The Impact of Training Data on MRS Metabolite Quantification with Deep Learning

Zien Ma[1], SM Shermer[2], Oktay Karakus[1], Frank C Langbein[1]

[1] School of Computer Science and Informatics, Cardiff University, UK {MaZ17,KarakusO,LangbeinFC}@cardiff.ac.uk

[2] Faculty of Science & Engineering (Physics), Swansea University, UK, s.m.shermer@gmail.com

https://qyber.black/

Realistic Noise Simulation Boosts Deep Learning Performance for MRS Metabolite Quantification

Deep learning models for MRS metabolite quantification are highly sensitive to the realism of training data. **Given the challenge of obtaining *in vivo* ground-truth metabolite concentrations, training and validation rely on simulations and phantom data.** Using a Y-shaped autoencoder model, this study investigated the impact of varying metabolite basis sets and noise models in these simulations. While basis variations had limited effect, we found that **training the autoencoder with realistic noise models estimated from experimental data (Generalized Gaussian: GG) significantly improved quantification accuracy on phantom spectra** compared to simpler additive Gaussian noise (ADC). This emphasises that accurate noise simulation is crucial for improving the transferability of simulation-trained models and developing more accurate and robust deep learning quantification methods for MRS.

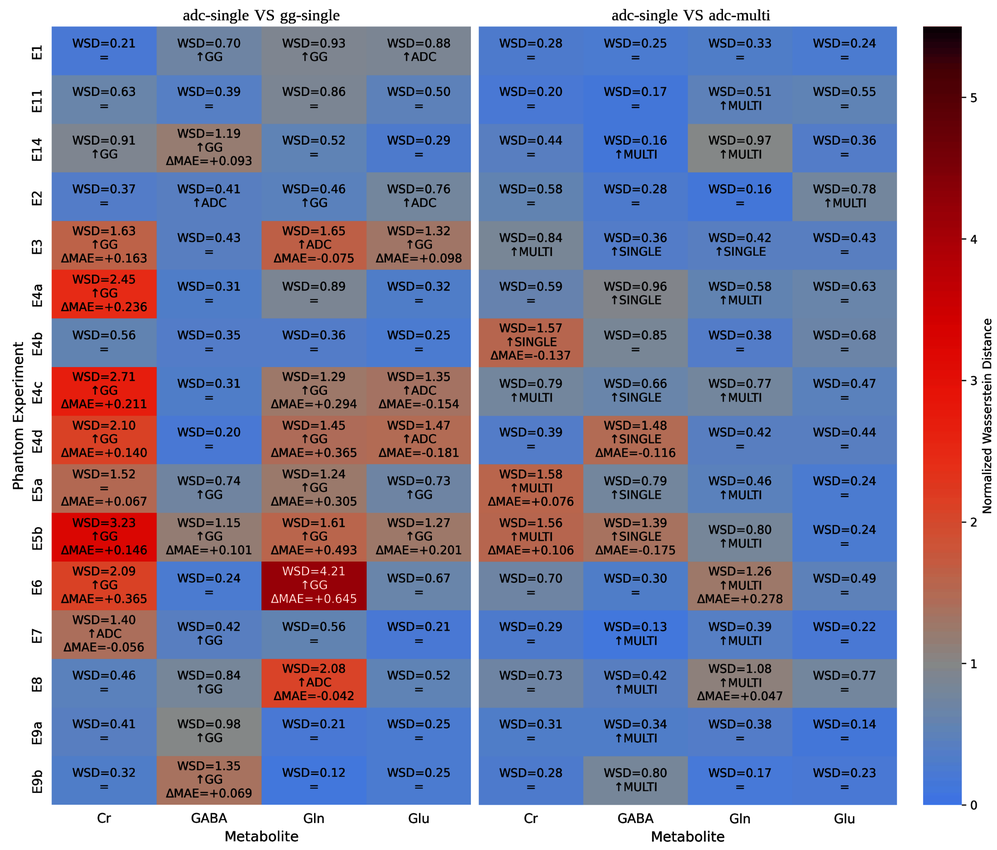

Figure 1: Comparison of quantification performance across different training datasets for different phantom sequences (E1-E14) and metabolites using the one-tailed Wilcoxon signed-rank test (WSD: Wasserstein Distance; ΔMAE: Mean Absolute Error Difference).

1. Introduction

Deep learning (DL) offers promising improvements in metabolite quantification for magnetic resonance spectroscopy (MRS) [1]. However, its performance often struggles to improve upon traditional approaches such as LCModel [2], especially in real-world settings [3]. This reliance on simulated training data stems from the **high cost of acquisition, scarcity of ground-truth data, and challenges in obtaining reliable *in vivo* measurements**. However, **models trained solely on simulations often perform poorly on phantom or *in vivo* spectra**, highlighting a generalisation gap caused by discrepancies between simulated and experimental data (Fig. 3).

This study investigates how simulation realism impacts the transferability and accuracy of **a state-of-the-art Y-shaped autoencoder model [4] (Fig. 2), designed specifically for quantification of metabolites in MEGAPRESS spectra,** when evaluated on experimental phantom data [5]. We focus on two key contributors to simulation realism:

- variability in metabolite basis sets (multi-model incorporating simulated variations in chemical shifts and J-couplings vs. single fixed model)
- the realism of added noise (estimated GG vs. simple ADC).

Our analysis illustrates the influence of these training data characteristics on DL quantification performance on phantom spectra.

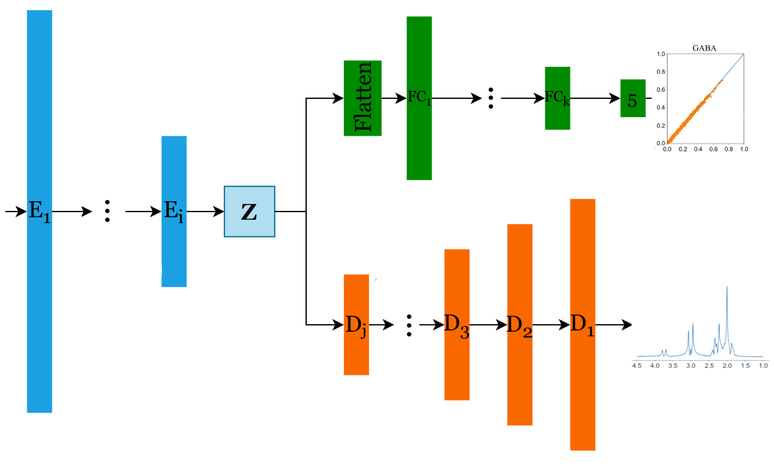

Figure 2: The Y-shaped dense autoencoder model architecture.

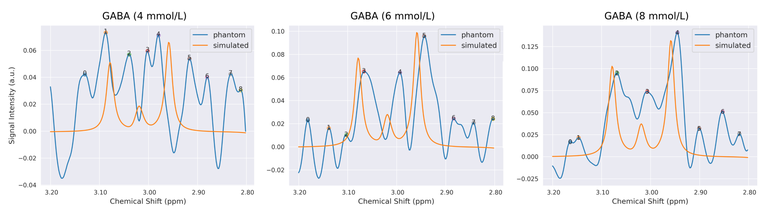

Figure 3: GABA phantom vs. simulated spectra at 4, 6, and 8 mmol/L. Each panel shows a different concentration

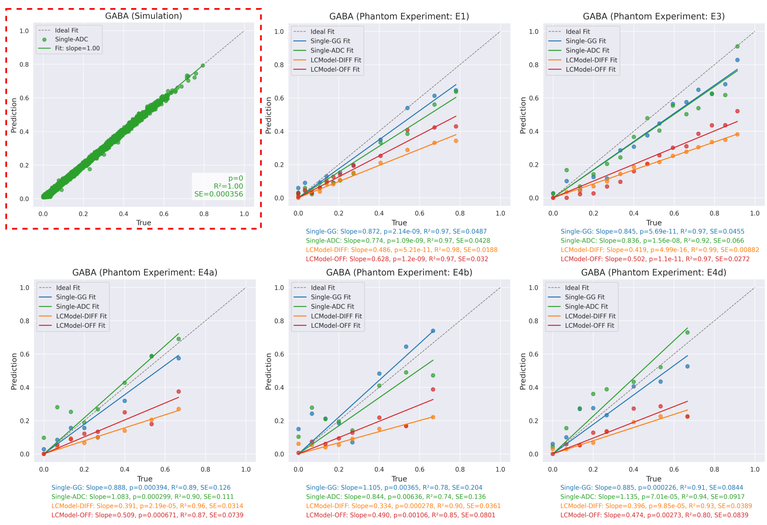

Figure 4: Performance comparison on phantom spectra between models trained on simulation (`single-ADC`, `single-GG`) datasets (red dashed box indicates performance on simulated validation data). LCModel performance is included for comparison.

2. Methods

Spectral Simulation. Basis sets were generated using FID-A [6] for MEGAPRESS edit-on/off spectra. Simulated spectra were constructed by linearly combining basis spectra, weighted by relative concentrations sampled uniformly in $[0, 1]$ using Sobol low-discrepancy sequences (10^5 spectra per dataset). Variations included:

- **Noise Models:**
  - `ADC`: Simple additive Gaussian noise applied in the time domain, with manually tuned variance to match phantom SNR levels.
  - `GG`: Generalized Gaussian noise added in the frequency domain, parameters estimated from phantom noise using MCMC fitting [7].
- **Basis Variations:**
  - `Single`: fixed, single basis set for all spectra.
  - `Multi`: randomly selected basis set from pool generated with varied chemical shifts and J-couplings based on models from literature.

Dataset Variants. Combining basis variability and noise models yielded 4 training datasets: `single-ADC`, `single-GG`, `multi-ADC`, `multi-GG`.

Model Architecture & Training. A parameterized Y-shaped dense autoencoder [4] was used. Architecture parameters were optimized using Bayesian hyperparameter search. The model was trained separately on 4 datasets and evaluated on phantom spectra.

3. Results

Evaluation Metrics. Normalised Wasserstein Distance (WSD >1 indicates significant distributional difference) was used to assess distributional differences between models trained under different data. One-tailed Wilcoxon signed-rank tests were applied to paired absolute errors (|prediction – ground truth|) for each phantom experiment to determine directional performance differences. Heatmap annotations indicate statistically significant comparisons between models ('=' denotes no significant difference, Fig. 1).

Summary. Models trained on single-basis data with generalized Gaussian (GG) noise exhibited lower errors compared to those trained with additive Gaussian (ADC) noise, particularly for Glutamine (Gln) and Creatine (Cr) (Fig. 1). Training on multi-basis datasets did not yield significant improvements over single-basis training data. Importantly, models trained with single-basis data and realistic GG noise outperformed or matched LCModel for key metabolite: GABA (Fig. 4), demonstrating their potential.

4. Conclusion

Overall, modeling realistic GG noise had a greater impact on performance than basis variability. Future work could further enhance accuracy by refining single-basis simulations to better replicate experimental spectra.

References

[1] van de Sande, DMJ, Merkofer, JP, Amirrajab, S, *et al.*: **A review of machine learning applications for the proton MR spectroscopy workflow.** *Magnetic Resonance in Medicine* **90**(4), 1253–1270 (2023)

[2] Provencher, SW: **Estimation of metabolite concentrations from localized in vivo proton NMR spectra.** *Magnetic Resonance in Medicine* **30**(6), 672–679 (1993)

[3] Rizzo, R, Dziadosz, M, Kyathanahally, SP, *et al.*: **Quantification of MR spectra by deep learning in an idealized setting: [...].** *Magnetic Resonance in Medicine* **89**(5), 1707–1727 (2024)

[4] Ma, Z, Karakus, O, Schermer, SM, Langbein, FC: **Metabolite quantification from edited magnetic resonance spectra with deep learning.** Preprint (2024). `https://mrs.qyber.dev/paper-mrsnet-autoencoder`

[5] Shermer, SM, Jenkins, C, Chandler, M, Langbein, FC: **Magnetic resonance spectroscopy data for GABA quantification using MEGAPRESS pulse sequence.** *IEEE Data Port* (2019). DOI: `10.21227/ak1d-3s20`

[6] Near, J, Simpson, R, Jezzard, P, *et al.*: **FID-A—Advanced MR spectroscopy processing and simulation.** Github, Version 1.2 (2018). `https://github.com/CIC-methods/FID-A/tree/V1.2`

[7] Karakus, O, Kuruoğlu, EE, Altınkaya, MA: **Beyond trans-dimensional RJMCMC with a case study in impulsive data modeling.** *Signal Processing* **153**, 396–410 (2018)