



# DEEP Picker1D and Voigt Fitter1D: a versatile tool set for the automated quantitative spectral deconvolution of complex 1D-NMR spectra

Da-Wei Li<sup>1</sup>, Lei Bruschweiler-Li<sup>1</sup>, Alexandar L. Hansen<sup>1</sup>, and Rafael Brüschweiler<sup>1,2,3</sup>

<sup>1</sup>Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, USA

<sup>2</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

<sup>3</sup>Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, USA

**Correspondence:** Da-Wei Li (lidawei@gmail.com) and Rafael Brüschweiler (bruschweiler.1@osu.edu)

Received: 7 November 2022 – Discussion started: 15 November 2022

Revised: 15 December 2022 – Accepted: 16 December 2022 – Published: 8 February 2023

**Abstract.** The quantitative deconvolution of 1D-NMR spectra into individual resonances or peaks is a key step in many modern NMR workflows as it critically affects downstream analysis and interpretation. Depending on the complexity of the NMR spectrum, spectral deconvolution can be a notable challenge. Based on the recent deep neural network DEEP Picker and Voigt Fitter for 2D NMR spectral deconvolution, we present here an accurate, fully automated solution for 1D-NMR spectral analysis, including peak picking, fitting, and reconstruction. The method is demonstrated for complex 1D solution NMR spectra showing excellent performance also for spectral regions with multiple strong overlaps and a large dynamic range whose analysis is challenging for current computational methods. The new tool will help streamline 1D-NMR spectral analysis for a wide range of applications and expand their reach toward ever more complex molecular systems and their mixtures.

## 1 Introduction

One of the major strengths of NMR spectroscopy is its broad applicability to a vast range of molecular systems in solution or in the solid state. Because the nuclei of many atoms in molecules are NMR-active, such as hydrogen atoms, the information content of NMR spectra is uniquely rich, allowing studies of molecular composition, interactions, structure and dynamics at atomic detail. Due to its quantitative nature, NMR is also highly suitable for the analysis of molecular mixtures for component identification and quantification with application in metabolomics (Markley et al., 2017) and for monitoring of industrial chemical and biochemical processes (Wang et al., 2021).

Despite enormous methodological progress made over many decades of NMR research that has resulted in a vast collection of different NMR experiments, in many NMR facilities the most popular choice remains the standard one-dimensional (1D) <sup>1</sup>H proton NMR experiment. This is the

result of several factors, such as good sensitivity, short measurement time (often associated with a low user fee), straightforward processing, and easy and dependable implementation on different types of NMR spectrometers. However, due to the richness of the resulting <sup>1</sup>H NMR spectrum in many samples, it is prone to various amounts of spectral overlaps, which is the overlap of two or more resonances, rendering the identification and quantification of the underlying resonances challenging (Giraudeau, 2017).

Because the first step of the analysis of almost every NMR spectrum consists of the identification of the individual resonances, spectral crowding often makes the process incomplete, ambiguous or even impossible. For many years, spectral analysis has been routinely assisted by computer software to perform useful tasks like peak picking and peak integration, thereby speeding up the analysis process by supporting human experts during this process (Nelson and Brown, 1989; Martin, 1994; Cobas et al., 2013). A number of commercial general-purpose software packages are

available for the analysis of 1D  $^1\text{H}$  NMR spectra, such as the ACD/NMR workbook suite (<https://www.acdlabs.com/>, last access: 6 January 2023), the AMIX software (<https://www.bruker.com>, last access: 6 January 2023), the Chemomx NMR suite (<https://www.chemomx.com>, last access: 6 January 2023), and MNova NMR (<https://www.mestrelab.com>, last access: 6 January 2023). Recent developments in NMR-based metabolomics, which oftentimes involve highly complex  $^1\text{H}$  NMR spectra, have led to a proliferation of academic software for the (semi-)automated analysis of such spectra, including MetaboLab (Ludwig and Gunther, 2011), BATMAN (Hao et al., 2014), Bayesil (Ravanbakhsh et al., 2015), AQuA (Rohnisch et al., 2018), ASICS (Lefort et al., 2019), rDolphin (Canueto et al., 2018) and MetaboDecon1D (Hackl et al., 2021). Some of these programs are suitable for untargeted compound identification, whereas others only map those spectral features that are contained in a pre-defined metabolite spectral database.

For a fully quantitative spectral analysis, numerical lineshape fitting has become the method of choice, using a parametric representation of each resonance in the spectrum (Higinbotham and Marshall, 2001; Smith, 2017; Sokolenko et al., 2019). Commonly used lineshapes are Lorentzian, Gaussian, and Voigt profiles that may explicitly include truncation or apodization effects, such as sinc wiggles (Dudley et al., 2020). Because essentially all fitting software relies on a local nonlinear least squares minimization between the model and the experimental spectrum, such as a Levenberg–Marquardt minimizer, accurate line position and linewidth for each resonance as input parameters are of paramount importance. Because such information is hard to obtain by automated computational approaches alone, lineshape fitting often requires significant interactive intervention by a human expert. This applies in particular to spectral regions with significant peak overlap, manifested, for example, by one or several shoulder peaks and a large dynamic range. Although sophisticated mathematical peak picking algorithms have been developed that identify realistic peak positions (Cobas et al., 2013), they work best for well-resolved peaks or peaks with moderate overlap but tend to fail in the case of strong overlaps and overlaps involving three or more peaks.

Recent applications of machine-learning methods, in particular of deep neural networks (DNNs), have shown qualitative progress in the ability to deconvolute complex multidimensional NMR spectra (Li et al., 2022b). In the case of DEEP Picker, training was exclusively based on a library containing 5000 synthetic 1D test spectra consisting of three to nine individual Voigt-shaped peaks with random amplitudes and positions amounting to a collection of training spectra with a wide range of spectral overlap (Li et al., 2021). The algorithm was then generalized to two-dimensional (2D) NMR spectra as encountered in many protein NMR and metabolomics applications.

In the present work, we introduce DEEP Picker for untargeted applications to complex 1D-NMR spectra, includ-

ing complex biological mixtures. The deconvolution power of DEEP Picker1D is demonstrated for spectra with various amounts of overlap and how it can be paired with the nonlinear least squares fitting software Voigt Fitter1D for a fully quantitative deconvolution of the input spectra. The computer codes of DEEP Picker1D and Voigt Fitter1D are made publicly available.

## 2 Materials and methods

### 2.1 Sample preparation

#### Glucose sample

2 mM glucose (from Sigma-Aldrich) was prepared in  $\text{D}_2\text{O}$  before 600  $\mu\text{L}$  was transferred to a 5 mm NMR tube for NMR data collection.

#### Mouse urine sample

A frozen mouse urine sample was thawed on ice. An aliquot of 178  $\mu\text{L}$  mouse urine was mixed with 20  $\mu\text{L}$  sodium phosphate buffer (500 mM) in  $\text{D}_2\text{O}$  and 2  $\mu\text{L}$  DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid from 10 mM stock solution prepared in  $\text{D}_2\text{O}$ ) with a final pH of 7.4; 200  $\mu\text{L}$  of the final sample was transferred to a 3 mm NMR tube for NMR data collection.

### 2.2 NMR experiments and processing

All NMR spectra were collected at 298 K on Bruker AVANCE III HD 850 MHz spectrometers equipped with a cryogenically cooled TCI probe. A 1D  $^1\text{H}$  NOESY glucose spectrum was recorded with a total of 32 768 complex data points and 64 scans. The relaxation delay between consecutive scans was 12 s, the spectral width was 13 ppm, and the transmitter frequency offset was set to 4.7 ppm. NMR data were zero-filled 4-fold, apodized using a cosine-squared window function, Fourier-transformed, and phase-corrected using the Bruker Topspin 4 software.

A 1D  $^1\text{H}$  mouse urine spectrum was recorded with the Bruker standard pulse sequence “zgesgppp”, which is a  $^1\text{H}$  perfect-echo 1D experiment with excitation-sculpting water suppression, with a total of 53 190 complex data points and 64 scans. The relaxation delay between consecutive scans was 4 s. The spectral width was 25 ppm, with the transmitter frequency offset set to 4.7 ppm. The NMR free induction decay was zero-filled 2-fold, apodized using a  $2\pi$ -Kaiser window function, Fourier-transformed, and phase-corrected using NMRPipe (Delaglio et al., 1995).

A 2D  $^{13}\text{C}$ - $^1\text{H}$  high-resolution HSQC spectrum of mouse urine was recorded with Bruker pulse program “hsqcetg-psisp2.2”, which is a sensitivity-enhanced  $^{13}\text{C}$ - $^1\text{H}$  HSQC with bi-level adiabatic decoupling; 3072 total complex data points in the  $^1\text{H}$   $t_2$  dimension and 512 total complex points

in the  $^{13}\text{C}$   $t_1$  dimension were recorded. For each  $t_1$  increment 16 scans were recorded, and the relaxation delay between consecutive scans was set to 1.5 s. The spectral widths along the  $^1\text{H}$  and  $^{13}\text{C}$  dimensions were 18 and 185 ppm, respectively. The transmitter frequency offsets were 4.7 and 82.5 ppm, respectively. NMR data were zero-filled 8-fold in both dimensions, apodized using a  $2\pi$ -Kaiser window function, Fourier-transformed, and phase-corrected using NMR-Pipe (Delaglio et al., 1995).

### 2.3 Deep neural network DEEP Picker1D and Voigt Fitter1D

DEEP Picker1D is a deep neural network that was trained on a library of 5000 synthetic 1D-NMR spectra containing between three and nine peaks with a Voigt lineshape and variable amounts of overlaps (Li et al., 2021). In the original work, DEEP Picker was specifically adapted for the analysis of 2D NMR spectra and subsequently combined with the Voigt Fitter software for the quantitative analysis of 2D NMR metabolomics spectra either as standalone software or incorporated into the public web server COLMARq (Li et al., 2022a). Briefly, DEEP Picker1D is a convolutional neural network, which was trained using TensorFlow v1.3 (Abadi et al., 2016), taking a 1D spectrum as input. It contains seven hidden convolutional layers, one hidden max-pooling layer, and two parallel output layers with a total of 8037 trainable parameters. A convolutional output classifier layer with SoftMax activation classifies every input data point by assigning an individual score for three peak classes (main peaks in class 2, shoulder peaks in class 1, no peak in class 0). The class with the maximal score is then chosen as the predicted class with the numerical score as a quantitative measure of confidence of the predicted class for each data point of the input spectrum. For any data point predicted to be a peak (class 2 or 1), DEEP Picker1D also predicts the sub-pixel peak position relative to the on-grid points, peak amplitude, peak width, and Lorentzian vs. Gaussian components to the Voigt shape using a convolutional output regressor layer. Although DEEP Picker1D is a rather accurate predictor of peak parameters in its own right, these values can be further refined by the Voigt Fitter1D software by performing a nonlinear least square fit of the original input 1D spectrum in terms of Voigt peak shapes using the DEEP Picker1D output peak parameters as input. Voigt Fitter1D is essentially a 1D version of the 2D Voigt Fitter software published previously (Li et al., 2022a). DEEP Picker1D paired with Voigt Fitter1D results in a fully quantitative representation of the input 1D-NMR spectrum in terms of a finite set of 1D Voigt-shaped peaks.

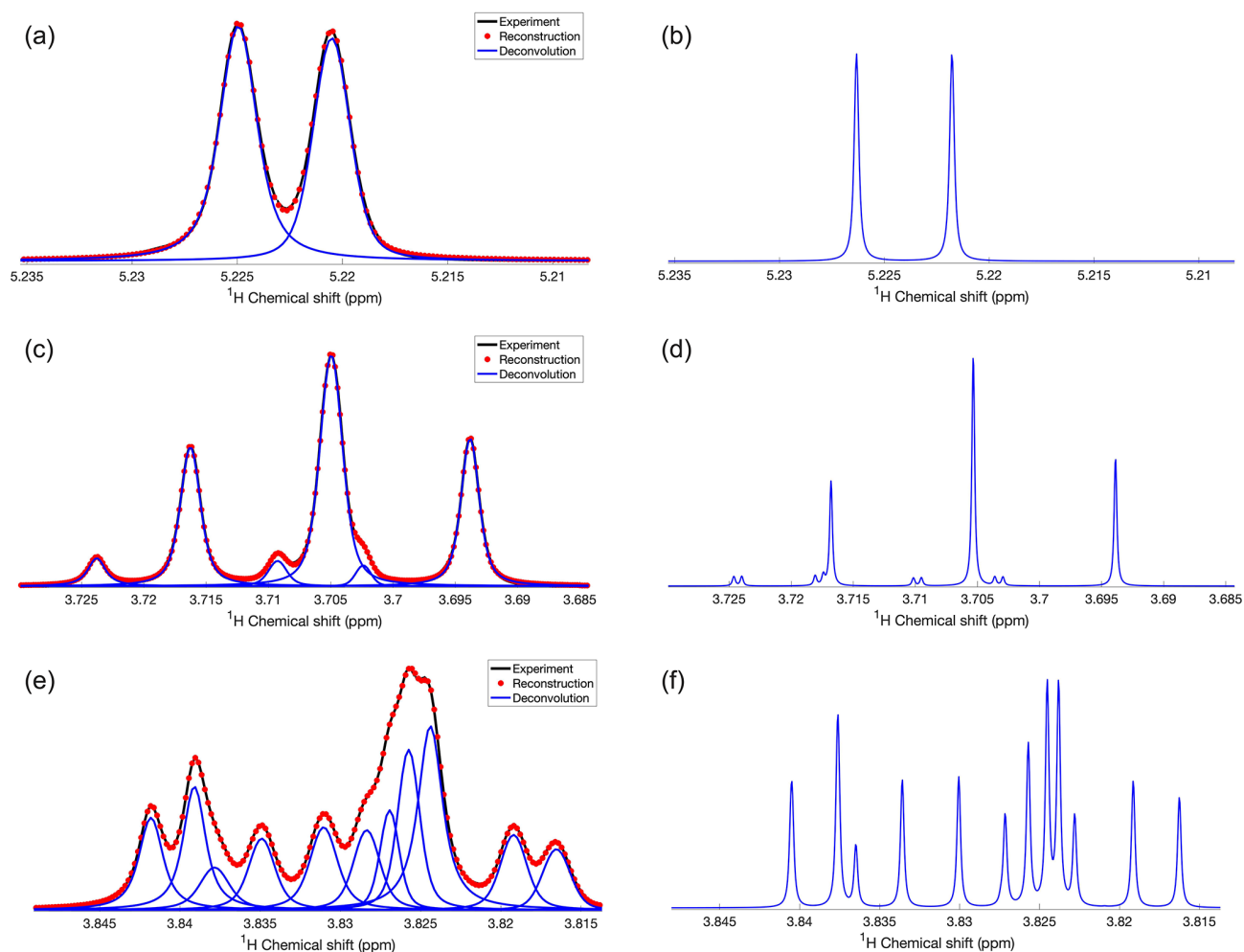
The input spectrum for DEEP Picker1D needs to be preprocessed in standard fashion, including phase correction, baseline correction, zero filling, apodization and Fourier transformation. DEEP Picker1D contains two models whereby model 1 (model 2) has optimal performance

when the digital resolution is sufficiently high around 12 (8) points per peak (PPP). Deep Picker1D performs best for peaks with a moderate to high signal-to-noise ratio ( $S/N$ ) and lineshapes that closely follow Voigt profiles with a  $S/N > 10$  where the noise level is defined as the standard deviation of the spectrum in a peak-free region. In the presence of significant amounts of noise, non-negligible lineshape distortions, such as those caused by temperature fluctuations, or suboptimal shimming during data collection, Deep Picker1D may pick some false peaks, for example, by interpreting lineshape distortions as shoulder peaks. Voigt Fitter1D has built-in tools to remove spectral features from its peak list when one of the following situations occurs: (i) a fitted peak is too wide; i.e., the peak width is larger than the fitting region, or it becomes too narrow, i.e., the peak width is less than 1 point; (ii) a fitted peak strongly overlaps with another peak, so that merging of two peaks into a single peak causes a minimal change in the fitting error. Deep Picker1D and Voigt Fitter1D together provide a self-sufficient spectral analysis tool set for the complete deconvolution of 1D spectra into individual peaks. Peak parameters such as peak position, peak height and peak volume can then be directly used for downstream analysis, such as compound identification and quantitative NMR applications when incorporated into a quantitative NMR workflow (qNMR). Because error estimation is an important part of any quantitative data analysis, Monte Carlo-based error propagation is implemented in Voigt Fitter1D as an option. It performs repetitive fitting of the reconstructed spectrum after adding random noise with the same standard deviation as that of the experimental input spectrum for each round of fitting. The output from this error estimation procedure contains the fitting parameters from each round from which the uncertainty of each peak parameter is obtained.

## 3 Results

DEEP Picker1D and Voigt Fitter1D performance is first demonstrated for glucose in  $\text{D}_2\text{O}$  (Fig. 1).

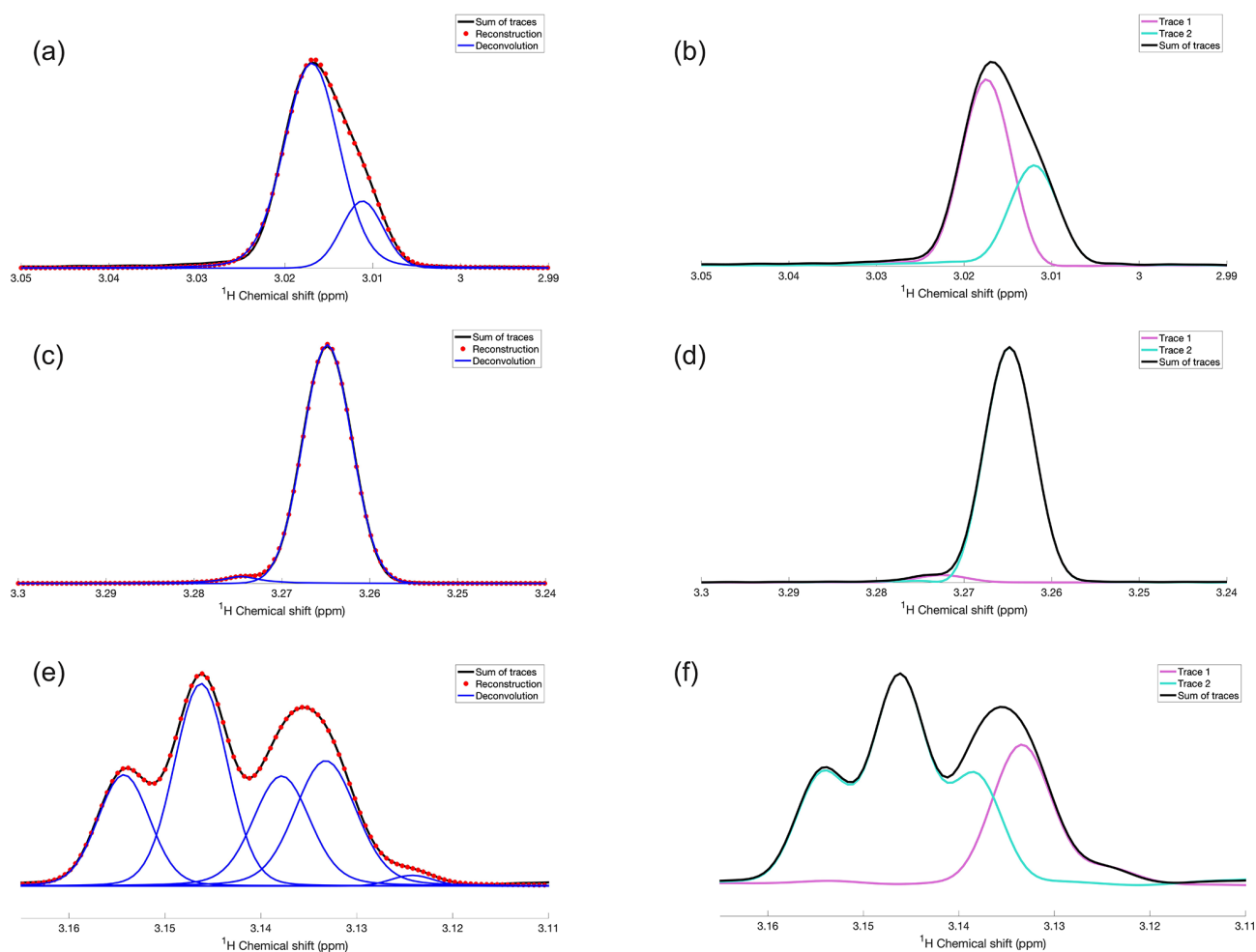
Because glucose populates two non-equivalent isomers  $\alpha$ -glucose and  $\beta$ -glucose with different relative populations that interconvert on a slow timescale and displays strong coupling effects even at a high magnetic field, the deconvolution of its 1D  $^1\text{H}$  NMR spectrum is notoriously difficult. Figure 1 shows selected regions of the 1D  $^1\text{H}$  NMR glucose spectrum with variable amounts of peak overlap. The experimental spectra (black) along with the deconvolution results (blue) are shown in the left column (Fig. 1a, c, and e). The right column (Fig. 1b, d, and f) shows the corresponding spectral regions derived from quantum-mechanical spin simulations using chemical shifts and scalar J-couplings obtained from the GISSMO library (Dashti et al., 2018). An artificially slow, uniform transverse  $R_2$  relaxation rate of  $0.6\text{ s}^{-1}$  was applied to the simulated free induction decays (FIDs) so that, after



**Figure 1.** Demonstration of DEEP Picker1D and Voigt Fitter1D for selected regions of a 1D  $^1\text{H}$  spectrum of glucose. (a, c, and e) Experimental and reconstructed spectra are depicted in black lines and red dots, respectively. Deconvoluted individual peaks are depicted as blue lines. (b, d, and f) Simulated spectra including strong coupling effects with the spin Hamiltonian constructed based on parameters taken from the GISSMO website at the same  $B_0$  field strength (850 MHz  $^1\text{H}$  frequency) as in the experiments. Transverse  $R_2$  relaxation rates were uniformly set to a low value of  $0.6\text{ s}^{-1}$  to obtain a very high-resolution spectrum for better comparison with DEEP Picker. Pairs of panels (a, b), (c, d), and (e, f) show the same 1D spectral regions. DEEP Picker1D and Voigt Fitter1D correctly deconvoluted the experimental spectra for the simple region (a) and more complex regions (c, e). A few peaks cannot be deconvoluted because of strong spectral overlap, such as the small peak around 3.717 ppm in (d) and the peak around 3.823 ppm in (f). The deconvolution by DEEP Picker1D was performed with model 2 with a PPP number of 9.

Fourier transformation, the resulting spectrum has sharp lines for easy recognition of the individual peaks and for the comparison with the automated deconvolution results. Figure 1a starts out with a symmetric doublet centered at 5.223 ppm, which is accurately picked and fitted by DEEP Picker1D and Voigt Fitter1D, in agreement with the simulation results in Fig. 1b. Figure 1c and d show a triplet centered around 3.705 ppm whereby the strong central peak overlaps with two much smaller peaks on each side, which are correctly picked and fitted. According to the simulation, there is another small peak around 3.718 ppm, which however strongly overlaps with a much stronger peak at 3.716 ppm and therefore could not be identified by DEEP Picker1D. This small

peak also cannot be discerned by visual inspection (note that the small J-splitting of the small peak in the simulated spectrum of Fig. 1d is not resolved in the experimental spectrum of Fig. 1c). The most complex region of the glucose spectrum (3.81–3.85 ppm) is depicted in Fig. 1e along with its deconvolution, which is in very good agreement with the simulated peaks (Fig. 1f). The neural network does a remarkable job in identifying the small peak at 3.838 ppm, which only gives rise to a very faint shoulder peak of its downfield-shifted larger neighbor. The broad, somewhat oddly shaped spectral feature from 3.82 to 3.83 ppm in the experiment is deconvoluted into four individual peaks whereby the small peak found in the simulation at 3.823 ppm was not deconvoluted



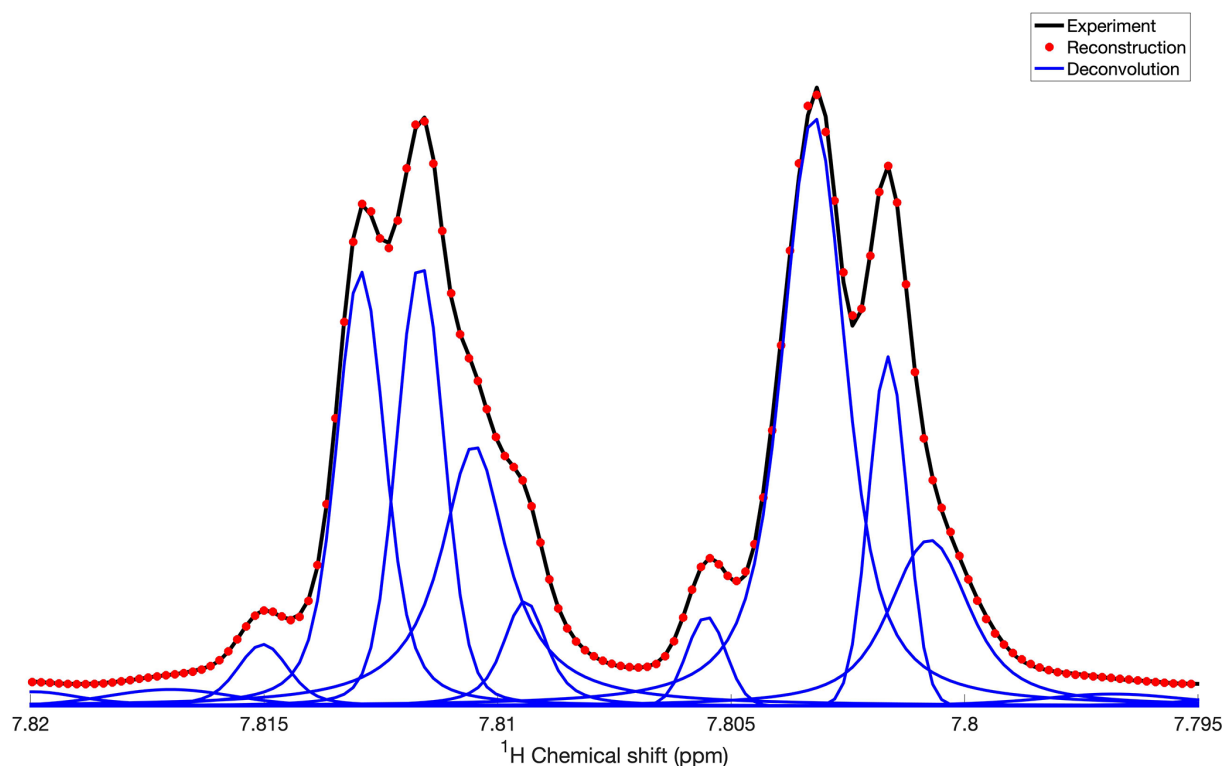
**Figure 2.** Application of DEEP Picker1D and Voigt Fitter1D to selected regions of 1D spectra, which were generated by adding two selected traces along the direct  $^1\text{H}$  dimension from the experimental 2D  $^{13}\text{C}$ - $^1\text{H}$  HSQC of a mouse urine sample. (a, c, and e) Experimental and reconstructed spectra are depicted as black lines and red dots, respectively. Deconvoluted individual peaks are depicted as blue lines. (b, d, and f) The two HSQC traces and their sum are depicted as purple, cyan, and black lines, respectively. Pairs of panels (a, b), (c, d), and (e, f) show the same 1D spectral regions for comparison. The deconvolution by DEEP Picker1D was performed with model 1 with a PPP number of 12.

by DEEP Picker1D because it overlaps too strongly with the main peak at 3.824 ppm. This is consistent with the general rule that two peaks whose positions differ within their linewidths are hard to deconvolute, especially when their amplitudes significantly differ from each other.

To assess the deconvolution accuracy of our tool, we constructed experimental spectra with overlaps from resolved spectra by co-adding traces of a  $^{13}\text{C}$ - $^1\text{H}$  HSQC spectrum of mouse urine along the direct  $^1\text{H}$  detection dimension at a fixed  $^{13}\text{C}$  chemical shift. Selected examples of overlapping peaks, both in isolation and as a superposition, are shown in Fig. 2. The left column (Fig. 2a, c, and e) shows the experimental superpositions (black) together with their deconvolution (blue) and the full spectral reconstruction (red), which can be directly compared with the individual traces (purple and cyan) in the right column (Fig. 2b, d, and f). Figure 2a

and b show two strongly overlapped peaks of different amplitudes giving rise to a sum peak with a noticeable protrusion on its right flank, which are accurately deconvoluted and fitted by DEEP Picker1D and Voigt Fitter1D. Figure 2c and d show a similar scenario, except that the amplitude ratio of the two peaks is around 35 : 1, which is much larger than in Fig. 2a and b. Again, deconvolution was achieved with high accuracy. Figure 2e and f demonstrate the deconvolution capacity for a challenging case of four moderately to strongly overlapped peaks. Although the peak at 3.139 ppm is wedged between two stronger peaks, it is successfully extracted by the peak picking and fitting algorithms. The final example (Fig. 3) shows a region of the mouse urine spectrum (black) along with the deconvolution (blue) and reconstruction result. The algorithm deconvolutes the spectrum by identifying not only the main peaks, but also all the minor peaks, includ-





**Figure 3.** Application of DEEP Picker1D and Voigt Fitter1D to a spectral region of the 1D  $^1\text{H}$  spectrum of mouse urine. Experimental and reconstructed spectra are depicted as black lines and red dots, respectively. Deconvoluted individual peaks are depicted as blue lines. The deconvolution by DEEP Picker1D was performed with model 2 with a PPP number of 8.

ing the peak at 7.809 ppm, with confidence, demonstrating the potential of the proposed deconvolution method in practice when encountering spectra with highly overlapped regions, such as those routinely collected for urine and other complex biofluids in the context of metabolomics.

#### 4 Discussion and conclusion

In the vast majority of modern NMR applications, one of most critical steps in NMR spectral analysis is the identification of individual peaks along with their quantitative parametrization by lineshape fitting. The result of this procedure often dictates the usefulness, and ultimately the success, of the collected experiment. Traditional peak picking methods rely on clearly defined mathematical criteria, such as the properties of the first and second derivatives of the spectrum, to identify individual peaks. These criteria are often too rigid to deal with spectral overlap scenarios encountered in practice. After proper training, a deep neural network like DEEP Picker1D, on the other hand, has a stunning ability to track major and minor spectral features, surpassing the capacity of most human NMR practitioners. Through the combination of advanced machine learning by the convolutional deep neural network DEEP Picker1D and a peak fitting routine Voigt Fitter1D, it was demonstrated how 1D-NMR spectral

features of variable complexity can be deconvoluted into individual resonances in a reliable and accurate manner. The success rate of the method depends on the quality of spectra that can be affected by sample preparation, NMR data acquisition, and pre-processing. This concerns the elimination or suppression of the solvent signal or of a prominent background caused, for example, by the presence of a macromolecular matrix in the sample. Although apodization, zero filling, and phase and baseline correction are standard steps during data processing, they need to be applied judiciously to prevent suboptimal performance of spectral deconvolution and fitting. Phase errors of up to about  $3^\circ$  can be tolerated, but for larger phase distortions, DEEP Picker1D may interpret asymmetries in the peak shapes as shoulder peaks. Similarly, poor shimming of higher-order shims, especially  $z^2$  and  $z^4$ , can lead to systematic peak asymmetries across the spectrum, which DEEP Picker1D may interpret as shoulder peaks. In order to accurately recognize peak shapes, DEEP Picker1D requires an adequate digital resolution, which is around 8 or 12 points across a single peak, depending also on the chosen DEEP Picker1D model. If needed, lower-resolution spectra can be easily subjected to appropriated zero filling to meet this criterion. Peak shapes should follow in good approximation Voigt profiles, which can be achieved by the application of common window functions such as those described for the

processing of the spectra in this work (cosine-squared and  $2\pi$ -Kaiser window functions; see the “Materials and methods” section). As discussed previously (Li et al., 2022a), the computational time of Voigt Fitter1D scales linearly with the number of peaks, allowing rapid fitting of complex 1D spectra with even thousands of peaks. The fitting of the 1D mouse urine spectrum with a total of 4500 Voigt-shaped peaks took about 1 min on a standard desktop computer. Like all nonlinear optimization software, Voigt Fitter1D cannot guarantee that the final solution is the global  $\chi^2$  minimum. Therefore, a nearly complete list of high-quality initial peaks returned by DEEP Picker1D that match the ground truth as closely as possible is key for the success of Voigt Fitter1D.

A surge in metabolomics research over recent years has spurred the development of advanced quantitative tools for the analysis of complex NMR spectra for both 1D and 2D spectra. Some metabolomics software (Hao et al., 2014; Ravanbakhsh et al., 2015) is specifically geared toward the quantification of specific metabolites with known reference spectra, limiting their application to specific samples only, such as serum. In the case of DEEP Picker1D and Voigt Fitter1D, the analysis is performed in a fully untargeted manner, i.e., without any molecular spectral templates, allowing its application to essentially any NMR spectrum that consists of resonances with Voigt lineshapes. The deconvolution results can then be further analyzed, for example, by querying against a spectral database or for quantitation of mixture component concentrations. In the case of a cohort of samples, the DEEP Picker1D and Voigt Fitter1D results can be used for univariate or multivariate statistical analysis for the assessment of statistically significant differences between cohorts. Metabolomics query capabilities for the analysis of the output of DEEP Picker1D and Voigt Fitter1D, which will also take into account peak shifts caused, for example, by pH differences between samples, are currently under development. The DEEP Picker1D and Voigt Fitter1D software can also be applied to a pseudo-2D series of 1D spectra for the extraction of longitudinal  $R_1$ , transverse  $R_2$  relaxation parameters or translational diffusion constants by diffusion-ordered NMR (Johnson, 1999). The unique strength of the combination of DEEP Picker1D with Voigt Fitter1D is their ability to accurately deconvolute and reconstruct NMR spectra of generic origin ranging from well-resolved to highly crowded, which should fulfill the growing needs in a wide range of contemporary NMR applications.

**Data availability.** The spectra shown in this paper are for illustration purposes and are not essential for using the software. Please add “Test spectra are available from the authors upon request.”

**Code availability.** The 1D versions of DEEP Picker and Voigt Fitter are implemented in C/C++ and are now part of the DEEP Picker package, which is freely available from <https://github.com/>

lidawei1975/deep (last access: 6 January 2023) under the GNU General Public License Agreement. They can also be accessed conveniently as a web server at <https://spin.ccic.osu.edu/index.php/deep1d> (last access: 6 January 2023), which in addition provides an intuitive interactive interface for the visual inspection of the results.

**Author contributions.** DWL and RB conceived and designed the project. DWL developed the machine learning framework and wrote the program. LBL prepared samples for NMR experiments. LBL and ALH performed NMR experiments. All the authors contributed to model development and data analysis. DWL and RB prepared the manuscript.

**Competing interests.** The authors have the following competing interests: we are considering filing a patent application for the licensing of the DEEP Picker1D software to commercial users.

**Disclaimer.** Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Financial support.** This work was supported by the National Institutes of Health (NIH) (grant NIH R35GM139482) and the National Science Foundation (NSF) (grant MCB-2103637). All NMR experiments were performed at the shared CCIC NMR facility at the Ohio State University.

**Review statement.** This paper was edited by Nicola Salvi and reviewed by Milo Westler and one anonymous referee.

## References

- Abadi, M., Barham, P., Chen, J. M., Chen, Z. F., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. Q.: TensorFlow: A system for large-scale machine learning, in: Proceedings of Osd’16: 12th Usenix Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 265–283, WOS:000569062400017, 2016.
- Canueto, D., Gomez, J., Salek, R. M., Correig, X., and Canellas, N.: rDolphin: a GUI R package for proficient automatic profiling of 1D (1)H-NMR spectra of study datasets, *Metabolomics*, 14, 24, <https://doi.org/10.1007/s11306-018-1319-y>, 2018.
- Cobas, C., Seoane, F., Vaz, E., Bernstein, M. A., Dominguez, S., Perez, M., and Sykora, S.: Automatic assignment of 1H-NMR spectra of small molecules, *Magn. Reson. Chem.*, 51, 649–654, <https://doi.org/10.1002/mrc.3995>, 2013.
- Dashti, H., Wedell, J. R., Westler, W. M., Tonelli, M., Aceti, D., Amarasinghe, G. K., Markley, J. L., and Eghbalnia, H. R.: Applications of Parametrized NMR Spin Systems of Small Molecules, *Anal. Chem.*, 90, 10646–10649, <https://doi.org/10.1021/acs.analchem.8b02660>, 2018.

- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A.: NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR*, 6, 277–293, <https://doi.org/10.1007/BF00197809>, 1995.
- Dudley, J. A., Park, S., MacDonald, M. E., Fetene, E., and Smith, C. A.: Resolving overlapped signals with automated Fit-NMR analytical peak modeling, *J. Magn. Reson.*, 318, 106773, <https://doi.org/10.1016/j.jmr.2020.106773>, 2020.
- Giraudeau, P.: Challenges and perspectives in quantitative NMR, *Magn. Reson. Chem.*, 55, 61–69, <https://doi.org/10.1002/mrc.4475>, 2017.
- Hackl, M., Tauber, P., Schweda, F., Zacharias, H. U., Altenbuchinger, M., Oefner, P. J., and Gronwald, W.: An R-Package for the Deconvolution and Integration of 1D-NMR Data: MetaboDecon1D, *Metabolites*, 11, 452, <https://doi.org/10.3390/metabo11070452>, 2021.
- Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., and Ebbels, T. M.: Bayesian deconvolution and quantification of metabolites in complex 1D-NMR spectra using BATMAN, *Nat. Protoc.*, 9, 1416–1427, <https://doi.org/10.1038/nprot.2014.090>, 2014.
- Higinbotham, J. and Marshall, I.: NMR lineshapes and lineshape fitting procedures, *Ann. R. NMR S.*, 43, 59–120, [https://doi.org/10.1016/s0066-4103\(01\)43009-2](https://doi.org/10.1016/s0066-4103(01)43009-2), 2001.
- Johnson, C. S.: Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications, *Prog. Nucl. Mag. Res. Sp.*, 34, 203–256, [https://doi.org/10.1016/s0079-6565\(99\)00003-5](https://doi.org/10.1016/s0079-6565(99)00003-5), 1999.
- Lefort, G., Liaubet, L., Canlet, C., Tardivel, P., Pere, M. C., Quesnel, H., Paris, A., Iannuccelli, N., Vialaneix, N., and Servien, R.: ASICS: an R package for a whole analysis workflow of 1D <sup>1</sup>H NMR spectra, *Bioinformatics*, 35, 4356–4363, <https://doi.org/10.1093/bioinformatics/btz248>, 2019.
- Li, D., Hansen, A. L., Yuan, C., Bruschweiler-Li, L., and Bruschweiler, R.: DEEP Picker is a Deep Neural Network for Accurate Deconvolution of Complex Two-Dimensional NMR Spectra, *Nat. Commun.*, 12, 5229, <https://doi.org/10.1038/s41467-021-25496-5>, 2021.
- Li, D. W., Leggett, A., Bruschweiler-Li, L., and Bruschweiler, R.: COLMARq: A Web Server for 2D NMR Peak Picking and Quantitative Comparative Analysis of Cohorts of Metabolomics Samples, *Anal. Chem.*, 94, 8674–8682, <https://doi.org/10.1021/acs.analchem.2c00891>, 2022a.
- Li, D. W., Hansen, A. L., Bruschweiler-Li, L., Yuan, C., and Bruschweiler, R.: Fundamental and practical aspects of machine learning for the peak picking of biomolecular NMR spectra, *J. Biomol. NMR*, 76, 49–57, <https://doi.org/10.1007/s10858-022-00393-1>, 2022b.
- Ludwig, C. and Gunther, U. L.: MetaboLab—advanced NMR data processing and analysis for metabolomics, *BMC Bioinformatics*, 12, 366, <https://doi.org/10.1186/1471-2105-12-366>, 2011.
- Markley, J. L., Bruschweiler, R., Edison, A. S., Eghbalian, H. R., Powers, R., Raftery, D., and Wishart, D. S.: The future of NMR-based metabolomics, *Curr. Opin. Biotech.*, 43, 34–40, <https://doi.org/10.1016/j.copbio.2016.08.001>, 2017.
- Martin, Y. L.: A Global Approach to Accurate and Automatic Quantitative Analysis of NMR Spectra by Complex Least-Squares Curve Fitting, *J. Magn. Reson. Ser. A*, 111, 1–10, <https://doi.org/10.1006/jmra.1994.1218>, 1994.
- Nelson, S. J. and Brown, T. R.: The accuracy of quantification from 1D-NMR spectra using the PIQABLE algorithm, *J. Magn. Reson.*, 84, 95–109, [https://doi.org/10.1016/0022-2364\(89\)90008-5](https://doi.org/10.1016/0022-2364(89)90008-5), 1989.
- Ravanbakhsh, S., Liu, P., Bjorndahl, T. C., Mandal, R., Grant, J. R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., and Wishart, D. S.: Accurate, fully-automated NMR spectral profiling for metabolomics, *PLOS ONE*, 10, e0124219, <https://doi.org/10.1371/journal.pone.0124219>, 2015.
- Rohnisch, H. E., Eriksson, J., Mullner, E., Agback, P., Sandstrom, C., and Moazzami, A. A.: AQUA: An Automated Quantification Algorithm for High-Throughput NMR-Based Metabolomics and Its Application in Human Plasma, *Anal. Chem.*, 90, 2095–2102, <https://doi.org/10.1021/acs.analchem.7b04324>, 2018.
- Smith, A. A.: INFOS: spectrum fitting software for NMR analysis, *J. Biomol. NMR*, 67, 77–94, <https://doi.org/10.1007/s10858-016-0085-2>, 2017.
- Sokolenko, S., Jezequel, T., Hajjar, G., Farjon, J., Akoka, S., and Giraudeau, P.: Robust 1D-NMR lineshape fitting using real and imaginary data in the frequency domain, *J. Magn. Reson.*, 298, 91–100, <https://doi.org/10.1016/j.jmr.2018.11.004>, 2019.
- Wang, R. C. C., Campbell, D. A., Green, J. R., and Cuperlovic-Culf, M.: Automatic 1D <sup>1</sup>H NMR Metabolite Quantification for Bioreactor Monitoring, *Metabolites*, 11, 157, <https://doi.org/10.3390/metabo11030157>, 2021.