Deep learning-based phase and baseline correction of 1D ¹H NMR Spectra

Simon Bruderer, Federico Paruzzo and Christine Bolliger

Bruker Switzerland AG, 8117 Fällanden, Switzerland.

Contents

Abstract

Phase and baseline corrections are important processing steps in the analysis of NMR spectra, and as a consequence, many different approaches have been developed in the past to carry out these tasks automatically. While these methods perform generally well, the performance of many of them suffers when applied to spectra with high signal densities such as e.g. proton spectra. Here, we introduce a deep learning-based method for phase and baseline correction of 1D ¹H NMR spectra. We show that this method represents a major step forward compared to previously available TopSpin solutions. The algorithm provides consistently better correction of phase and baseline both for low- and high-field spectra, even reaching human-level quality results in phase correction accuracy. The new method is available in TopSpin as command *apbk* starting from version 4.1.3, and it marks a further step towards the fully automated analysis of NMR spectra.

Introduction

The correction of phase and baseline distortions are typically among the first processing steps in the analysis of raw NMR spectra. A phase corrected spectrum has the real spectrum in absorption mode and the imaginary spectrum in dispersion mode. Baseline corrected real spectra have all regions not containing signals close to zero amplitude, within the noise level. Carrying out these corrections is important not only to obtain better looking spectra that are easier to analyze, but also to ensure correct results in applications such as e.g. quantification.

Both phase and baseline distortions are caused by physical limitations in the acquisition of the NMR data. We can split the main causes of phase distortions into two categories. First, we have differences between the reference phase and the receiver detector phase, related to signal propagation during the acquisition of the experiment (which depends for example on lengths and impedances of cables and coils). This yields a frequency independent contribution known as 0th order phase distortion. The second contribution is related to the fact that acquisition cannot start at time zero, because the magnetization starts to precess during the radiofrequency pulse. This causes a frequency dependent phase distortion known as 1st order phase distortion. Baseline distortions are caused, for example, by ring-down of probes and other electronic components, pulse breakthrough, or background signals originating from various components in the probe and samples. Although both phase and baseline distortions are considerably less severe in modern spectrometers and consoles, a certain amount lies in inherent technical properties of NMR acquisition and cannot be prevented.

Automated approaches to replace manual corrections have been suggested as early as 1969,[1] with many others following.[2-10] These approaches often correct phase and baseline separately. However, since it is more difficult to correct the phase in spectra with baseline distortion, or to estimate the baseline in spectra with residual phase distortion, the separate correction of phase and baseline often leads to inaccurate results. To overcome this limitation, Bruker has released in TopSpin 4.0.6 the command *apbk*, which performs simultaneous phase and baseline correction of X-nuclei spectra. For proton spectra, the automated correction is more challenging

than for X-nuclei due to the fact that the spectra are often more crowded and have fewer regions that can be used to estimate the baseline. This is especially true for low field NMR spectra, where the broader spectral features leave even fewer regions signal-free. Baseline detection for these edge cases is a challenging recognition task, well suited for deep learning.

Over the past few years, deep learning has found a growing number of applications in NMR spectroscopy. It has shown great flexibility and to be able to perform various tasks ranging from signal processing to structure verification (see [11] and [12] for reviews). Examples include peak picking, [13] spectral reconstruction,[14-15] and chemical shift prediction,[16] and show superior results compared to previous techniques. Bruker has recently deployed the deep learning-based command *sigreg* for signal region detection to TopSpin.[17]

Here, we introduce a deep learning-based method for automatic phase and baseline correction of 1D ¹H spectra. In this method, which is available in TopSpin and extends the existing command *apbk* to proton spectra, we use artificial intelligence - in particular, deep learning - to detect the baseline of crowded spectra. We show that it provides substantially better results compared to the frequently used *apk / abs* combination in TopSpin. The new method shows also higher consistency between the results obtained for low- and high-field NMR spectra, making it interesting both for the Fourier 80 benchtop NMR system and high-field Bruker spectrometers.

Methods

The method presented here uses a deep neural network to identify and fit the baseline, in combination with a classical approach for phase correction, akin to the methods presented in [18] and [8]. After an initial rough phase correction, the solution is iteratively improved by deriving the baseline using the deep neural network and fitting a linear phase relation. We found such a combination of deep learning and classical fitting techniques to achieve better results than a pure deep learning approach. The neural network for baseline identification and fitting has a total of around 50k weights implemented in a combination of convolutional, recurrent and dense layers. The convolutional layers are set up as inception layers similar to GoogLeNet.[19]

The deep neural network was trained using 100,000 artificially generated spectra. These synthetic spectra spread a base frequency range between 80 MHz and 800 MHz, and have features (e.g. line widths, multiplicities, and J-couplings) that match the typical values observed experimentally. Signal-to-noise ratios of the synthetic spectra are between 10 and 10,000. A baseline was added to each spectrum and used as target for supervised learning.

To test the method, we have used both experimental and synthetic data. Experimental data is the target of our method, but it is only available in limited quantity and testing relies on the manual correction of phase and baseline distortions. Synthetic spectra, on the other hand, are available in large amounts, and the exact degree of baseline and phase distortions are known. The synthetic data might however not have all the features observed experimentally, and at times contain unrealistic baseline shape. The experimental test set contains 100 experimental spectra at base frequencies between 80 MHz and 700 MHz. Thereof 22 were acquired using the Bruker Fourier 80 spectrometer. The spectra have been manually phase and baseline corrected by Bruker NMR experts. The synthetic test set contains 500 spectra and has been generated in the same way as the training data, but with stronger baseline distortion to reduce the bias in the results.

To evaluate the quality of the phase and baseline corrections separately, we have defined custom **phase and baseline scores**. We assessed the phase correction by comparing the applied correction with the target phase at each peak position. The **phase score** is then defined as the average of the absolute phase differences. For the baseline assessment, the spectrum is first corrected for the residual phase yielding a perfectly phased spectrum, which might however still contain a baseline contribution. The averaged absolute difference of that spectrum and the spectrum without baseline then yields the baseline score value. The **baseline score** is provided relative to the RMS noise. Thus, if the entire baseline is offset by one RMS noise, a baseline score of 1 is obtained. The baseline score is calculated across the whole spectrum and thus measures the quality of the baseline correction also in signal regions, where finding the baseline is more challenging, but important to obtain accurate integrals. Examples of phase and baseline scores for spectra with different degrees of phase and baseline distortions are shown in **Figure 1**.
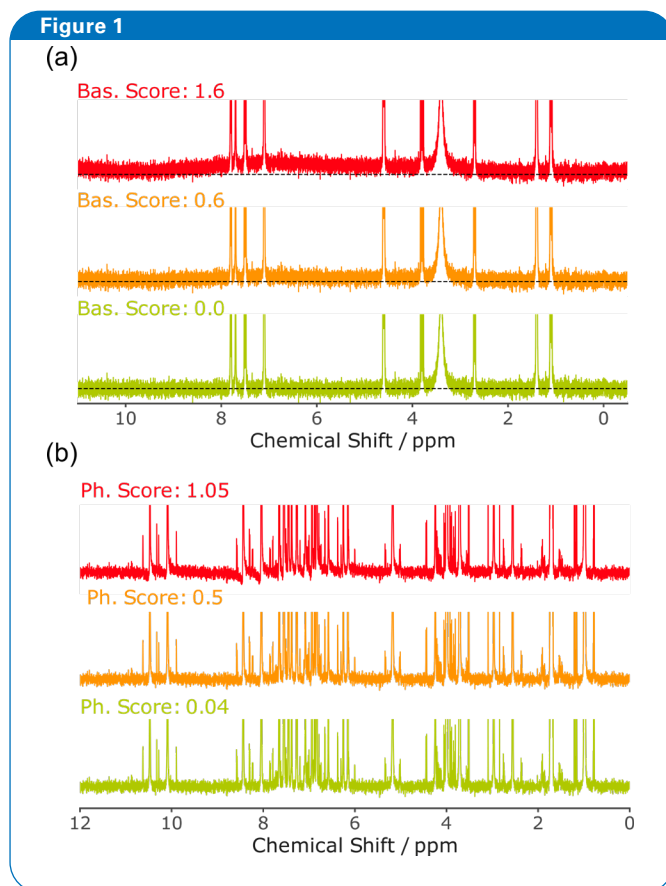

Figure 1

Figure 1: Examples of baseline and phase scores. (a) Baseline scores for a synthetic spectrum with different degrees of baseline distortion. The spectrum has no phase distortion. (b) Phase scores obtained with manual phase correction performed by three NMR experts on the same spectrum.

Results and discussion

Figure 2 compares the results of the new method (*apbk*) to the commands previously implemented in TopSpin for phase and baseline correction (*apk* and *abs*, respectively) on three experimental spectra. The new method provides a better correction of both phase and baseline distortions: the improvement in baseline correction can be seen in signal free regions, where *apbk* provides a flat spectrum, while the results by *apk / abs* show wiggles. The baseline correction by *apbk* is also better at the edges of the spectra, where *apk / abs* sometimes fails to correct the so-called smiles, a digital filter artifact. An improvement in phase correction can be seen by more symmetric lines compared to the slightly asymmetric lines in the spectra processed by *apk / abs*.

While the examples in **Figure 2** give a good visual representation of the performance of *apbk* and *apk / abs*, for a more comprehensive comparison we have evaluated the results obtained on 500 synthetic and 100 experimental ^{1}H spectra using the phase and baseline scores. To better interpret the results we have first estimated the typical accuracy obtained with manual phase correction. We have done this by distributing 35 synthetic NMR spectra with known phase distortion to 8 of our internal NMR experts for manual correction. Examples of different manual corrections and the corresponding phase scores are given in **Figure 1b**. We found that the average phase score obtained by the experts for spectra with signal-to-noise > 1,000 is 0.55°, which we will consider in the following as representative of the average manual phase correction.

The comparison of the results obtained with *apbk* and with *apk / abs* on 500 synthetic spectra is shown in **Figure 3**. The baseline scores in **Figure 3a** indicate that *apbk* provides better baseline correction. The median baseline score achieved by *apbk* is 0.27 compared to 1.29 for *apk / abs*. Both first and third quartiles are lower for *apbk* (0.14 and 0.81 for *apbk* and 0.35 and 7.7 for *apk / abs*). The phase scores are given in **Figure 3b**. The median phase score of 0.19° for *apbk* compared to 0.66° for *apk / abs* confirms improved phase correction capabilities of *apbk*. Both first and third quartiles are lower for *apbk* (0.10° and 0.48°) compared to *apk / abs* (0.26° and 3.8°). We conclude that *apbk* performs better on the synthetic data set than
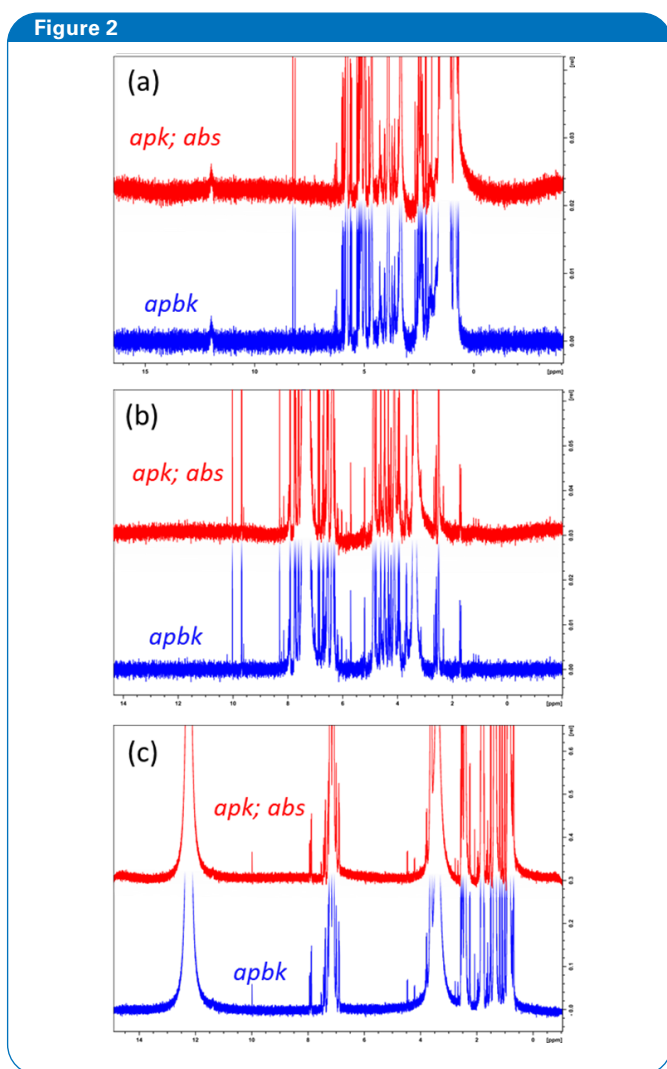
Figure 2. Examples of three experimental spectra with phase and baseline distortions corrected using *apbk* (blue) and *apk / abs* (red).
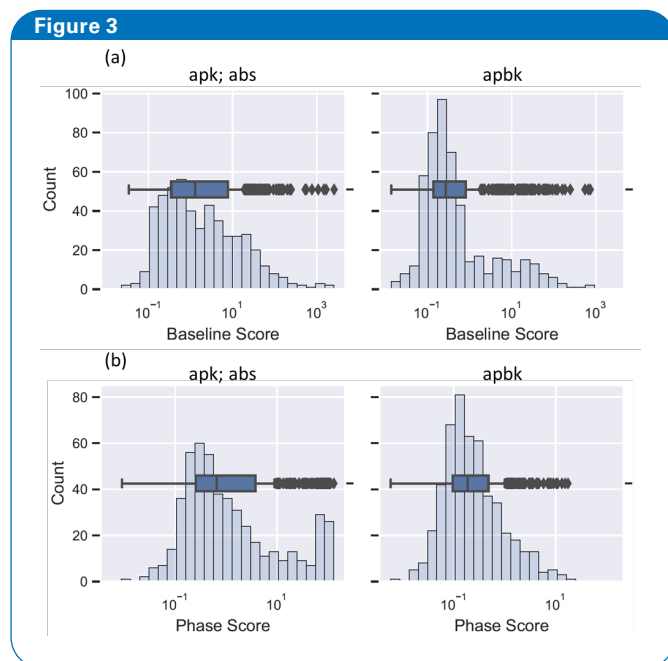
Figure 3. Baseline (a) and phase (b) scores obtained correcting phase and baseline distortions of 500 synthetic ^{1}H NMR spectra using *apk / abs* (left column) and *apbk* (right column). The box plot vertically centered in each of the plots shows the quartiles extended with 1.5 interquartile ranges on either side. Data points that fall outside that range are indicated by diamonds.

apk / abs, both for baseline and phase correction. Note that, due to the way the synthetic data was created, some of the baseline distortions contained in this test set are larger than the typical baseline distortions observed experimentally. As a consequence, this result might not be representative of the accuracy obtained experimentally, but it still shows the higher robustness of *apbk*.

To further compare the performance of the two methods, we have performed the same test on 100 experimental ¹H spectra. The results are shown in **Figure 4**. The baseline scores in **Figure 4a** have a median of 0.51 for *apbk* and 0.64 for *apk / abs*. First and third quartiles are 0.26 and 1.3 for *apbk* and 0.28 and 2.6 for *apk / abs*. Analogously to what we found for the synthetic datasets, *apbk* provides a better baseline correction than *apk / abs*. It shows a slightly lower median value and more consistent results across the dataset, indicated by a significantly lower 3rd quartile. The phase scores are provided in **Figure 4b**. The median is 0.31° for *apbk* and 0.55° for *apk / abs*, while first and third quartiles are 0.17° and 0.64° for *apbk* and 0.24° and 1.2° for *apk / abs,* respectively. The median phase score of the new method is far below the average accuracy achieved by the experts (0.55°), and the third quartile is just above this value. This indicates that *apbk* achieves human level quality results on a large fraction of the experimental spectra.

In **Figure 5,** we compare the results of *apbk* on low-field (80 MHz) with those on high-field (>80 MHz) spectra. The experimental dataset is the same that was used to create **Figure 4**. The baseline scores obtained with *apbk* are 0.89 and 0.42 for low- and high-field spectra respectively (**Figure 5a**), compared to 2.0 and 0.57 obtained with *apk / abs*. The new method performs better both on low- and high-field data, but the quality gain is especially evident on the data acquired using the Fourier 80 spectrometer. The phase scores shown in **Figure 5b** have a median of 0.36° and 0.25° for low- and high-field spectra corrected by *apbk* and 1.2° and 0.41° for *apk / abs*. This is in line with what we have observed for baseline correction and shows that *apbk* performs generally better for both low- and high-fields spectra.
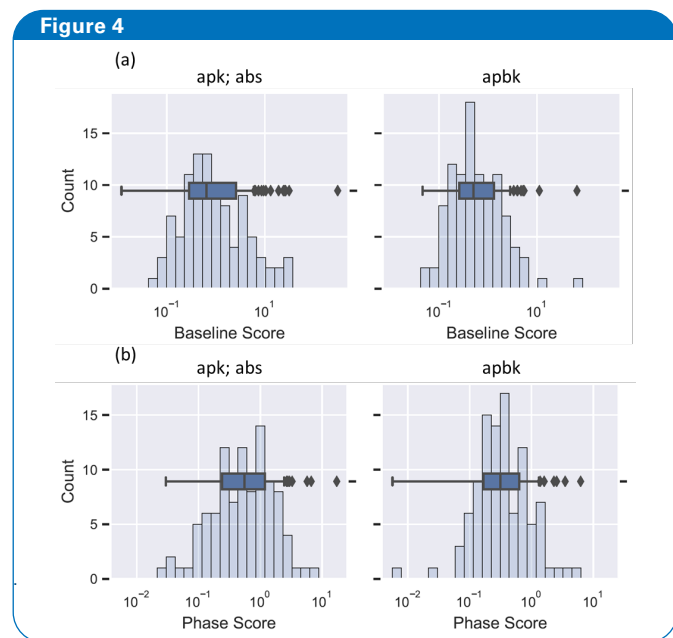

Figure 5. Baseline (a) and phase (b) scores obtained on high-field (blue) and low-field (orange) ¹H NMR spectra. The corrections were carried out using *apk / abs* (left column) and *apbk* (right column), on the 100 experimental spectra shown in Figure 4.

Figure 4. Baseline (a) and phase (b) scores obtained correcting phase and baseline distortions of 100 experimental ¹H NMR spectra using *apk / abs* (left column) and *apbk* (right column). The box plot vertically centered in each of the plots shows the quartiles extended with 1.5 interquartile ranges on either side. Data points that fall outside that range are indicated by diamonds.
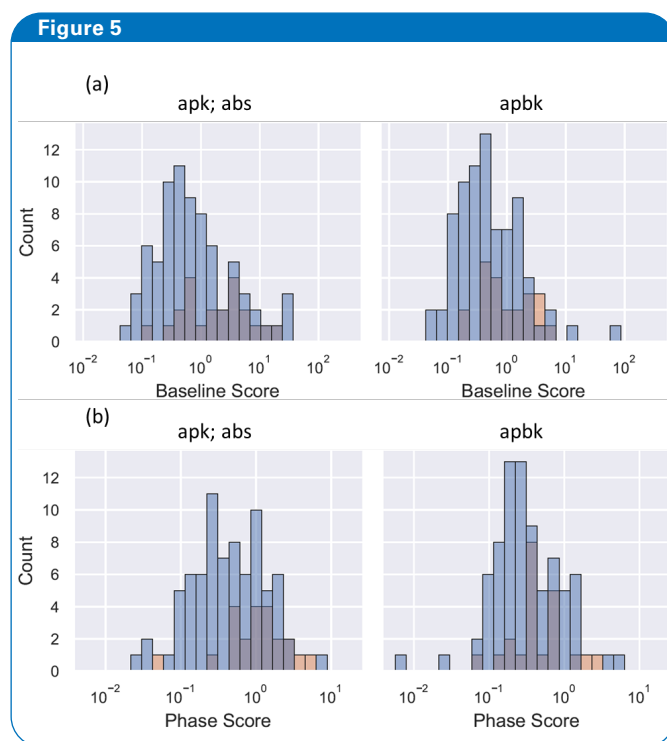
Conclusions

We have introduced a deep learning-based method for simultaneous phase and baseline correction of 1D ^{1}H spectra. The method uses a deep neural network to estimate the baseline together with a classical algorithm to determine the phase correction. The new method is fully automated and does not require any user input. It is available as TopSpin command *apbk* starting from TopSpin 4.1.3.

Comparison on synthetic and experimental spectra show significantly better performance of *apbk* compared to previously available TopSpin commands for phase and baseline corrections (*apk / abs*). In particular, *apbk* achieves human-level quality results on the phase correction of 100 experimental NMR spectra. We also find that *apbk* performs consistently better both for low- and high-field spectra, providing a more robust phase and baseline correction method that is suited both for the new Fourier 80 and high-field Bruker spectrometers.

Practical tips

- Use the commands *apbk -bo* and *apbk -po* to correct respectively only baseline and phase.

- To obtain the best results for both phase and baseline corrections, you can define your signal region manually and then use the command *apbk -intrng -n*.

- You can significantly reduce baseline distortions and remove 1st order phase distortions acquiring your spectrum using the digitation mode *baseopt*.

- You restrict the phase correction to 0th order only by using the command *apbk -apk0*.

- The deep learning algorithm described in this paper will not run by default on ^{1}H spectra that were acquired using pulse sequences with solvent suppression (the command falls back to *apk / abs*). If you want to force the use of the machine learning algorithm on these datasets use the command *apbk -f*.

- You can find a description of all the options available for the *apbk* command in the TopSpin processing manual.[21]

Acknowledgement

References

[1] R. R. Ernst, «Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy,» Journal of Magnetic Resonance (1969), Bd. 1, Nr. 1, pp. 7-26, 1969.

[2] G. A. Pearson, «A general baseline-recognition and baseline-flattening algorithm,» Journal of Magnetic Resonance (1969), Bd. 27, Nr. 2, pp. 265-272, 1977.

[3] L. Chen, Z. Weng, L. Goh, and M. Garland, «An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization,» Journal of Magnetic Resonance, Bd. 158, Nr. 1-2, pp. 164-168, 2002.

[4] J. C. Cobas, M. A. Bernstein, M. Martín-Pastor und P. G. Tahoces, «A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data,» Journal of Magnetic Resonance, Bd. 183, Nr. 1, pp. 145-151, 2006.

[5] Y. Xi und D. M. Rocke, «Baseline correction for NMR spectroscopic metabolomics data analysis,» BMC bioinformatics, Bd. 9, Nr. 1, pp. 1-10, 2008.

[6] Z.-M. Zhang, S. Chen, and Y.-Z. Liang, «Baseline correction using adaptive iteratively reweighted penalized least squares,» Analyst, Bd. 135, Nr. 5, pp. 1138-1146, 2010.

[7] Q. Bao, J. Feng, F. Chen, W. Mao, Z. Liu, K. Liu, and C. Liu, «A new automatic baseline correction method based on iterative method,» Journal of Magnetic Resonance, Bd. 218, pp. 35-43, 2012.

[8] V. Zorin, M. A. Bernstein und C. Cobas, «A robust, general automatic phase correction algorithm for high-resolution NMR data,» Magnetic Resonance in Chemistry, Bd. 55, Nr. 8, pp. 738-746, 2017.

[9] M. Sawall, E. von Harbou, A. Moog, R. Behrens, H. Schröder, J. Simoneau, E. Steimers und K. Neymeyr, «Multi-objective optimization for an automated and simultaneous phase and baseline correction of NMR spectral data,» Journal of Magnetic Resonance, Bd. 289, pp. 132-141, 2018.

[10] E. Steimers, M. Sawall, R. Behrens, D. Meinhardt, J. Simoneau, K. Münnemann, K. Neymeyr und E. von Harbou, «Application of a new method for simultaneous phase and baseline correction of NMR signals (SINC),» Magnetic Resonance in Chemistry, Bd. 58, Nr. 3, pp. 260-270, 2020.

[11] C. Cobas, «NMR signal processing, prediction, and structure verification with machine learning techniques,» Magnetic Resonance in Chemistry, Bd. 58, Nr. 6, pp. 512-519, 2020.

[12] D. Chen, Z. Wang, D. Guo, V. Orekhov und X. Qu, «Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy,» 2020.

[13] P. Klukowski, M. Augoff, M. Zięba, M. Drwal, A. Gonczarek und M. J. Walczak, «NMRNet: a deep learning approach to automated peak picking of protein NMR spectra,» Bioinformatics, Bd. 34, Nr. 15, pp. 2590-2597, 2018.

[14] X. Qu, Y. Huang, H. Lu, T. Qiu, D. Guo, T. Agback, V. Orekhov und Z. Chen, «Accelerated nuclear magnetic resonance spectroscopy with deep learning,» Angewandte Chemie, Bd. 132, Nr. 26, pp. 10383-10386, 2020.

[15] D. F. Hansen, «Using deep neural networks to reconstruct non-uniformly sampled NMR spectra,» Springer, 2019.

[16] J. Li, K. C. Bennett, Y. Liu, M. V. Martin und T. Head-Gordon, «Accurate prediction of chemical shifts for aqueous protein structure on "Real World" data,» Chemical Science, Bd. 11, Nr. 12, pp. 3180-3191, 2020.

[17] F. Paruzzo, S. Bruderer, Y. Janjar, B. Heitmann und C. Bolliger, «Automatic Signal Region Detection in ^{1}H NMR Spectra Using Deep Learning,» Bruker Whitepaper, 2020.

[18] H. De Brouwer, «Evaluation of algorithms for automated phase correction of NMR spectra,» Journal of magnetic resonance, Bd. 201, Nr. 2, pp. 230-238, 2009.

[19] [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke und A. Rabinovich, «Going deeper with convolutions,» Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015.

[20] R. B. Williams, «Quantitative Determination of Organic Structures by Nuclear Magnetic Resonance: Intensity Measurements,» Wiley Online Library, 1958.

[21] «Processing Commands and Parameters - TopSpin User Manual,» Bruker, 2021.

● **Bruker BioSpin**

info@bruker.com
www.bruker.com