RESEARCH ARTICLE

Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals

Amirmohammad Shamaei1,2 | Jana Starcukova1 | Iveta Pavlova1 | Zenon StarcukJr.1

1Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic
2Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic

Correspondence
Amirmohammad Shamaei, Magnetic Resonance Group, Institute of Scientific Instruments of the Czech Academy of Sciences, Kralovopolska 147, CZ 61264, Brno, Czech Republic.
Email: amirshamaei@isibrno.cz

Funding information
European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie, Grant/Award Number: 813120; Czech Academy of Sciences, Institute of Scientific Instruments, RVO, Grant/Award Number: 68081731

Purpose: A supervised deep learning (DL) approach for frequency and phase correction (FPC) of MRS data recently showed encouraging results, but obtaining transients with labels for supervised learning is challenging. This work investigates the feasibility and efficiency of unsupervised deep learning–based FPC.

Methods: Two novel deep learning–based FPC methods (deep learning–based Cr referencing and deep learning–based spectral registration), which use a priori physics domain knowledge, are presented. The proposed networks were trained, validated, and evaluated using simulated, phantom, and publicly accessible in vivo MEGA-edited MRS data. The performance of our proposed FPC methods was compared with other generally used FPC methods, in terms of precision and time efficiency. A new measure was proposed in this study to evaluate the FPC method performance. The ability of each of our methods to carry out FPC at varying SNR levels was evaluated. A Monte Carlo study was carried out to investigate the performance of our proposed methods.

Results: The validation using low-SNR manipulated simulated data demonstrated that the proposed methods could perform FPC comparably with other methods. The evaluation showed that the deep learning–based spectral registration over a limited frequency range method achieved the highest performance in phantom data. The applicability of the proposed method for FPC of GABA-edited in vivo MRS data was demonstrated. Our proposed networks have the potential to reduce computation time significantly.

Conclusions: The proposed physics-informed deep neural networks trained in an unsupervised manner with complex data can offer efficient FPC of large MRS data in a shorter time.

KEYWORDS
deep learning, edited MRS, frequency correction, MR spectroscopy, phase correction

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 Ústav přístrojové techniky AV ČR, v. v. i.. Magnetic Resonance in Medicine published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.
1 | INTRODUCTION

In MRS, typically more transients are acquired and averaged to increase the low SNR. However, individual transients might have different frequency and phase shifts because of hardware imperfections, physiologic processes, or other instabilities. Averaging transients without frequency and phase correction (FPC) would result in line-broadening and lineshape imperfection of the combined MRS signal. Thus, FPC should be performed for each transient before averaging. It is even more critical to use accurate FPC while using spectral-edited MRS methods to prevent artifacts caused by subtraction. Thus, FPC is a consensus-recommended and effective step in MRS signal processing.

Several FPC approaches have been developed. The FPC methods can be classified into absolute and relative methods. Absolute approaches correct each individual transient absolutely, whereas relative methods align the transients to a reference signal. A commonly used approach for FPC is to use the water peak and read the phase and frequency from it. Another approach is to fit a certain metabolite peak to a model and then estimate the frequency and phase shifts from the model. One approach that has been proposed and evolved recently is spectral registration (SR). Spectral registration fits each signal to a reference signal in the time domain through the adjustment of frequency and phase terms. Even though SR works very well for small shifts, it struggles with larger shifts and signals with low SNR. Modified versions of SR successfully address some of the mentioned problems. Most of these approaches are time-consuming for large data sets, such as high-resolution MRSI data sets, which may have thousands of spectra.

The recent success of deep learning (DL), one of the latest machine learning (ML) approaches, in a wide range of tasks, including the MR field, suggests that it could also handle FPC. Recently, DL-based solutions have been proposed for metabolite quantification in the frequency domain, detecting and removing ghosting artifacts, as well as supervised deep-learning approaches in terms of precision and time efficiency.
2 | METHODS

2.1 | Data normalization

In contrast to the application of DL in machine vision or speech recognition, where the input data can be normalized by a nonlinear transform, MRS signals must be normalized by a linear transform. In this study, each complex signal $S(t)$ was rescaled as

$$S(t)_{\text{normalized}} = \frac{S(t)}{\max_r(|S(r)|)}.$$ (1)

A similar approach could be dividing the signal by the absolute value of the first point of the signal, but it would be less generally applicable because the initial point can be influenced by the filtering processes in the receive chain, or the maximum may occur later in an echo or with coupled resonances.

2.2 | Data augmentation

It is known that the sufficient size and diversity of data are important factors for the effectiveness of most DL models. However, having rich and sufficient datasets is rare in the field of MRS and MRSI. Data augmentation is a viable option, which simulates credible data by minor alterations to data in a small existing data set. Data augmentation used in computer vision applications to reduce the generalization error of models applies to flips, translations, and rotations, which would be meaningless in spectroscopy. To generate credibly varied FID signals, we chose a set of physics-informed alterations, simulating the practical data variability:

1. Frequency shift
2. Phase shift
3. Apodization (line broadening)
4. Amplitude change
5. Adding noise
6. Adding a nuisance peak (residual water and lipids)

2.3 | Data sets

2.3.1 | Simulated data set

A simulated data set was used in silico ground-truth information for evaluating the performance of our proposed networks and for comparison with the commonly used FPC methods. The simulated data set was obtained by alternating a single MR signal acquired from a rat brain as described subsequently.

A single-voxel spectroscopy (SVS) MR in vivo signal was acquired from a rat’s right hippocampus (256 transients, voxel size $= 1.5 \times 1.5 \times 4$ mm$^3$) in a 9.4T small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany) using a PRESS sequence (spectral width $= 4400$ Hz, 2048 points, TE = 16.5 ms, TR = 2500 ms) with water and outer-volume suppression by VAPOR. The signal (further referred to as the basis signal $S_{\text{basis}}(t)$) was created after transients were corrected for $B_0$ instability due to eddy currents as well as $B_0$ drift, and averaged using Bruker proprietary software, Paravision.

All experiments were approved by the Czech Governmental Animal Care Committee, in compliance with Czech Animal Protection.

The simulated data set, containing 24,000 artificial signals, was generated from $S_{\text{basis}}(t)$ by an augmentation procedure. The basis signal was multiplied by factors drawn from a normal distribution with a mean of 1 and a SD of 0.1. Then a set of lipid nuisance peak and a set of unstable residual water nuisance peak, generated using Equation 2 (parameters are listed in Supporting Information Table S1), were added to signals, randomly and independently:

$$S(t)_{\text{nuisance peak}} = A_a e^{-d_a t} e^{-(2\pi f_a t + \varphi_a)},$$ (2)

where $i = \sqrt{-1}$; $t$ is a vector of time points; and $A_a$, $d_a$, $f_a$, and $\varphi_a$ are the amplitude, the damping factor, the precession frequency, and the phases of the nuisance peak, respectively. Signals containing the lipid peak, the unstable residual water peak, and both peaks were labeled as LC, UW, and UW & LC, respectively.

All artificial signals were further apodized by normally distributed random dampings corresponding to Lorentzian linewidths with a mean of 2 Hz and a SD of 0.2 Hz. Then, uniformly distributed artificial frequency and phase offsets in the range of $-20$ to $20$ Hz and $-90°$ to $90°$, respectively, were applied to signals. Before the normalizing step, the SNR of signals (time origin magnitude to noise SD) was set in the range of approximately 9 to 27 by introducing random complex Gaussian white noise. Signals in the simulated data set were shuffled randomly, and 90% of the data set was allocated to the training subset; 9% for the validation subset, and the remaining 1% for the test subset.

2.3.2 | Phantom data set

Phantom data were used to assess the performance of our methods in the absence of a large number of training data measured during temperature-dependent changes in the $B_0$ field.

An SVS MR signal was acquired in a phantom of known metabolite concentrations (N-acetylaspartate...
17.5 mmol/L, glutamate 26.2 mmol/L, myo-Inositol 17.5 mmol/L, creatine [CR] 12.7 mmol/L, taurine 10.1 mmol/L; 2048 transients, voxel size $= 3 \times 3 \times 3 \text{ mm}^3$) in a 9.4 T small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany), while the temperature of the phantom was altered between 35°C and 40°C and the frequency adjustment of the scanner was switched off, using PRESS sequence (spectral width $= 4400 \text{ Hz}$, 2048 points, TE $= 16.5 \text{ ms}$, TR $= 2500 \text{ ms}$) with water and outer-volume suppression by VAPOR.31

The phantom data set, containing 24000 artificial signals, was generated as follows: We selected 400 (out of 2048) of the acquired transients as basis signals, randomly; then, a subset of 60 signals was generated from each basis signal by the same augmentation procedure as used for the simulated data set except that the basis signal was not multiplied by factors, the nuisance peaks were not added, and the SNR of signals was set in the range of about 7 to 70. Finally, all subsets were stacked together to create the final training data set. The rest of the 1648 transients were used as an unseen test subset.

2.3.3 | Big GABA dataset

Data from the public repository Big GABA26,27 were used to demonstrate the applicability of the proposed method on FPC of edited in vivo signals.

We selected 48 GABA-edited MEGA-PRESS subsets (subjects) acquired on Siemens scanners from four different sites (S1, S5, S6, and S8 [all that we found public and readable]; 3T field strength; spectral width $= 4000 \text{ Hz}$, 4096 points, TE $= 68 \text{ ms}$; ON/OFF editing pulses $= 1.9/7.46 \text{ ppm}$; editing pulse duration $= 15 \text{ ms}$, TR $= 2000 \text{ ms}$; 320 averages; $30 \times 30 \times 30 \text{ mm}^3$; and medial parietal lobe voxel) because Siemens data3 had greater median within-participant SD of estimated phase offsets and larger variance of lipid contamination than data from Philips and GE, and relatively high average frequency offsets, which are undesirable conditions for conventional FPC methods.5

We allocated 40 of 48 selected in vivo subsets (15360 transients) to the training subset (12800 transients), and the rest of the subsets (2560 transients) were used as an unseen test subset.

2.4 | Deep model

2.4.1 | The DAE proposed for deep learning–based peak referencing

The DAE is a type of deep artificial neural network that is created to learn the coding of data in an unsupervised manner. The fundamental underlying concept of autoencoders is to use the input data as the target (ie, attempting to reconstruct the input data in the output layer).13 Typically, a DAE consists of two parts: an encoder and a decoder.

Figure 1 illustrates the most common architecture of a DAE. The encoder function $h = f(x)$ maps the $n$-dimensional input vector ($x \in \mathbb{R}^n$) to the $n'$-dimensional latent vector ($h \in \mathbb{R}^{n'}$), while the decoder function $\hat{x} = g(h)$ aims to reconstruct the $n$-dimensional output vector ($\hat{x} \in \mathbb{R}^n$) from the latent space representation. The mathematical expression of a DAE can be written as follows:

$$\hat{x} = g(f(x; \theta_e); \theta_d),$$

where $\theta_e$ and $\theta_d$ are the parameters set of encoder and decoder layers, respectively.

For FPC, the latent space representation must be interpretable parameters, such as frequency and phase shifts. Frequency and phase shifts can be estimated by fitting a certain metabolite peak to a model, such as a Lorentzian lineshape. Then, the frequency and phase shifts can be read from the model. To this end, we proposed a convolutional encoder/model-decoder15 architecture. Our proposed DAE has a conventional encoder consisting of a pipeline of a dropout layer32 convolutional layers,33,34 and rectified linear unit layers,34 which encode a complex input signal into a latent space, and a decoder that reconstructs a Lorentzian lineshape of a certain peak in the input signal using the latent space parameters. The proposed DAE architecture is depicted in Figure 2. Because it has been shown that using complex-valued data improves the robustness and
SHAMAEI ET AL.

FIGURE 2 Illustration of the proposed convolutional encoder/model-decoder for the deep learning–based Creatine referencing (dCrR) method. The network’s input is a complex signal in the time domain, which is fed to the encoder. The encoder consists of a dropout layer, six convolutional blocks, and an FC layer (see details in Supporting Information Table S2). A convolutional block (dashed square) consists of a one-dimensional convolution (Conv1d) layer followed by a ReLU layer. The model-decoder of the dCrR (Equation 4) reconstructs the output signal. The DAE was trained to encode the input vector in the time domain into parameters that can be used to reconstruct the output vector in the time domain. The proposed network is trained by minimizing the mean square error (MSE) between the input and the output. The input and output signals are depicted in the frequency domain for the sake of easier understanding. Abbreviations: $A_M$, amplitude; $d_M$, damping factor; $f_M$, resonance frequency; $\varphi_M$, zero-order phase of the Lorentzian lineshape.

The DAE proposed for deep learning–based SR

With a simple modification, the proposed approach could estimate the relative frequency and phase shifts by fitting each signal to a reference signal and be applicable to various MRS experiments. In other words, the SR method could be used in our proposed encoder/model-decoder network. This modification is referred to as deep learning–based spectral registration (dSR). For dSR, the efficiency of fitting MRS data, the input and output of the proposed DAE were set to be complex signals in the time domain.

Time-domain fitting of a single Lorentzian lineshape to a signal with several peaks using our proposed network is a challenging task in which the optimization algorithm aims to increase the linewidth to decrease the error. Previous studies addressed these problems by fitting a signal in the frequency domain over a limited range and including a linear baseline in their model. We found that adding a rough estimate of a baseline, obtained by apodizing the input signal with Lorentzian kernel with a large linewidth, into the reconstruction function improves our fitting. Hence, the decoder part combines a mathematical model (Lorentzian lineshape) and the input signal, $x$, apodized with the Lorentzian kernel. The mathematical expression of the decoder can be written as follows:

$$g(x; A_M, d_M, f_M, \varphi_M, t) = A_M e^{-d_M t} e^{-i(2\pi f_M t + \varphi_M)} + x(t) e^{-d_L t},$$

where $A_M$, $d_M$, $f_M$, and $\varphi_M$ are the amplitude, the damping factor, the resonance frequency, and the zero-order phase of the selected Lorentzian lineshape, respectively, and $d_L$ is the linewidth of the Lorentzian kernel. Experimentally, $d_L$ was set to 500 Hz in this study.

Training our proposed network is an unsupervised learning task that does not require ground-truth frequency and phase shifts and can be done by minimizing the differences between the original input and the consequent reconstruction. In each iteration step of training, the parameters of the encoders are adjusted according to the gradient of the loss function with respect to the given parameters of the Lorentzian lineshape ($A_M, d_M, f_M$, and $\varphi_M$).

In this study, Cr peak at 3.027 ppm was selected to be fitted by a Lorentzian line shape in the decoder. The FIDs were truncated to the initial 512 points for limiting the contribution of noise, which typically predominates in the later part of FIDs. Then, the truncated FIDs were used as inputs to the network. After training, the encoder of the proposed DAE (deep learning–based creatine referencing [dCrR]) was detached from the network and used to estimate the frequency and phase of the Cr peak in a test transient. Then the estimates were used for frequency and phase correction of the test transient. The pipeline of FPC of GABA-edited MEGA-PRESS transients is provided in Supporting Information Figure S4.

2.4.2 The DAE proposed for deep learning–based SR

With a simple modification, the proposed approach could estimate the relative frequency and phase shifts by fitting each signal to a reference signal and be applicable to various MRS experiments. In other words, the SR method could be used in our proposed encoder/model-decoder network. This modification is referred to as deep learning–based spectral registration (dSR). For dSR, the
mathematical expression of the decoder can be written as follows:

$$g(f_M, \varphi_M, t) = R(t)e^{-i(2\pi f_M t + \varphi_M)},$$

(5)

where $f_M$ and $\varphi_M$ are the frequency and the phase shifts of the input signal with respect to the reference scan $R(t)$. Thus, in the dSR, the encoder estimates two parameters (relative phase and frequency shifts) instead of estimating the four parameters of a lineshape. The proposed DAE architecture for dSR is depicted in Supporting Information Figure S5. We proposed a new ML-based algorithm for finding the reference signal in which the k-means algorithm was used to cluster magnitude-valued signals in the training set. The number of clusters was set to two: one for samples with nuisance peaks and another for samples without nuisance peaks. k-Means algorithm is an unsupervised method; thus, the clusters are not identified. Based on prior testing, samples containing nuisance peaks have a higher intensity than samples without nuisance peaks in the initial points of their mean of the samples in the cluster. Therefore, the first 10 points of each mean of samples in the cluster were averaged, and the cluster with a lower average was identified as the cluster containing samples without nuisance peaks; then, the signal with the highest SNR from this cluster was selected as the reference scan (see details in Supporting Information Text S2 and Figure S6).

### 2.4.3 Training the proposed DAEs over a limited frequency range

The MRS signals may have unstable frequency components. Because all frequency components are present at all time points, unstable frequency components may bring errors into our proposed method. To avert this situation, the proposed architectures (dCrR and dSR) were trained over a limited frequency range (2.5 to 3.5 ppm) (referred to as dCrRF [deep learning-based creatine referencing over a limited frequency range] and dSRF [deep learning-based spectral registration over a limited frequency range]). Note that the input of the encoder and the output of the model-decoder were still in the time domain, and the discrepancy between fast Fourier transformation of the input and output was calculated over a limited frequency range.

### 2.5 Implementation details and training

All steps were run on a computer with a dual EPYC 7742 (2 × 64 cores) processor and one graphics processing unit (NVIDIA A100 40 GB). Moreover, all steps were run on Google Colaboratory (free-to-use hardware). The DAE was implemented in Python with the help of the Pytorch lightning interface. The architecture of the network and training parameters were optimized using the Bayesian Optimization HyperBand algorithm with the help of the Tune framework. The details of the optimization are given in Supporting Information Text S1. All training was performed using the mean-squared error loss and an Adam optimizer with a batch size of 16, a learning rate of $4 \times 10^{-5}$, and 150 epochs. The training progress for the simulated data set is provided in Supporting Information Figures S2 and S3.

An early-stopping strategy was performed by monitoring the average error of the validation subset at the end of every epoch and stopping the training when no improvement was observed in 10 epochs. The SR and the SRF methods were tested using the FID-A toolbox, the CrR and the CrRF methods with the Gannet toolbox, and the frequency domain correlation methods with a limited frequency range methods, and previous studies (Tapper et al and Ma et al) with our in-house code.

All proposed methods (dCrR, dCrRF, dSR, and dSRF) were trained, validated, and tested using the simulated and the phantom data set, and the dCrR method was trained and tested using the big GABA data sets.

### 2.6 Statistics and quality evaluation

#### 2.6.1 Performance analysis

For the simulated data set, in which the true shifts from the basis signal were known, the error was defined as the difference between the estimated and the true shifts. The accuracy and precision of an FPC method were established as the average and the SD of the error, respectively. In addition, the performance of the dCrR method trained with the simulated data set was investigated beyond the trained range of frequency and phase (−40 Hz and −180°, respectively).

For the phantom and the Big GABA data set, in which true shifts were not known, the quality of alignment was measured by comparing the similarity index (SI), which is the sum of all elements of a similarity matrix. Each element of the similarity matrix is the normalized scalar product (the equation as implemented is provided in Supporting Information Equation S1) for each pair of spectra (fast Fourier transformation of a FID) over a limited frequency range (from 2.5 to 3.5 ppm). Note that higher SI denotes superior performance.
2.6.2 Performance against noise

The stability of the dCrR, dSR, and SR method against noise was evaluated. A set of transients was generated using the following procedure. First, a frequency shift of 5 Hz and a phase shift of 45° were added to \( S_{\text{basis}}(t) \). Second, 20 realizations of a random complex Gaussian noise with a linearly increasing SD were introduced to the shifted signal such that SNR was in the range of approximately 8 to 110. The frequency and phase shifts of the generated set were estimated using the networks trained with the simulated data set. The FPC performance was evaluated as a function of SNR.

2.6.3 Monte Carlo analysis

Monte Carlo studies were carried out to investigate the performance of the dCrR, dSR, and SR methods. A set of transients was generated using the following procedure. First, a frequency shift of 5 Hz and a phase shift of 45° were added to \( S_{\text{basis}}(t) \). Second, 256 realizations of a random complex Gaussian noise with the same SD were introduced to the shifted signal such that SNR was approximately 15. The frequency and phase shifts of the generated set were estimated using the networks trained with the simulated data set. The FPC performance of dCrR, dSR, and SR methods was compared.

3 RESULTS

The training and processing time of each proposed network for the simulated data set is listed in Table 1. Approximately, frequency and phase estimation of one transient requires approximately 3 ms. The evaluation of the performance of different methods (precision, SI, linewidth, the processing time of FPC, and training time)

| TABLE 1 | Comparison of our proposed method with existing commonly used FPC methods for the simulated data set |
|-------------------------------|-----|-----------|-----|-----|-----|-----|-----|-----|-----|
|                              | dSR | dSRF      | SR  | SRF | dCr | dCrF | Cr  | Corr| CorrF|
| Precision of frequency estimation (Hz) | All samples | 1.02 | 0.73 | 1.57 | 11.46 | 0.94 | 0.95 | 7.92 | 10.70 | 11.33 | **0.71** | 0.99 |
|                              | Samples without nuisance peaks | 1.00 | 0.69 | 1.01 | 11.02 | 0.93 | 0.90 | 9.62 | 9.98 | 9.89 | **0.56** | 0.85 |
|                              | Samples with nuisance peaks | 0.97 | **0.73** | 1.57 | 11.68 | 0.94 | 0.97 | 6.51 | 11.07 | 11.96 | 0.78 | 1.06 |
| Precision of phase estimation (°) | All samples | 5.87 | 4.89 | 7.44 | 15.92 | 6.62 | 6.79 | 32.34 | 16.87 | 17.46 | 6.78 | **4.29** |
|                              | Samples without nuisance peaks | 4.16 | 3.34 | **2.61** | 16.75 | 6.89 | 7.00 | 38.80 | 16.62 | 16.46 | 4.64 | 4.24 |
|                              | Samples with nuisance peaks | 5.75 | 5.52 | 8.50 | 15.34 | 6.35 | 6.52 | 26.68 | 17.02 | 18.07 | 5.89 | **4.32** |
| SI All samples | 0.44 | **0.44** | 0.44 | 0.32 | 0.42 | 0.42 | 0.36 | 0.34 | 0.32 | 0.44 | 0.44 |
| Cr linewidth (Hz) Averaged signal | 10.38 | 10.22 | 10.77 | 15.34 | **10.14** | 10.17 | 10.71 | 19.14 | 20.55 | 10.45 | 10.53 |
| Training time (min) CPU All samples | **127** | 138 | — | — | 328 | 354 | — | — | — | — |
| Training time (min) GPU All samples | **2.26** | 7.8 | — | — | 8.4 | 7.3 | — | — | — | 20.8 | 29.64 |
| Training time (min) Google Colaboratory All samples | **4.53** | 10.6 | — | — | 11.1 | 11.2 | — | — | — | — |
| Processing time (ms) CPU Per signal | **3** | **3** | 57 | 71 | **3** | **3** | 53 | 33 | 29 | 15 | 15 |
| Processing time (ms) GPU Per signal | **0.1** | **0.1** | — | — | **0.2** | **0.2** | — | — | — | 0.6 | 0.6 |

Note: Bold text indicates the best performance in each metric. The SI of the test subset without FPC was 0.14. Precision: SD of the difference between the estimated and the true shift.

Abbreviations: Corr, frequency domain correlation; CorrF, frequency domain correlation over a limited frequency range; CPU, central processing unit; Cr linewidth, linewidth of Cr peak at 3 ppm; CrR, creatine referencing; GPU, graphics processing unit; SI, similarity index; SRF, SR over a limited frequency range.
for DAE) for the test signals of the simulated data set and of the phantom are summarized in Tables 1 and 2, respectively.

The Bland–Altman plots in Figure 3A,B show the accuracy and biases in the estimation of frequency and phase shifts using dCrR, dCrRF, and dCr. The highest precision (0.94 Hz and 6.62°) among the absolute FPC methods was achieved with the proposed dCrR. Although dCrRF performed similarly to dCrR, CrR showed the lowest performance (7.92 Hz and 32.34°). All of the proposed methods performed well in spectra with and without the simulated nuisance peaks. The agreement, estimated by R² value, was high for dCrR (R² frequency = 0.99 and R² phase = 0.99) and dCrRF (R² frequency = 0.99 and R² phase = 0.99) and moderate for CrR (R² frequency = 0.61 and R² phase = 0.75). Figure 3C,D illustrates the results of dCrR on unseen test signals in which offsets are beyond trained bound (−40 to 40 Hz and −180° to 180°). The network showed poor performance beyond its trained bound (precisions for frequency and phase estimation were 8.69 Hz and 60.04°, respectively). Figure 3E,F shows the spectra and the similarity matrix heatmaps obtained before and after the FPC tested, respectively. The dCrR method increased the SI in the visualized test signals from 0.14 to 0.42.

Figure 4 shows a pairwise correlational comparison of the relative FPC methods (our proposed dSR and dSRF with SR and SRF). The R² value of each method pair and the true value of shifts are reported in the corresponding axes. The R² indicates that there is a high degree of agreement between the frequency and phase estimations between methods except for SRF. The agreement between the estimations of methods and true values (R²) was high for dSR, dSRF, and SR and was low for SRF (R² frequency = 0.03 and R² phase = 0.94).

Table 2: Comparison of our proposed method with existing commonly used FPC methods for the phantom data set

| Method | SI  | Cr linewidth (Hz) |
|--------|-----|--------------------|
| dCrR  | 0.66 | 6.02               |
| dCrRF | 0.66 | **6.00**           |
| dSR   | 0.60 | 6.80               |
| dSRF  | 0.67 | 6.03               |
| SR    | 0.58 | 6.75               |
| SRF   | 0.60 | 6.46               |
| CrR   | 0.58 | 7.03               |

Note: Bold text indicates the best performance in each metric. The SI of the test subset without FPC was 0.50.

Figure 5 shows spectra from the test signals of the phantom data set and the corresponding heatmaps of similarity matrices before and after correction using the dCrR method. The dCrR method achieved the highest performance and increased the similarity among transients from 0.50 to 0.66, and decreased the linewidth of Cr peak at 3 ppm in the averaged spectrum by 1.5 Hz (Table 2). We observed that changing temperature during measurement altered the amplitude and linewidth of peaks, which resulted in a decreased similarity index. The dSR method failed to increase the SI and reduce the Cr linewidth because dSR tends to correct frequency and phase shifts of large frequency components, such as the residual water. However, dSRF overcame this issue and demonstrated superior performance by limiting the frequency range.

The link between the absolute error in estimates and the SNR is shown in Figure 6A,B using a scatter plot. The comparison of the proposed dSR and dCrR with SR methods is presented. In low SNR, dSR and SR methods outperformed dCrR in terms of the precision of phase shift estimation, but dCrR demonstrated more resilience in terms of the precision of frequency estimation.

Figure 6C shows a comparison of the dSR, dCrR, and SR methods in the Monte Carlo analysis using a scatter-plot visualization of the joint distribution of frequency and phase. For the simulated data set, the mean error of dSR (1.36 ± 0.69 Hz and 0.51 ± 2.33°) and SR (−0.67 ± 0.81 and 3.15 ± 2.53°) showed similar performance, whereas dCrR performed less precise phase-shift estimation (4.69 ± 0.77 Hz and −41.27° ± 5.827°). The true values of the frequency and the phase shifts were 5 Hz and 45°.

Figure 7 illustrates an unseen test subset of a GABA-edited in vivo data set (site = 1, subject = 3; 160 edited transients; 160 unedited transients) and a heatmap of their similarity matrix before and after FPC by the dCrR method. Our method increased the SI in the visualized test subset from 0.80 and 0.82 to 0.92 for edited (ON) and unedited (OFF) spectra, respectively. In all test subsets, the SI was increased from 0.88 ± 0.05 to 0.93 ± 0.02.

Figure 8 shows the comparison of the results of dCrR-based and SR-based correction of ON and OFF transients of the test subsets. On average, dCrR performed better than SR, as indicated by the mean SI (the mean SI was increased from 0.88 ± 0.05 [ON: 0.89 ± 0.04, OFF: 0.87 ± 0.05] to 0.93 ± 0.02 [ON: 0.93 ± 0.02, OFF: 0.93 ± 0.02] and 0.90 ± 0.04 [ON: 0.91 ± 0.03 OFF: 0.89 ± 0.04] by dCrR-based correction and SR-based correction, respectively).
FIGURE 3  A, B, Testing results of the simulated data set. A, B, Plots of the error in deep learning (DL)–estimated phase (A) and frequency (B) shifts against the actual shifts, respectively. The $R^2$ values of the method types are color-coded. C, D, Testing the dCrR method beyond the trained range of phase and frequency (−180° to 180° and −40 to 40 Hz, respectively). Plots of the error in DL-estimated phase (C) and frequency (D) shifts against the actual shifts, respectively, using dCrR. E, Uncorrected spectra and their similarity matrix. F, The same spectra after frequency and phase correction (FPC) using dCrR and their similarity matrix. Dark blue and green spectra show the average uncorrected and corrected spectra, respectively. Abbreviations: CrR, creatine referencing; dCrRF, dCrR over a limited frequency range; Free, without any nuisance peak; LC, with lipid peak; UW, with unstable water peak.

4 | DISCUSSION

In this study, the combination of DL and mathematical modeling was demonstrated to be able to provide FPC in simulated, phantom, and in vivo MRS data. We compared our method with five previously published methods: SR, SRF, frequency domain correlation, frequency domain correlation over a limited frequency range, and CrR, as well
FIGURE 4  Pairwise comparison of the results of the spectral registration (SR)–based FPC methods for the test subset of the simulated dataset. The upper triangular shows the correlations of frequency estimations, and the lower triangular shows the correlations of phase estimations. The fitted lines represent the linear regression model and the 95% confidence interval. The type of simulated nuisance peak is color-coded. The total $R^2$ values are calculated along the corresponding axes. The dashed gray lines are identity lines. Note that the reference signals of relative FPC methods were different, which resulted in shifting their scatter plot from the identity line. Abbreviations: dSR, deep learning–based SR; dSRF, dSR over a limited frequency range; SRF, SR over a limited frequency range as supervised DL-based methods. We evaluated the ability of each of these methods to estimate the frequency and phase shifts in the simulated MRS data set with known shifts and at varying SNR levels. Our results (Table 1) indicated an improvement in performance in terms of precision and the SI, the latter of which is a new measure proposed in this study for evaluating the FPC performance. Additionally, we compared our results with those obtained using other methodologies in terms of the linewidth of the Cr peak of the averaged signal obtained by summing corrected transients, and we discovered that the unsupervised DL-based methodology performed comparably to others.

We found that the traditional FPC over a limited frequency range performed poorly in spectra with low SNR (Table 1), whereas our proposed FPC methods over a limited frequency range (dSRF and dCrRF) can perform equally (Table 1) to our proposed
methods operating in the time domain (dCrR and dSR). Our methods also performed well in a Monte Carlo study, in which the phase-shift and frequency-shift estimate precisions were found to be very good and in general accordance with the SR method (Figure 6C).

While the performance of conventional SR-based FPC methods can depend on the presence of nuisance peaks...
An example of FPC using dCrR for a test set in the GABA-edited in vivo data set. Unedited spectra (A) and their similarity matrix (B) before FPC. Edited spectra (C) and their similarity matrix (D) before FPC. Unedited spectra (E) and their similarity matrix (F) after FPC. Edited spectra (G) and their similarity matrix (H) after FPC. (I) Average uncorrected spectra (blue, unedited; red, edited) and their difference (dark green). (J) Average corrected spectra using dCrR (blue, unedited; red, edited) and their difference (dark green)

(Table 1 and Refs 3, 7), our result demonstrated that our methods functioned effectively in the simulated data set regardless of the presence of nuisance peaks (Table 1).

Our proposed method can be trained in a few minutes (Table 1) because of using a one-dimensional signal as the input and a relatively tiny network. In the case of large MRS data sets or repeated measurements with the same conditions, the training time will be compensated by the FPC processing time, which is significantly shorter than in methods based on nonlinear least squares (Table 1). Moreover, the training time can be reduced by using well-established methods such as few-shot transfer learning, in which a network pretrained on large simulated or in vivo data sets can be trained in a few iterations using a few samples.

When the SNR of the test signals was lowered, the performance of our proposed methods (dCrR and dSR) and SR was reduced. When the SNR was decreased below the SNR of the training set, the performance deteriorated even further, as the signal is dominated by noise. We observed that in low-SNR signals, the dSR and SR methods performed better in phase-shift estimation, whereas dCrR worked much better in frequency estimation (Figure 6).
Regardless of performance, the encoder/model-decoder provides a unique flexibility advantage. It leverages the underlying prior knowledge, which can be beneficial for estimating frequency and phase, independent of the kind of MRS data. Therefore, this approach may be used immediately to almost any sort of MRS data with little or no modification.

Contrary to the previous applications\textsuperscript{7,22} of DL in FPC that used a supervised way utilizing simulated data, our proposed network was trained in an unsupervised way. This is advantageous, as most of the MRS data are unlabeled, and simulated data sets may not accurately reflect all in vivo circumstances, such as macromolecules and artifacts. Additionally, a single network was trained in this work to deliver both frequency and phase estimations by including prior knowledge into the decoder, whereas earlier work failed to train a single network.

It has been demonstrated that traditional FPC methods can benefit from additional information in a data set,\textsuperscript{10} addressing the problem of selecting a reference signal in SR by using a weighted average reference determined by mutual information in data. The encoder/model-decoder may assist in extracting patterns and information from data by introducing more complex models.

Overfitting is a common pitfall in DL.\textsuperscript{13} We implemented a dropout layer in the input to remove a part of the input randomly in every training step, which is a computationally inexpensive and very effective regularization strategy for decreasing overfitting and increasing the generalization of the network.\textsuperscript{32}

Training neural networks for regression problems necessitates a well-calibrated estimation. The results revealed a significant linear link between the true and estimated values (Figures 3 and 4), indicating a well-calibrated estimation, although additional examination of the results is necessary.

Along with demonstrating the performance of the proposed approach on simulated and phantom data, the method was used to carry out FPC and enhance the similarity of signals in publicly accessible GABA-edited in vivo MRS data. It should be emphasized that the proposed network was fed with both edited and unedited signals and trained simultaneously. The result shows the same performance for the edited and the unedited input.

In general, DL-based techniques are restricted in terms of generalizability,\textsuperscript{13} especially DL algorithms in the MR domain due to the fact that their training data might be confined to a single scanner, a single sequence, and/or a single vendor.\textsuperscript{45} In this study, only phantom data and in vivo data sets gathered from four distinct locations utilizing a single vendor and a single sequence were used to demonstrate the applicability of our method. A crucial step toward the generalizability and clinical use of our method is training and testing using multicenter and multivendor data.

One caveat in this study is that comparing processing time per transient between algorithms might not be widely valid, as it might be affected by the parameters and conditions of algorithms. Using FLOPS (floating-point operations per second) to assess the computing cost\textsuperscript{46} can help for a better comparison.

In the present study, we focused primarily on the validation of our proposed methods using in silico ground-truth knowledge and showed its application in GABA-edited in vivo MRS data. We are aware that further evaluation and comparison with a more robust method\textsuperscript{10} are needed. In addition, Cr referencing-based approaches, such as dCrR, are not the optimal way for GABA-edited
in vivo MRS data alignment, and dSR or a more sophisticated DL-based method should be investigated in future researches.

5 | CONCLUSIONS

In general, our proposed time-domain FPC method, which is based on DL networks trained in an unsupervised way with complex data, may yield results comparable to previous FPC methods. The proposed approaches can perform absolute and relative FPC on extensively manipulated data in a shorter amount of time once the network is trained. Thus, our proposed approach could aid in the acceleration of analyzing large MRS data sets. Further study is needed to evaluate the generalizability of the proposed methods for multivendor data.

ACKNOWLEDGMENT

The authors thank Radim Kořínek, Ph.D. (Czech Academy of Sciences, Institute of Scientific Instruments, Czech Republic), for his valuable technical support.

FUNDING INFORMATION

This work is part of the project that has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 813120 (INSPIRE-MED) and has been also supported with the institutional support RVO:68081731 - Czech Academy of Sciences, Institute of Scientific Instruments.

DATA AVAILABILITY STATEMENT

The source code is freely available at https://github.com/isi-nmr/DeepFPC. For questions, please contact the authors. For testing the frequency domain correlation and frequency domain correlation over a limited frequency range, we developed a MATLAB script, which is publicly accessible at https://github.com/isi-nmr/Frequency-and-Phase-Correction-of-MRS-signals-Using-Cross-Correlation.

ORCID

Amirmohammad Shamaei https://orcid.org/0000-0001-8342-3284
Jana Starcukova https://orcid.org/0000-0003-0337-7893
Iveta Pavlova https://orcid.org/0000-0003-0961-8272
Zenon StarcukJr. https://orcid.org/0000-0002-1218-0585

TWITTER

Amirmohammad Shamaei @AmirLand6

REFERENCES

1. de Graaf RA. In vivo NMR spectroscopy. Vivo NMR Spectrosc. 2019;560:375-437.
2. Hui SCN, Mikkelsen M, Zöllner HJ, et al. Frequency drift in MR spectroscopy at 3 T. Neuroimage. 2021;241:118430.
3. Near J, Edden R, Evans CJ, Harris A, Jezzard P. Frequency and phase drift correction of magnetic resonance spectroscopy data by spectral registration in the time domain. Magn Reson Med. 2015;50:44-50.
4. Near J, Harris AD, Juchem C, et al. Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts’ consensus recommendations. NMR Biomed. 2021;34:e4257.
5. Wiegars EC, Philips BWJ, Heerschap A, van der Graaf M. Automatic frequency and phase alignment of in vivo J-difference-edited MR spectra by frequency domain correlation. MAGMA. 2017;30:537-544.
6. Keating B, Deng W, Roddey JC, et al. Prospective motion correction for single-voxel 1H MR spectroscopy. Magn Reson Med. 2010;64:672-679.
7. Tapper S, Mikkelsen M, Dewey BE, et al. Frequency and phase correction of J-difference edited MR spectra using deep learning. Magn Reson Med. 2021;85:1755-1765.
8. Helms G, Piringer A. Restoration of motion-related signal loss and line-shape deterioration of proton MR spectra using the residual water as intrinsic reference. Magn Reson Med. 2001;46:395-400.
9. Wilson M. Robust retrospective frequency and phase correction for single-voxel MR spectroscopy. Magn Reson Med. 2019;81:2878-2886.
10. Mikkelsen M, Tapper S, Near J, Mostofsky SH, Edden RAE. Correcting frequency and phase offsets in MRS data using robust spectral registration. NMR Biomed. 2020;33:e4368.
11. Ernst T, Li J. A novel phase and frequency navigator for proton magnetic resonance spectroscopy using water-suppression cycling. Magn Reson Med. 2011;65:13-17.
12. Waddell KW, Avison MJ, Joers JM, Gore JC. A practical guide to robust detection of GABA in human brain by J-difference spectroscopy at 3 T using a standard volume coil. Magn Reson Imaging. 2007;25:1032-1038.
13. Goodfellow I, Bengio Y. Courville A. Deep Learning. MIT Press; 2016.
14. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29:102–27.
15. Gurbani SS, Sheriff S, Maudsley AA, Shim H, Cooper LAD. Incorporation of a spectral model in a convolutional neural network for accelerated spectral fitting. Magn Reson Med. 2019;81:3346-3357.
16. Lee HH. Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. Magn Reson Med. 2019;82:33-48.
17. Kreis R, Kyathanahally SP. Deep learning approaches for detection and removal of ghosting artifacts in MR spectroscopy. Magn Reson Med. 2018;863:851-863.
18. Lee H, Lee HH, Kim H. Reconstruction of spectra from truncated free induction decays by deep learning in proton
magnetic resonance spectroscopy. *Magn Reson Med.* 2020;84:559-568.

19. Klukowski P, Augoff M, Zierba M, Drwal M, Gonzalez A, Walczak MJ. NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics.* 2018;34:2590-2597.

20. Iqbal Z, Nguyen D, Hangel G, Motyka S, Bogner W, Jiang S. Super-resolution 1H magnetic resonance spectroscopic imaging utilizing deep learning. *Front Oncol.* 2019;9:1-13.

21. Gurbani SS, Schreibmann E, Maudsley AA, et al. A convolutional neural network to filter artifacts in spectroscopic MRI. *Magn Reson Med.* 2018;80:1765-1775.

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

23. Zhang Z, Zha H. Linear low-rank approximation and non-linear dimensionality reduction. *Sci China, Ser A Math.* 2004;47:908-920.

24. Melas-Kyriazi L. The Mathematical Foundations of Manifold Learning. arXiv:2011.03107, 2020.

25. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006;313:504-507.

26. Mikkelsen M, Barker PB, Bhattacharyya PK, et al. Big GABA: edited MR spectroscopy at 24 research sites. *Neuroimage.* 2017;159:32-45.

27. Mikkelsen M, Rimbault DL, Barker PB, et al. Big GABA II: water-referenced edited MR spectroscopy at 25 research sites. *Neuroimage.* 2019;191:537-548.

28. Edden RAE, Puts NAJ, Harris AD, Barker PB, Evans CJ. Gannet: a batch-processing tool for the quantitative analysis of gamma-aminobutyric acid-edited MR spectroscopy spectra. *J Magn Reson Imaging.* 2014;40:1445-1452.

29. Kreis R, Cudalbu C, De GRA, et al. Terminology and concepts for the characterization of in vivo MR spectroscopy methods and MR spectra: background and experts’ consensus recommendations. *NMR Biomed.* 2021;34:e4347.

30. Lam F, Li Y, Peng X. Constrained magnetic resonance spectroscopic imaging by learning nonlinear low-dimensional models. *IEEE Trans Med Imaging.* 2020;39:545-555.

31. Tkáč I, Starčuk Z, Choi I-Y, Gruetter R. In vivo 1 H NMR spectroscopy of rat brain at 1 ms echo time. *Magn Reson Med.* 1999;41:649-656.

32. Srivastava N, Hinton G, Krizhevsky A, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929-1958.

33. Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern.* 1980;36:193–202.

34. Schmidhuber, J. (2014). Deep learning in neural networks: An overview. * Neural Networks, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

35. Sokolenko S, Jézéquel T, Hajjar G, Farjon J, Akoka S, Giraud Peut. Robust 1D NMR lineshape fitting using real and imaginary data in the frequency domain. *J Magn Reson.* 2019;298:91-100.

36. Brunton SL, Kutz IN. *Data-driven science and engineering: Cambridge University Press; 2019.

37. Bisong E. Google Colaboratory. New York: Cambridge University Press; 2019.

38. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst.* 2019;32:1-12.

39. Falcon W. PyTorch lightning. GitHub. https://github.com/PyTorchLightning/pytorch-lightning.

40. Falkner S, Klein A, Hutter F. BOHB: Robust and Efficient Hyper-parameter Optimization at Scale. In: 35th International Conference on Machine Learning, PMLR; 2018:1437-1446. Accessed March 30, 2019.

41. Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stocia I. Tune: a research platform for distributed model selection and training. 2018. https://arxiv.org/abs/1807.05118v1. Accessed January 25, 2022.

42. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: Proceedings of the Third International Conference on Learning, San Diego, California, 2015.

43. Simpson R, Devenyi GA, Jezzard P, Hennessy TJ, Near J. Advanced processing and simulation of MRS data using the FID appliance (FID-A)—an open source MATLAB-based toolkit. *Magn Reson Med.* 2017;77:23-33.

44. Sun Q, Liu Y, Chua T-S, Schiele B. Meta-transfer learning for few-shot learning. https://github.com/y2l/meta-transfer-learning-tensorflow. Accessed August 15, 2022.

45. Grövik E, Yi D, Iv M, et al. Handling missing MRI sequences in deep learning segmentation of brain metastases: a multicenter study. *npj Digit Med.* 2021;4:1-7.

46. Liu Z, Mao H, Wu C-Y, et al. A ConvNet for the 2020s. 2022. https://arxiv.org/abs/2201.03545v1. Accessed February 11, 2022.

47. Evans CJ, Puts NAJ, Robson SE, et al. Subtraction artifacts and frequency (Mis-)alignment in J-difference GABA editing. *J Magn Reson Imaging.* 2013;38:970-975.

**SUPPORTING INFORMATION**

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

**Figure S1** Bin-based visualization of the result of the Monte Carlo analysis of the simulated data set

**Figure S2** Online monitoring of precision of a set of transients in the validation subset during training

**Figure S3** Validation loss (mean square error [MSE]) versus training steps

**Figure S4** Process flow of one-shot frequency and phase correction of J-difference-edited MR spectra using a single deep neural network

**Figure S5** Illustration of the proposed convolutional encoder–model decoder for deep learning–based spectral registration (dSR) method

**Figure S6** Reference scan selected by our proposed machine learning–based algorithm for the deep spectral registration method

**Table S1** Parameters for generating the simulated nuisance peak

**Table S2** Summary of the encoder’s network
Text S1 Bayesian hyper-parameterization

Text S2 A novel machine learning–based algorithm for finding the reference scan for the dSR method

Equation S1 Equation for deriving each element of the similarity matrix

How to cite this article: Shamaei A, Starcukova J, Pavlova I, Starcuk Z. Model-informed unsupervised deep learning approaches to frequency and phase correction of MRS signals. Magn Reson Med. 2023;89:1221-1236. doi: 10.1002/mrm.29498