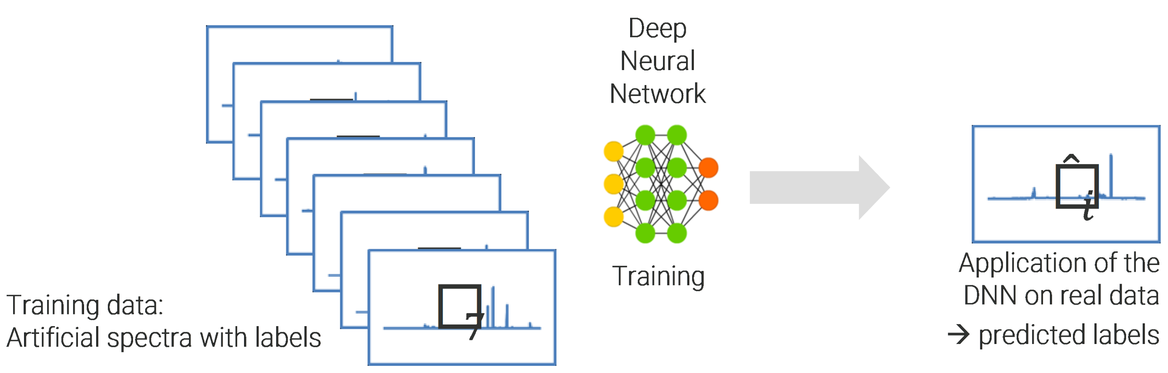
TopSpin Uses Machine Learning to Process and Analyze Data

Machine Learning

Machine learning, a subset of artificial intelligence, is revolutionizing data analysis by enabling computers to learn patterns and make decisions on their own without explicit programming.

In the context of TopSpin, neural network-based tools are used to interpret 1D proton spectra. These neural networks are extensively trained on millions of artificially generated spectra containing a wide variety of features. This synthetic data approach allows for the exploration of nuanced features that may not be readily apparent from experimental spectra alone.

At Bruker we carefully label the simulated data with specific attributes, such as peak position or signal presence, to create high quality data sets for the training of our neural networks. Then, when presented with real data, these trained neural networks can accurately identify and label corresponding features in the input, facilitating robust spectral analysis.



Baseline and Phase Correction

Despite advances in data acquisition in NMR spectroscopy, the baseline and phase correction remain fundamental steps in data processing. In order to avoid biasing the results due to manual processing by – especially non-expert – operators, it is nowadays accepted as best-practice in the qNMR community to rely on robust automatic processing routines.

To address this complexity, Bruker has introduced a new machine learning-based command called **apbk** that streamlines the process by performing both baseline and phase correction in a single step.

An important aspect of **apbk** is its commitment to delivering the highest quality results by ensuring that the processed data matches or exceeds the manual corrections of experienced analysts. In addition, this algorithm is available in automation. This results in a significant improvement in data quality.

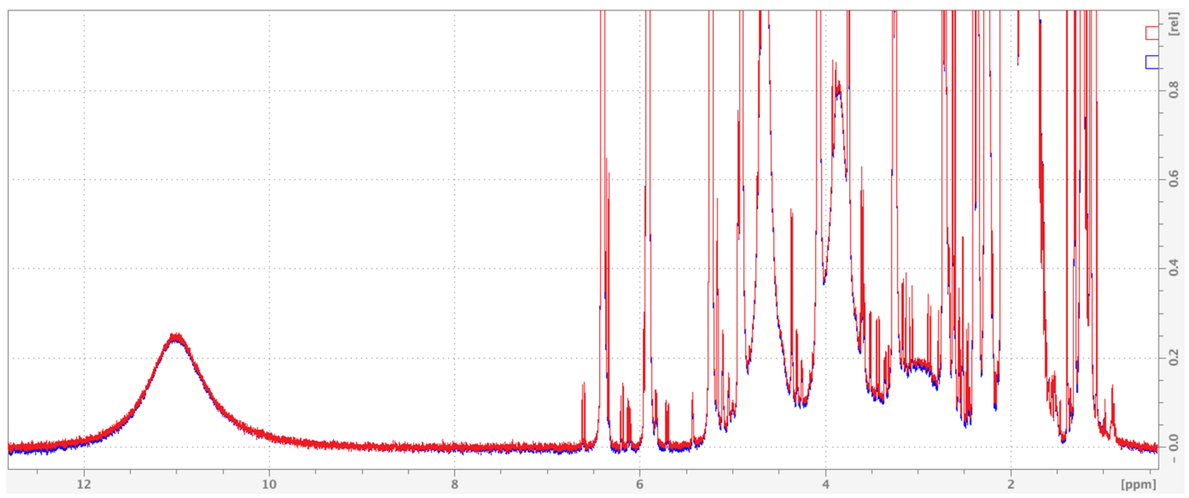
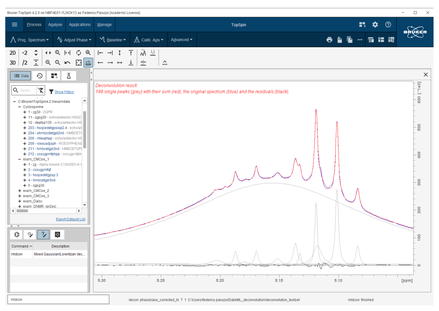


Fig. 1: Overlay of spectra processed by expert and apbk

Want to Learn More About Deconvolution?



Deconvolution

In NMR spectroscopy, deconvolution refers to the process of separating overlapping signals in a spectrum to reveal the underlying individual components or peaks.
Here, we present **mldcon**, Bruker's new fully automatic, parameter-free, machine learning based deconvolution of ^{1}H NMR spectra (for more details scan the QR code in the lower left corner).

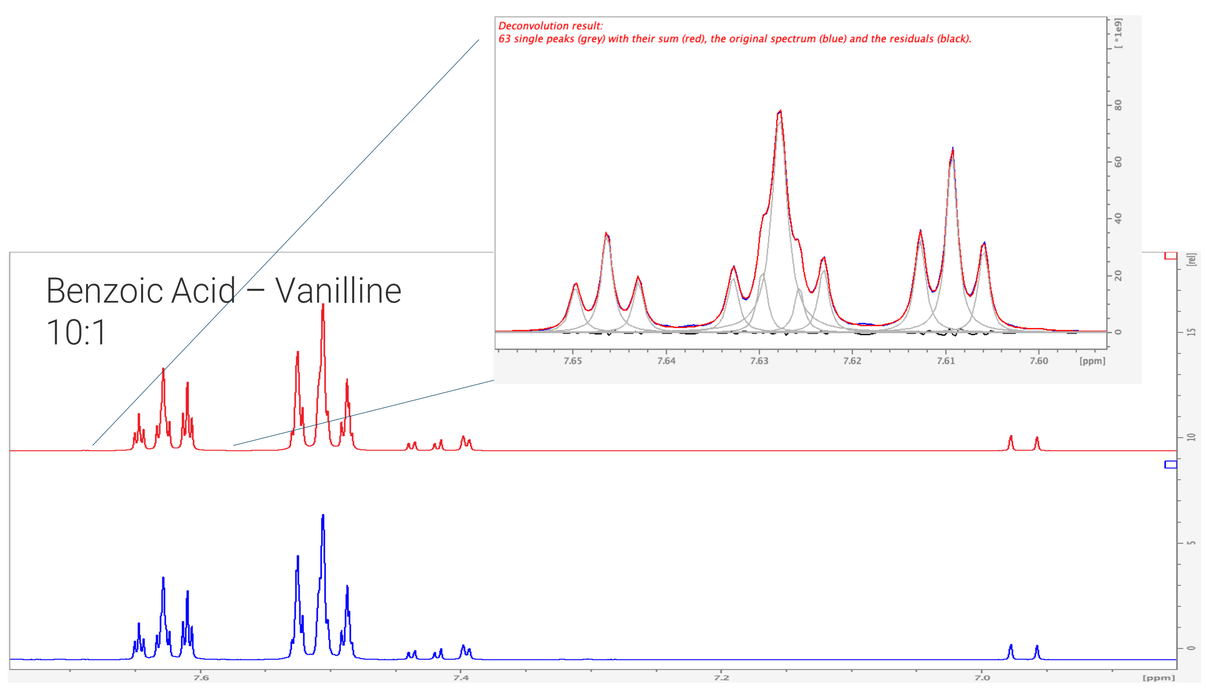


Fig. 2: The comparison between experimental spectra and the sum of peaks generated from a deconvolution algorithm provides valuable insights into the accuracy and reliability of the deconvolution performance.

To evaluate the effectiveness of the deconvolution algorithm, a comparative analysis was conducted to estimate the relative concentrations of two components: benzoic acid and vanillin, at a 10.04:1 ratio. Across varying concentrations (9 mM and 100 mM), the deconvolution-based estimates consistently exhibit better precision and accuracy compared to integration methods.

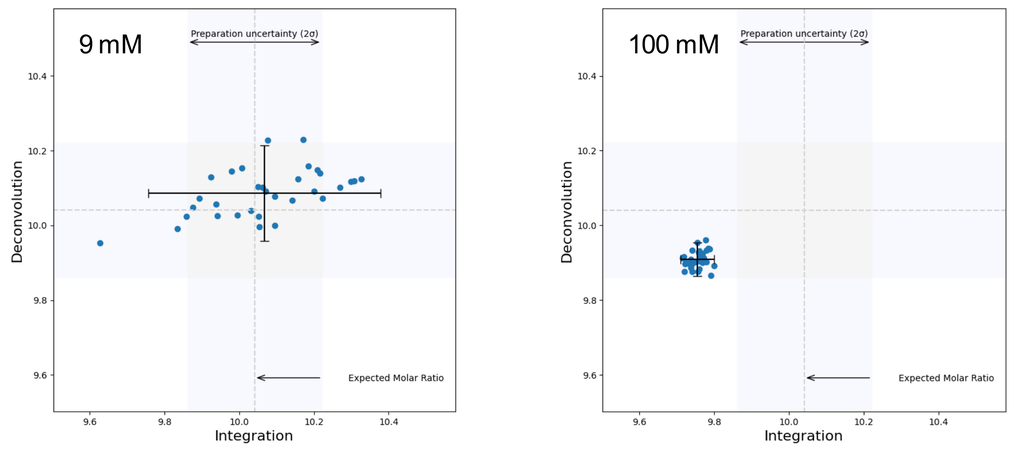


Fig. 3: Compares relative concentration estimates derived from the deconvolution algorithm (vertical axis) and integration methods (horizontal axis). For each concentration, a series of 32 spectra was acquired. The grey lines in the graph represent the expected concentration ratio, while the blue ranges indicate the expanded uncertainty (k=2, 95 % confidence) of the preparation process.

Conclusion

Machine learning transforms NMR data processing and analysis, ushering in a new era of automation.

In Topspin 4, new machine learning-based commands operate autonomously, increasing efficiency and accuracy in critical tasks such as concentration estimation:

- **apbk** automates phase and baseline correction with expert precision
- **mldcon** introduces a groundbreaking, fully automated deconvolution algorithm.