Journal of Magnetic Resonance 336 (2022) 107151

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

Deep regression with ensembles enables fast, first-order shimming in low-field NMR

Moritz Becker, Mazin Jouda, Anastasiya Kolchinskaya, Jan G. Korvink *

Karlsruhe Institute of Technology (KIT), Institute of Microstructure Technology, Karlsruhe 76131, Germany

ARTICLE INFO

Article history:
Received 3 November 2021
Revised 21 January 2022
Accepted 24 January 2022
Available online 9 February 2022

Keywords:
Deep learning
Shimming database (ShimDB)
Nuclear magnetic resonance
Automated magnet shimming

ABSTRACT

Shimming in the context of nuclear magnetic resonance aims to achieve a uniform magnetic field distribution, as perfect as possible, and is crucial for useful spectroscopy and imaging. Currently, shimming precedes most acquisition procedures in the laboratory, and this mostly semi-automatic procedure often needs to be repeated, which can be cumbersome and time-consuming. The paper investigates the feasibility of completely automating and accelerating the shimming procedure by applying deep learning (DL). We show that DL can relate measured spectral shape to shim current specifications and thus rapidly predict three shim currents simultaneously, given only four input spectra. Due to the lack of accessible data for developing shimming algorithms, we also introduce a database that served as our DL training set, and allows inference of changes to ^{1}H NMR signals depending on shim offsets. *In situ* experiments of deep regression with ensembles demonstrate a high success rate in spectral quality improvement for random shim distortions over different neural architectures and chemical substances. This paper presents a proof-of-concept that machine learning can simplify and accelerate the shimming problem, either as a stand-alone method, or in combination with traditional shimming methods. Our database and code are publicly available.

1. Introduction

In recent decades, deep learning (DL) [1] has shown unprecedented achievements in various industrial and scientific fields. Although areas of research such as computer vision or natural language processing have become commonplace, many fields of science have only recently begun to exploit the vast possibilities that DL can provide. One such area is nuclear magnetic resonance (NMR) spectroscopy, a non-destructive technique widely used in chemistry, physics and medicine to study the properties of liquid or solid samples. The most notable contributions from DL currently employed in NMR include, but are not limited to, methods for reconstruction of non-uniformly sampled (NUS) spectra [2] and truncated free induction decays (FIDs) [3], chemical shift prediction [4], or denoising and segmentation of magnetic resonance images [5,6]. Other possible applications of DL for solving challenges in NMR are also discussed and suggested in the community [7].

Most of these approaches and suggestions are being applied at the post-processing stages of NMR spectroscopy, taking the hardware setup for the NMR measurement as granted. In contrast, we propose to optimize the preparation preceding an NMR measurement with deep learning methods.

One crucial parameter of the NMR setup, which requires the most careful and precise tuning for subsequent successful measurements, is the homogeneity of the magnetic field. The strength of the magnetic field B_0 directly influences the precession frequency f of each stationary spin with its gyromagnetic ratio γ, as described by the well-known Larmor equation $2\pi f = -\gamma B_0$. For a perfectly uniform field and a chemically homogeneous sample, the observable signal in the frequency domain yields a single Lorentzian peak centered at f. However, in an inhomogeneous magnetic field B_0, the spins experience different local fields due to various effects (e.g. susceptibility differences between material inside the field, manufacturing inaccuracies of the coil, or intramolecular shielding effects in the sample itself), and thus possess frequencies shifted around the central Larmor frequency. This means that inhomogeneities broaden the line shapes of the spectrum, decrease their amplitude and consequently decrease the signal-to-noise ratio (SNR). Furthermore, the non-bijective mapping of the FID from a three-dimensional volume to a one-dimensional spectrum introduces ambiguities, i.e., the connection between the location of each distortion and its direct impact on

* Corresponding author.
E-mail address: jan.korvink@kit.edu (J.G. Korvink).

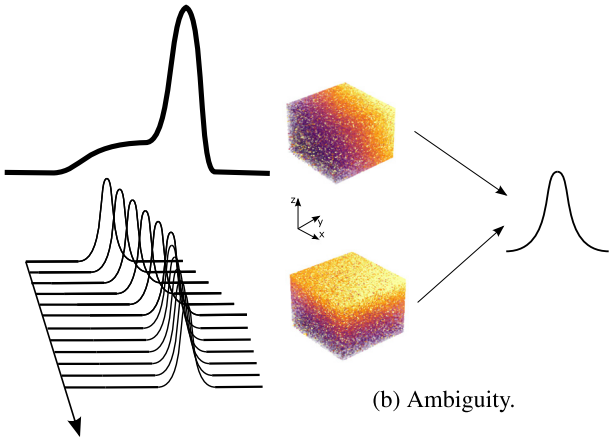
https://doi.org/10.1016/j.jmr.2022.107151
1090-7807/

the spectrum is lost during the acquisition process (Fig. 1). Especially with recent efforts in miniaturizing NMR technologies [8], signal sensitivity increases, but field inhomogeneities remain hard to eliminate.

To overcome distortions in the measured spectrum caused by a magnetic field inhomogeneity, a method referred to as "shimming" is used. Modern shimming is best described as the procedure of superimposing a secondary correction magnetic field by adjusting the currents in a finite set of field-orthogonal coils (the so-called shim coils) to correct for inhomogeneities in the magnetic field B_0. Modern NMR spectrometers usually require the magnetic field to be uniform in the range of parts per billion (ppb).

However, the shimming procedure is often a tedious and painstaking process that demands extensive time and experience from the operator [9]. This is due to the large number of shim coils required for spectroscopy, the inter-dependence of shim field patterns (violations of orthogonality), and the lack of a straightforward solution for the correct shim coil values in an (expected) non-convex solution space. This situation makes it a challenge to provide a correct, fast, and reliable shimming procedure for the magnetic field, especially on the order of a few seconds.

Several approaches already exist to solve this problem and automate the shimming procedure to increase spectral quality. The most robust approaches to date utilize the downhill simplex (or Nelder-Mead) method [10,11], and adaptations thereof [12]. Also, automated shimming based on a lock channel is widely used, e.g. for continuously adapted shimming during lengthy NMR measurements. Unfortunately, with large initial field inhomogeneities, locking may be impossible due to its inherently low SNR. A significant contribution to shimming utilizes gradient shimming [13,14]. For this, ideally, rapidly switchable gradient coils would be necessary, which may not be available. Also, the B_0 field maps could be distorted [15], requiring appropriate corrections and potentially repetitions. Despite emerging solutions [16–18], signal-based methods remain significant and can automatically improve spectral quality if sufficient runtime is provided. Nevertheless, the achieved improvements are often not perfect, and manual refinement is necessary due to imperfect hardware, non-orthogonal or dependent shims [19], and sensitivity to starting values and step sizes for regular shimming methods.



(a) Lineshape. (b) Ambiguity.

Fig. 1. Illustration of the fundamental problems in NMR shimming based on spectral data. (a) Field inhomogeneities induce differing Larmor frequencies, which causes line broadening, and thus reduces spectral quality. (b) The non-bijectivity between the one-dimensional spectrum and its three-dimensional origin introduces additional ambiguities. The two cubes indicate differing spatially varying field strengths in the region of interest, as caused by inhomogeneities, that nevertheless give cause to similarly shaped spectra.

We propose to fill this gap by utilizing a deep learning (DL) algorithm to guide the shimming process. As a tool, DL has demonstrated great success over various domains by end-to-end learning, i.e., an algorithm learns, given only input and target, to automatically detect features. In general, the feature detection is enabled by combining representations in multiple layers based on representations from the previous layer, where each layer represent more abstract feature levels [1]. With sufficient capacity, DL can thus learn arbitrary complex functions of high-dimensional input, and still allows for fast inference. Therefore, we hypothesize that it is possible for DL to learn shim values given 1D signals drawn from 3D space, even when the shim values do not directly correspond to visible features in the signal. Accordingly, we utilize supervised deep regression due to its capability to automatically detect non-linear relations between high-dimensional input data and numerical targets with high performance. We furthermore merge deep regression with the idea of ensembles, by combining multiple weak models to reduce prediction variance.

We focus on a non-iterative method for initial shimming to rapidly reach a state near the global optimum. Our scenario assumes shimming a probe from scratch with first-order shims, and beneficial use cases are the acceleration or improvement of existing automated shimming methods, focusing on high-throughput NMR alone or in conjunction with miniaturized hardware [20], without the use of gradients or even a lock channel. For this, we have generated a publicly available database for first-order NMR shimming that allows inference of spectral changes depending on shim offsets. We utilize the database to train a set of deep regression models (so-called weak learners) that can simultaneously predict three first-order shim currents given four distinct NMR measurements: the current unshimmed spectrum, and three spectra with individually modified shim values. The weak learners are then combined in an ensemble, via a *meta*-model, to increase prediction stability, and the performance is analyzed *in situ* for different evaluation metrics. Furthermore, we conduct limited comparison with regular shimming based on the downhill simplex method.

In summary, our paper makes the following contributions:

- Creation of the first spectral database (ShimDB) dedicated to low-field NMR shimming[1];
- Proof-of-concept for utilizing deep learning for shimming;
- Establishment of a method for rapid shimming based on deep regression with ensembles (DRE)[1], shown in Fig. 2.

Note that we only use information learned from data, i.e., we employ a knowledge-based approach. We also do not require priors, or mathematical formulations of spatial shim functions, as would be required for gradient shimming.

The rest of the paper is organized as follows. Section 2 provides an overview of related work on automated shimming, deep regression, and ensembles. In Section 3, we introduce our database, and in Section 4, our method. In Section 5, we describe our experiments for both DL training and *in situ* deployment. We discuss our approach in Section 6 and conclude in Section 7.

2. Fundamentals and related work

We start by unfolding the shimming problem and highlighting some of the successful or recent approaches. We also describe relevant advances in deep learning that we adopted in our method.

[1] Database publicly available at https://github.com/mobecks/ShimDB and code at https://github.com/mobecks/dre-nmr-shim.

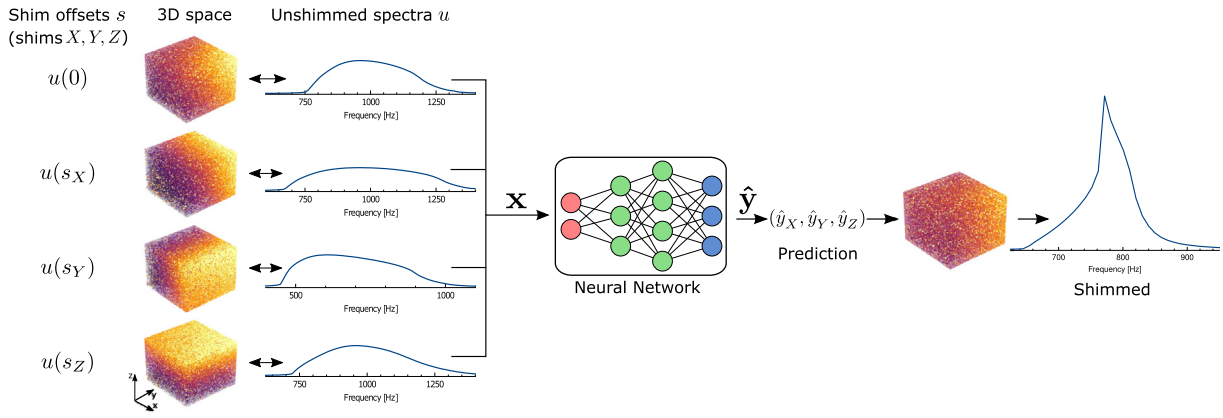


Fig. 2. Sketch of deep regression with ensembles (DRE). Three-dimensional distortions (illustrated as a 3D "inhomogeneity cube") of the sample volume collapse to a one-dimensional signal. By systematic offsets of the available shim currents, a batch of distinguishable spectra is obtained, which serves as input to a deep neural network. The prediction contains the shim values to achieve a more homogeneous field and thus a spectrum of higher quality.

2.1. Automated shimming

The shimming problem. The nuclear spin ensemble precession frequencies differ depending on the locally experienced and spatially varying field strengths of an inhomogeneous magnetic field $\vec{B}_0\left(\vec{r}\right)$ at a 3D coordinate $\vec{r} = (r_1, r_2, r_3)$. Mathematically, the shimming algorithm finds the scalar weights $(w_1, w_2, \ldots, w_n)$ for n shim currents such that the corrected magnetic field $\vec{B}_0^*$ is as uniform as possible:

$$\vec{B}_0^*\left(\vec{r}\right) = B_0\left(\vec{r}\right) + \sum_{i=1}^{n} w_i \vec{S}_i\left(\vec{r}\right). \tag{1}$$

The basis set of (ideally orthogonal) spatial functions $\vec{S}_i, i \leqslant n \in \mathbb{N}$, and their real scalar weights w_i are adjusted to reproduce and thus cancel the field inhomogeneities $\Delta\vec{B}_0$. Each $\vec{S}_i$ represents a specific shim coil (or spatial shim coil function) with its current w_i. An example of the shimming procedure for an arbitrary axis in space is given in Fig. 3. Magnetic field inhomogeneities cause distortions in the spectrum, which can be canceled by sequentially adding spatial shim functions $\vec{S}_i$ with the correct weights w_i. The concept of homogenization by superimposing correction fields was developed by Golay in 1958 [22].

In general, according to their optimization objective, automated shimming methods can be separated into signal-based and field-based methods. Field-based methods, or gradient shimming [13], use B_0 field maps acquired by gradient-echo imaging sequences to calculate an optimal combination of basis functions to cancel inhomogeneities. We focus on iterative signal-based shimming, which optimizes a scalar quality criterion based on signals such as the FID, lock channel, or spectral lineshapes. Moreover, we can neglect theoretical limitations of spatial functions and directly optimize for the shim currents.

1-D signal-based shimming. These are based on 1-D search algorithms, covered by the Tuning [23] or Coggins [24] algorithms, and are characterized by optimizing one variable at a time. The methods all share the procedure of repeatedly comparing three spectra until the minimum quality criterion of choice can be approximated by fitting a parameterized parabolic curve. Adjusting one shim at a time is simple and often faster than other methods, but does not incorporate dependencies between shim values. This is why it often has to be iterated.

N-D signal-based shimming. In n-dimensional optimization, a group of n variables are adjusted simultaneously. For this, the downhill simplex method [11], or its modifications [12], are usually used. The methods commonly use a geometrical polytope (a "simplex") of $n+1$ vertexes, where each vertex is represented by the quality criterion corresponding to specific shim settings. The simplex then evolves through solution space using geometrical

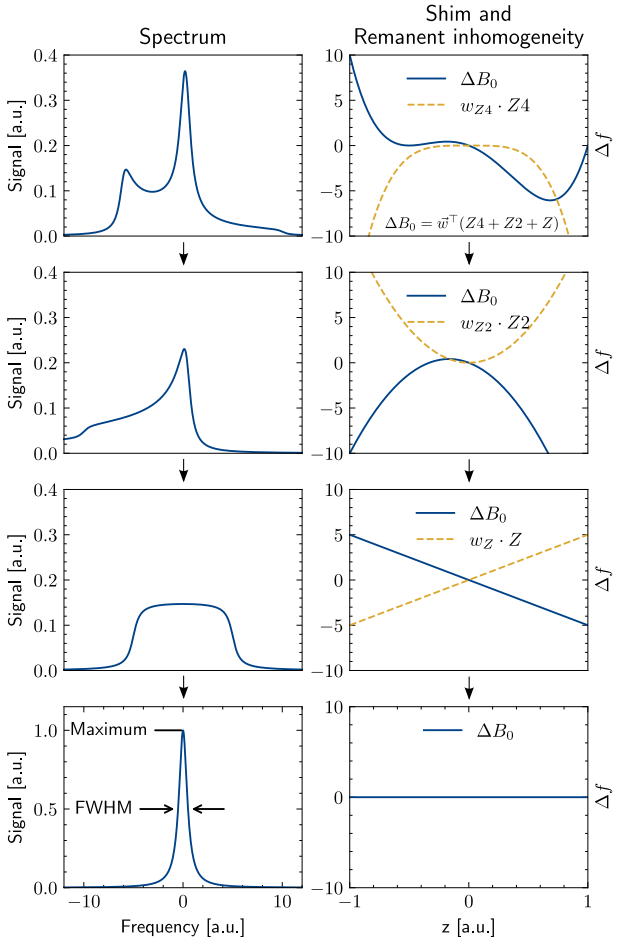


Fig. 3. Principle of shimming along the z-axis using higher-order shims. Starting from an unshimmed spectrum with unknown field inhomogeneities ΔB_0, correction fields induced by the shim coils (dashed lines) and their correct weights w_i must be selected sequentially such that the inhomogeneities are canceled. The quality of the final spectrum is usually specified by the full width at half maximum (FWHM). Layout inspired by [21].

operations such as reflection, expansion, and contraction of the simplex, based on the worst, average and best quality criterion, until a local minimum is reached, evidenced by vertices of similar quality criterium. Note that the Nelder-Mead simplex method [10] applicable in shimming should not be confused with the simplex algorithm of Dantzig for linear programming [25].

The major limitation of the downhill simplex method is its slow convergence speed [26]. Several improvements, such as quasi-gradient methods [27], adaptive shrinking coefficients [28] or perturbed centroids [29], are combined in [12] to enhance NMR shimming.

Other methods, such as modified steepest descent [11], or the rapid, modified simplex method proposed by Webb et al. [30], also enable n-dimensional shim optimization.

Other approaches. The main idea behind the method introduced by Michal [19] is to orthogonalize the shim coil gradients such that the optimization of one shim becomes independent of the others. With truly-orthogonal "composite shims", finding a global minimum becomes a one-dimensional calculation. The method brings various limitations: gradients show non-linear behaviour due to current supply, heating, or other components, and thus affect the symmetry required to calculate composite shims.

2.2. Deep learning methodologies

We now consider deep learning methods, a branch of machine learning, which learns representations from a set of prior data by abstraction at multiple levels.

Deep learning overview. Deep neural networks are most commonly implemented as stacked layers of artificial neurons, where the weight of each neuron is updated (or learned) with backpropagation [31] and gradient descent, to minimize a loss function on given training data. However, generalisation is judged by the prediction performance on previously unseen (out-of-sample) data. Therefore, different regularization techniques can prevent overfitting or unwanted memorization of the training samples. For example, early stopping of the training process is used when the training and test errors diverge, dropout [32] randomly shuts down neurons during weight update, and augmentation is used to increase the amount and variance of data. The usage of non-linear activation functions, such as the rectified linear unit (ReLU) [33], with a sufficient number of processing layers (depth), allows neural networks to approximate any arbitrary function. Furthermore, different connection patterns, e.g. each-to-each (fully-connected), of the neurons in and between each layer, allow for efficient prediction on different data structures (e.g. sequential or image-like). We refer to [1] for a more detailed description of the entire DL idea.

Deep Regression. The generalization of standard regression (i.e. prediction of a single continuous variable) to multiple, possibly interdependent variables is often referred to as multi-target regression [34]. When deep neural networks replace the function approximation, one commonly refers to deep regression. In [35], vanilla deep regression is coined to refer to convolutional neural networks (CNNs) with a linear last layer. Spectral reconstruction [2] and denoising methods [5] in NMR can be seen as regression methods, where input and output shapes are similar. Regression without convolutions is used to predict chemical shifts in [36] or [37]. However, in both cases, it only predicts a single target.

We also note with interest all the advances made for successful deep regression. Nevertheless, to allow for exploration of DL in the shimming problem, we first investigate vanilla models that require fewer assumptions, which were found to yield comparable results for complex regression models in computer vision experiments [35].

1D Convolutions. From [38], we adopt the idea of interpreting NMR spectra in the frequency domain as 1D-images, in order to apply CNNs and developments from computer vision. The concept behind convolutional layers, inspired by the virtual cortex, is to sweep filter kernels over a grid-like input to generate representations of the next layer, instead of direct links used in fully-connected layers. CNNs incorporate parameter sharing and sparse connectivity to decrease memory requirements and allow predictions independent of the features' locations [39]. We also extend this idea by using multiple input spectra as a batched input to our model, similar to RGB channels of an image. A comprehensive overview of the capabilities of one-dimensional CNNs, which are applied to NMR spectra in this paper, is given in [40]. The possibility to visually relate specific distortions in the NMR spectrum to specific shim currents [41] supports our goal of automating the shimming process with DL.

Ensemble methods. Ensemble methods in machine learning combine multiple models to construct a more powerful model to achieve higher accuracy or lower variance in predictions [42,43]. In general, ensembles consist of two levels: multiple weak learners (level-0), and a combination of their predictions (level-1), often represented by a *meta*-model. Several forms, such as bagging [42], boosting [44], or stacking [45] can be distinguished, and they differ in data handling or training of the different levels' models.

Some applications of ensembles-to-regression are summarized in [46]. In the field of NMR, ensembles have already been used to predict fish sizes from metabolomic profiles using spectral data [47].

3. First-order shimming dataset

The success of deep learning algorithms strongly depends on the quantity and quality of available data, which should represent the task as accurate as possible without introducing biases. Thus, the first step in the development of DL-assisted automatic shimming was the creation of a database for algorithm training. The shimming database (ShimDB)[2] is a collection of proton NMR signals recorded under the application of shim coil fields and so far contains a subset called LinearShimDB, a small-scale dataset containing over 9000 instances with only linear shim offsets. It allows inference of changes to the NMR spectrum or free induction decay (FID) depending on linear shim offsets. Each data instance includes the following information: A binary file containing the raw ^{1}H-FID with dimensions 1×32768; the shim values $\in \left[-2^{15}, 2^{15}\right]$ for n shims; the acquisition parameters; and the processing parameters. We pretend that the spectrometer only has $n = 3$ first-order shims, thereby only X, Y, and Z shim values are non-zero, but this is easily extended.

The measured sample consists of distilled water mixed with copper sulfate to reduce spin–lattice or longitudinal relaxation time T_1, allowing for faster database acquisition. In analogy to the study [48], the 50 ml H_2O is mixed with 0,062 g $CuSO_4$+5H_2O (CAS No. 7758–99-8), resulting in a concentration of 5 mmol/L $CuSO_4$. With inversion recovery experiments we find that $T_1 \approx 290$ ms.

The data was acquired on the low-field Magritek Spinsolve 80 Carbon spectrometer (Magritek GmbH, Aachen Germany, [49]) with a ^{1}H frequency of 80 MHz using standard 5 mm sample tubes and the Spinsolve-Expert software. Experimental parameters, and the dataset's characteristics, are summarized in Table 1. A large reception bandwidth was chosen because the initial line shape is unknown when using large shim offsets. Also, the frequency lock is not activated so that the signal potentially could leave the field of view.

[2] It is intended to continuously extend the database with new subsets.

Table 1
Characteristics and acquisition parameters of the first-order shimming dataset (LinearShimDB).

Characteristics	Nr. spectra	9261
	Shim range R	$\pm$10000
	Step size s	1000
	Shims	X, Y, Z
Acquisition parameters	Nucleus	^{1}H
	Bandwidth	20 kHz
	Points	32768
	Repetition time	2000 ms
	Filtering	-
	Phase correction	ϕ_0

The acquisition procedure of the LinearShimDB subset was as follows. The manufacturer's automated shimming technique, based on the downhill simplex method [50], was used to obtain a reference spectrum of decent quality. Then, all shim values except the three linear shims X, Y and Z were set to a current of zero Ampére. The resulting spectrum and corresponding shim settings $\mathbf{y}_r$ were used as the reference values. The database parameters were obtained by relative, systematic offsets $\mathbf{y} = \mathbf{y}_r + (\alpha s_X, \beta s_Y, \gamma s_Z)$ from the reference shim values in a range R with step size s, where $\alpha, \beta, \gamma \in \left[-\frac{R}{s}, \frac{R}{s}\right]$ change in a grid-like manner. R is chosen large enough to mimic shimming a probe from scratch. For each combination, the raw FID, acquisition parameters, and shim values were stored. [3]

We believe that our data can successfully be reused for training ML models on different setups and scenarios by transfer learning or domain adaptation techniques [51].

4. Deep regression with ensembles for shimming

In this section, we define the mathematical problem and introduce our network structures for deep regression, where we differentiate between weak learners (level-0), and the *meta*-model (level-1). Furthermore, we introduce our performance metrics.

4.1. Problem definition in terms of DL and NMR

Let $\mathcal{D} = \{(\mathbf{x}, \mathbf{y})_i\}_{i=1}^{|\mathcal{D}|}$ be our database, where $(\mathbf{x}, \mathbf{y})_i$ is a pair including the input $\mathbf{x} \in \mathbb{R}^{W \times 4}$ with dimensions $W \times 4$, where W is the input's width, and the associated target $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, defined as a real-valued vector of n elements, with n being the number of separate shim coils. Also, consider the regression model $\mathbb{F}_\theta(\cdot)$, represented by a deep convolutional neural network with parameters θ. The network parameters θ are learned in a supervised manner using the database $\mathcal{D}$ in order to minimize the mean squared error (MSE) between the prediction $\hat{\mathbf{y}} = \mathbb{F}_\theta(\mathbf{x})$ and the target $\mathbf{y}$.

In terms of NMR, the predictions $\hat{\mathbf{y}}$ translate/map to the shim values w_i and should solve Eq. 1. The inputs $\mathbf{x}$ are defined as $\mathbf{x} = [u(0), u(s_X), u(s_Y), u(s_Z)]$, where the unshimmed spectrum u changes as a function of systematic shim offsets s. The DL model predicts the shim correction terms $\mathbb{F}_\theta(\mathbf{x}) = (\hat{y}_X, \hat{y}_Y, \hat{y}_Z)$, such that $\mathbf{y}_i - \hat{\mathbf{y}}_i \approx 0$. Note that we do not have access to either S_i, nor the magnetic field $\vec{B}_0$. We also compare the mean absolute error (*MAE*) to measure *in situ* shimming performance between the predicted and relative offset values of the reference shims (see Section 3) with $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$.

[3] The dataset is publicly available at https://github.com/mobecks/ShimDB under the CC BY-SA 4.0 licence.

Our approach stems from the concept of deep regression in a multi-input and multi-output setting. In our scenario, the model $\mathbb{F}$ is an ensemble of weak learners combined with a multi-layer perceptron (MLP) as the *meta*-model, as proposed in subsection 4.2.

4.2. Deep learning architectures and pipeline

Level-0. Unlike in computer vision, in NMR there is no bijective mapping between the input to the model and its output (i.e. prediction at the pixel level of the input). Indeed, the shimming problem starts with a field inhomogeneity in 3D space, translates this into a 1D NMR signal, after which there is no straightforward link back to the correct shim currents. Therefore, we provide the weak learners with a batch of four NMR spectra $\mathbf{x}$ as input and predict a three-valued vector of continuous values $\hat{\mathbf{y}}$, i.e., the shim values.

Each architecture begins with $3-5$ blocks of one-dimensional convolutional layers, with varying kernel sizes, followed by two fully connected layers with 32 nodes. Heterogeneous architectures allow for higher variance in predictions that are useful for the ensemble model. Additionally, we include dropout layers after convolution and fully-connected layers to prevent overfitting [32]. The last layer uses linear activation for regression of the targets. The architecture is illustrated in Fig. 4a.

Level-1. The *meta*-model combines features of its m heterogeneous weak learners, and represents a mixture of stacking and boosting (see subsection 2.2). We investigate the following forms:

- Simple average over all weak learner predictions.
- Non-linear combination, with a fully-connected layer, of the weak learner's regression layer ($\in \mathbb{R}^{m \times n}$).
- A two-layer multi-layer perceptron (MLP) based on the second-to-last fully-connected layer of the level-0 models. The MLP has $m \times 32$ nodes in its first, and 32 nodes in its second layer.

Also refer to Fig. 4b for a visualization of the ensemble with an MLP-based *meta*-model.

4.3. Spectral quality and performance metrics

To judge the quality of spectra, we introduce a criterion c that can be used for global or local quality judgements of single peaks. For a spectrum g of interest and a reference spectrum r, it is defined as:

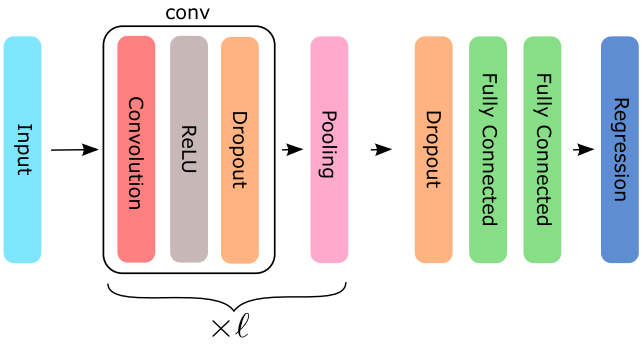
$$c(g, r) = \frac{1}{2}\left(\lambda_1 \cdot \frac{\mathrm{FWHM}(r)}{\mathrm{FWHM}(g)} + \lambda_2 \cdot \frac{\max(g)}{\max(r)}\right), \tag{2}$$

where $\max(\cdot)$ is the maximum peak height and $\mathrm{FWHM}(\cdot)$ is the full width at half maximum of that peak. λ_i can be used to increase or decrease the impact of each term. The quality parameter c indicates whether spectrum g is worse ($0 < c < 1$) or better ($c > 1$) than the reference r.

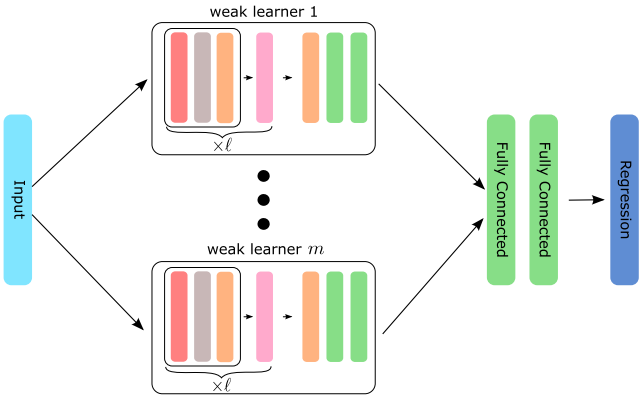
Other spectral properties, such as peak symmetry, can easily extend the criterion. But since linear shims cannot lead to symmetrical line shapes, we have refrained from more precise quality measures such as the envelope of a spectral peak [52]. An extension to multiple peaks could be realized via virtual peaks [53].

Furthermore, the following metrics are introduced to analyze the performance of our method in laboratory experiments for the $n = 3$ first-order shims:

- Success rate $SR \in [0, 1]$. We defined the SR w.r.t. the criterion c. If the predicted shim setting yielded c higher than all c in the input batch, then the method was deemed successful. For a single experiment, it was defined as:



(a) Weak learner architecture.



(b) Ensemble architecture with MLP *meta*-model.

Fig. 4. Visualization of neural network architectures. Each weak learner of the ensemble consists of ℓ layers followed by two fully-connected layers. The *meta*-model shown here is a two-layer MLP trained based on weak learners on the transfer set $\mathcal{T}$.

$$SR = \begin{cases} 1, & \text{if } c_{sh} > [c_{init}, c_X, c_Y, c_Z] \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $[c_{init}, c_X, c_Y, c_Z]$ are the quality values for the input batch and c_{sh} is the criterion after shimming with DRE.

- Correct direction ratio $DiR \in [0, 1]$. Indicator whether the method pointed towards the global minimum and equals to 1 if the predicted signs matched the distortion's signs:

$$DiR = \frac{1}{n}\sum_{i=1}^{n}(sgn(\hat{y}_i) == sgn(y_i)), \tag{4}$$

where $sgn(\cdot)$ is the sign function, $\hat{y}_i$ is the model's prediction and y_i is the true distortion.

- Mean improvement of criterion c (see Eq. 2) with $\lambda_1 = \lambda_2$. Given in percentage for $c(g, r)$, where g is the spectrum with predicted correction and r is the initial spectrum.
- Averaged *MAE* between predictions and random distortion.
- Generalization to other substances.

5. Experiments

In this section, we describe prerequisites in pre-processing, hardware, and DL training, followed by our evaluation protocol, and the results for offline and *in situ* (online) experiments.

5.1. Setup or implementation details

Dataset creation and pre-processing. Each member of a batch of input spectra $\mathbf{x}$ consists of one spectrum corresponding to a unique target value $\hat{\mathbf{y}}$, and three spectra with offsets of s in the X, Y, and Z shims, respectively. The database $\mathcal{D}$ with input-target pairs $(\mathbf{x}, \mathbf{y})_i$ is constructed by mining the dataset LinearShimDB from Section 3. Each raw FID is fast-Fourier-transformed and phase-corrected to yield a 1D spectrum, using `nmrglue` [54] and the same phase correction values given by the system's auto-phase method. Note that the unique target values $\mathbf{y}$ for each input $\mathbf{x}$ is not represented by its absolute shim values, but is defined by its relative distortion w.r.t. the reference spectrum of $\mathcal{D}$. This prevents the model from learning absolute shim currents that would depend on the hardware that the database was acquired on, and forces the model to learn the relative shim offsets that will improve a given spectral shape.

In order to achieve faster convergence and generalization, our data is further pre-processed. Each spectrum is normalized by a constant normalization factor of $1e5$ (the maximum intensity for perfect shims) such that the spectral values lie within the range $[0, 1]$. To meet resource constraints, all spectra are downsampled from 32768 to 2048 data points. The regression targets (stored as int16 integers) are divided by 2^{15} to avoid exploding gradients, and then multiplied by 100 to avoid vanishing gradients during DL model training. We exclude dataset samples where offsets are incomplete. Thus, the final subsets for training, validation, and test are of size 6400/800/801.

Due to time constraints during the acquisition of $\mathcal{D}$ in a grid-like manner (Section 3), we expect that the spectra of each instance $(x, y)_i$ inhibit some reality gap and temporal equalization. Therefore, we introduce an additional transfer database $\mathcal{T}$ with $|\mathcal{T}| = 100$ to differentiate from the systematic nature of data collection. $\mathcal{T}$ is obtained under the same conditions as $\mathcal{D}$, but each spectrum of $\mathbf{x}$ is jointly acquired and is used to either fine-tune the weak learner or the *meta*-model.

Hardware interface. Communication with the Magritek Spinsolve spectrometer was enabled through a custom interface between Python and the python-like programming language `Prospa`, upon which the Spinsolve-Expert software is built. With this interface, it is possible to benefit from open-source python libraries for NMR data processing (`nmrglue` [54]) and deep learning frameworks (`PyTorch` [55], and `ray tune` [56]). The standard shimming routines used by the spectrometer's software are based on parabolic interpolation and the downhill simplex method, and the latter is adopted for comparison.

DL training details. Level-0 base models underwent a limited neural architecture search (NAS), a technique for automating the neural network architecture design [57], using `ray tune` [56], and random search over a variable number of layers ℓ, kernel sizes, and other design choices. All weak learners were trained with a learning rate of 0.001, batch size of 32, and the Adam optimizer [58] for a maximum of 150 epochs utilizing early stopping. We selected the top-50 architectures among 300 runs w.r.t. validation error. The level-1 *meta*-models were trained with hyperparameter optimization (HPO) using random search and early stopping. We manually selected the best fully-connected and MLP-based networks w.r.t. their validation loss over 500 run. For a detailed description, see subsection S1.2.

Hardware requirements. The DL training was performed with an AMD Ryzen 5900X equipped with 64 GB RAM, and a graphics processing unit NVIDIA GeForce RTX 3090. The LinearShimDB roughly requires 3.4GB of disc space.

5.2. In situ evaluation protocol

We demonstrate *in situ* functionality by testing the method on a set of 100 random distortions $y_X, y_Y, y_Z \in [-10000, 10000]$ of X, Y, Z shims, drawn from a uniform distribution. We deliver success rate *SR*, direction rate *DiR*, mean improvement of our quality parameter *c*, and *MAE* over five different model types, including single weak

learners and diverse ensembles. Furthermore, we investigate generalizability towards multiple samples, i.e., pure water with $CuSO_4$ as in Section 3, ethanol (CAS No. 67–66-3) dissolved in water with a molar fraction of $\chi = [0.1, 0.5]$, and isopropanol (CAS No. 67–63-0) in water with $\chi = 0.5$.

5.3. Level-0 training results

Offline experiments, i.e., training on static data, show that it is possible to predict three distinct variables from an input of four 1-D signals with no apparent correlation between the input and output dimensions. The best weak learner achieved an *MAE* of 596 ± 769 (mean±standard deviation) for a step size of $s = 1000$ on the test set. As the possible precision was limited by the sampling resolution of the underlying training data, results near $1/2$ of the step size s indicated good performance. Detailed results are given in Table S1. The entire network, including the *meta*-model, was tested *in situ* only.

5.4. In situ results

The most crucial aspect of the method is not merely its training performance, but also its applicability. Thus, the performance of our method is evaluated in laboratory experiments, and a selection of graphical results is given in Table 3 for our most promising methods (single model and MLP-based) and different substances (H_2O, ethanol and isopropanol).

Interpretation of different model types. The results seem to indicate that our method works in practise. Even weak learners achieved an *SR* of 93% and a large improvement (mean of +435%) on the spectral quality for water. However, the variance in criterion improvement and error remained high. Here, the ensemble method with an MLP-based *meta*-model achieved more robust but conservative predictions. Overall, a single model and the MLP-based ensemble yield comparable results for all measured metrics.

Absent improvement by averaging the top 50 untuned models confirmed the need to train a *meta*-model. Furthermore, the advantage of a two-layer MLP with non-linear connections of the second-to-last features is shown over a simple non-linear combination of the weak learner's last layer. We also fine-tuned a single model to the transfer dataset $\mathcal{T}$, yielding worse *in situ* results. Overall, a transfer set seems to be unnecessary.

The robustness problem of a single model compared to the MLP-based ensemble is visualized in Fig. 5 based on the experiments that yield Table 2 for water. The single model can often achieve narrower linewidths but shows higher variance.

Interpretation of generalizability. Through experiments on ethanol with $\chi = [0.1, 0.5]$ and isopropanol with $\chi = 0.5$, we strive to show the generalization of the method to samples other than H_2O. Indeed, the training sample in both datasets $\mathcal{D}$ and $\mathcal{T}$ is water, of course, with only a single peak. Surprisingly, we achieve success rates above 91% for samples with more than one peak, despite the risk of confusion among the peaks.

Hardware requirements. The most resource-intense MLP-based ensemble model required 190 ms on average for prediction, using an Intel Core i5-8500 and 200 MB of RAM. The required disc space was between 0.5 and 2.5MB for each weak learner, and 200 KB for the *meta*-model. Compared to acquisition times for the NMR measurement, and storage space available on recent computers, these were negligible requirements. One complete cycle of DRE, including spectra acquisition, takes 31s without special time-saving efforts being made.

5.5. Comparison

We compare our automated shimming method to Magritek's built-in implementation of the downhill simplex method, which in turn is based on the algorithm in [50].

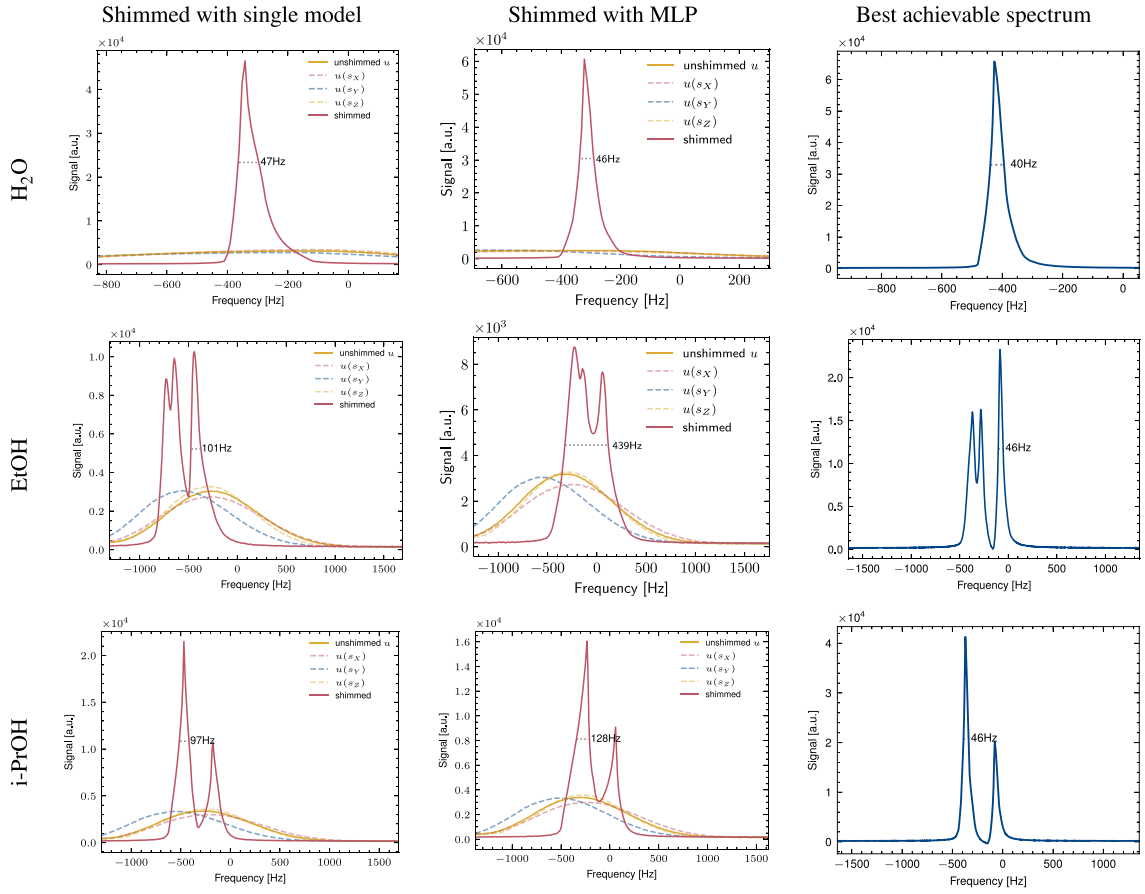
We expect faster convergence of the simplex method when using the improvements as proposed by [12]. Nevertheless, Yao et al. state that their method behaves similarly to the simplex method when fewer shim coils are used, so we restricted our considerations to the Magritek implementation of the regular downhill simplex method. This implementation requires at least $n + 1$ measurements for n shims to initialize its simplex structure, and one to four (average of two) function evaluations per iteration [11]. Our method consistently needs four spectra for one iteration, and one spectrum to check the results. We used a single model and DRE with MLP as the *meta*-model for comparison.

Although the downhill simplex is known to have little influence on the initial simplex' size and shape [11], and tends to produce rapid drops of initial values [59], we were able to accelerate the shimming process. We compared results obtained from the standard Nelder-Mead method with i iterations and step size $s = 1000$, versus the results for simplex with $i - 4$ iterations while

Table 2
In situ results of automated shimming of the DRE method using a single model and different ensemble types over different substances, with molar fraction (χ). Values are reported as mean ± standard deviation over 100 random distortions drawn from a uniform distribution. The best values are marked in bold. Abbreviations: *FC* = fully-connected, *MLP* = multi-layer perceptron, $\mathcal{T}$ = transfer database, *SR* = success rate, *DiR* = direction ratio, *MAE* = mean absolute error, *c* = criterion.

		Single model		Ensemble		
		Untuned	Tuned to $\mathcal{T}$	Average	FC	**MLP**
H_2O	Success Rate *SR*	0.93	0.91	0.61	0.76	**0.94**
	Direction Ratio *DiR*	**0.90**	0.89	0.77	0.78	0.89
	Mean improvement of *c*	+**435**%±559	+394%±471	+125%±244	+127%±124	+399%±560
	Averaged *MAE*	1878±1271	1703±969	3989±2529	3191±1459	**1756**±1016
Ethanol (0.5)	*SR*	0.92	0.89	0.55	0.73	0.93
	DiR	0.86	0.86	0.79	0.78	0.85
	Mean improvement of *c*	157%±130	136%±99	28%±14	73%±63	123%±34
	Averaged *MAE*	2431±1287	2492±1303	4510±1999	3441±1602	2460±1217
Ethanol (0.1)	*SR*	0.91	0.90	0.68	0.71	0.89
	DiR	0.87	0.85	0.77	0.78	0.85
	Mean improvement of *c*	+265%±214	+205%±158	+77%±38	+91%±66	+257%±314
	Averaged *MAE*	2170±1292	2478±1334	3995±2344	3462±1701	2250±1217
Isopropanol (0.5)	*SR*	0.91	0.95	0.65	0.80	0.90
	DiR	0.89	0.90	0.77	0.82	0.87
	Mean improvement of *c*	+299%±320	+257%±278	+74%±154	+91%±107	+201%±156
	Averaged *MAE*	1967±1177	2102±1229	4199±2533	3393±1561	2097±956

Table 3
Exemplary selected results of shimmed spectra for water, ethanol ($\chi = 0.5$), and isopropanol. A single model's performance is compared to ensembles with an MLP-based *meta*-model. Additionally, the optimal spectrum by applying the simplex method only to the first-order shims is reported. Abbreviations: EtOH = ethanol, i-PrOH = isopropanol.



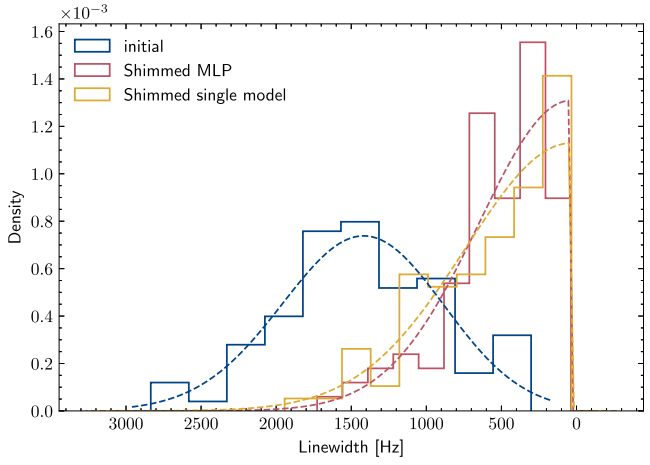


Fig. 5. Probability density function (PDF) indicating the distribution of initial and shimmed linewidths (FWHM) of our 100 *in situ* runs. The results are given for a single model and an MLP-based ensemble on a sample of H_2O. The dashed lines visualize skewed normal distributions that are fitted to the histogram (continuous line).

initialized with our method (see Table 4). We reduce iterations by four because one iteration of DRE needs four spectra for its prediction.

Furthermore, we compared the number of function evaluations necessary for the simplex to reach a criterion equivalent to the DRE prediction, as shown in Table 5. The procedure is as follows: First, our method "deep regression with ensembles" (DRE) predicts shim settings for a random distortion drawn from a uniform distribution. Then, the downhill simplex method is started from the same distortion with a step size of $s = 1000$ (as used for the database in Section 3). The algorithm is stopped when it reaches a linewidth equivalent to the one achieved with DRE prediction, and the number of acquisitions (function evaluations) are reported. If the simplex is not able to find an equivalent linewidth within 50 iterations, it is stopped.

The results in Tables 4 and 5 demonstrate that, using our method as stand-alone or in combination with regular shimming methods will provide an advantage in either the number of necessary acquisitions, or the achieved spectral quality. Note that deep regression (DR) apparently yields similar performance to deep regression with ensembles (DRE) in quality improvement, but

Table 4
Comparison of criterion improvement w.r.t. initial spectrum between the default downhill simplex and simplex initialized with our method. Default simplex is run for i iterations and DR/DRE + simplex for $i-4$ iterations because one iteration of DR/DRE requires four measurements. Here $i = 10$. Values are reported as $c_{initial}\times$ improvement $\pm$ std. Best values are marked in bold. Abbreviations: DR = Deep Regression, DRE = DR with ensembles.

Method	Iterations i	Criterion after DR/ DRE	Criterion after simplex
simplex	10	-	$\times 5.2 \pm 3.96$
DR + simplex	$1+6$	$\times 4.6 \pm 4.9$	$\times 15.4 \pm 11.1$
DRE + simplex	$1+6$	$\times$**5.0** ± 6.3	$\times$**15.7** ± 11.7

Table 5
Comparison of necessary NMR acquisitions in the downhill simplex method to achieve a similar criterion as obtained with our method. Lower is better and best values are marked in bold. Abbreviations: fe = function evaluations, FWHM = full width at half maximum, DR = Deep Regression, DRE = DR with ensembles.

Method	$FWHM_{initial}$	$FWHM_{shimmed}$	fe
DR	1432±532	516±330	**4 + 1**
simplex	"	552±411	16.3±11.1
DRE	"	486±342	**4 + 1**
simplex	"	**473**±414	21.1±16.6

ensembles demonstrate advantage in the number of necessary NMR acquisitions as compared to the simplex method.

6. Discussion

Discovering new methods for fast and efficient NMR shim coil calibration remains a challenge, but we see great potential in a fusion of traditional and modern algorithmic methods to achieve both faster and more precise shimming.

Assumptions of our approach. Currently, our approach mimics a scenario where only the first-order shims are available to shim a probe from scratch. Thus, the best achievable FWHM is of the order of tens of Hz, but our method shows that very broad initial lineshapes can be improved, very nearly reaching the optimum. Furthermore, using linear shims is the first step for DL to advance shimming for more complex and non-standard cases (e.g. high-throughput parallel spectroscopy where some samples may be located off-center with respect to the shim system, shimming of micro coils). Although first-order shims generally have a prominent influence on reducing inhomogeneity, dealing with higher-order shims should be considered in future work. We are also convinced that localizable information, e.g. about axial magnetization [17], with special hardware additions could help the DL method and allow to reduce the number of required input spectra. Our method contrasts with traditional shimming approaches, in that it is not iterative, i.e., the result emerges after a single step. Therefore, we cannot guarantee that iterating DR or DRE will converge to better results under other circumstances.

Deep learning considerations. Also, it is unclear how or whether supervised learning can be scaled for shimming of higher-order shims, especially w.r.t. exponentially increasing requirements in the data acquisition task. One approach to counteract this limitation is to employ a sufficiently realistic digital twin (simulation) of the shimming problem to generate synthetic data.

In any case, this study has not yet unlocked the full potential of deep learning, as applied to the shimming problem in NMR. Especially, recent advances in explainable artificial intelligence (XAI) [60] can help to understand decisions made by neural networks, or new neural architectures such as transformers [61,62] are creating a big fuss and yield excellent performance in various tasks. However, to stay grounded in the complexity of our method, we expect higher stability of a single model by using benchmark architectures such as ResNet [63], and increased efficiency by realizing other deep learning techniques [64].

Despite the limitation described above, we could prove that the concept of using DL for NMR shimming is a promising approach to accelerate the entire shimming process.

In general, we also wish to draw attention to the lack of published code for related shimming methods, and the difficulty of comparing published results obtained on diverse spectrometers and hardware setups.

7. Conclusion

In this paper, we have conclusively shown that deep learning can be used for fast, first-order shimming in a low-field NMR setup, i.e., DL can tackle the ambiguity and lineshape problems inherited by shimming. With our dataset for shimming, we furthermore established a basis for continued research in this area. The applicability of DL was demonstrated both offline and *in situ*: First, by training neural networks to simultaneously predict three linear-field shim currents necessary to partially cancel an arbitrary distortion. And second, the deployment on a spectrometer of a *meta*-model based on an ensemble of weak learners revealed that our non-iterative method generally points towards the solution and shows a high success rate in improving spectral quality. Despite the models being trained entirely on water, we experienced generalizability to more complex samples with more than one spectral peak.

Author contributions

J.K. acquired the funding, and together with M.J., conceived the idea, and performed supervision. M.B. generated the database and designed the DL algorithm. M.B. conceived and performed the numerical experiments. M.B. prepared the data tables and plots, and wrote the initial manuscript. M.B.,A.K.,M.J. and J.K. iterated the manuscript, and proof read the manuscript. All authors participated in the discussions.

Financial disclosure

J.G.K. declares financial interest in the startup company Voxalytic GmbH, which develops and sells NMR equipment. The other authors declare no interest.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Craig Eccles of Magritek for his great support. Plots in Fig. 3, Table 3 and Fig. 5 where made with the Python library SciencePlots [65], and we sincerely thank John Garrett for making the code available online. We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

Appendix A. Supplementary material

The following supporting information is available as part of the online article: Supplementary at the end of this document, data-

base at https://github.com/mobecks/ShimDB and code at https://github.com/mobecks/dre-nmr-shim. Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmr.2022.107151.

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, https://doi.org/10.1038/nature14539, URL: http://www.nature.com/articles/nature14539.

[2] D.F. Hansen, Using Deep Neural Networks to Reconstruct Non-uniformly Sampled NMR Spectra, J. Biomol. NMR 73 (10–11) (2019) 577–585, https://doi.org/10.1007/s10858-019-00265-1, URL: http://link.springer.com/10.1007/s10858-019-00265-1.

[3] H. Lee, H.H. Lee, H. Kim, Reconstruction of spectra from truncated free induction decays by deep learning in proton magnetic resonance spectroscopy, Magn. Reson. Med. 84 (2) (2020) 559–568, https://doi.org/10.1002/mrm.28164, URL: https://onlinelibrary.wiley.com/doi/10.1002/mrm.28164.

[4] S. Liu, J. Li, K.C. Bennett, B. Ganoe, T. Stauch, M. Head-Gordon, A. Hexemer, D. Ushizima, T. Head-Gordon, Multiresolution 3D-DenseNet for Chemical Shift Prediction in NMR Crystallography, J. Phys. Chemi. Lett. 10 (16) (2019) 4558–4565, https://doi.org/10.1021/acs.jpclett.9b01570, URL: https://pubs.acs.org/doi/10.1021/acs.jpclett.9b01570.

[5] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, T. Tan, Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network, Japanese J. Radiol. 36 (9) (2018) 566–574, https://doi.org/10.1007/s11604-018-0758-8, URL: http://link.springer.com/10.1007/s11604-018-0758-8.

[6] Z. Akkus, A. Galimzianova, A. Hoogi, D.L. Rubin, B.J. Erickson, Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions, J. Digit. Imaging 30 (4) (2017) 449–459, https://doi.org/10.1007/s10278-017-9983-4, URL: http://link.springer.com/10.1007/s10278-017-9983-4.

[7] D. Chen, Z. Wang, D. Guo, V. Orekhov, X. Qu, Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy, Chemistry – A Eur. J.s 26 (46) (2020) 10391–10401, https://doi.org/10.1002/chem.202000246, URL https://onlinelibrary.wiley.com/doi/10.1002/chem.202000246.

[8] J.G. Korvink, N. MacKinnon, V. Badilita, M. Jouda, "Small is beautiful" in NMR, J. Magn. Reson. 306 (2019) 112–117, https://doi.org/10.1016/j.jmr.2019.07.012, URL: http://arxiv.org/abs/1908.03262, https://linkinghub.elsevier.com/retrieve/pii/S1090780719301326.

[9] G.N. Chmurny, D.I. Hoult, The Ancient and Honourable Art of Shimming, Concepts Magnetic Resonance 2 (3) (1990) 131–149, https://doi.org/10.1002/cmr.1820020303, URL: https://onlinelibrary.wiley.com/doi/10.1002/cmr.1820020303.

[10] J.A. Nelder, R. Mead, A Simplex Method for Function Minimization, Comput. J. 7 (4) (1965) 308–313, https://doi.org/10.1093/comjnl/7.4.308, URL: https://academic.oup.com/comjnl/article-lookup/doi/10.1093/comjnl/7.4.308.

[11] R.R. Ernst, Measurement and Control of Magnetic Field Homogeneity, Rev. Sci. Instrum. 39 (7) (1968) 998–1012, https://doi.org/10.1063/1.1683586, URL: http://aip.scitation.org/doi/10.1063/1.1683586.

[12] K. Yao, M. Liu, Z. Zheng, T. Shih, J. Xie, H. Sun, Z. Chen, Automatic Shimming Method Using Compensation of Magnetic Susceptibilities and Adaptive Simplex for Low-Field NMR, IEEE Trans. Instrum. Meas. 70 (2021) 1–12, https://doi.org/10.1109/TIM.2021.3074951, URL: https://ieeexplore.ieee.org/document/9410624/.

[13] M.G. Prammer, J.C. Haselgrove, M. Shinnar, J.S. Leigh, A new approach to automatic shimming, J. Magnetic Resonance (1969) 77 (1) (1988) 40–52, https://doi.org/10.1016/0022-2364(88)90030-3, URL https://linkinghub.elsevier.com/retrieve/pii/0022236488900303.

[14] R. Gruetter, Fast, noniterative shimming of spatially localized signals. In vivo analysis of the magnetic field along axes, J. Magn. Reson. (1969) 96 (2) (1992) 323–334, https://doi.org/10.1016/0022-2364(92)90085-L, URL https://linkinghub.elsevier.com/retrieve/pii/002223649290085L.

[15] Q. Bao, F. Chen, L. Chen, K. Song, Z. Liu, C. Liu, A new gradient shimming method based on undistorted field map of B0 inhomogeneity, J. Magn. Reson. 265 (2016) 25–32, https://doi.org/10.1016/j.jmr.2015.12.008, URL: https://linkinghub.elsevier.com/retrieve/pii/S1090780715003079.

[16] G. Liu, X. Qu, S. Cai, Z. Zhang, Z. Chen, C. Cai, Z. Chen, Fast 3D gradient shimming by only 22 pixels in XY plane for NMR-solution samples, J. Magn. Reson. 248 (2014) 13–18, https://doi.org/10.1016/j.jmr.2014.09.002, URL: https://linkinghub.elsevier.com/retrieve/pii/S1090780714002390.

[17] S. Gross, C. Barmet, B.E. Dietrich, D.O. Brunner, T. Schmid, K.P. Pruessmann, Dynamic nuclear magnetic resonance field sensing with part-per-trillion resolution, Nature Commun. 7 (1) (2016) 13702, https://doi.org/10.1038/ncomms13702, URL: https://doi.org/10.1038/ncomms13702, http://www.nature.com/articles/ncomms13702.

[18] H. Barjat, P.B. Chilvers, B.K. Fetler, T.J. Horne, G.A. Morris, A Practical Method for Automated Shimming with Normal Spectrometer Hardware, J. Magn. Reson. 125 (1) (1997) 197–201, https://doi.org/10.1006/jmre.1996.1094, URL: https://linkinghub.elsevier.com/retrieve/pii/S1090780796910940.

[19] C.A. Michal, Magnetic field homogeneity: A new approach to orthogonalizing and optimizing shim gradients, J. Magn. Reson. 185 (1) (2007) 110–117, https://doi.org/10.1016/j.jmr.2006.12.006, URL: https://linkinghub.elsevier.com/retrieve/pii/S1090780706003995.

[20] Y. Li, A.M. Wolters, P.V. Malawey, J.V. Sweedler, A.G. Webb, Multiple Solenoidal Microcoil Probes for High-Sensitivity, High-Throughput Nuclear Magnetic Resonance Spectroscopy, Anal. Chem. 71 (21) (1999) 4815–4820, https://doi.org/10.1021/ac990855y, URL: https://pubs.acs.org/doi/10.1021/ac990855y.

[21] M. Weiger, T. Speck, Shimming for High-Resolution NMR Spectroscopy, in: Encyclopedia of Magnetic Resonance, John Wiley & Sons Ltd, Chichester, UK, 2011, https://doi.org/10.1002/9780470034590.emrstm1228. URL: http://doi.wiley.com/10.1002/9780470034590.emrstm1228.

[22] M.J.E. Golay, Field Homogenizing Coils for Nuclear Spin Resonance Instrumentation, Rev. Sci. Instrum. 29 (4) (1958) 313–315, https://doi.org/10.1063/1.1716184, URL: http://aip.scitation.org/doi/10.1063/1.1716184.

[23] W.E. Hull, NMR Tips for Shimming, Part II. Computerized Shimming with the Tuning Algorithm, Bruker Spin Report 152/153.

[24] D. Holz, D. Jensen, R. Proksa, M. Tochtrop, W. Vollmann, Automatic shimming for localized spectroscopy, Med. Phys. 15 (6) (1988) 898–903, https://doi.org/10.1118/1.596173, URL: http://doi.wiley.com/10.1118/1.596173.

[25] G. Dantzig, Linear Programming and Extensions, RAND Corporation, Santa Monica, 1963. https://doi.org/10.7249/R366. URL: https://www.rand.org/pubs/reports/R366.html.

[26] W.E. Hull, NMR Tips for Shimming, Part II. Computerized shimming with the Simplex algorithm, Bruker Spin Report 154/155.

[27] N. Pham, A. Malinowski, T. Bartczak, Comparative Study of Derivative Free Optimization Algorithms, IEEE Trans. Industr. Inf. 7 (4) (2011) 592–600, https://doi.org/10.1109/TII.2011.2166799, URL: http://ieeexplore.ieee.org/document/6011694/.

[28] F. Gao, L. Han, Implementing the Nelder-Mead simplex algorithm with adaptive parameters, Comput. Optim. Appl. 51 (1) (2012) 259–277, https://doi.org/10.1007/s10589-010-9329-3, URL: http://link.springer.com/10.1007/s10589-010-9329-3.

[29] I. Fajfar, A. Buermen, J. Puhan, The Nelder-Mead simplex algorithm with perturbed centroid for high-dimensional function optimization, Optim. Lett. 13 (5) (2019) 1011–1025, https://doi.org/10.1007/s11590-018-1306-2, URL: http://link.springer.com/10.1007/s11590-018-1306-2.

[30] P. Webb, A. Macovski, Rapid, fully automatic, arbitrary-volumein vivo shimming, Magn. Reson. Med. 20 (1) (1991) 113–122, https://doi.org/10.1002/mrm.1910200112, URL: https://onlinelibrary.wiley.com/doi/10.1002/mrm.1910200112.

[31] Hecht-Nielsen, Theory of the backpropagation neural network, in: International Joint Conference on Neural Networks, IEEE, 1989, pp. 593–605. https://doi.org/10.1109/IJCNN.1989.118638. URL: http://ieeexplore.ieee.org/document/118638/.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Machine Learn. Res. 15 (56) (2014) 1929–1958, URL http://jmlr.org/papers/v15/srivastava14a.html.

[33] V. Nair, G.E. Hinton, Rectified Linear Units Improve Restricted Boltzmann Machines Vinod, in: Proceedings of the 27th International Confer- ence on Machine Learning, 2010, https://doi.org/10.5555/3104322.3104425.

[34] H. Borchani, G. Varando, C. Bielza, P. Larra naga, A survey on multi-output regression, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 5 (5) (2015) 216–233. https://doi.org/10.1002/widm.1157. URL: https://onlinelibrary.wiley.com/doi/10.1002/widm.1157.

[35] S. Lathuiliere, P. Mesejo, X. Alameda-Pineda, R. Horaud, A Comprehensive Analysis of Deep Regression, IEEE Trans. Pattern Anal. Mach. Intell. 42 (9) (2020) 2065–2081, https://doi.org/10.1109/TPAMI.2019.2910523, URL: https://team.inria.fr/perception/deep-regression/, https://ieeexplore.ieee.org/document/8686063/.

[36] Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, J. Biomol. NMR 48 (1) (2010) 13–22, https://doi.org/10.1007/s10858-010-9433-9, URL: http://link.springer.com/10.1007/s10858-010-9433-9.

[37] D. Li, R. Brüschweiler, PPM_One: a static protein structure based chemical shift predictor, J. Biomol. NMR 62 (3) (2015) 403–409, https://doi.org/10.1007/s10858-015-9958-z, URL: http://link.springer.com/10.1007/s10858-015-9958-z.

[38] X. Qu, Y. Huang, H. Lu, T. Qiu, D. Guo, T. Agback, V. Orekhov, Z. Chen, Accelerated Nuclear Magnetic Resonance Spectroscopy with Deep Learning, Angew. Chem. Int. Ed. 59 (26) (2020) 10297–10300, https://doi.org/10.1002/anie.201908162, URL: https://onlinelibrary.wiley.com/doi/10.1002/anie.201908162.

[39] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, URL: http://www.deeplearningbook.org.

[40] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: A survey, Mech. Syst. Signal Process. 151 (2021) 107398, https://doi.org/10.1016/j.ymssp.2020.107398, URL: https://doi.org/10.1016/j.ymssp.2020.107398, https://linkinghub.elsevier.com/retrieve/pii/S0888327020307846.

[41] T.D. Claridge, High-Resolution NMR Techniques in Organic Chemistry, Elsevier, 2016. https://doi.org/10.1016/C2015-0-04654-8. URL: https://linkinghub.elsevier.com/retrieve/pii/C20150046548.

[42] L. Breiman, Bagging predictors, Machine Learn. 24 (2) (1996) 123–140, https://doi.org/10.1007/BF00058655, URL: http://link.springer.com/10.1007/BF00058655.

[43] T.G. Dietterich, Ensemble Methods in Machine Learning, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 1857 LNCS, Springer, 2000, pp. 1–15.

https://doi.org/10.1007/3-540-45014-9_1. URL: http://link.springer.com/10.1007/3-540-45014-9_1.
[44] H. Schwenk, Y. Bengio, Boosting Neural Networks, Neural Comput. 12 (8) (2000) 1869–1887, https://doi.org/10.1162/089976600300015178, URL: https://direct.mit.edu/neco/article/12/8/1869-1887/6403.
[45] T.G. Dietterich, Machine Learning Research: Four Current Directions, AI Magazine (1990) 1–47, https://doi.org/10.1609/aimag.v18i4.1324.
[46] J. Mendes-Moreira, C. Soares, A.M. Jorge, J.F.D. Sousa, Ensemble approaches for regression, ACM Comput. Surv. 45 (1) (2012) 1–40, https://doi.org/10.1145/2379776.2379786, URL: https://dl.acm.org/doi/10.1145/2379776.2379786.
[47] T. Asakura, Y. Date, J. Kikuchi, Application of ensemble deep neural network to metabolomics studies, Anal. Chim. Acta 1037 (2018) 230–236, https://doi.org/10.1016/j.aca.2018.02.045, URL: https://linkinghub.elsevier.com/retrieve/pii/S0003267018302605.
[48] L. Kjær, C. Thomsen, O. Henriksen, P. Ring, M. Stubgaard, E.J. Pedersen, Evaluation of Relaxation Time Measurements by Magnetic Resonance Imaging, Acta Radiol. 28 (3) (1987) 345–351, https://doi.org/10.3109/02841858709177361, URL: http://informahealthcare.com/doi/abs/10.3109/02841858709177361.
[49] Magritek GmbH (2021). URL: www.magritek.com.
[50] W.H. Press, Numerical Recipes, 3rd ed., Cambridge University Press, Cambridge, 2014.
[51] S.J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191, URL: http://ieeexplore.ieee.org/document/5288526/.
[52] M. Weiger, T. Speck, M. Fey, Gradient shimming with spectrum optimisation, J. Magn. Reson. 182 (1) (2006) 38–48, https://doi.org/10.1016/j.jmr.2006.06.006, URL: https://linkinghub.elsevier.com/retrieve/pii/S1090780706001583.
[53] J. Schlenke, L. Hildebrand, S. Felsinger, B. Reusch, L. Brecker, Automated Signal Detection as Tool to Evaluate Magnetic Field Homogeneity from Fourier Transformed Proton NMR Spectra, Appl. Magn. Reson. 44 (6) (2013) 745–758, https://doi.org/10.1007/s00723-013-0442-1, URL: http://link.springer.com/10.1007/s00723-013-0442-1.
[54] J.J. Helmus, C.P. Jaroniec, Nmrglue: an open source Python package for the analysis of multidimensional NMR data, J. Biomol. NMR 55 (4) (2013) 355–367, https://doi.org/10.1007/s10858-013-9718-x, URL: http://link.springer.com/10.1007/s10858-013-9718-x.
[55] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, Adv. Neural Informat. Process. Syst. 32 (2019) 8024–8035. URL: https://arxiv.org/pdf/1912.01703.pdf, http://arxiv.org/abs/1912.01703.
[56] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J.E. Gonzalez, I. Stoica, Tune: A Research Platform for Distributed Model Selection and Training, ICML. URL: http://ray.readthedocs.io/en/latest/tune.html. http://arxiv.org/abs/1807.05118.
[57] T. Elsken, J.H. Metzen, F. Hutter, Neural Architecture Search: A Survey, J. Machine Learn. Res. 20 (2018) 1–21, URL http://arxiv.org/abs/1808.05377.
[58] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. URL: http://arxiv.org/abs/1412.6980.
[59] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions, SIAM J. Optim. 9 (1) (1998) 112–147, https://doi.org/10.1137/S1052623496303470, URL: https://epubs.siam.org/page/terms, http://epubs.siam.org/doi/10.1137/S1052623496303470.
[60] A. Adadi, M. Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE Access 6 (2018) 52138–52160, https://doi.org/10.1109/ACCESS.2018.2870052, URL: https://ieeexplore.ieee.org/document/8466590/.
[61] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, Adv. Neural Informat. Process. Syst. (2017- (2017)) 5999–6009, URL: http://arxiv.org/abs/1706.03762.
[62] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, M. Shah, Transformers in Vision: A Survey, ACM Comput. Surv., 2021. https://doi.org/10.1145/3505244. URL: http://arxiv.org/abs/2101.01169.
[63] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2016-Decem, IEEE, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90. URL: http://ieeexplore.ieee.org/document/7780459/.
[64] G. Menghani, Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better, ACM Comput. Surv., 2021. URL: http://arxiv.org/abs/2106.08962.
[65] J.D. Garrett, garrettj403/SciencePlots (2021). https://doi.org/10.5281/zenodo.4106649.