TECHNICAL NOTE

Chemical shift and relaxation regularisation improve the accuracy of ^{1}H MR spectroscopy analysis

Martin Wilson

Centre for Human Brain Health and School of Psychology, University of Birmingham, Birmingham, UK

Correspondence
Martin Wilson, Centre for Human Brain Health, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.
Email: wilsonmp@bham.ac.uk

Abstract

Purpose: Accurate analysis of metabolite levels from ^{1}H MRS data is a significant challenge, typically requiring the estimation of approximately 100 parameters from a single spectrum. Signal overlap, spectral noise and common artefacts further complicate analysis, leading to instability and reports of poor agreement between different analysis approaches. One inconsistently used method to improve analysis stability is known as regularisation, where poorly determined parameters are partially constrained to take a predefined value. In this study we examine how regularisation of frequency and linewidth parameters influences analysis accuracy.

Methods: The accuracy of three MRS analysis methods was compaired: 1) ABfit, 2) ABfit-reg and 3) LCModel, where ABfit-reg is a modified version of ABfit incorporating regularisation. Accuracy was assessed on synthetic MRS data generated with random variability in the frequency shift and linewidth parameters applied to each basis signal. Spectra (N=1000) were generated across a range of SNR values (10, 30, 60, 100) to evaluate the impact of variable data quality.

Results: Comparison between ABfit and ABfit-reg demonstrates a statistically significant ($p < 0.0005$) improvement in accuracy associated with regularisation for each SNR regime. An approximately 10% reduction in the mean squared metabolite errors were found for ABfit-reg compared to LCModel for SNR >10 ($p < 0.0005$). Furthermore, Bland-Altman analysis shows that incorporating regularisation into ABfit enhances its agreement with LCModel.

Conclusion: Regularisation is beneficial for MRS fitting and accurate characterisation of the frequency and linewidth variability *in vivo* may yield further improvements.

KEYWORDS:
ABfit-reg, ABfit, LCModel, open-source, spectral fitting, regularisation

WORD COUNT: 3344

Received 1 April 2016; | Revised 1 June 2016; | Accepted 1 June 2016

DOI: xxx/xxxx

1 | INTRODUCTION

^{1}H MR spectroscopy (MRS) is a unique tool for measuring brain metabolite levels with numerous applications in both the clinical[1] and neuroscience domains[2,3]. One particularly challenging aspect of MRS is the accurate estimation of metabolite levels amidst confounding factors, such as noise, residual water, outer-volume artefacts and lineshape variability[4]. Furthermore, these challenges are exacerbated by strong spectral overlap between some metabolite signals. While a number of approaches may be used to extract meaningful information from MRS data[5], spectral fitting is currently the most popular and recommended approach for MRS analysis[6], and a number of methods have been developed based on this strategy.

The LCModel analysis algorithm has become the most widely used spectral fitting method for MRS research, since its introduction in 1993[7]. The method is based on performing a "least-squares" fit of a "linear-combination" of predefined molecular signals (known as a basis-set) to the experimentally acquired spectral data. The fitting model estimates individual amplitude parameters for each element in the basis set, and following appropriate scaling, these correspond directly to the metabolite levels in the region of interest. In addition to simple scaling amplitudes, a number of non-linear parameters need to be estimated during the fitting process to ensure accurate metabolite level estimates. These include parameters that influence all basis-set signals equally, such as phase and lineshape, and those that are unique to each basis-set element, such as a chemical shift (frequency) and T_2 relaxation adjustments. These parameters are known to be influenced by experimental conditions, for instance B_0 inhomogeneity will have a significant influence on all metabolite lineshapes, and physiological changes in temperature and pH are likely to influence individual chemical shifts[8,9]. Finally, spectral fitting also requires the estimation of the baseline signal, which is characteristically smooth compared to the basis-set signals, and typically arises from incomplete water suppression and scalp lipid contamination. In LCModel, the baseline signal is modelled using smoothing splines[10].

More recently, a range of alternative fitting methods have been proposed, which share the same basic principle of the parametric adjustment of a basis set of metabolite signals to match the acquired data in a least-squares sense. These approaches may be broadly characterised based on their baseline modelling strategy, altough we note these methods also differ in a number of other ways. QUEST[11] and TARQUIN[12] both perform fitting in the time-domain, where the influence of rapidly decaying baseline signals is significantly suppressed by evaluating the fitting residual after a short temporal delay. FSL-MRS[13] and OSPREY[14] incorporate frequency-domain polynomial and spline baselines respectively, where the baseline smoothness is controlled by adjusting the spectral density of spline knots or highest degree of polynomial function. AQSES[15], ABfit[16] and ProFit-1D[17] all use penalised splines[18] for baseline modelling, where baseline smoothness is controlled by a penalty factor parameter, which is automatically determined in ABfit and ProFit-1D. While other MRS fitting approaches have been developed, including those optimised for edited[19] and 2D-MRS analysis[20], we limit our discussion here to those focused on the analysis of ^{1}H MRS data acquired with single-voxel or MRSI localisation schemes — resulting in one spectrum for each spatially encoded location.

MRS fitting algorithm accuracy is typically assessed in two ways: (1) direct comparison with a gold-standard reference method based on experimentally acquired data, or (2) the analysis of synthetic MRS data. Both methods have their relative advantages, with experimental data being better suited to the realistic evaluation of degraded data quality and artefacts on metabolite estimates. However, the "ground-truth" metabolite levels are not available for experimentally acquired data, presenting a significant challenge in judging the relative performance of each method. Knowledge of the genuine metabolite levels is readily available for synthetic MRS, however this comes at the risk of being contrived due to unrealistic synthesis assumptions.

One further strategy for assessing analysis accuracy involves the preparation of a phantom containing metabolites dissolved in solution. Whilst this approach benefits from knowledge of the "ground-truth" concentrations, in practice it is less commonly used than those described above. The preparation of the mixture of around 20 metabolites, required to approximate a typical brain spectrum, can become prohibitively expensive when high accuracy (and therefore chemical purity) is required. Additional complications arise from attempting to match: 1) typical water and metabolite relaxation properties; 2) pH; 3) temperature and 4) macromolecular contributions, to those typically observed *in vivo*.

Larger studies comparing various MRS fitting method have been performed, with Zöllner at al reporting results from a multi-centre dataset comprised of 277 short TE PRESS spectra acquired from healthy participants at 3 Tesla[21]. Only a weak to moderate agreement was found between approaches, despite relatively high data quality compaired to clinical MRS. A second study, based on synthetic conventional MRS data, also found relatively poor agreement between methods[22]. Weak

agreement between fitting approaches was also found for GABA edited MRS[23], a surprising finding considering the relative spectral simplicity of GABA edited spectra compared to conventional short TE MRS.

A typical MRS fitting model, with a basis set of around 25 to 30 signals, requires the estimation of approximately 75 to 100 parameters, in addition to those associated with the baseline model. The stable and accurate estimation of these parameters in the presence of noise, signal overlap and spectral artefacts presents a significant challenge, and overfitting is one likely cause of disagreement between fitting approaches. Restricting parameter estimates to feasible values, for example constraining metabolite amplitude estimates to be greater than or equal to zero, known as "hard-constraints", is one way to reduce overfitting. Another approach, known as "soft-constraints" or regularisation, encourage parameters to take an expected value. Unlike hard-constraints, soft-constraints may be violated, provided a clear deviation is supported by the data. For example, the default behaviour of LCModel enforces a soft-constraint on the level of GABA to take the value of 0.04 times the sum of total-NAA, total-creatine and 3×total-choline. Whilst soft-constraints of this type improve fitting stability for signals with comparable intensities to the noise level, they have also been shown to introduce unwanted bias, and are generally not advised[22]. Furthermore, bias introduced by these soft-constraints is likely exacerbated in pathology — where metabolite levels are more variable, e.g. low or absent total-NAA in brain tumour spectra.

Regularisation may also be applied to the parameters used to model the small frequency and linewidth differences between the basis-set signals and acquired data. Unlike amplitude regularisation, as described above, the relationship between fitting accuracy and regularisation of these non-linear parameters has not yet been reported. In this study, the ABfit fitting method is compared to a new version, incorporating regularisation of the frequency and relaxation parameters (ABfit-reg), and LCModel using synthetic MRS data. Regularisation is shown to significantly improve accuracy for the expected variations in frequency and relaxation, and ABfit-reg is shown to provide either comparable, or improved accuracy relative to LCModel.

2 | METHOD

2.1 | Regularisation implementation

Support for the regularisation of the individual basis signal frequency and relaxation parameters is added to ABfit algorithm[16], available as part of the spant[24] MRS analysis package. To differentiate between the closely related approaches we will use the terms "ABfit" to denote the method as described previously[16], and "ABfit-reg" to denote the updated version with regularisation. All aspects of the ABfit and ABfit-reg methods are identical except for those explicitly stated below.

During the final phase (step 4) of the ABfit algorithm[16] small individual frequency shift ($\mathbf{f}_i$) and Lorentzian line-broadening adjustments, due to T_2 relaxation, ($\mathbf{d}_i$) are applied in the time-domain to each signal in the basis-set $\mathbf{M}^{\mathrm{TD}}$:

$$\underline{\mathbf{M}}^{\mathrm{TD}} = \mathbf{M}^{\mathrm{TD}} * \mathbf{l}^{\mathrm{TD}} * e^{(2i\pi\mathbf{f}_i - \pi\mathbf{d}_i)\mathbf{t}}. \tag{1}$$

A "global" lineshape adjustment term ($\mathbf{l}^{\mathrm{TD}}$) is also applied equally to all signals in the basis-set to model inhomogeneity in the static magnetic field. The global lineshape model is Gaussian, modified with an asymmetry parameter (a_g), as described by Stancik and Brauns[25]. Transforming the modified basis to the frequency domain ($\underline{\mathbf{M}}_s^{\mathrm{FD}}$) and combining with a basis of spline functions ($\mathbf{B}$), to model smooth baseline features, leads to the modelled spectrum $\hat{\mathbf{y}}$:

$$\begin{bmatrix}\mathbf{B} & \underline{\mathbf{M}}_s^{\mathrm{FD}}\end{bmatrix} \hat{\mathbf{a}} = \hat{\mathbf{y}}, \tag{2}$$

where $\hat{\mathbf{a}}$ represents the combined vector of amplitudes for each signal in the basis and spline function. The objective function is defined as the least-squares difference between the modelled spectrum ($\hat{\mathbf{y}}$) and the acquired data ($\underline{\mathbf{y}}_s^{\mathrm{FD}}$), and non-linear fitting parameters, e.g. frequency and linewidth, are adjusted to minimise this function using the Levenberg-Marquardt algorithm[26]. In contrast to ABfit, terms to regularise the frequency shifts ($\mathbf{f}_i^r$), Lorentzian line-broadening ($\mathbf{d}_i^r$) and global lineshape asymmetry (a_g^r) parameter are appended to the objective function in ABfit-reg as follows:

$$\left\| \begin{bmatrix} \underline{\mathbf{y}}_s^{\mathrm{FD}} \\ 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{y}} \\ \mathbf{f}_i^r \\ \mathbf{d}_i^r \\ a_g^r \end{bmatrix} \right\|^2 \tag{3}$$

where the double vertical bars represent the l^2 norm.

Equation 4 defines the non-regularised objective function used by ABfit for comparison:

$$\left\| \underline{\mathbf{y}}_s^{\mathrm{FD}} - \hat{\mathbf{y}} \right\|^2. \tag{4}$$

Each of the penalty terms added to equation 3 follow the basic form of scaling the regularised fitting parameters by the standard deviation of the noise (σ), estimated from a signal free region of the acquired spectrum, and scalar values determining the regularisation strength for frequency shifts (f_{reg}), Lorentzian line-broadening (d_{reg}) and global lineshape asymmetry (a_{reg}):

$$\mathbf{f}_i^r = \frac{\sigma\,\mathbf{f}_i}{f_{\text{reg}}\,F_0 \times 10^{-6}}, \tag{5}$$

$$\mathbf{d}_i^r = \frac{\sigma\,(\mathbf{d}_i - d_{\text{exp}})}{d_{\text{reg}}}, \tag{6}$$

$$a_g^r = \frac{\sigma\,a_g}{a_{\text{reg}}}. \tag{7}$$

This specific form of regularisation was chosen to be compatible with the approach taken by LCModel, where the regularisation strength is scaled to have comparable magnitude to the spectral noise. Frequency shift regularisation is scaled by $F_0 \times 10^{-6}$ to convert to ppm units, where F_0 is the transmitter frequency in Hz. The expected Lorentzian line-broadening in Hz parameter (d_{exp}) is introduced to penalise deviations from this common value, whereas values deviating from zero are penalised for the frequency shift and global lineshape asymmetry parameters.

2.2 | Synthetic MRS

Synthetic MRS data were generated using the spant analysis package[24] to establish the influence of regularisation on MRS fitting accuracy. Spectra were generated from a basis set comprised of 29 signals, and corresponding concentrations, listed in Table 1. Metabolite concentrations were based on approximate ranges for healthy brain tissue as listed in de Graaf[27]. Metabolite simulations were based on chemical shift and J-coupling values from Govindaraju et al[28], whereas lipid and macromolecular resonances were based on parameters in Table 1 from Wilson et al[12]. An additional inverted singlet resonance (-CrCH2) at 3.913 ppm was included in the basis set, as it is often used to model the effect of water suppression on the downfield creatine and phosophocreatine CH2 resonances. The basis set was simulated for the semi-LASER pulse sequence[29] at a field strength of 3 Tesla ($F_0 = 127.8 \times 10^6$ Hz) and an echo-time of 28 ms (TE1=8 ms, TE2=11 ms, TE3=9 ms) with ideal RF pulses. Each basis signal was simulated over 1024 complex points sampled at a temporal frequency of 2000 Hz.

Prior to summation of the basis signals, normally distributed random Lorentzian linebroadening and frequency shifts were applied. Assumptions about the expected variability in linebroadening and frequency shifts were derived from the LCModel manual[30] to ensure a fair comparison between methods. The standard deviation of frequency shifts was 0.004 ppm, with a mean of zero, equivalent to the LCModel parameter `DESDSH`. The expected Lorentzian broadening is derived from the `DEEXT2` parameter as follows:

TABLE 1 Basis set signals and synthetic MRS data concentrations.

Name (abbreviation)	Conc. (mM)
Alanine (Ala)	0.2
Aspartate (Asp)	1.5
Creatine (Cr)	5.0
Gamma-aminobutyric acid (GABA)	1.0
Glucose (Glc)	1.0
Glutamate (Glu)	10.0
Glutamine (Gln)	3.0
Glutathione (GSH)	2.0
Glycerophosphocholine (GPC)	1.2
Glycine (Gly)	1.0
Lactate (Lac)	0.5
Myo-inositol (Ins)	5.0
N-acetyl aspartate (NAA)	10.0
N-acetyl aspartyl glutamate (NAAG)	2.0
Phosphocholine (PCho)	0.5
Phosphocreatine (PCr)	4.5
Phosphoethanolamine (PEth)	0.8
Scyllo-inositol (sIns)	0.4
Taurine (Tau)	1.5
-CrCH2	0.0
Lip09	0.0
Lip13a	0.0
Lip13b	0.0
Lip20	0.0
MM09	4.0
MM12	4.0
MM14	4.0
MM17	4.0
MM20	4.0

$$d_{\text{exp}} = \frac{\texttt{DEEXT2}\sqrt{\texttt{HZPPPM}/85.15}}{\pi} = 0.78\,\text{Hz}, \tag{8}$$

where $\texttt{HZPPPM} = F_0 \times 10^{-6}$ and $\texttt{DEEXT2} = 2.0$. Similarly, the standard deviation of the Lorentzian broadening is determined as:

$$\frac{\texttt{DESDT2}\sqrt{\texttt{HZPPPM}/85.15}}{\pi} = 0.156\,\text{Hz}, \tag{9}$$

where $\texttt{DESDT2} = 0.4$. Following the application of random linebroadening and frequency shifts, basis signals were scaled according to the concentrations listed in Table 1, summed, and dampened with 4 Hz Gaussian linebroadening to model inhomogeneity in the static magnetic field typically observed for good quality MRS data acquired at 3 Tesla. Normally distributed random complex noise was added to generate sets of spectra with different SNRs.

1000 spectra with differing noise samples, frequency shifts ($\mathbf{f}_i$) and linebroadening parameters ($\mathbf{d}_i$) were generated across 4 spectral SNR (10, 30, 60 and 100) regimes. SNR was calculated as the maximum spectral

value divided by the standard deviation of the noise, measured from the real component of the spectra. A second set of data were also generated in the same way across the 4 SNR regimes, except frequency shifts and linebroadening parameters followed a uniform (rather than normal) statistical distribution. Minimum and maximum limits for the uniformly distributed parameters were set to 95% confidence intervals of the standard deviations used for the normally distributed dataset: $\pm 0.004 \times 1.96$ ppm for the frequency shifts and $0.78 \pm 0.156 \times 1.96$ Hz for linebroadening.

Water reference data was also generated, consisting of a singlet resonance at 4.65 ppm with a Gaussian linewidth of 4 Hz and an amplitude (25116) corresponding to the default assumptions for LCModel water-scaling[30] (`WCONC` = 35880, `ATTH2O=0.7`).

The 8 synthetic MRS datasets, comprised of 1000 spectra each, basis-set and water reference data, are available from Zenodo https://doi.org/10.5281/zenodo.14165737 in NIfTI MRS format[31]. Code to generate the synthetic data, and reproduce all results presented in the following sections, is available from GitHub https://github.com/martin3141/abfit_reg_paper.

2.3 | Fitting method details and comparison

Synthetic MRS data was analysed with three different methods: 1) ABfit as previously described[16]; 2) ABfit-reg as described above and 3) LCModel[7]. A comparison between ABfit and ABfit-reg was performed to isolate the specific influence of regularisation, since all other aspects are identical. LCModel was also included to investigate if regularisation improved the agreement between fitting methods.

The default LCModel method was applied with the `NRATIO` fitting parameter set to zero, disabling soft-constraints on metabolite ratios, as recommended by Marjańska et al[22] to avoid the potential for bias. The following regularisation parameters were used with ABfit-reg to allow a direct comparison with the assumptions made by LCModel: $f_{\text{reg}} = 0.004$ ppm, $d_{\text{exp}} = 0.78$ Hz, $d_{\text{reg}} = 0.156$ Hz. The lineshape regularisation parameter was set as: $a_{\text{reg}} = 0.1$, approximately equal to the standard deviation of the 95% confidence interval of ± 0.25 used in ABfit. Upper and lower constraints for $\mathbf{f}_i$ and a_g were not enforced in ABfit-reg and non-negativity was the only constraint for $\mathbf{d}_i$. Finally, constraints on the the global frequency offset parameter (f_0), in the step 4 of the algorithm, were increased from a negligibly small value in ABfit to 0.05 ppm in ABfit-reg.

A fourth exploratory analysis was performed where the ABfit-reg method was used as described above, but with an additional hard constraint applied to the individual frequency shifts ($10^{-6} > \mathbf{f}_i > -10^{-6}$), effectively eliminating these parameters from the fitting model. This, "fixed frequency", fitting approach will be referred to as ABfit-reg-ff.

Fitting accuracy was evaluated as the sum of squared differences from the known concentrations across the 19 metabolite signals listed in Table 1. Agreement between ABfit-reg and LCModel was assessed using Bland-Altman plots[32] and all statistical tests were performed in the R software environment[33].

3 | RESULTS

Example plots of fitting results for three example spectra (SNR=10, 30, 100) for ABfit, ABfit-reg and LCModel are shown in Figure 1. The consistent absence of spectral features above the noise level in the fit residual and flat baselines suggest a similarly high level of analysis quality between the approaches.

Figure 2 shows a clear difference between the accuracy of ABfit and ABfit-reg, despite the similar appearance in fitting results (Figure 1). The addition of regularisation results in a statistically significant improvement in accuracy of ABfit across all of the four SNR regimes investigated. The relative reduction in the mean squared metabolite error between ABfit and ABfit-reg was 22%, 31%, 45% and 60% for SNRs of 10, 30, 60 and 100 respectively. The reducing relative error associated with increasing SNR suggests that errors associated with local-minima may be dominant at higher SNR, and that regularisation guides the optimisation procedure to improved solutions.

ABfit-reg is shown to have a statistically significant improvement in accuracy over LCModel for SNRs of 30, 60 and 100, and that the methods have effectively the same accuracy for an SNR of 10 (Figure 2). The relative reduction in the mean squared metabolite error between LCModel and ABfit-reg was 8%, 10% and 11% for SNRs of 30, 60 and 100 respectively.

Influence of the statistical distribution of random shifts and damping factors is very minor (Figure 2 vs S1), suggesting the advantages of regularisation are not heavily dependant on the assumption of normality.

Figure S2 shows that individual frequency shift regularisation has comparable accuracy to simply fixing these shifts to zero ($\mathbf{f}_i = 0$) at SNR = 10. This is likely due to the errors associated with spectral noise being comparable, or greater, than the errors associated with inaccurate individual frequency estimation. For SNRs greater than

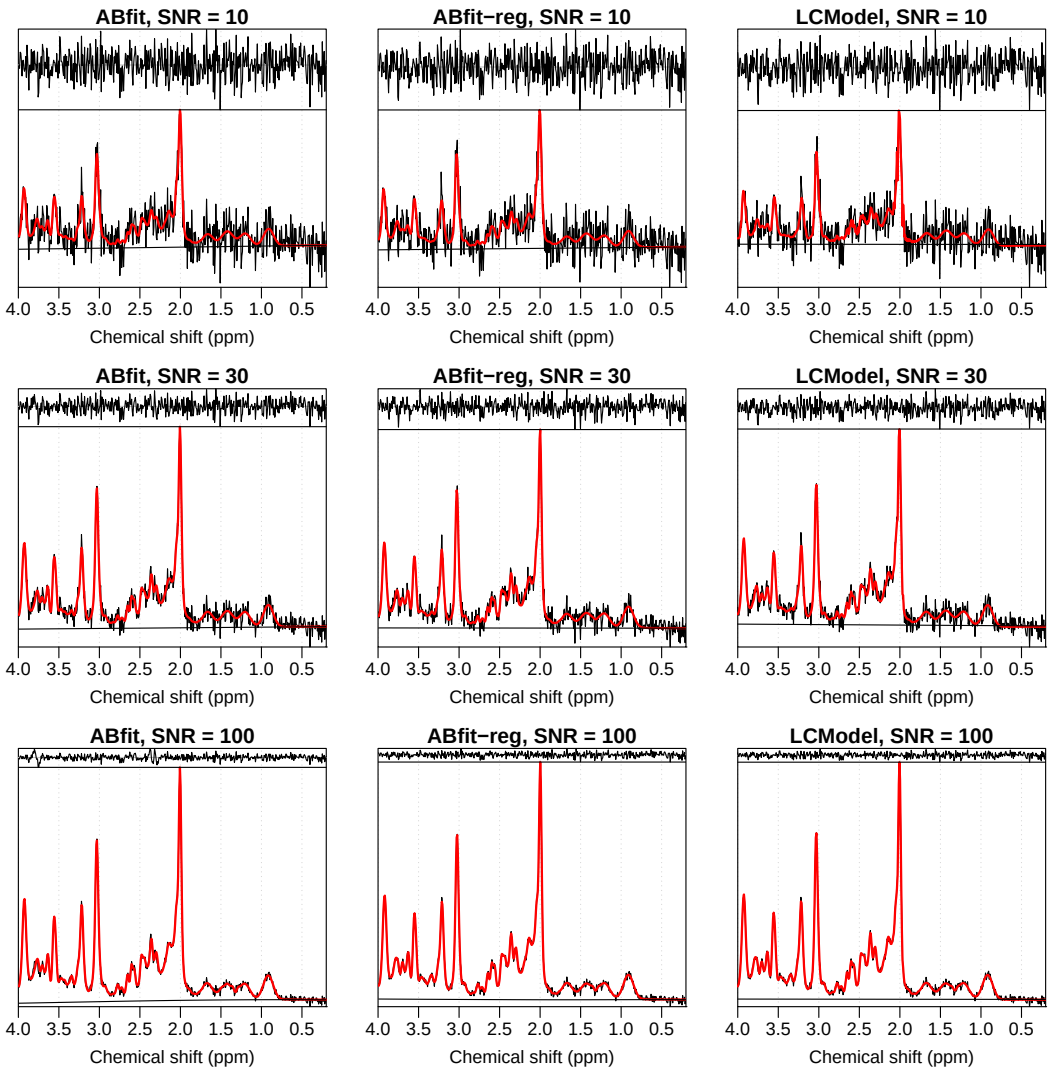

FIGURE 1 Fitting result plots of three example spectra (SNR=10, 30, 100) analysed with ABfit, ABfit-reg and LCModel. Acquired spectral data is shown in black with the fit shown in red. The black traces above and below the acquired spectrum represent the fitting residual and estimated baseline respectively.

10, the errors associated with poor frequency alignment become dominant, with ABfit-reg demonstating improved accuracy over ABfit-reg-ff ($p < 0.0005$).

Bland-Altman analysis was performed to establish if regularisation improved agreement between fitting methods. Figure 3 shows the agreement between LCModel and ABfit / ABfit-reg for GABA, Glu and tNAA (NAA + NAAG), chosen due to their differing contributions to typical healthy brain MRS. Strong reductions in both bias and variability are observed for ABfit-reg vs LCModel compaired to ABfit vs LCModel, demonstrating that regularisation is a significant factor to consider when comparing fitting results from experimentally acquired data - where the true accuracy is unknown.

4 | DISCUSSION

Regularisation of the non-linear fitting parameters for small adjustments to the frequency and linewidth of each basis signal is an integral part of the widely-used LCModel fitting method. However, to the best of the author's knowledge, a direct evaluation of the benefits of this approach have not been reported. In this study, we adapt the ABfit method to incorporate regularisation (ABfit-reg) and demonstrate an improvement in fitting accuracy across a range of spectral SNRs. Furthermore, we show that incorporating regularisation into ABfit enhances its agreement with LCModel, offering a partial explanation for the generally poor agreement between different fitting approaches[21,22,23].

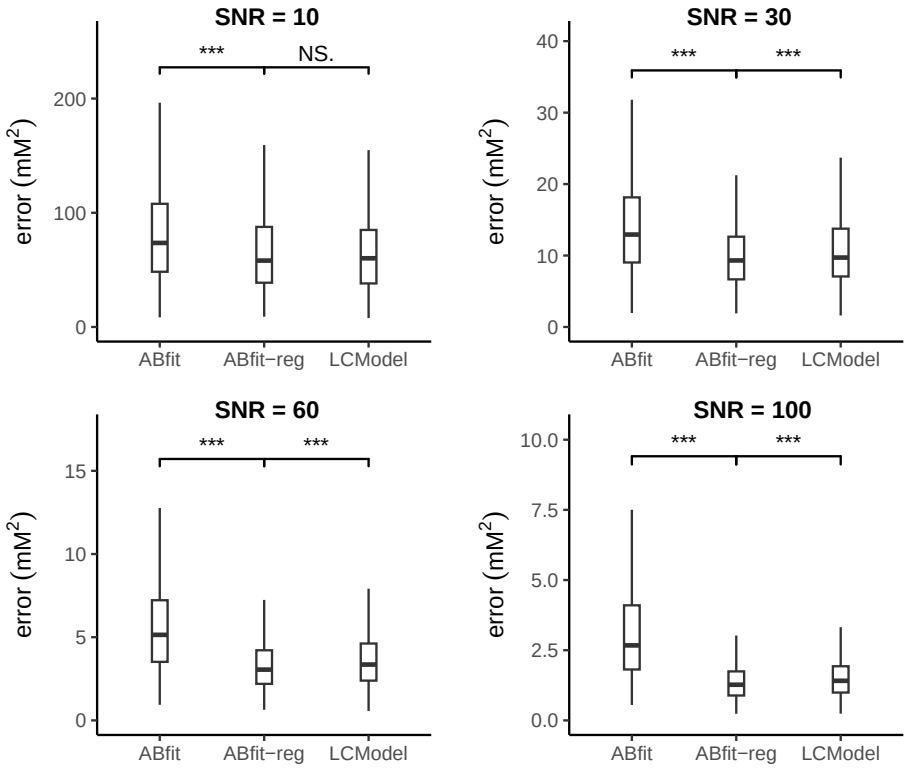

FIGURE 2 Box and whisker plots comparing metabolite estimate errors (sum of squared differences) for ABfit, ABfit-reg and LCModel from 1000 spectral fits per spectral SNR regime. Normally distributed random frequency shift and linebroadening parameters were applied to individual basis signals. Outliers are not plotted to aid visual comparison between the median values. Statistical significance labels represent a t-test between the two fitting methods: NS. not significant; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$.

We also note a small, but statistically significant, improvement in accuracy of ABfit-reg compaired to LCModel for spectral SNRs greater than 10. One potential explanation for this contrast in performance may be due to the different approaches to estimate baseline smoothness, however baseline distortions were intentionally not included in the synthetic data model to reduce bias from this aspect. Another possibility is the global lineshape model in LCModel has a greater level of flexibility compaired to ABfit-reg. Since the simulated data model assumes a simple Gaussian lineshape, we may expect ABfit-reg to have improved accuracy due to less degrees of freedom (two lineshape parameters), whereas LCModel may be expected to perform better for more heavily distorted lineshapes due to its "model-free" approach.

In this simulation study we make the same assumptions about random frequency and linewidth changes as LCModel to ensure a fair comparison. However the LCModel manual[30] states these assumptions are based on expected differences between *in vivo* and *in vitro* conditions. Historically, metabolite basis sets for MRS analysis were experimentally acquired from spherical phantoms containing solutions of individual metabolites *in vitro*. However, since the development of accurate numerical spectral simulation[34,35] and published chemical shift and J-coupling metabolite values[28], the use of simulated basis has become the standard approach. Experimentally acquired basis sets are typically acquired at room temperature, whereas simulation parameters are derived from experiments acquired at the more useful physiological temperature of (37°C). Since metabolite chemical shifts have a known dependence on temperature[8], it is possible that the range of frequency deviations assumed by LCModel are overestimated for simulated basis sets. Furthermore, chemical shifts are also known to depend on pH[9], therefore we might expect lower variability for normally functioning brain tissue compaired to ischemic or cancerous tissue. Further work could be undertaken to measure the true variability of metabolite signals *in vivo*, for healthy and diseased tissue, to optimise regularisation strength — potentially resulting in more accurate metabolite measures.

In this study, we focused on the most commonly used *in vivo* MRS spectral fitting model, where each

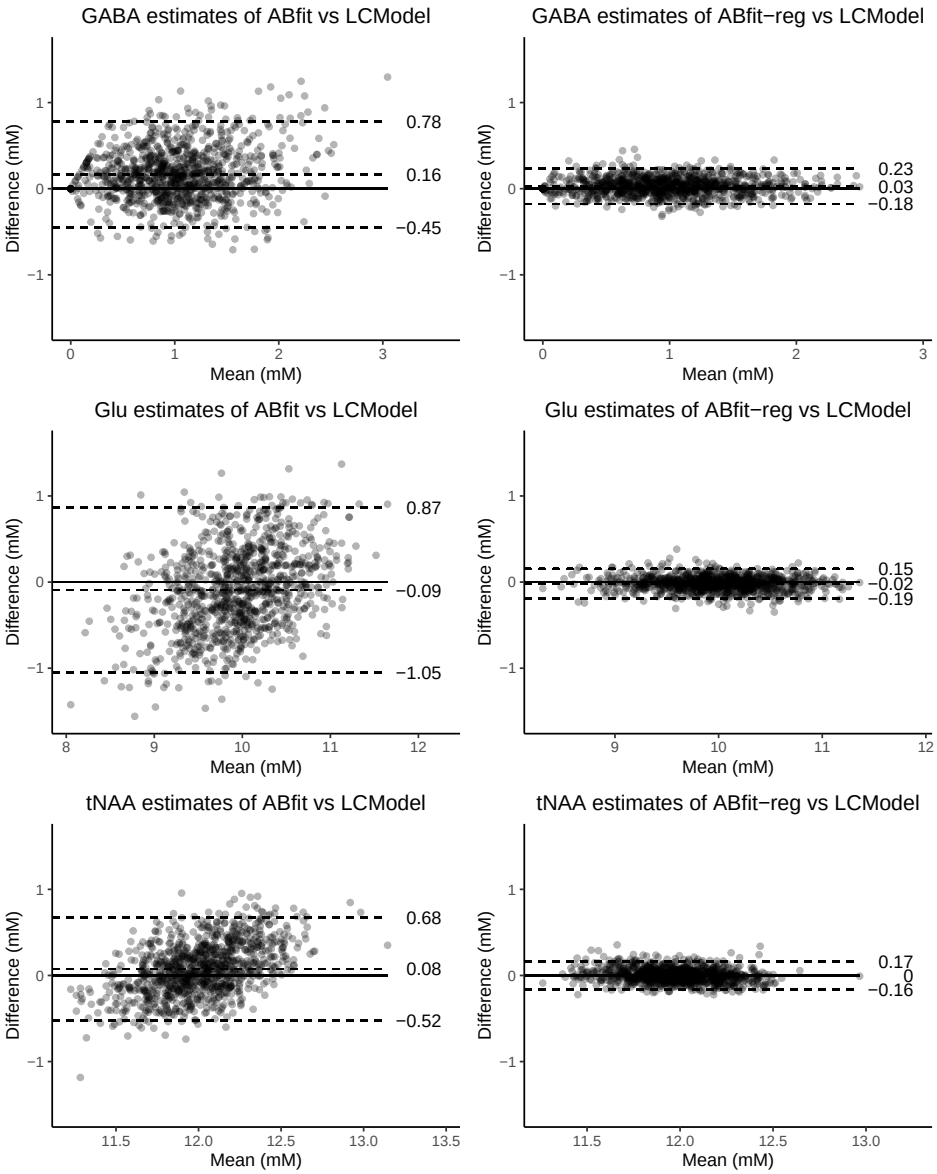

FIGURE 3 Bland-Altman plots comparing agreement between regularised (ABfit-reg, LCModel) and non-regularised (ABfit) fitting methods. The central horizontal dashed line represents the mean difference, with lower and upper dashed lines representing the 95% confidence intervals of mean difference.

metabolite corresponds to one signal, modified by a single frequency shift parameter and single linebroadening parameter. Whilst commonly used, this approach is a simplification of the potential variability in metabolite resonances, where each set of chemically equivalent spins (in the same molecule) has the potential for independent frequency and linebroadening terms. These, more complex, models have been developed primarily for *in vitro* analysis of diseased tissue[36], where larger variations in pH necessitate greater model freedom, and higher SNR reduces the risk of overfitting. The use of such models *in vivo* is more controversial, since the evidence for significant "proton-group" independence is less established and lower SNR and spectral resolution increase the potential for overfitting. Suggested future work involves estimating the magnitude of these "proton-group" changes from high quality *in vivo* data, and the use of regularisation to reduce fitting instability when applying these more complex models to poorer quality data.

In conclusion, we have demonstrated the benefits of regularisation for MRS analysis, and highlight the potential of characterising frequency and linewidth variability *in vivo* for further improvements in accuracy.

Conflict of interest

The author declares no potential conflict of interests.

Data availability statement

Synthetic MRS datasets are available from Zenodo https://doi.org/10.5281/zenodo.14165737. Code to generate the synthetic data and reproduce all figures and results is available from GitHub https://github.com/martin3141/abfit_reg_paper.

ORCID
Martin Wilson 0000-0002-2089-3956

REFERENCES

1. Oz Gülin, Alger Jeffry R., Barker Peter B., et al. Clinical proton MR spectroscopy in central nervous system disorders. *Radiology.* 2014;270(3):658–679.
2. Mullins Paul G., McGonigle David J., O'Gorman Ruth L., et al. Current practice in the use of MEGA-PRESS spectroscopy for the detection of GABA. *NeuroImage.* 2014;86:43–52.
3. Mullins Paul G.. Towards a theory of functional magnetic resonance spectroscopy (fMRS): A meta-analysis and discussion of using MRS to measure changes in neurotransmitters in real time. *Scandinavian Journal of Psychology.* 2018;59(1):91–103. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjop.12411.
4. Kreis Roland. Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts. *NMR in biomedicine.* 2004;17(6):361–381.
5. Poullet Jean-Baptiste, Sima Diana M., Van Huffel Sabine. MRS signal quantitation: A review of time- and frequency-domain methods. *Journal of Magnetic Resonance.* 2008;195(2):134–144.
6. Near Jamie, Harris Ashley D., Juchem Christoph, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR in biomedicine.* 2021;34(5):e4257.
7. Provencher S. W.. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic Resonance in Medicine.* 1993;30(6):672–679.
8. Wermter Felizitas C., Mitschke Nico, Bock Christian, Dreher Wolfgang. Temperature dependence of 1H NMR chemical shifts and its influence on estimated metabolite concentrations. *Magnetic Resonance Materials in Physics, Biology and Medicine.* 2017;30(6):579–590.
9. Moore G. J., Sillerud L. O.. The pH dependence of chemical shift and spin-spin coupling for citrate. *Journal of Magnetic Resonance. Series B.* 1994;103(1):87–88.
10. O'Sullivan Finbarr. A Statistical Perspective on Ill-Posed Inverse Problems. *Statistical Science.* 1986;1(4):502–518. Publisher: Institute of Mathematical Statistics.
11. Ratiney H., Sdika M., Coenradie Y., Cavassila S., Ormondt D., Graveron-Demilly D.. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR in biomedicine.* 2005;18(1):1–13.
12. Wilson Martin, Reynolds Greg, Kauppinen Risto A., Arvanitis Theodoros N., Peet Andrew C.. A constrained least-squares approach to the automated quantitation of in vivo ^{1}H magnetic resonance spectroscopy data. *Magnetic Resonance in Medicine.* 2011;65(1):1–12.
13. Clarke William T., Ligneul Clémence, Cottaar Michiel, Ip I. Betina, Jbabdi Saad. Universal dynamic fitting of magnetic resonance spectroscopy. *Magnetic Resonance in Medicine.* 2024;91(6):2229–2246. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.30001.
14. Oeltzschner Georg, Zöllner Helge J., Hui Steve C. N., et al. Osprey: Open-source processing, reconstruction & estimation of magnetic resonance spectroscopy data. *Journal of Neuroscience Methods.* 2020;343:108827.
15. Poullet Jean-Baptiste, Sima Diana M., Simonetti Arjan W., et al. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. *NMR in Biomedicine.* 2007;20(5):493–504. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nbm.1112.
16. Wilson Martin. Adaptive baseline fitting for 1H MR spectroscopy analysis. *Magnetic Resonance in Medicine.* 2021;85(1):13–29.
17. Borbath Tamas, Murali-Manohar Saipavitra, Dorst Johanna, Wright Andrew Martin, Henning Anke. ProFit-1D-A 1D fitting software and open-source validation data sets. *Magnetic Resonance in Medicine.* 2021;86(6):2910–2929.
18. Eilers Paul H. C., Marx Brian D.. Flexible smoothing with B-splines and penalties. *Statistical Science.* 1996;11(2):89–121. Publisher: Institute of Mathematical Statistics.
19. Edden Richard A.E., Puts Nicolaas A.J., Harris Ashley D., Barker Peter B., Evans C. John. Gannet: A batch-processing tool for the quantitative analysis of gamma-aminobutyric acid–edited MR spectroscopy spectra. *Journal of Magnetic Resonance Imaging.* 2014;40(6):1445–1452. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.24478.
20. Chong Daniel G. Q., Kreis Roland, Bolliger Christine S., Boesch Chris, Slotboom Johannes. 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. *Magnetic Resonance Materials in Physics, Biology and Medicine.* 2011;24(3):147–164.
21. Zöllner Helge J., Považan Michal, Hui Steve C.N., Tapper Sofie, Edden Richard A.E., Oeltzschner Georg. Comparison of different linear-combination modeling algorithms for short-TE proton spectra. *NMR in biomedicine.* 2021;34(4):e4482.
22. Marjańska Małgorzata, Deelchand Dinesh K., Kreis Roland. Results and interpretation of a fitting challenge for magnetic resonance spectroscopy set up by the MRS study group of ISMRM. *Magnetic resonance in medicine.* 2022;87(1):11–32.

23. Craven Alexander R., Bhattacharyya Pallab K., Clarke William T., et al. Comparison of seven modelling algorithms for gamma-aminobutyric acid-edited proton magnetic resonance spectroscopy. *NMR in biomedicine.* 2022;35(7):e4702.
24. Wilson Martin. spant: An R package for magnetic resonance spectroscopy analysis. *Journal of Open Source Software.* 2021;6(67):3646.
25. Stancik Aaron L., Brauns Eric B.. A simple asymmetric lineshape for fitting infrared absorption spectra. *Vibrational Spectroscopy.* 2008;47(1):66–69.
26. Levenberg Kenneth. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics.* 1944;2(2):164–168.
27. De Graaf Robin A. *In Vivo NMR Spectroscopy: Principles and Techniques, 3rd Edition.* Wiley; 2018.
28. Govind Varan, Young Karl, Maudsley Andrew A.. Corrigendum: proton NMR chemical shifts and coupling constants for brain metabolites. Govindaraju V, Young K, Maudsley AA, NMR Biomed. 2000; 13: 129-153. *NMR in biomedicine.* 2015;28(7):923–924.
29. Oz Gülin, Tkáč Ivan. Short-echo, single-shot, full-intensity proton magnetic resonance spectroscopy for neurochemical profiling at 4 T: validation in the cerebellum and brainstem. *Magnetic Resonance in Medicine.* 2011;65(4):901–910.
30. Provencher Stephen. *LCModel & LCMgui User's Manual.* 2021.
31. Clarke William T., Bell Tiffany K., Emir Uzay E., et al. NIfTI-MRS: A standard data format for magnetic resonance spectroscopy. *Magnetic Resonance in Medicine.* 2022;.
32. Bland J. M., Altman D. G.. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England).* 1986;1(8476):307–310.
33. Team R Core. R: A Language and Environment for Statistical Computing. 2021;.
34. Smith S. A., Levante T. O., Meier B. H., Ernst R. R.. Computer Simulations in Magnetic Resonance. An Object-Oriented Programming Approach. *Journal of Magnetic Resonance, Series A.* 1994;106(1):75–105.
35. Zhang Yan, An Li, Shen Jun. Fast computation of full density matrix of multispin systems for spatially localized in vivo magnetic resonance spectroscopy. *Medical Physics.* 2017;44(8):4169–4178.
36. Reynolds Greg, Wilson Martin, Peet Andrew, Arvanitis Theodoros N.. An algorithm for the automated quantitation of metabolites in in vitro NMR signals. *Magnetic Resonance in Medicine.* 2006;56(6):1211–1219.

List of Figures

1 Fitting result plots of three example spectra (SNR=10, 30, 100) analysed with ABfit, ABfit-reg and LCModel. Acquired spectral data is shown in black with the fit shown in red. The black traces above and below the acquired spectrum represent the fitting residual and estimated baseline respectively. 6

2 Box and whisker plots comparing metabolite estimate errors (sum of squared differences) for ABfit, ABfit-reg and LCModel from 1000 spectral fits per spectral SNR regime. Normally distributed random frequency shift and linebroadening parameters were applied to individual basis signals. Outliers are not plotted to aid visual comparison between the median values. Statistical significance labels represent a t-test between the two fitting methods: NS. not significant; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$. 7

3 Bland-Altman plots comparing agreement between regularised (ABfit-reg, LCModel) and non-regularised (ABfit) fitting methods. The central horizontal dashed line represents the mean difference, with lower and upper dashed lines representing the 95% confidence intervals of mean difference. 8

LIST OF FIGURES FOR SUPPORTING INFORMATION

Figure S1 Box and whisker plots comparing metabolite estimate errors for ABfit, ABfit-reg and LCModel from 1000 spectral fits per spectral SNR regime. Uniformly distributed random frequency shift and linebroadening parameters were applied to individual basis signals. Outliers are not plotted to aid visual comparison between the median values. Statistical significance labels represent a t-test between the two fitting methods: NS. not significant; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$.

Figure S2 Box and whisker plots comparing metabolite estimate errors for ABfit-reg and ABfit-reg with fixed ($\mathbf{f}_i = 0$) individual frequency parameters (ABfit-reg-ff). Plots are derived from 1000 spectral fits per spectral SNR regime. Normally distributed random frequency shift and linebroadening parameters were applied to individual basis signals. Outliers are not plotted to aid visual comparison between the median values. Statistical significance labels represent a t-test between the two fitting

methods: NS. not significant; * $p < 0.05$; ** $p < 0.005$;
*** $p < 0.0005$.

□