et al., 2018). In addition, constraints on the spatial distribution of the signal with specific regularization, such as total generalized variation (TGV), has shown to further enhance the SNR in MRSI reconstruction (Klauser et al., 2018). TGV regularization aims to denoise by enforcing smooth spatial variations, however with known limitations in terms of detecting focal pathology (Knoll et al., 2011). Other approaches based on smoothing using splines, sliding windows or Gaussian windows lead to a deterioration of spectral/temporal resolution as well as artefactual auto-correlation (Rowland et al., 2021). Finally, deep learning approaches have been very recently suggested (Lei et al., 2021; Dziadosz et al., 2021), but likely require more investigation to become robust.

The main challenge of sparse representations such as SVD resides in the determination of the appropriate thresholds that separate the noise from the signal. In MRS, this arbitrary threshold can lead to possible elimination of spectral features that are on the same order of magnitude as noise components. One solution to choosing a threshold in a sparse domain has been proposed recently, with the initial aim to denoise diffusion MRI data (Veraart et al., 2016). It is based on the Marchenko-Pastur principal component analysis (MP-PCA) technique, which exploits the fact that noise eigenvalues follow the asymptotic universal Marchenko-Pastur distribution, a result of the random matrix theory for noisy covariance matrices. This method thus provides a data-driven (more specifically, noise-driven) approach to distinguish noise from the signal components in SVD, since the cut-off is obtained by iteratively fitting the MP distribution to the tail of eigenvalues, and has shown its superiority to TGV for instance (Veraart et al., 2016). In practice, MP-PCA is suitable for the denoising of data with a high level of redundancy and a constant noise level across them. In the case of a diffusion MRI dataset for example, this could correspond to images acquired with different diffusion-weightings and directions. Since its initial development for diffusion MRI, its applications have been extended to functional MRI (Ades-Aron et al., 2021, Diao et al., 2021), T_2 relaxometry (Does et al., 2019), preclinical ^{1}H-MRSI (Simicic et al., 2021) and ^{31}P-MRSI (Froeling et al., 2021). More recently, the NORDIC method (Moeller et al., 2021) has been introduced and addresses issues that are largely related to clinical diffusion MRI data, namely the use of multichannel coils for image acquisition acceleration, whose recombination results in a spatially varying and non-Gaussian noise distribution (cf. g-factor maps), and the fact that most data are retrieved and processed in magnitude space, further skewing the noise distribution. In the field of MRS, these two issues are in general not problematic since the multiple coil data featuring Gaussian noise are linearly combined maintaining the Gaussian characteristics, and since complex-valued data is used. In the broader context of matrix denoising, soft thresholding and optimal shrinkage of singular values (Gavish and Donoho, 2017, Johnstone and Paul, 2018, Ma et al., 2020) have shown to outperform hard thresholding like MP, especially in the case of low SNR input matrices.

The aim of the present study was to implement and test the potential of MP-PCA for denoising ^{1}H DW-MRS. The performance of MP-PCA was tested using Monte Carlo simulations and *in vivo* experiments in rat brain at 9.4 T and in human brain at 3 T.

2. Methods

The following terminology will be used throughout the manuscript. The SNR referred to as *time-domain SNR* in simulations is defined as the magnitude (absolute value) of the first complex point of the free induction decay (FID) over one standard deviation (SD) of noise, taken on the real part of the FID tail (time points 1500 to 2048) (Kreis et al., 2021). The SNR referred to as *spectral SNR* or *SNR* corresponds to the SNR of the NAA singlet at 2.01 ppm, defined as the NAA peak height taken on magnitude spectra to avoid phasing and linewidth issues, over one standard deviation of noise taken in a noise-only region of the real part of the spectra (from 8.2 to 10.9 ppm for simulations and rodent data, and 13.0 to 20.1 ppm for human data).

The term *apparent SNR* will be used to refer to the SNR after denoising. The term *shot* will be used to refer to every complex FID in each shell, i.e. of a row of matrix Z, according to a recent consensus on terminology in MRS (Kreis et al., 2021). The terms *shell* will be used to designate a set of 100 (simulations), 128 (*in vivo* – rodent), and 32 (*in vivo* – human) shots for a given b-value. The term *estimated spectral fit uncertainty (ESFU)* (Marjańska et al., 2022) will be used to refer to the estimated lower error bounds for the concentration estimates determined by LCModel, for which the term *Cramer Rao Lower Bounds (CRLB)* may not apply after denoising.

2.1. Theory

Let Z be an initial noisy matrix in the temporal domain, $Z \in \mathcal{M}_{n\times m}(\mathbb{C})$, where n is the number of shots, and m is the number of time points in the FID signal:

$$Z = \tilde{Z} + \varepsilon$$

where $\tilde{Z} \in \mathcal{M}_{n\times m}(\mathbb{C})$ is the signal information and $\varepsilon \in \mathcal{M}_{n\times m}(\mathbb{C})$ the Gaussian, uncorrelated noise. For this section, we will assume that $2n < m$ and $2n \gg 1$ (asymptotic condition of the MP law). The real and imaginary parts of Z are concatenated on the first dimension (n), and the resulting matrix $Y \in \mathcal{M}_{2n\times m}(\mathbb{R})$ is centered, such that:

$$X = Y - \mathbf{1}_{2n}^{\mathrm{T}}\overline{Y}$$

where $X \in \mathcal{M}_{2n\times m}(\mathbb{R})$, $\overline{Y} \in \mathcal{M}_{1\times m}(\mathbb{R})$ is the column-wise mean of Y and $\mathbf{1}_{2n}^{\mathrm{T}}$ is a column vector of $2n$ ones. Matrix X is then decomposed using the singular value decomposition:

$$X = USV^T$$

where $U \in \mathcal{M}_{2n\times 2n}(\mathbb{R})$, $S \in \mathcal{M}_{2n\times m}(\mathbb{R})$ and $V \in \mathcal{M}_{m\times m}(\mathbb{R})$. Columns of U are singular vectors of the first dimension (shots), columns of V are singular vectors of the second dimension (time points) and S contains the singular values of X, arranged in descending order, which are also the square root of the eigenvalues of $X^T X$. Since $X = Y - \mathbf{1}_{2n}^{\mathrm{T}}\overline{Y}$, $\frac{1}{2n}X^T X$ is the covariance matrix of Y. The Marchenko-Pastur distribution is then fitted to the smallest non-zeros eigenvalues λ of $\frac{1}{2n}X^T X$:

$$p(\lambda|\sigma,(2n-P)/m) = \begin{cases} \frac{\sqrt{(\lambda_+-\lambda)(\lambda-\lambda_-)}}{2\pi\lambda\sigma^2(2n-P)/m} & \textit{if } \lambda_- \le \lambda \le \lambda_+ \\ 0 & \textit{otherwise} \end{cases}$$

where σ is the noise level estimated from the input matrix X, P is the number of signal-carrying eigenvalues, λ_- the smallest noise-related eigenvalue and λ_+ the largest. P corresponds to the number of values λ such that $\lambda \ge \lambda_+$, with $\lambda_+ = \sigma^2(1+\sqrt{\frac{2n-P}{m}})^2$. The matrix Y can then be approximated by:

$$\hat{Y} = US_PV^T + \mathbf{1}_{2n}^{\mathrm{T}}\overline{Y}$$

where S has been truncated at rank P.

2.2. Monte Carlo simulations

Synthetic ^{1}H MR spectra were created (Matlab, MathWorks, Natick, MA, USA) to mimic experimental conditions in the rat brain (see Section 2.3 below). 19 metabolites, listed with their corresponding concentrations in **Table 1**, were simulated using NMRSCOPE-B from jMRUI (Starčuk and Starčuková, 2017), with published J-coupling and chemical shifts constants (Govindaraju et al., 2000, Govind et al., 2015) and

Table 1
Simulated metabolites with their respective concentrations and diffusion coefficients used in the MC simulations.

Metabolite	Concentration (mM)	D_{intra} (μm^2/ms)	Metabolite	Concentration (mM)	D_{intra} (μm^2/ms)
Alanine (Ala)	0.8	0.2695	Lactate (Lac)	0.8	0.65
Ascorbate (Asc)	1.5	0.3115	N-acetylaspartate (NAA)	9	0.4
Aspartate (Asp)	2	0.67	scyllo-Inositol (Scyllo)	0.1	0.3805
Creatine (Cr)	4	0.5	Taurine (Tau)	4.5	0.55
Phosphocreatine (PCr)	4.5	0.5	Glucose (Glc)	1.7	0.57
gamma-Aminobutyric acid (GABA)	1.6	0.378	N- acetylaspartylglutamate (NAAG)	0.3	0.4
Glutamine (Gln)	3	0.384	Phosphatidylethanolamine (PE)	0.5	0.318
Glutamate (Glu)	10	0.5	Glycerophosphocholine (GPC)	0.8	0.45
Glutathione (GSH)	1.5	0.2655	Phosphocholine (PCho)	0.2	0.45
myo-Inositol (mIns)	6.5	0.45			

the SPECIAL sequence (9.4 T, echo time (TE) 2.8 ms). The lineshapes of the individual signals were constructed using a sum of 0.2 Hz Lorentzian and 1.8 Hz Gaussian apodizations, and a full macromolecule spectrum acquired *in vivo* (MM, 1.3 mM) was included (Simicic et al., 2021).

The FID were generated with 2048 points. Diffusion weighting was simulated using Callaghan's model of diffusion in randomly oriented sticks (Callaghan et al., 1979), with 10 b-values: 0.4, 1.5, 3.4, 6, 7.6, 13.4, 15.7, 20.8, 25.2, 33.3 ms/μm^2. Intra-stick free diffusion coefficients ranging from 0.265 to 0.67 μm^2/ms (**Table 1**) were attributed to the 19 metabolites and 0.005 μm^2/ms to the MM. Metabolite intra-stick free diffusivities were set to be five times the apparent diffusion coefficient (ADC) of the ensemble of randomly-oriented cellular processes in the rodent brain from literature (Ligneul et al., 2019). These values were retrospectively found to be in the same range as the intra-stick free diffusion coefficients estimated *in vivo* in the present work. A residual water signal was added to each spectrum (16 Hz Lorentzian linewidth, mono-exponential decay with apparent diffusivity of 0.2 μm^2/ms, random phase). An additional 5 Hz Lorentzian line broadening was finally applied to all spectra. To simulate the full dataset for MP-PCA denoising (matrix Z), Gaussian noise was added to the real and imaginary parts of the FID, with a single shot time-domain SNR of 13. One hundred noisy FID were generated for each b-value (constituting a "shell") and B_0 drifts (random drift in [-15,+15] Hz) and phase distortions (random phase in [0,30]°) were added on individual shots, mimicking high SNR experimental *in vivo* rodent DW-MRS data. The initial matrix Z thus consisted of 1000 rows (10 b-values, 100 shots per b-value) and 2048 columns (FID time points).

Finally, the matrix Z was generated 100 times with different noise realizations, water residual signal, B_0 drifts and phase distortions, and the effect of denoising on simulations was assessed in terms of variations across the MC iterations.

2.3. In vivo *rodent experiments*

All experiments were approved by The Committee on Animal Experimentation for the Canton de Vaud, Switzerland. ^{1}H DW-MRS acquisitions were performed on a horizontal actively shielded 9.4 Tesla system (Magnex Scientific, Oxford, UK) interfaced to a Varian Direct Drive console (Palo Alto, CA, USA), equipped with 400 mT/m gradients and using a home-built 14 mm diameter surface ^{1}H-quadrature transceiver.

Four adult male Wistar rats were scanned under isoflurane anesthesia (~1.5%). During the DW-MRS experiments, animals were placed in an in-house-built cradle, and their head was fixed in a stereotaxic system (bite bar and a pair of ear bars). The respiration rate and body temperature were monitored using a small-animal monitor system (SA Instruments, New York, NY, USA). Body temperature was measured with a rectal thermosensor and maintained at 37.7±0.2°C by warm water circulation.

First- and second-order shims were adjusted using FASTMAP (Gruetter and Tkác, 2000), achieving water linewidths of 18-21 Hz in the volume of interest (VOI). DW-MRS data were acquired using a diffusion-weighted STEAM sequence (Callaghan, 1993, Frahm et al., 1989, Kunz et al., 2010) (TE/mixing time (TM)/repetition time (TR)=15/112/4000 ms) in a VOI of 162 to 245 μl depending on the animal. The water signal was suppressed by using the VAPOR module interleaved with outer volume suppression blocks (Tkác et al., 1999). Diffusion gradients were applied simultaneously along three orthogonal directions (δ=6 ms, Δ=120 ms). A total of eleven b-values with 128 shots were acquired: 0.4, 1.5, 3.4, 6.0, 7.6, 9.3, 13.4, 15.7, 20.8, 25.2 and 33.3 ms/μm^2.

2.4. In vivo *human experiments*

Human ^{1}H DW-MRS acquisitions were performed on a 3 Tesla Magnetom Skyra Connectom-A system (Siemens Healthineers Erlangen, Germany), equipped with 300 mT/m gradients and using a 32-channel head coil.

Four healthy volunteers (3 males/1 female) out of the twelve in the cohort of ref (Şimşek et al., 2022), featuring the highest water SNR at b=1.4 ms/μm^2 and the least drop for the 0.9 ppm MM signal, were selected to test the denoising procedure. All experiments had been approved by the competent ethical review board. A voxel (23±2 cm^3) was positioned in the occipito-parietal cortex and DW-MRS acquisitions were performed with an ECG-triggered diffusion-weighted STEAM sequence ($TE/TM/TR_{min}$=30/65/1800 ms). Eleven b-values were acquired (0.37, 1.4, 2.7, 5.4, 8.2, 10.9, 15.5, 18.4, 21.6, 23.3 and 25.1 ms/μm^2) using metabolite cycling, where some of the b-values had multiple sets of 64 shots. Thirty-two metabolite and thirty-two water spectra (4000 Hz spectral width, 4096 complex points) were constructed from 64 shots per b-value by difference and summation, respectively.

A table summarizing the minimum reporting standards in MRS is presented in supplementary materials (**Table S 5**), for simulations and *in vivo* data.

2.5. *MP-PCA denoising*

The raw data individual spectra in matrix Z were first eddy-current- and phase-corrected (maximization of the area of the metabolite region for each spectrum). The details of the processing steps specific to the human data are presented in supplementary materials (**Text S 1**). The resulting complex-valued FID were split into real and imaginary parts and organized into a matrix Y where the second dimension contained the time domain sampling and the first dimension a concatenation of all shots/b-values/real and imaginary parts. This was done in order to balance the number of rows with the number of columns and to increase the smallest dimension of X. The matrix Y was centered column-wise and assigned to matrix X. A summary of the denoising strategies and of the study design is presented in **Figure 1**.

MP-PCA denoising performances were first tested on shots with no diffusion weighting ("single-shell") and different noise generations on the MC simulations, and compared to summation of the individual shots. A matrix X of size 200×2048 was made of 100 single shots of the same shell (here: b=0), for multiple noise levels (SNR 13 – **Figure 2** and SNR

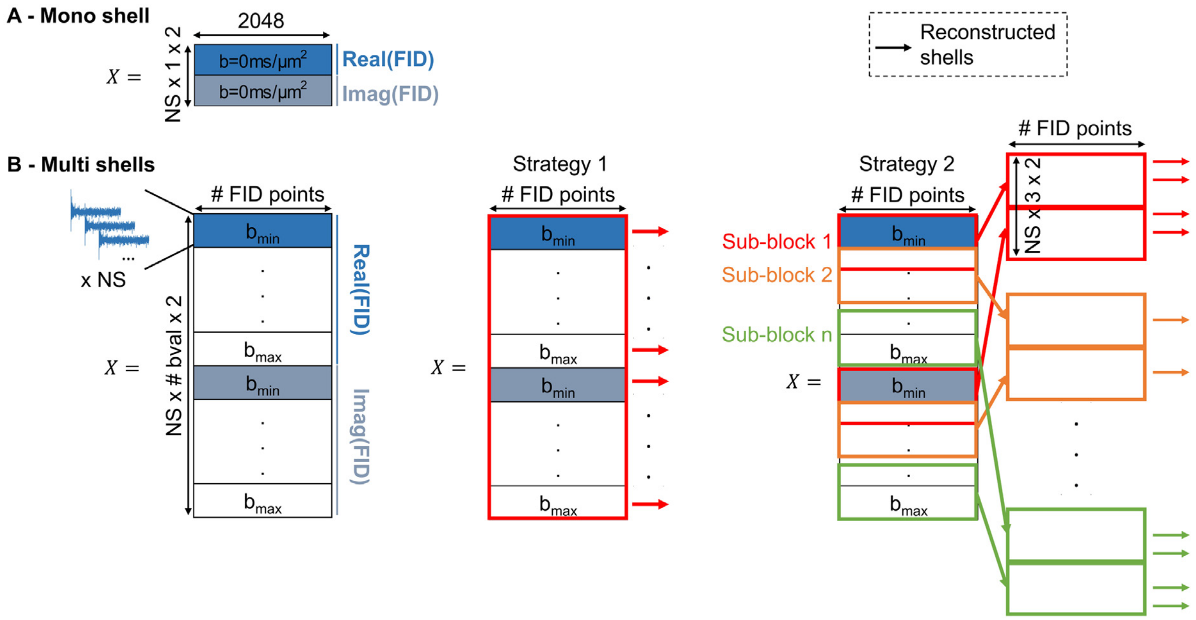


Figure 1. Study design and denoising strategies. **A:** Matrix organization for denoising a single-shell. This approach led to a similar result as summation of the shots on MC simulations (see Figure 2). **B:** Matrix organization for multi-shell full matrix (strategy 1) and sliding window denoising (strategy 2), the latter showing a reduced noise heterogeneity across shells. Strategy 3 (identical to strategy 2 with half the number of shots) is not displayed, showing similar results as strategy 2, yet with an increased number of outliers in the diffusion decay estimates. NS: number of shots, N_{bval}: number of b-values.

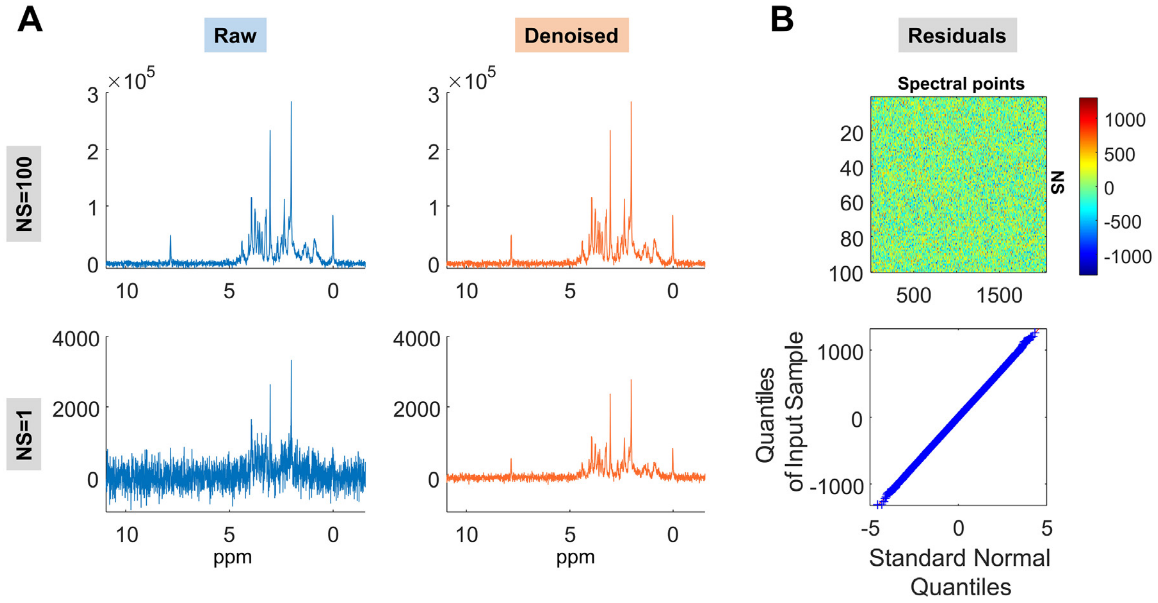


Figure 2. MP-PCA denoising performance on NS=100 shots of the same shell (with different noise realizations and no phase/frequency drifts). **A:** Raw (blue) and denoised (orange) spectra, of the summed 100 shots (top) and of a single shot (bottom): $SNR_{raw,100}$=101.9, $SNR_{raw,1}$=11.0, $appSNR_{dn,100}$=101.8, $appSNR_{dn,1}$=51.2. **B top:** Residuals (denoised minus raw matrix) for the real part of the spectra. **B bottom:** Quantile-quantile (Q-Q) plot of the spectral residuals. Denoising a single-shell performs similarly to the summation of single shots (rank $P = 1$ selected by the MP fit) and yields a Gaussian distribution of residuals.

1, 2 and 5 – **Figure S 1**), as well as without or with phase/frequency drifts in the original spectrum (**Figure 2** and **Figure S 1**, respectively).

For denoising heterogeneous matrices X composed of all b-values ("multi-shells"), two strategies were compared, both on MC simulations and on *in vivo* data:

1) **Multi-shell full matrix denoising - strategy 1**: MP-PCA denoising was performed on the full matrix. For MC simulations, the matrix X to denoise was of size 2000 × 2048: 10 shells with 100 shots (i.e. noise realisations), and recreated 100 times. For *in vivo* rodent data, the matrix X to denoise was of size 2816 × 2048: 11 shells with 128 shots. For *in vivo* human data, the matrix X to denoise was of size 832 to 960 × 3481: 13 to 15 shells with 32 shots.
2) **Multi-shell sliding window (sw) denoising - strategy 2**: MP-PCA denoising was performed on a subset of the full matrix, using a sliding window of sub-blocks of three shells among all shells, and the denoised spectra output is selected when the shell is the middle of the 3-shell sub-block (similarly to the dMRI procedure (Veraart et al., 2016)). The first and last shells were selected together with the second shell and one before last from the first and last sub-blocks, respectively (**Figure 1**).

The same denoising procedure as strategy 2 but using only half of available shots (**strategy 3**) was tested on simulations and on *in vivo* rodent data, with 50 and 64 shots per b-value, respectively. This strategy aimed at assessing whether a reduction in scan time for the same data quality could be achieved, comparing datasets with fewer and denoised shots to the original complete sets.

2.6. Quantification and modelling

Raw and denoised simulations and rodent spectra were further corrected for B_0 drifts (alignment of the tCr peak at 3.03 ppm or NAA at 2.01 ppm in each spectrum to its position in the first spectrum after 8 Hz apodization) and summed (for each b-value).

Metabolite concentrations were quantified using LCModel. The metabolite basis set was composed of the noiseless simulated signals for the MC study, and of spectra simulated using the acquisition parameters for the *in vivo* acquisitions, all basis sets containing an *in vivo*-acquired macromolecule signal. In addition, for the *in vivo* rodent data, separately simulated MM and lipid components from LCModel were included to compensate for possible lipid contamination due to the large size of the voxel and its position close to the scalp (Oz et al., 2005, Cudalbu et al., 2021). The LCModel parameter controlling the baseline stiffness, DKNTMN, was set to 0.25.

The randomly oriented stick model was fitted to the decay of each metabolite concentration as a function of b-value using a non-linear least squares algorithm in Matlab (*fit* function, *Trust-Region* method). The concentration decays as a function of b-value were fitted for each of the 100 MC iterations for simulations, and for each rat or volunteer individually for the *in vivo* data. The median estimated diffusion coefficients D_{intra} with SD (across the 100 MC iterations or across the animals/volunteers) were extracted. Percentage bias is reported for the concentrations and D_{intra} ($(Value_{method}$-$Value_{noiseless})/Value_{noiseless}$).

Statistical tests were performed in RStudio (RStudio, PBC, Boston, MA). For simulations, D_{intra} estimates based on raw and denoised data (from each denoising strategy) were compared to the D_{intra} estimate from the noiseless data using a repeated-measures one-way ANOVA, and p-values were corrected for multiple comparisons with Dunnett's post-hoc test. For *in vivo* data, D_{intra} estimates based on raw and denoised data (from each denoising strategy) were compared using a repeated-measures one-way ANOVA, and pairwise p-values were corrected for multiple comparisons with Tukey's post-hoc test. The following statistical significance values were used: * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001.

3. Results

The performance of the denoising strategies was assessed in terms of apparent SNR, spectral residuals (denoised summed spectra minus raw summed spectra for a given shell), rank selection, noise correlation within and across shells, as well as precision and accuracy of metabolite quantification for each b-value and of resulting diffusivity estimation.

3.1. Monte Carlo simulations

This section aims to study the effect of MP-PCA denoising on simulated DW-MRS data while having access to the ground truth.

3.1.1. Single-shell: MP-PCA denoising versus summation

Figure 2 shows the performance of MP-PCA denoising on a single-shell matrix, i.e. NS=100 shots of a spectrum with no diffusion weighting. Since summation (accumulation of spectra with different noise realisations but the same signal content) is a very efficient denoising strategy, it will be compared to MP-PCA. For a single-shell, denoising performs similarly to averaging on the summed spectra (NS=100, Figure 2A, top) and a rank $P = 1$ is selected by the MP fit. Single shots are also strongly denoised (NS=1, Figure 2A, bottom) but this representation should be handled with care since single shots are reconstructed from the entire denoised matrix and thus are not an equivalent representation of single shot raw data. The spectral residuals (100 shots x 2048 real spectral points) follow a Gaussian distribution and no structure in the metabolites' region was observed (Figure 2B). When phase and frequency drifts are applied across shots on the simulated spectrum, and at sufficiently high SNR, a rank $P > 1$ is retained by the MP fit (Supplementary **Figure S 1**B).

3.1.2. Multi-shell - strategy 1: MP-PCA denoising on the entire diffusion-weighted matrix

Figure 3A shows simulated diffusion-weighted spectra at 10 b-values. 11 principal components were retained by the MP fit (Figure 3B and **Figure S 3**C for their representation). The raw and denoised spectra for the two extreme b-values are shown in Figure 3C, for a single shot (bottom) and for the sum of the 100 shots (top). Denoising yields an improved spectral apparent SNR, on individual shots and on their sum. The central panel of Figure 4A shows that the noise level is non-uniform across shells after denoising with strategy 1, the shell containing the higher b-value experiencing a stronger denoising effect, as evidenced by the ratio of spectral noise variances at b_{min} and b_{max}. Although the noise level is shell-dependent, its distribution in a noise-only spectral region within one shell remained Gaussian after denoising.

3.1.3. Multi-shell - strategy 2 versus strategy 1

An alternative strategy of denoising using a sliding window of 3 shells, denoted as a "sub-block", is proposed (strategy 2), and aims at reducing the non-uniform noise level across shells introduced by strategy 1. This resulted in minimal SNR heterogeneity within each sub-block on which the denoising was applied and is similar to what is used in dMRI (Lei et al., 2021) where the columns of matrix X are composed of a sliding spatial kernel of voxels. However, here we strive to reduce heterogeneity in the diffusion dimension (row-wise).

Although strategy 2 shows smaller noise reductions versus raw compared to strategy 1 (at b_{min}, 2.3 apparent SNR increase for strategy 2 versus 2.7 for strategy 1, at b_{max}, 3.6 apparent SNR increase for strategy 2 versus 6.8 for strategy 1, Figure 4C), it reduced the non-uniform noise levels across shells (Figure 4A). On the summed spectra: $\frac{\sigma_{bmin}}{\sigma_{bmax}} = 2.24$ for strategy 1 and $\frac{\sigma_{bmin}}{\sigma_{bmax}} = 1.65$ for strategy 2, whereas this ratio before denoising was close to 1 since single shots were created with the same noise level in each shell. The excessive noise reduction at high b-values (and potential wiping of signal) is also manifest, yet reduced with strategy 2. Noise levels on single shots display the same overall pattern as on the sum (Figure 4A, bottom). These observations suggest that some correlation is introduced in the noise, also shown in Figure 5.

The decreasing number of signal-carrying components retained by the MP fit as a function of sub-block number *(*Figure 4B) highlights that, at low SNR (i.e. the noisiest sub-matrix, containing the highest b-values), less meaningful information can be separated from the noise (also shown in **Figure S 1**). In strategy 2, the apparent spectral SNR (Figure 4C) increases by a factor of 2.3 at b_{min} and 3.6 at b_{max} and follows a similar trend as in the raw data. However, it reaches a maximum for central b-values in strategy 1, possibly resulting from a "decay" of the apparent noise levels, as detailed in **Figure S 2**. The term SNR after denoising should be used carefully in the light of the noise correlations described below (Figure 5).

The summed residuals across shots (Figure 4D) show hardly any structure in the metabolites' region. The weak residuals around the NAA and Cr peaks may be caused by differences between the phase and frequency drift correction factors estimated from the raw or denoised data, or a change in linewidth after denoising, leading to spectral misalignment before summation (Stoyanova and Brown, 2002).

We further analysed the correlations of the NAA peak amplitude and of the noise between single shots and the sum of NS=100 introduced by

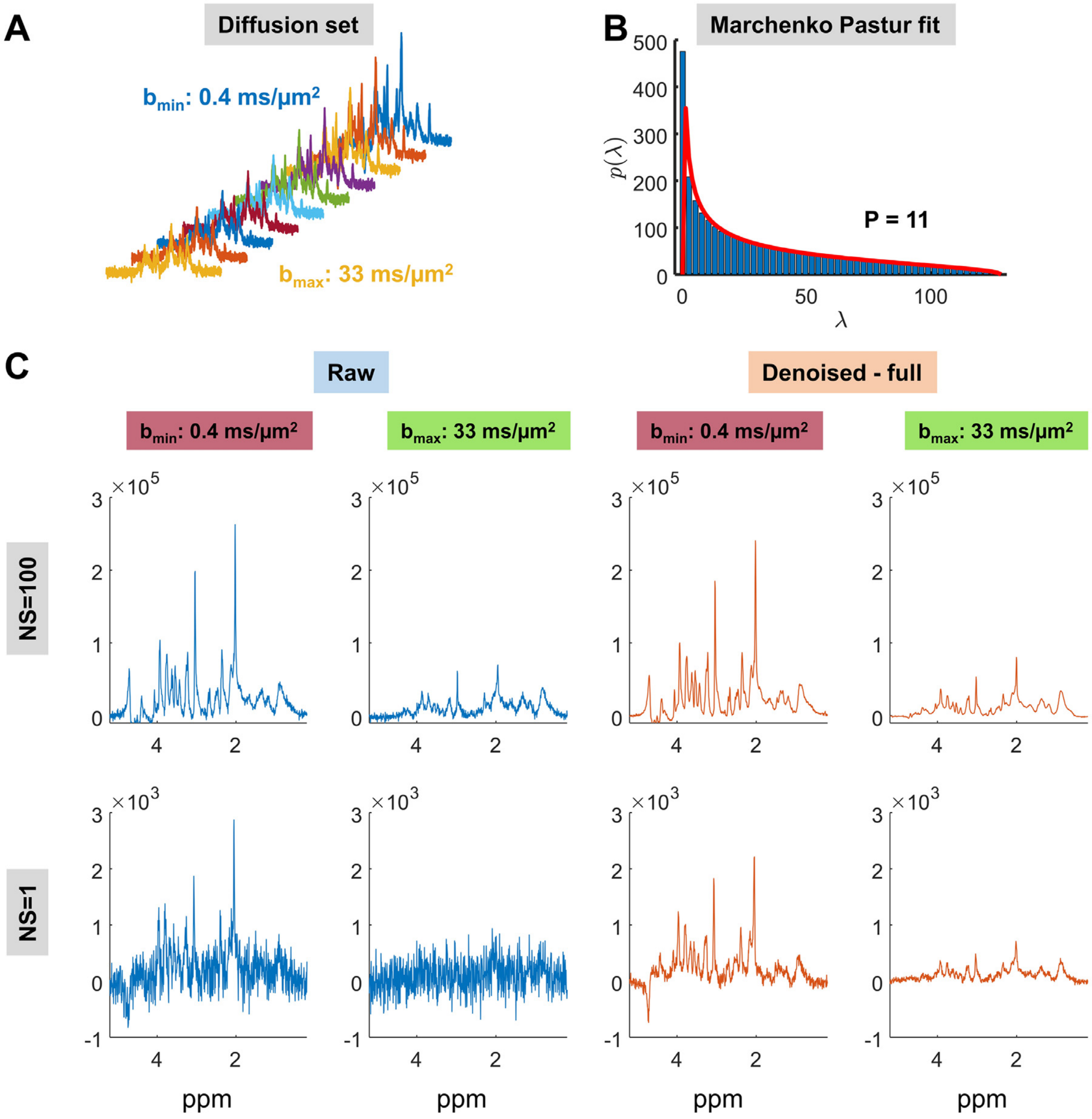


Figure 3. MP-PCA denoising performance on the full diffusion-weighted matrix X made up of 10 shells, with 100 shots (NS) each (10×100×2 × 2048 FID points – '×2' is for the concatenation of real and imaginary parts of the FID) – strategy 1. **A:** Simulated diffusion-weighted spectra at each b-value. **B:** Example MP fit on matrix X for strategy 1 for one MC iteration. **C:** Example raw and denoised spectra, at low and high b-value, of the sum of the 100 shots (top) and of a single shot (bottom). $\mathrm{SNR}_{\mathrm{raw,100,bmin}}$=90.4, $\mathrm{SNR}_{\mathrm{raw,1,bmin}}$=11.9, $\mathrm{SNR}_{\mathrm{raw,100,bmax}}$=24.0, $\mathrm{SNR}_{\mathrm{raw,1,bmax}}$=4.3, $\mathrm{appSNR}_{\mathrm{dn,100,bmin}}$=229.3, $\mathrm{appSNR}_{\mathrm{dn,1,bmin}}$=41.8, $\mathrm{appSNR}_{\mathrm{dn,100,bmax}}$=155.4, $\mathrm{appSNR}_{\mathrm{dn,1,bmax}}$=27.7. Denoising improves apparent spectral SNR.

MP-PCA (Figure 5). The NAA peak amplitude at b_{min} (top left) scales with the number of shots (NS=100), as expected, both for raw and denoised data. At lower SNR (b_{max}, bottom left), the NAA peak amplitude on the raw data does not exactly scale with NS because of possible artefacts in the summation, such as improper frequency/phase drifts correction, leading to partially incoherent summation. For the denoised data, the coherent summation property seems to be restored (amplitude ratio close to 100), which can be due to an improved frequency/phase drifts correction after denoising (Figure 6) and/or to the creation of more self-similar spectra after rank truncation. The noise level in the raw data displays a ratio that scales with $\sqrt{NS}$, as expected, both for b_{max} and b_{min}. For the denoised data, some correlation in the noise across shots is introduced by both denoising strategies, leading to a noise ratio scaling with a factor greater than $\sqrt{NS}$.

3.1.4. Estimation of metabolite concentrations as a function of b-value

In our post-processing pipeline, denoising was performed before B_0 drift correction. This allowed for a more accurate realignment of spectra within each b-value, most noticeably at b_{max} (Figure 6): the correction factors derived from the denoised data were closer to ground truth (RMSE: 2.7 Hz) compared to the ones derived from the raw data (RMSE: 6.0 Hz), although the latter yielded a higher amplitude of the summed signal. Metabolite concentrations at b_{min} and b_{max} for all denoising strategies, together with relative ESFU, are presented in **Table S 1**. They highlighted an overall stronger bias introduced by the denoising strategies with respect to the one of the raw data for low-concentrated metabolites, but a weaker one for high-concentrated metabolites, even with strategy 3. Fit precision (ESFU) is strongly improved after denoising for all metabolites. When comparing strategies 1 and 2 on concentration

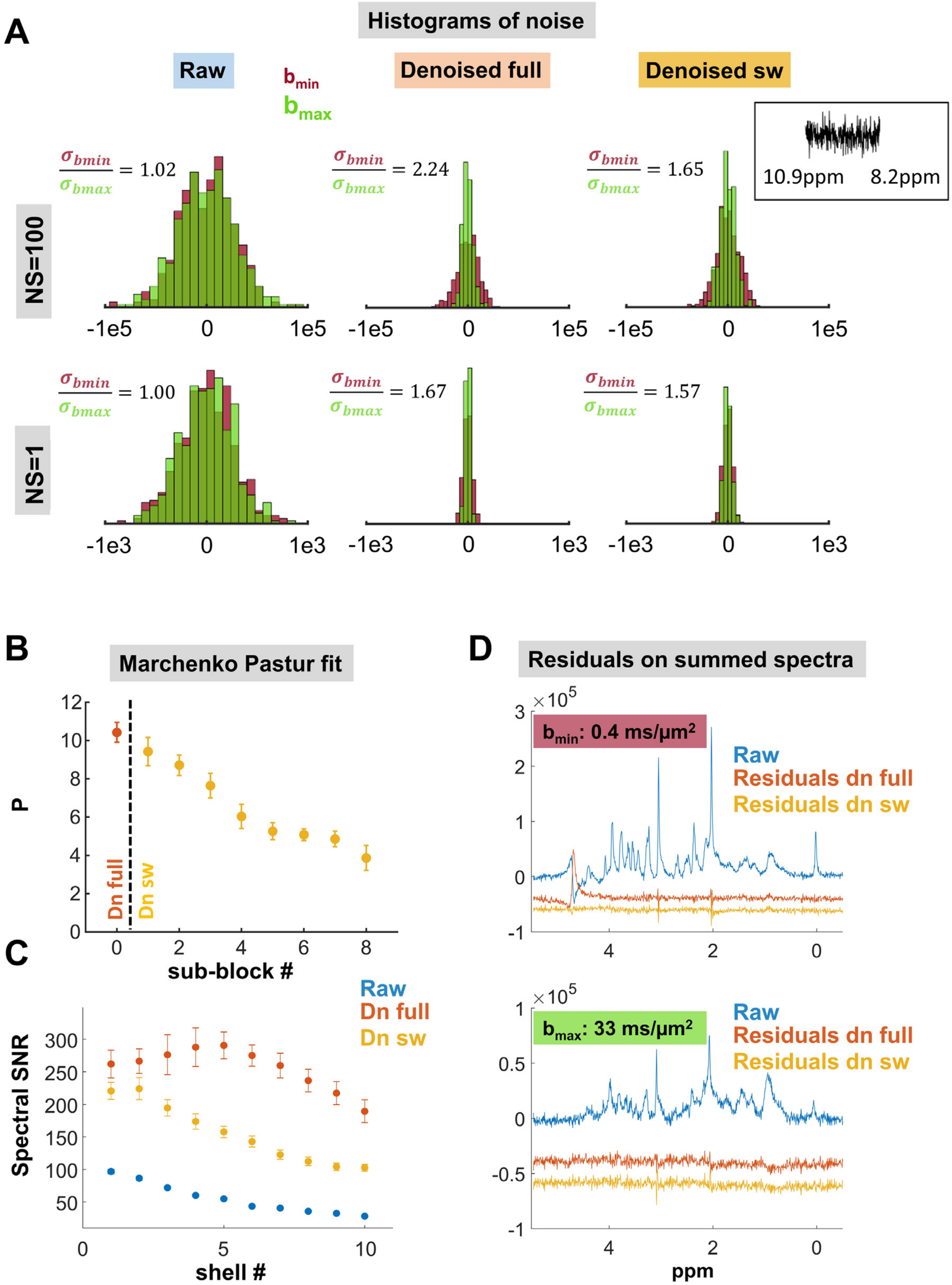


Figure 4. Comparison of MP-PCA denoising performance on the full matrix (strategy 1) or using a sliding window of 3 shells over the diffusion-weighted matrices (strategy 2). **A:** Histograms of spectral noise between 8.2 and 10.9 ppm (a region with no signals), for a single shot (bottom) and for the sum of the 100 shots (top), before and after denoising using strategies 1 and 2, for the lowest (red) and highest (green) b-values. The mean ratio across MC iterations of the noise level at b_{min} over the one at b_{max} is displayed in each case. Standard deviations across MC associated to the mean ratios displayed: for $NS = 100$, 0.04 (raw), 0.24 (dn full), 0.15 (dn sw), and for $NS = 1$, 0.05 (raw), 0.33 (dn full), 0.40 (dn sw). **B:** Number of principal components retained as signals (i.e. the rank P) by the MP fit, in strategy 1 (orange) and for each sub-block in strategy 2 (yellow), as mean and SD across MC iterations. **C:** Spectral (apparent) SNR on the summed spectra for each shell of raw and denoised data (strategy 1 & 2), as mean and SD across MC iterations. **D:** Spectral residuals on the summed spectra for the two denoising strategies at low (top) and high (bottom) b-values, shifted downwards for display. Both denoising strategies gave heterogeneous noise levels and increases in apparent SNR with no structure in spectral residuals. Strategy 2 mitigates some effects of strategy 1, namely the non-uniform SNR gain and variance across shells.

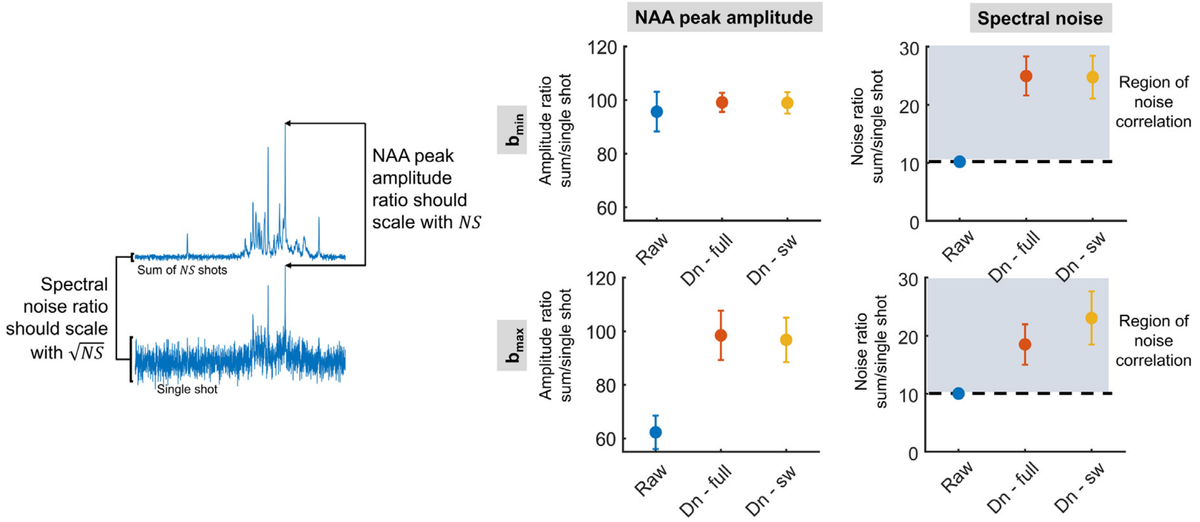


Figure 5. Increased correlation in NAA peak amplitude and noise level after denoising, for low and high diffusion weighting, between one shot and the sum of NS shots within a shell. Mean and SD across MC are displayed. The region of noise correlation is shaded in grey.

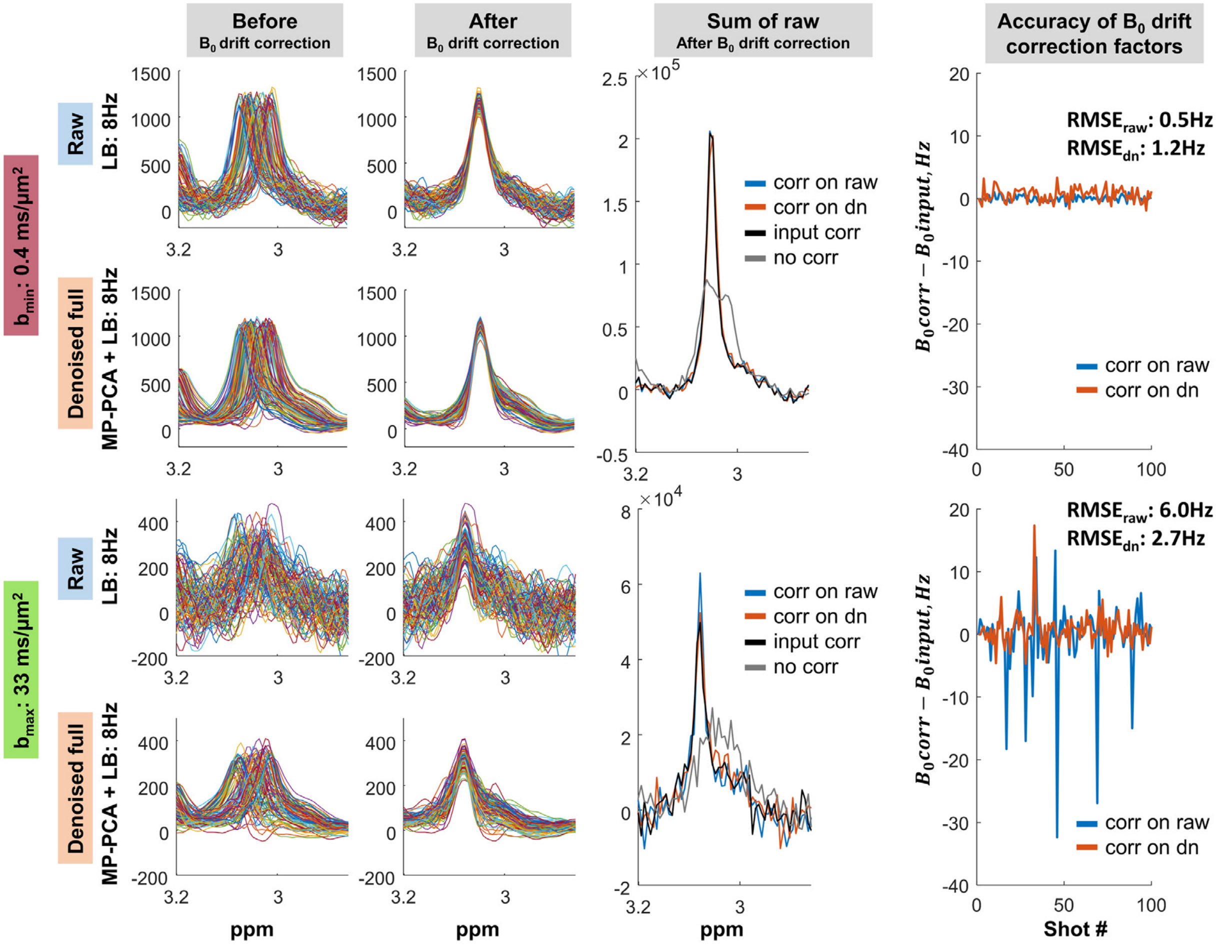
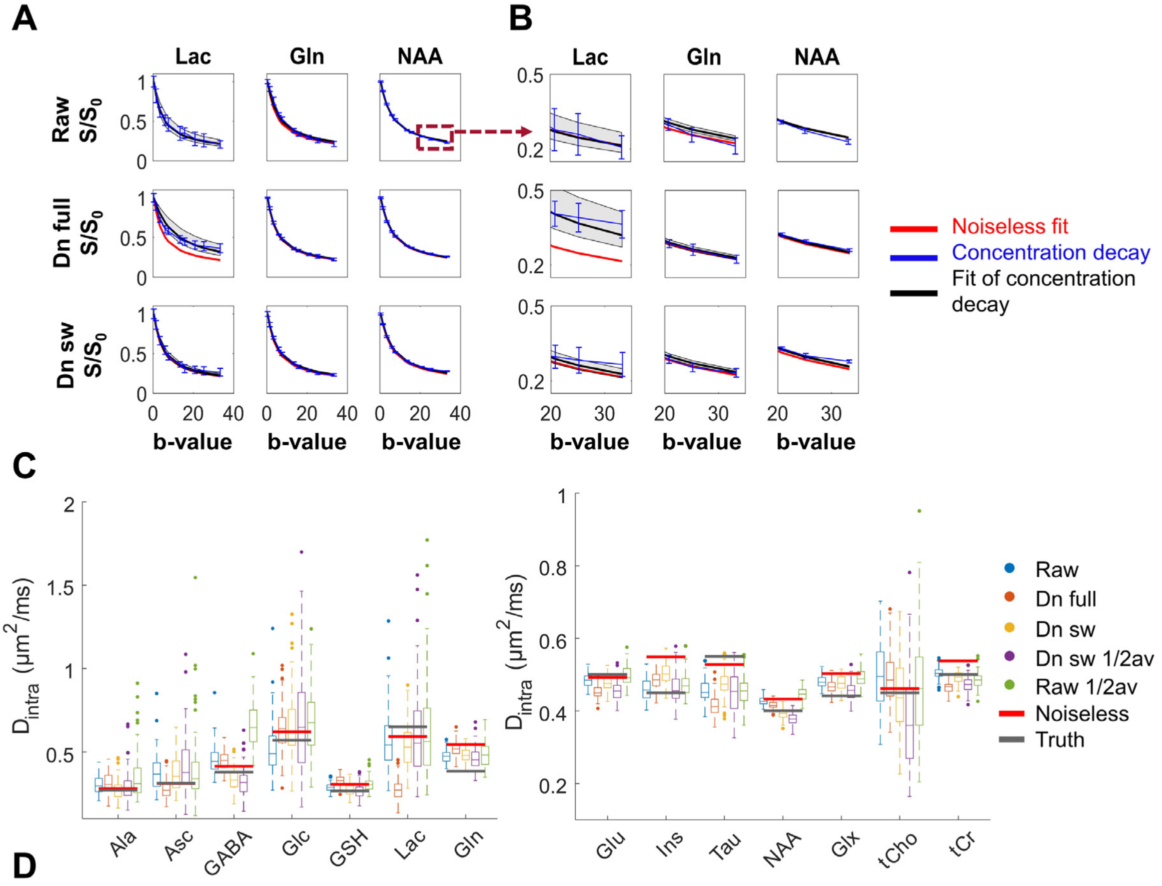


Figure 6. Spectral realignment (B_0 drift correction) after denoising. The B_0 drift correction was performed by aligning the frequency-domain position of the tCr peak to its position on the first spectrum, using a Lorentzian apodization of 8 Hz, on raw and denoised data, for 1 MC iteration (left panels, before/after, for b_{min} and b_{max}). The central panel shows the summed raw spectra with corrections derived either from the raw or the MP-PCA data, as compared to the summed raw spectra where the negative input B_0 drifts have been applied. Denoising yields no benefit on B_0 drift correction in the case of sufficient SNR (e.g. at b_{min}). At low SNR (e.g. b_{max}), the summed raw spectra with corrections derived from MP-PCA is closely matching to the one reconstructed from the input B_0 drift values, yet with a smaller amplitude than the summed spectra with corrections from the raw data. Denoising before B_0 drift correction led to a better accuracy of the B0 drift estimates with respect to the input drifts (right panel) at b_{max}, and a worse accuracy at b_{min}. $RMSE_{method} = \sqrt{\frac{1}{NS}\sum_{i=1}^{NS}(B_0corr, method - B_0, input)^2}$.



D

Mean bias vs noiseless	Ala	Asc	GABA	Glc	GSH	Lac	Gln	Glu	Ins	Tau	NAA	Glx	tCho	tCr
Raw	8,1	21,2	9,3	-16,8	-6,2	-3,7	-13	-1,7	-16,1	-13,8	-1,2	-4,4	7,8	-6,3
Dn – full	9,9	-10,2	7,2	3,8	7,4	-53,8	-4,7	-8,1	-11,7	-21,4	-3,9	-7,5	6,2	-13,5
Dn - sw	-2,4	18,6	-19,1	5,6	-10,2	-10,6	-11,8	-3,3	-8,5	-10	-9,2	-5,3	-4,3	-8,3
Dn – sw 1/2av	5,3	35,4	-23,3	16,5	-13,1	1,1	-15,5	-7,7	-16,1	-14,4	-12,6	-9,6	-20,2	-12,1
Raw – 1/2av	25,7	30	59,1	12,4	-0,5	5,8	-11,3	0,9	-13,6	-13,9	3,3	-2,3	0,2	-10

# outliers	Ala	Asc	GABA	Glc	GSH	Lac	Gln	Glu	Ins	Tau	NAA	Glx	tCho	tCr
Raw	0	2	0	2	1	0	2	0	0	3	1	2	3	2
Dn – full	2	1	0	0	1	1	4	0	1	2	0	5	3	0
Dn - sw	1	0	0	4	0	0	0	5	3	3	0	7	0	0
Dn – sw 1/2av	2	0	6	1	2	4	1	1	2	2	1	1	1	1
Raw – 1/2av	3	1	0	3	1	0	0	0	5	2	1	1	2	3

Figure 7. Concentration decays as a function of b-value and D_{intra} estimates in simulations. **A:** Representative concentration decay curves for three metabolites: Lac, Gln, NAA, normalized to the concentration at the lowest b-value. Overlaid curves are: mean and SD of concentrations across MC iterations (blue), Callaghan model fit using the mean D_{intra} estimated across MC iterations (black) and Callaghan model fit of the quantified noiseless concentration decay (red). **B:** Zoom-in of panel A for b-values between 20 and 33 ms/µm^2. **C:** Estimated metabolite D_{intra} from Callaghan's model using raw or denoised data, for various denoising strategies. The values labelled as "truth" represent the diffusion coefficients given as input in the simulations, and the values labelled as "noiseless" represent the LCModel concentrations fit from the noiseless data. **D:** % bias on D_{intra} ((D_{method}-$D_{noiseless}$)/$D_{noiseless}$) and number of outliers between all methods and the noiseless fit. The D_{intra} that differ from the noiseless values by more than ±10% bias are highlighted in orange. The ouliers are defined as elements being above the upper quartile or below the lower quartile by a quantity greater than 1.5 times the interquartile range. The cases where denoising reduces or equalizes the number of outliers found with their raw data counterparts (raw or raw $\frac{1}{2}$ av) are highlighted in green. Some metabolite-dependant bias on the concentrations and on D_{intra} estimates is either introduced or reduced compared to the raw data after denoising.

decay curves, their impact was metabolite-dependent (Figure 7A-B). In the case of Lac, strategy 1 introduced a systematic bias (overestimated concentration) with respect to the noiseless fit, an effect largely mitigated using strategy 2. For Gln, however, both strategies (1 and 2) improved the decay curve accuracy, while no benefit was brought by any of the strategies for NAA.

In terms of D_{intra} estimation (Figure 7C-D), although p-values highlighted a systematic bias (versus noiseless data, **Table S 4**), strategy 2 led to an improvement in accuracy for some metabolites compared to the raw data and strategy 1 (Ala, tCho, mIns, Tau), a deterioration for some low concentrated metabolites (GABA, GSH, Lac) and similar accuracy for the remaining ones. The number of outliers was slightly reduced by all the denoising strategies.

Unfortunately, strategy 3, using half the data (i.e. NS=50) to assess if the total duration of the scan could be reduced without a significant compromise in accuracy and precision of metabolite concentration and

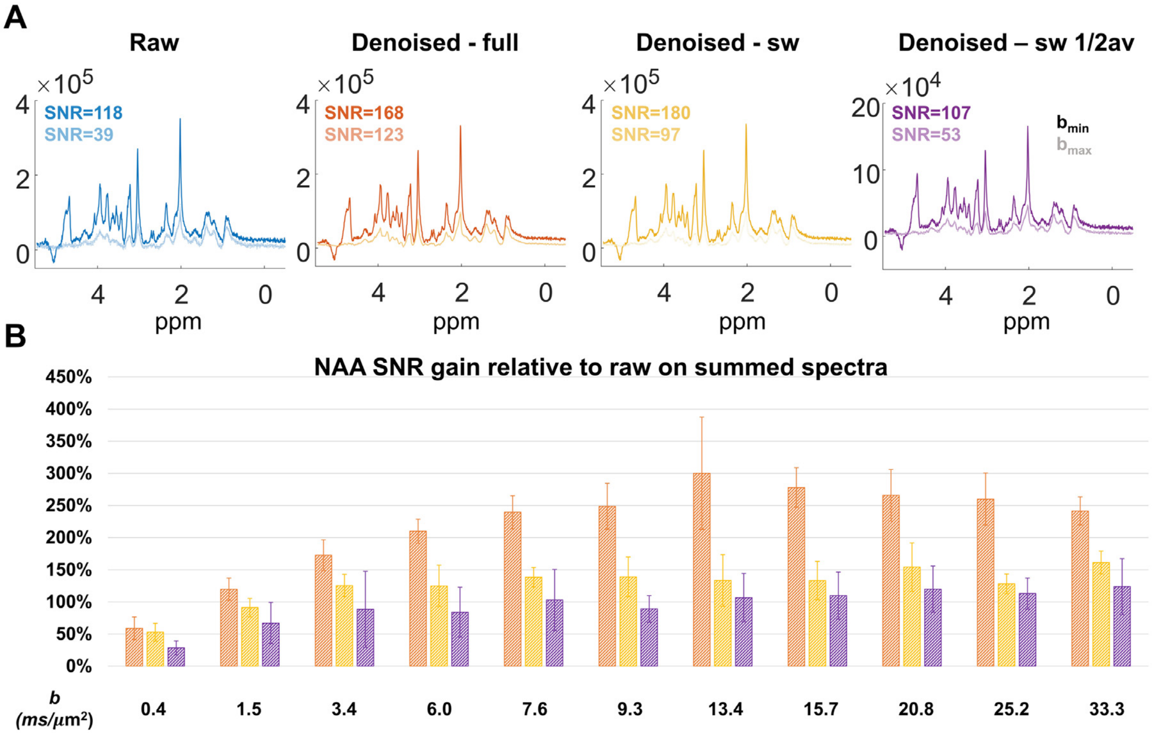


Figure 8. *In vivo* rodent data - spectral quality and apparent SNR gain, before versus after denoising. **A**: Representative summed spectra for one animal, at low (dark colors) and high (light colors) b-values, based on raw and denoised data, with strategies 1 to 3. NAA singlet SNR is displayed for each case. **B**: Relative apparent SNR gain from the denoising strategies 1 to 3, expressed in % increase compared to the raw data SNR, with mean and SD across animals. Orange: strategy 1, yellow: strategy 2, purple: strategy 3. Increased apparent spectral SNR was observed *in vivo*.

D_{intra}, yielded worse or at best similar accuracy and precision for D_{intra} as the full raw data (NS=100) but also as half the raw data (NS=50) depending on the metabolite.

3.2. In vivo *rodent data*

The same analyses were performed on *in vivo* data from four animals and MP-PCA denoising effects were compared to the ones observed in simulations.

3.2.1. *Apparent SNR*

The summed spectra for the two extreme b-values before and after denoising using all three strategies are shown in Figure 8A. Denoising improved the apparent SNR at all b-values, yet to a smaller extent compared to simulations (Figure 8B): on average, the SNR gain is % at b_{min} and 241% at b_{max} for strategy 1 and 53% at b_{min} and 61% at b_{max} for strategy 2. The apparent SNR gain follows a similar b-value dependence to the one in simulations, with a maximum for a central b-value for strategy 1 and a constant gain for strategy 2.

3.2.2. *Noise properties*

For strategies 1 and 2, the noise level on *in vivo* data after denoising was non-uniform across shells, both on the sum and on the single shots (Figure 9A), and strategy 2 attenuated this effect: on the summed spectra: $\frac{\sigma_{bmin}}{\sigma_{bmax}} = 2.49$ for strategy 1 and $\frac{\sigma_{bmin}}{\sigma_{bmax}} = 1.87$ for strategy 2. A rank $P = 12$ for strategy 1 and $P \in [4, 12]$ for strategy 2 was selected by the MP fit (Figure 9B), which was consistent among rats (strategy 1: $P = 11.5 \pm 0.58$, strategy 2: $P_{bmin} = 11.25 \pm 0.5$ and $P_{bmin} = 3.5 \pm 1$) and similar to the ranks found in simulations (Figure 3B). Similarly to the effect observed in simulations, the noise level distribution in a noise-only spectral region within one shell remained Gaussian after denoising.

The spectral residuals for both strategies showed no distinct structure around metabolite frequencies (Figure 9C), suggestive of a homogeneous denoising in the spectra.

3.2.3. *Estimation of metabolite concentrations as a function of b-value*

All denoising strategies yielded similar concentrations and reduced ESFU compared to the raw data for the six quantified metabolites at b_{min} and b_{max} (**Table S 2**). Similar trends to those identified in simulations are observed between estimates of D_{intra} from raw and denoised data (Figure 10). In the multiple comparison post-hoc test, only tCr D_{intra} showed a significant difference between strategy 1 and 3. For the high-concentrated metabolites (Glu, NAA and tCr), strategy 2 reduced the variability of D_{intra} estimates across animals, as compared to that from the raw data.

3.3. In vivo *human data*

The Connectom gradients allowed to reach strong diffusion weighting, making the human DW-MRS b-value range and the data quality comparable to the rodent ones. Figure 11A shows representative spectra at increasing b-values for one volunteer's dataset. The NAA spectral SNR decays with increasing b-values for the raw data, as expected (Figure 11B). In the case of strategy 1, the apparent spectral SNR is higher after denoising at all b-values and reaches a maximum for $b \sim 3$ ms/μm^2, an effect which is mitigated by strategy 2. There is hardly any structure in the spectral residuals for either strategy (Figure 11C). As observed in the simulations and in the rodent data, the denoising effect is stronger at b_{max} compared to b_{min}. The concentrations (**Table S 3**) and the D_{intra} estimates (Figure 11E) show no significant difference between the raw and denoised data, confirming the observation made in rodent DW-MRS data. The fit error (root mean square error between the fit and the experimental decay) is reduced after denoising for tNAA, tCho and tCr, with strategies 1 and 2.

4. Discussion

The aim of this work was to evaluate the performance of MP-PCA denoising on synthetic and experimental datasets of single-voxel diffusion-

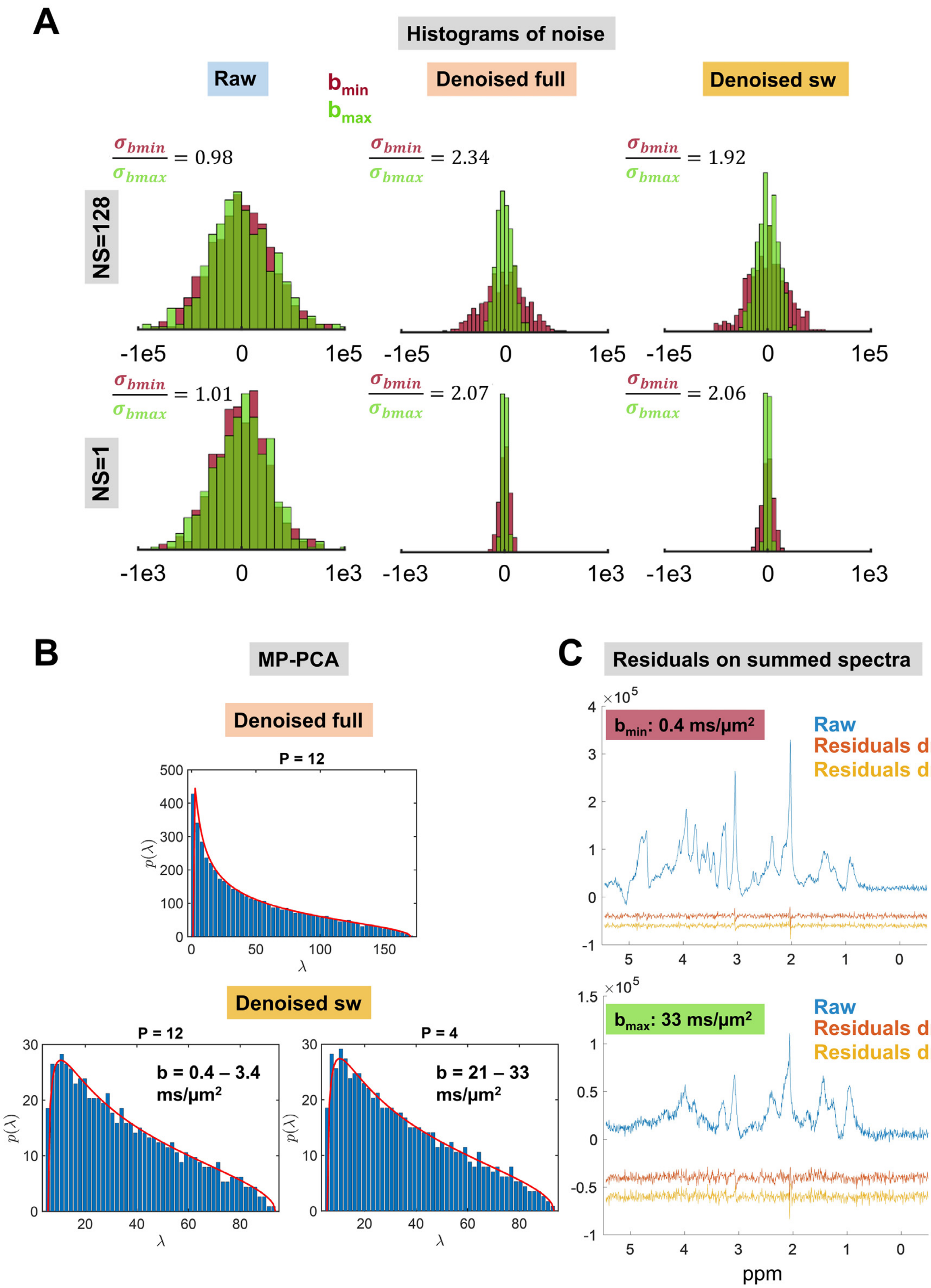


Figure 9. *In vivo* rodent data - MP-PCA denoising performance using strategy 1 and strategy 2. **A:** Histograms of spectral noise for one example animal in the 8.2-10.9 ppm noise-only region, for a single shot (bottom) and for the sum of the 128 shots (top), before and after each denoising strategy, for the smallest (red) and highest (green) b-values. The ratio of the experimental noise level at b_{min} over b_{max} is displayed in each case, averaged over the four animals. Standard deviations across animals associated to the mean ratios displayed: for $NS = 100$, 0.04 (raw), 0.28 (dn full), 0.25 (dn sw), and for $NS = 1$, 0.06 (raw), 0.44 (dn full), 0.65 (dn sw). **B:** MP fit for both strategies. **C:** Residuals between the denoised and raw spectra at the two extreme b-values, after summation of the 128 shots available, shifted downwards for display. The same trends as the ones for simulations can be observed: heterogeneous noise level across shells, increase in apparent SNR with no structure in spectral residuals after denoising, with strategy 2 mitigating some effects of strategy 1.

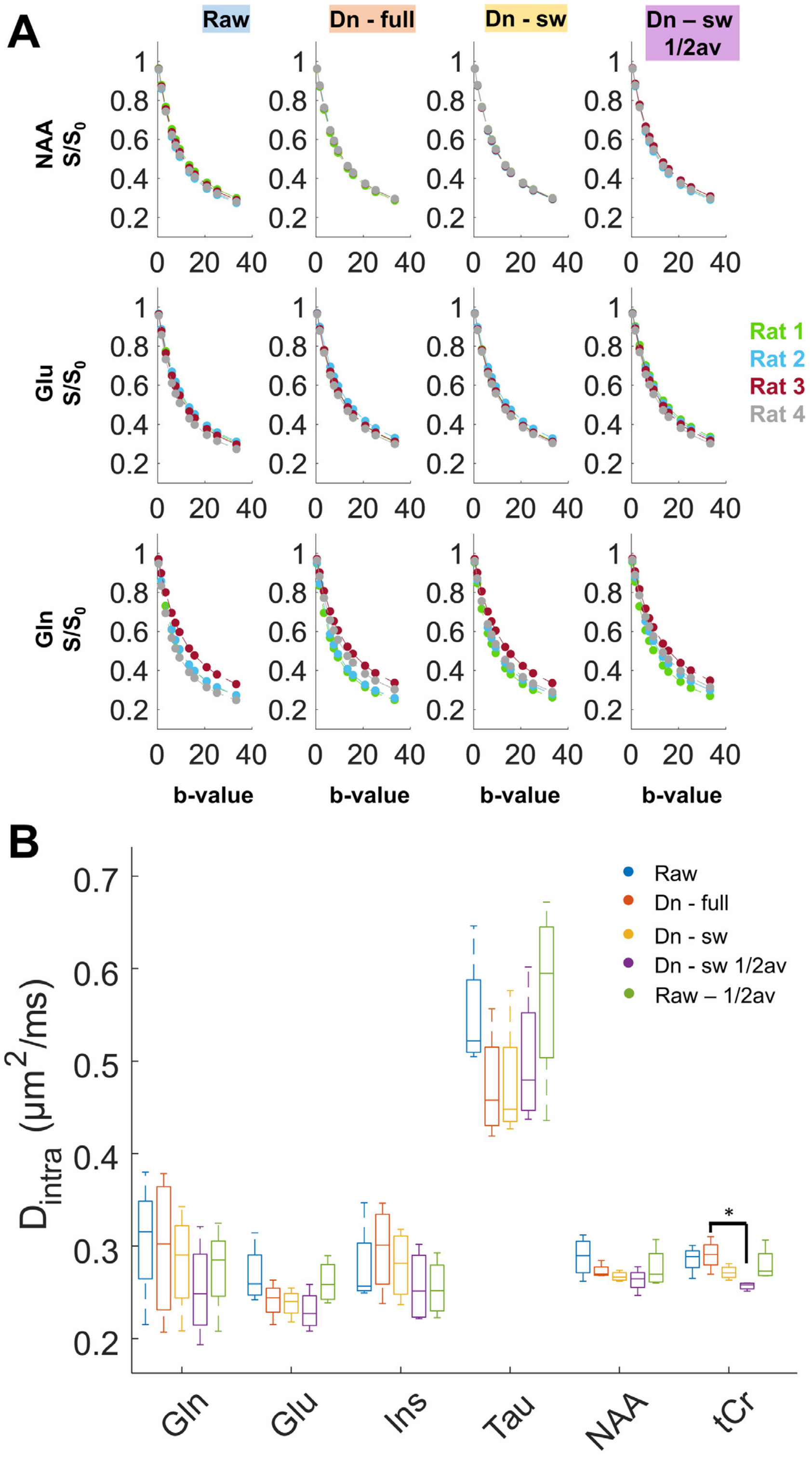


Figure 10. *In vivo* rodent data - concentration decays after quantification with LCModel, and resulting D_{intra} fit, for raw and denoised data with the three strategies. **A:** Representative decays across b-values for three metabolites: NAA, Glu, Gln, for each animal (circles), with concentrations normalized to the lowest b-value, and individual fits of Callaghan's model (solid line). **B:** Estimated D_{intra} from Callaghan's model for a few metabolites, for all strategies. Raw and denoised data provide similar estimates for most metabolites.

weighted ^{1}H-MRS comprising spectra at multiple diffusion-weightings (b-values), as compared to conventional averaging across each b-value. We investigated three denoising strategies, comparing their impact on the data structure (apparent SNR increase, spectral residuals, noise correlation), and evaluating their potential for improved diffusion coefficient estimates. Similar characteristics of the denoised spectra were observed between simulations and *in vivo* data (similar SNR on the raw data between simulations, rodent and human data, noise ratios between first and last shells, apparent SNR gain and evolution as a function of b-value, spectral residuals) thus ensuring that conclusions drawn from simulations with respect to the ground truth are relevant for the *in vivo* datasets.

4.1. Increased apparent spectral SNR

Simulations revealed that denoising all DW-spectra together significantly improved apparent spectral SNR for each b-value compared to averaging (Figure 3 and Figure 4 for simulations and Figure 8 for *in*

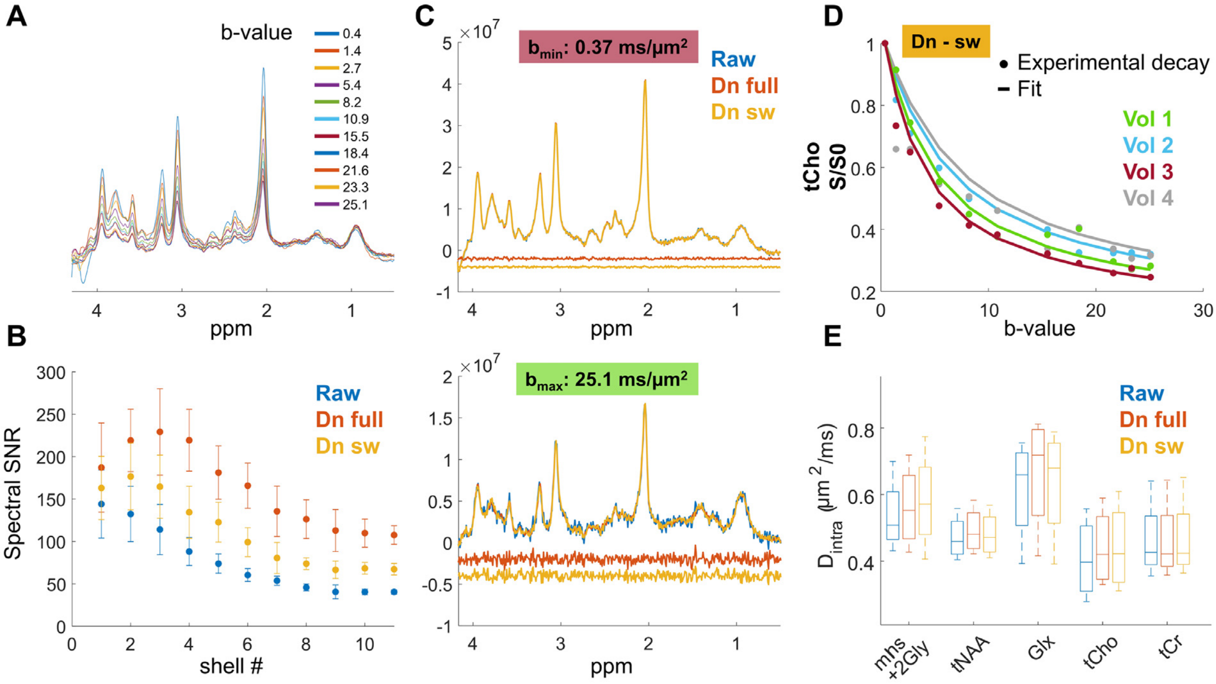


Figure 11. *In vivo* human data - effects of denoising in terms of apparent spectral SNR, residuals, fit and D_{intra} estimation. **A:** Representative diffusion-weighted raw spectra for one volunteer (LB=2 Hz). **B:** Evolution of (apparent) spectral SNR as a function of the b-value, mean and standard deviation across the 4 volunteers. **C:** Overlap of raw, denoised full (strategy 1) and denoised with a sliding window (strategy 2) spectra and residues shifted downwards for display. **D:** Representative experimental decays across b-values for tCho with strategy 2 (circles), with concentrations normalized to the lowest b-value, and individual fits of Callaghan's model for all volunteers (solid line). **E:** Estimated median and population variance for D_{intra} from the Callaghan's model for a few metabolites for the raw data and the two denoising strategies. The fit error (root mean square error between the fit and the experimental decay) was reduced after denoising for tNAA, tCho and tCr, with both strategy 1 (median RMSE normalized to the one of raw, for tNAA: 0.91 [range 0.79-0.93], for tCho: 0.84 [range 0.64-0.99] and for tCr: 0.90 [range 0.88-1.01]) and strategy 2 (for tNAA: 0.90 [range 0.87-1.02], for tCho: 0.92 [range 0.73-1.09] and for tCr: 0.89 [range 0.78-0.95]). Denoising improves apparent SNR, yields no spectral residual and the same D_{intra} estimates for raw and denoised.

vivo rodent data and Figure 11 for human data). Remarkably, denoising also provided the following two valuable features vs averaging.

First, the correction for B_0 drifts between individual shots of high b-value shells was more reliable after denoising especially at low SNR (Figure 6). Whether a stronger apodization or spectral registration at high b-values could mimic the benefit of denoising prior to B_0 drift correction should be further tested with multiple datasets. Interestingly, the correction for phase drifts did not improve after denoising. From this perspective, denoising could be used to determine the optimal frequency drift corrections on individual spectra, and apply it to raw spectra, as previously described in a simpler spectral pattern (Brown and Stoyanova, 1996).

Second, the individual spectra after denoising displayed dramatically higher apparent SNR than raw spectra, even at the highest b-value. This single-shot SNR increase, however, results from a correlation with spectra from other shells. Whether this improvement on single shots may benefit other applications where averaging multiple spectra is detrimental, such as functional MRS, where it could provide a boost in temporal resolution, should be the subject of future work.

Figure S 1 also provides a perspective: a reduction in voxel size could be acceptable (reduced SNR by a factor 2 to 3 while preserving the number of shots yielding to a similar rank and fit RMSE). This could help improving spatial resolution with little penalty in quantification.

4.2. Strategy 1 versus strategy 2

For large heterogeneities across the dataset to be denoised, such as the extreme case of very low and very high b-values (in our case $b \sim 0.4$ and up to $b \sim 30\ ms/\mu m^2$), more leakage from high to low-SNR data is expected after denoising, which may bias high b-values concentration estimates.

Two approaches can, however, mitigate this effect.

The first approach is to denoise using a sliding-window along b-values, so that the spectra used in each denoising matrix are more similar to each other in terms of SNR. Here we tried a sliding window of three b-values (effectively leading to 3 × 2 × NS rows, accounting for real and imaginary parts of the signal, where NS is the number of shots acquired for each b-value). While this approach resulted in a more limited noise reduction, especially at high b-values (apparent SNR increase compared to raw of 575% for strategy 1 and 265% for strategy 2 at b_{max} in simulations, 241% for strategy 1 and 161% for strategy 2 at b_{max} in rodents and 166% for strategy 1 and 66% for strategy 2 at b_{max} in humans), it preserved noise variance better across b-values. The apparent SNR increase is higher in simulations compared to *in vivo* data, possibly owing to sources of non-Gaussian noise or distortions present in raw *in vivo* spectra and absent in simulations. Additionally, in human data, motion and spurious echoe artefacts at low b-values (even after cutting out the water region) are more prominent than in rodent data. These variations will not be captured in the noise principal components, thus increasing the selected rank and leading to a smaller denoising effect. The number of components retained (P) as signal-carrying decreased across blocks, both in simulations and *in vivo* (Figure 4C for simulations, Figure 8B for rodent data and from $P_{bmin} = 71.0 \pm 9.6$ to $P_{bmax} = 7.3 \pm 2.1$ for human data). One reason is that at high b-values, the variance created by the actual (low SNR) signal is close to the noise floor. An additional hypothesis is that sources of structural/physiological noise in the spectra (e.g. frequency drifts) are more discernible at low b-value than at high b-value and contribute to signal-carrying components. Possibly also, in the shells containing the highest b-values, the variations in the input data are more Gaussian-distributed than the ones observed in the shells containing the lowest b-values. This may further improve the separation of the signal from the artefacts and the noise.

The second mitigating approach, which remains to be tested, could be to diversify the DW-MRS acquisition scheme not only into multiple b-values, but also directions and diffusion times instead of plain repetitions. While working with only a small range of b-values (with similar SNR) – as for the sliding window above – the denoising matrix construct could nonetheless collect multiple directions and diffusion times. This would also enable to generate large matrices, improving the MP-PCA performance by fulfilling the asymptotic condition of the random matrix theory.

4.3. Assessment of denoising quality

One important aspect of MP-PCA denoising is the assumption of Gaussian, constant and uncorrelated entry noise. This assumption can be easily violated for MR imaging in clinical setups where multi-channel receiver coils are recombined using sum-of-squares algorithms, following which the magnitude of the complex signal is retained. In contrast, our *in vivo* preclinical setup was ideal to fulfil this criterion, as the receiver coil was a quadrature circuit whose signals were recombined physically prior to amplification. Each channel (real and imaginary) of the complex signal retained Gaussian noise properties. Despite the apparent SNR increase and homogenous residuals within each shot, some noise correlation within (**Figure 5**) and across shells ("noise decay", **Figure 4**A and **Figure 9**A) introduced by MP-PCA were identified in the current study. Consequently, noise estimation with a prior of uncorrelated Gaussian noise should be avoided in denoised spectra, as well as quality assessment based on noise amplitude, such as CRLB or the fit quality number (FQN) (Kreis et al., 2021). A bootstrapping approach for the estimation of metabolites concentration uncertainty has been recently proposed (Clarke and Chiew, 2022), where multiple fits of the same spectrum corrupted by correlated noise estimated from the denoised data, are performed.

4.4. Noise properties

Spectral residuals: When comparing denoising to averaging on a single-shell without distortions (**Figure 2**), we observed no patterns in the spectral residuals and their distribution was Gaussian. For the multi-shell case, hardly any pattern was observed with the exception of small artefacts (spikes), likely due to a B_0-drift correction mismatch between raw and denoised data (**Figure 4**D for simulations, **Figure 9**C). Remarkably, these artefacts are not present in human data (**Figure 11**C) where the B_0-drift correction was performed before denoising.

Uniform noise level across spectral points: In the entire study (single-shell, multi-shell, simulations and *in vivo* data), the noise in a noise-only region on single shots was Gaussian-distributed after denoising. When investigating the variance on each spectral point across MC iterations, the authors of ref. (Clarke and Chiew, 2022), who investigated different low-rank denoising methods for MRSI data, reported a non-uniform variance. Concretely, the standard deviation of the spectrum across MC iterations is higher in the metabolite region of the spectrum and smaller in the noise region. In the present study, in the case of an input matrix with only one signal information (single-shell), which is not centered (see **Theory** section), and when a rank $P = 1$ is manually selected, a non-uniform variance on spectral points is also observed. When the matrix is not centered, the only singular value selected will be an estimate of the mean of the input matrix, which might be biased. Remarkably, no non-uniform variance across spectral points was observed in our work when denoising a matrix comprised of multi-shell data, even without centering, with any of the strategies. The high number of principal components selected, $P \sim 11 - 12$, probably mitigates this effect.

Non-uniform noise level across b-values: In the present work, the noise level was b-value dependent after MP-PCA denoising (less noise in the high b-value spectra), an effect which was reduced by using a sliding-window across b-values. The evolution of the spectral SNR after denoising with strategy 1 is very similar between simulations, rodent data and human data, reaching a maximum value for intermediate b-values, $b \sim 7.6\ ms/\mu m^2$ for the simulations (**Figure 4**C), $b \sim 13.4\ ms/\mu m^2$ for *in vivo* rodent data (**Figure 8**B), and $b \sim 3\ ms/\mu m^2$ for *in vivo* human data (**Figure 11**B). **Figure S 2** gives a tentative explanation of this effect. After MP-PCA, the time evolution of the spectral points in a noise-only region will be reconstructed from one of the first signal-carrying singular vectors in the shot dimension (U_1 **Figure S 2**B), representing the overall decay of metabolites across b-values (strongest contribution to the variance). Consequently, the noise points will decay with increasing b-values. Due to the initial positive/negative distribution of these noise points (**Figure S 2**D-E), the noise level (standard deviation of the noise points across a spectral region) will decrease at intermediate b-values and increase again at higher b-values. Meanwhile, the NAA concentration decay is similar for raw and denoised data, which results in a maximum apparent SNR at intermediate b-values. With a similar argument in the other dimension, the first signal-carrying singular vectors in the spectral dimension will represent high SNR spectra (V_1 in **Figure S 2**B). The closer the metabolite information to the noise level, the more likely it will be reconstructed from a linear combination of high SNR spectral information, and even more so when P is small. This observation challenges the use of MP-PCA denoising for extracting low-concentrated metabolites information from the noise floor using the entire range of b-values. The sliding window approach can however mitigate these effects, as shown throughout the present work.

The number of principal components retained with strategy 1 was $\sim 11 - 12$ for simulations and *in vivo* data, which was also the rank found when using optimal shrinkage of the principal components (Gavish and Donoho, 2017). The high number of components was mostly due to the B_0 drift distortions which were not corrected for prior to denoising, to the random water residual, and to possible sources of non-Gaussian noise in *in vivo* data. Structural noise, retained as signal component, which has a larger impact on low-b spectra (in particular the water residual which is completely suppressed at high b-values) may therefore serendipitously limit the impact of noise reduction across shells.

4.5. Estimation of diffusion coefficients

From the perspective of metabolite quantification, MP-PCA denoising reduced the concentration ESFU (**Table S 1, Table S 2** and **Table S 3**). After denoising, the ESFU are not equivalent to the Cramer Rao Lower Bounds, representing the lower bounds of the fitting error, and which are based on a correct model and Gaussian uncorrelated noise, a prior which may be violated after denoising. Simulations showed that denoising based on the full range of b-values could also introduce bias for some metabolite concentration decays, such as lactate (**Figure 7**A-B), and an over-estimation of the concentrations at high b-values compared to the same concentrations on the raw data (**Table S 1**). Interestingly, this over-estimation is not systematic anymore when comparing denoised vs noiseless data: although beyond the scope of this work, this observation highlights some systematic underestimation of concentrations with LCModel for raw data with realistic SNR and Lorentzian broadening, as shown in a MRS fitting challenge (Marjańska et al., 2022). The sliding-window approach (strategy 2) introduced less bias on metabolite concentrations at high b-values than the full-range denoising and the raw data for high-concentrated metabolites, in addition to better preserving the noise structure. The observations made on the accuracy and precision of metabolite quantification could not be directly transposed to the estimations of the free diffusion coefficients D_{intra}. Overall, the sliding window-denoising followed the raw data estimates for most metabolites: whether or not bias (>10%) existed in the raw data estimates, the same was observed for strategy 2. The only exceptions are GABA, GSH, Lac, for which more bias was introduced with strategy 2 and Glc, mIns, Tau for which less bias was introduced with strategy 2.

However, this performance may depend on the underlying diffusivity values chosen in our simulations.

In simulations, the variability across MC iterations was also reduced after denoising (when compared to the raw data with the same number of shots) for all metabolites (**Figure 7**C). Remarkably, in rodent data, MP-PCA denoising also contributed to reducing the variability in metabolite concentration decay curves across the different rats (which were all part of a homogeneous control group) for some metabolites (NAA, tCr, Glu on **Figure 10**). The estimated metabolite diffusivities were systematically lower with MP-PCA denoising vs raw data, though the ground truth is not known in this case. This could reflect the systematic under-estimation of the raw data concentrations found in simulations (mentioned above), yielding lower concentration values at the tail of the curve and thus a higher estimated diffusivity. For human data, the datasets were not fitted individually in ref Şimşek et al., 2022, owing to their low SNR, but after doing a cohort average. The present study shows the feasibility of individual volunteer fitting, yet no difference in D_{intra} between the raw and denoised data with any strategy and no reduction in the group variability across volunteers was observed, possibly owing to the high P. Though, the difference between the diffusion decay fitted with Callaghan's model and the experimental data tended to be reduced after denoising.

It should also be noted that Callaghan's model of randomly-oriented sticks may not be well-suited to describe the diffusion of certain metabolites *in vivo*, e.g. if they are also extracellular and/or if the radius of the dendrites cannot be assumed to be effectively zero. For human data, in addition, there is substantial contribution of white matter where the randomly oriented stick model may not apply for the current case where only one arbitrary diffusion direction was acquired.

Simultaneous spectral and diffusion modelling (Şimşek et al., 2022, Adalid et al., 2017, Najac et al., 2022, Clarke et al., 2022) – though possibly more challenging - may also offer increased fit stability, but may not readily be combined with MP-PCA denoising given the noise correlation between shells. A sliding-window approach along the spectral dimension instead of the diffusion-weighting dimension could also be considered. This would however lead to issues in LCModel quantification, and multiple resonances of the same metabolite (thus sharing common features) could be denoised separately, thereby decreasing the redundancy. Finally, the strong spectral overlap of some metabolite resonances also prevents the selection of a denoising window which could contain only one metabolite.

5. Conclusion

Overall, we have shown that MP-PCA denoising improves apparent SNR and B_0 drift correction and thus spectral averaging. For highly concentrated metabolites, which are the ones typically considered in DW-MRS studies, we have shown that denoising improves the within-group homogeneity of estimated diffusivities with little penalty to the diffusivity estimates – future work could focus on testing whether the between-group differences are thus reinforced by comparing a control to a patient group. However, for low-concentrated metabolites, we have also shown that denoising biases their estimated diffusivity due to signal leakage from the high-concentrated metabolites.

In agreement with previous studies, we suggest that PCA-denoising for diffusion MRS should be used with caution and we recommend that all effects should be tested in simulations prior to drawing conclusions on *in vivo* data. Uniform variance along the spectrum was preserved due to the matrix centering and the selection of a high rank P by the MP fit (with uncorrected B_0 drift prior to denoising), but noise correlation across rows were introduced as a consequence of the rank truncation, which should prevent the use of the term CRLB after denoising. We recommend the use of an across-shell sliding window denoising approach (i.e. denoising more self-similar matrices) to mitigate the b-value dependent noise level post-denoising.

For DW-MRS acquisitions that include multiple diffusion times and diffusion directions, it remains to be established whether signal can be separated from noise more efficiently due to higher self-similarity of different measures.

Data availability statement

The MP-PCA Matlab code had been made publicly available by the authors of ref. (Veraart et al., 2016), on the following repository: https://github.com/NYU-DiffusionMRI/mppca_denoise. The Matlab code used to generate the simulation data is available on the following repository: https://github.com/jessie-mosso/DWMRS-MPPCA and linked to the MRSHub (https://mrshub.org). Rodent experimental data used in the present manuscript are available upon reasonable request to the Corresponding Author.

Declaration of Competing Interest

The authors have no conflict of interest to declare.

Credit authorship contribution statement

Jessie Mosso: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Dunja Simicic:** Conceptualization, Methodology, Writing – review & editing. **Kadir Şimşek:** Methodology, Software, Investigation, Writing – review & editing. **Roland Kreis:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing, Project administration, Funding acquisition. **Cristina Cudalbu:** Conceptualization, Methodology, Software, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Ileana O. Jelescu:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

Acknowledgements

This work was supported by the Center for Biomedical Imaging of the UNIL, UNIGE, HUG, CHUV, EPFL, the Leenaards and Jeantet Foundations and the SNSF projects no 310030_173222, 310030_201218, 320030-175984, the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 813120 (INSPiRE-MED). I.O.J. is supported by the SNSF under an Eccellenza grant PCEFP2_194260.

The authors would like to thank Dr. Jelle Veraart, Dr. William Clarke, Dr. Bernard Lanz and Dr. Julien Valette for many interesting discussions on the topic.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119634.

References

Pfeuffer, J., Tkác, I., Gruetter, R., 2000. Extracellular-intracellular distribution of glucose and lactate in the rat brain assessed noninvasively by diffusion-weighted 1H nuclear magnetic resonance spectroscopy *in vivo*. J. Cereb. Blood Flow Metab. 20, 736–746.

Nicolay, K., Braun, K.P., Graaf, R.A., Dijkhuizen, R.M., Kruiskamp, M.J, 2001. Diffusion NMR spectroscopy. NMR Biomed. 14, 94–111.

de Graaf, R.A., Braun, K.P.J., Nicolay, K., 2001. Single-shot diffusion trace 1H NMR spectroscopy. Magn. Reson. Med. 45, 741–748.

Palombo, M., Shemesh, N., Ronen, I., Valette, J., 2018. Insights into brain microstructure from *in vivo* DW-MRS. Neuroimage 182, 97–116.

Ligneul, C., et al., 2019. Diffusion-weighted magnetic resonance spectroscopy enables cell-specific monitoring of astrocyte reactivity *in vivo*. Neuroimage 191, 457–469.

Najac, C., Branzoli, F., Ronen, I., Valette, J., 2016. Brain intracellular metabolites are freely diffusing along cell fibers in grey and white matter, as measured by diffusion-weighted MR spectroscopy in the human brain at 7 T. Brain Struct. Funct. 221, 1245–1254.

Ronen, I., et al., 2014. Microstructural organization of axons in the human corpus callosum quantified by diffusion-weighted magnetic resonance spectroscopy of N-acetylaspartate and post-mortem histology. Brain Struct. Funct. 219, 1773–1785.

Genovese, G., et al., 2021. Inflammation-driven glial alterations in the cuprizone mouse model probed with diffusion-weighted magnetic resonance spectroscopy at 11.7 T. NMR Biomed. 34, e4480.

Brand, A., Richter-Landsberg, C., Leibfritz, D., 1993. Multinuclear NMR studies on the energy metabolism of glial and neuronal cells. Dev. Neurosci. 15, 289–298.

Urenjak, J., Williams, S.R., Gadian, D.G., Noble, M., 1993. Proton nuclear magnetic resonance spectroscopy unambiguously identifies different neural cell types. J. Neurosci. 13, 981–989.

Harris, J.L., Choi, I.-Y., Brooks, W.M., 2015. Probing astrocyte metabolism *in vivo*: proton magnetic resonance spectroscopy in the injured and aging brain. Front. Aging Neurosci. 7, 202.

Rackayova, V., Cudalbu, C., Pouwels, P.J.W., Braissant, O., 2017. Creatine in the central nervous system: from magnetic resonance spectroscopy to creatine deficiencies. Anal. Biochem. 529, 144–157.

Ebel, A., Dreher, W., Leibfritz, D., 2006. Effects of zero-filling and apodization on spectral integrals in discrete Fourier-transform spectroscopy of noisy data. J. Magn. Reson. 182, 330–338.

Brender, J.R., et al., 2019. Dynamic imaging of glucose and lactate metabolism by 13C-MRS without hyperpolarization. Sci. Rep. 9, 3410.

Ahmed, O.A., 2005. New denoising scheme for magnetic resonance spectroscopy signals. IEEE Trans. Med. Imaging 24, 809–816.

Goryawala, M., Sullivan, M., Maudsley, A.A., 2020. Effects of apodization smoothing and denoising on spectral fitting. Magn. Reson. Imaging 70, 108–114.

Abdoli, A., Stoyanova, R., Maudsley, A.A., 2016. Denoising of MR spectroscopic imaging data using statistical selection of principal components. MAGMA 29, 811–822.

Nguyen, H.M., Peng, X., Do, M.N., Liang, Z.-P., 2013. Denoising MR spectroscopic imaging data with low-rank approximations. IEEE Trans. Biomed. Eng. 60, 78–89.

Clarke, W.T., Chiew, M., 2022. Uncertainty in denoising of MRSI using low-rank methods. Magn. Reson. Med. 87, 574–588.

Klauser, A., et al., 2018. Fast high-resolution brain metabolite mapping on a clinical 3T MRI by accelerated 1H-FID-MRSI and low-rank constrained reconstruction. Magn. Reson. Med. 81.

Simicic, D., Lê, T.P., van Heeswijk, R., Jelescu, I.O., Cudalbu, C., 2021. The impact of Marchenko-Pastur PCA denoising on high resolution MRSI in the rat brain at 9.4T. In: Proc. Intl. Soc. Mag. Reson. Med., 29.

Knoll, F., Bredies, K., Pock, T., Stollberger, R., 2011. Second Order Total Generalized Variation (TGV) for MRI. Magn. Reson. Med. 65, 480–491.

Pijnappel, W.W.F., van den Boogaart, A., de Beer, R., van Ormondt, D., 1992. SVD-based quantification of magnetic resonance signals. J. Magn. Reson. 97, 122–134.

Rowland, B.C., Sreepada, L. & Lin, A.P. A comparison of denoising methods in dynamic MRS using pseudo-synthetic data. 2021.02.23.21252282 https://www.medrxiv.org/content/10.1101/2021.02.23.21252282v1 (2021) doi:10.1101/2021.02.23.21252282.

Doyle, M., Chapman, B.L., Blackwell, G., Walsh, E.G., Pohost, G.M., 1994. Adaptive Fourier threshold filtering: a method to reduce noise and incoherent artifacts in high resolution cardiac images. Magn. Reson. Med. 31, 546–550.

Lei, Y., et al., 2021. Deep learning-based denoising for magnetic resonance spectroscopy signals. In: Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging vol. 11600. SPIE, pp. 16–21.

Dziadosz, M., Rizzo, R., Kyathanahally, S.P., Kreis, R., 2021. Denoising 1H MR spectra in a time–frequency representation by deep learning. ESMRMB 38th Annual Scientific Meeting.

Veraart, J., et al., 2016. Denoising of diffusion MRI using random matrix theory. Neuroimage 142, 394–406.

Ades-Aron, B., et al., 2021. Improved task-based functional MRI language mapping in patients with brain tumors through marchenko-pastur principal component analysis denoising. Radiology 298, 365–373.

Diao, Y., Yin, T., Gruetter, R., Jelescu, I.O., 2021. PIRACY: an optimized pipeline for functional connectivity analysis in the rat brain. Front. Neurosci. 15, 285.

Does, M.D., et al., 2019. Evaluation of principal component analysis image denoising on multi-exponential MRI relaxometry. Magn. Reson. Med. 81, 3503–3514.

Froeling, M., Prompers, J., Klomp, D., Van der Velden, T., 2021. PCA denoising and Wiener deconvolution of 31P 3D CSI data to enhance effective SNR and improve point spread function. Magn. Reson. Med. 85.

Moeller, S., et al., 2021. NOise reduction with DIstribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing. Neuroimage 226, 117539.

Gavish, M., Donoho, D.L., 2017. Optimal shrinkage of singular values. IEEE Trans. Inf. Theory 63, 2137–2152.

Johnstone, I.M., Paul, D., 2018. PCA in high dimensions: an orientation. Proc IEEE Inst. Electr. Electron. Eng. 106, 1277–1292.

Ma, X., Uğurbil, K., Wu, X., 2020. Denoise magnitude diffusion magnetic resonance images via variance-stabilizing transformation and optimal singular-value manipulation. Neuroimage 215, 116852.

Kreis, R., et al., 2021. Terminology and concepts for the characterization of *in vivo* MR spectroscopy methods and MR spectra: background and experts' consensus recommendations. NMR Biomed. 34, e4347.

Marjańska, M., Deelchand, D.K., Kreis, R.Team, the 2016 I. M. S. G. F. C., 2022. Results and interpretation of a fitting challenge for MR spectroscopy set up by the MRS study group of ISMRM. Magn. Reson. Med. 87, 11–32.

Starčuk, Z., Starčuková, J., 2017. Quantum-mechanical simulations for *in vivo* MR spectroscopy: principles and possibilities demonstrated with the program NMRScopeB. Anal. Biochem. 529, 79–97.

Govindaraju, V., Young, K., Maudsley, A.A., 2000. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR Biomed. 13, 129–153.

Govind, V., Young, K., Maudsley, A.A., 2015. Corrigendum: proton NMR chemical shifts and coupling constants for brain metabolites. Govindaraju V, Young K, Maudsley AA, NMR Biomed. 2000; 13: 129-153. NMR Biomed. 28, 923–924.

Simicic, D., et al., 2021. *In vivo* macromolecule signals in rat brain 1H-MR spectra at 9.4T: parametrization, spline baseline estimation, and T2 relaxation times. Magn. Reson. Med. 86, 2384–2401.

Callaghan, P.T., Jolley, K.W., Lelievre, J, 1979. Diffusion of water in the endosperm tissue of wheat grains as studied by pulsed field gradient nuclear magnetic resonance. Biophys. J. 28, 133–141.

Gruetter, R., Tkác, I., 2000. Field mapping without reference scan using asymmetric echo–planar techniques. Magn. Reson. Med. 43, 319–323.

Callaghan, P.T., 1993. Principles of Nuclear Magnetic Resonance Microscopy. Clarendon Press.

Frahm, J., et al., 1989. Localized proton NMR spectroscopy in different regions of the human brain *in vivo*. Relaxation times and concentrations of cerebral metabolites. Magn. Reson. Med. 11, 47–63.

Kunz, N., et al., 2010. Diffusion-weighted spectroscopy: a novel approach to determine macromolecule resonances in short-echo time 1H-MRS. Magn. Reson. Med. 64, 939–946.

Tkác, I., Starcuk, Z., Choi, I.Y., Gruetter, R., 1999. *In vivo* 1H NMR spectroscopy of rat brain at 1 ms echo time. Magn. Reson. Med. 41, 649–656.

Şimşek, K., Döring, A., Pampel, A., Möller, H.E., Kreis, R., 2022. Macromolecular background signal and non-Gaussian metabolite diffusion determined in human brain using ultra-high diffusion weighting. Magn. Reson. Med. 88, 1962–1977.

Oz, G., et al., 2005. Assessment of adrenoleukodystrophy lesions by high field MRS in non-sedated pediatric patients. Neurology 64, 434–441.

Cudalbu, C., et al., 2021. Contribution of macromolecules to brain 1 H MR spectra: experts' consensus recommendations. NMR Biomed. 34, e4393.

Stoyanova, R., Brown, T., 2002. NMR spectral quantitation by principal component analysis - III. A generalized procedure for determination of lineshape variations. J. Magn. Reson. (San Diego, Calif. : 1997) 154, 163–175.

Brown, T., Stoyanova, R., 1996. NMR spectral quantitation by principal-component analysis. II. Determination of frequency and phase shifts. J. Magn. Reson. Ser. B doi:10.1006/JMRB.1996.0106.

Adalid, V., et al., 2017. Fitting interrelated datasets: metabolite diffusion and general lineshapes. MAGMA 30, 429–448.

Najac, C., et al., 2022. Differences in diffusion-weighted MRS processing and fitting pipelines, and their effect on tissue modeling: results from a workshop challenge. Proc. Intl. Soc. Mag. Reson. Med. 30.

Clarke, W.T., Ligneul, C., Cottaar, M., Jbabdi, S., 2022. Dynamic fitting of functional MRS, diffusion weighted MRS, and edited MRS using a single interface. Proc. Intl. Soc. Mag. Reson. Med. 30.