Received: 1 February 2024 | Revised: 20 May 2024 | Accepted: 30 May 2024 | Published online: 18 June 2024

REVIEW

Journal of Data Science and Intelligent Systems
2024, Vol. 00(00) 1–17
DOI: 10.47852/bonviewJDSIS42022556

Insights into Nuclear Magnetic Resonance Data Preprocessing: A Comprehensive Review

Aixiang Jiang[1,2,3,*]

[1]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada*

[2]*Department of Pathology and Laboratory Medicine, University of British Columbia, Canada*

[3]*Centre for Lymphoid Cancer, British Columbia Cancer, Canada*

Abstract: Nuclear magnetic resonance (NMR) and its derivatives play a pivotal role in molecular analysis across research and clinical domains. However, the intricate nature of NMR data preprocessing, which is integral for accurate analysis, is not easily understood despite the availability of numerous software tools. This comprehensive review aims to unravel the complexities of preprocessing algorithms in both the time and frequency domains. It covers essential steps such as direct current offset removal, eddy current correction, shift and linear prediction, weighting, zero filling, domain transformation, phase error correction, baseline correction, solvent filtering, calibration and alignment, reference deconvolution, binning/bucketing, peak picking, peak fitting/deconvolution, compound identification, integration and quantification, normalization, and transformation. The review uses plain language to enhance accessibility and understanding. By demystifying the algorithms behind these preprocessing steps, we seek to help researchers and practitioners in navigating the nuances of NMR data preprocessing, ultimately fostering better understanding and practical application in molecular analysis.

Keywords: nuclear magnetic resonance, preprocessing, time domain, frequency domain

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a highly effective analytical tool for providing intricate details about the molecular structure, composition, and dynamics of a sample. It has also given rise to many new techniques, including multinuclear magnetic resonance, quantitative nuclear magnetic resonance, nuclear magnetic resonance imaging (MRI), functional MRI (fMRI), diffusion MRI, and diffusion tensor imaging [1–8]. These NMR techniques find widespread applications in various fields such as chemistry, biology, agriculture, and medicine [9–14]. Due to their ability to analyze metabolites, detect the structures of DNA, RNA, and proteins, and visualize human internal organs/serum without ionizing radiation, these techniques have proved to be versatile and valuable [15–19].

NMR spectroscopy utilizes powerful magnetic fields to analyze samples. When exposed to the radiofrequency radiation produced by an NMR spectrometer, the nuclei in molecules absorb the energy and transition to higher energy levels when possible. This phenomenon is known as excitation. After the radiofrequency pulses are turned off, the nuclei undergo relaxation, releasing the absorbed energy and returning to their original energy levels. The decaying signal resulting from this relaxation process is captured by a receiver coil surrounding the sample tube. The weak, energy-varying currents induced by the relaxation are detected as raw signals from the molecules.

Figure 1 uses a proton as an example nucleus to illustrate how a proton signal is generated.

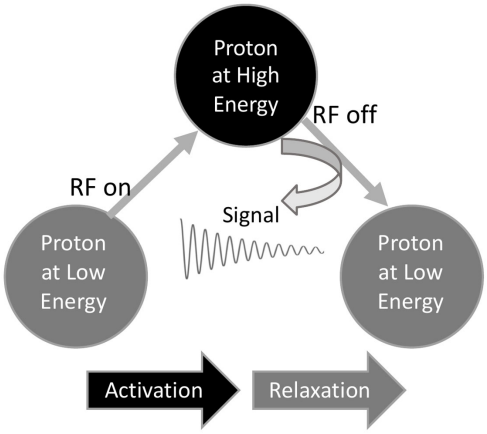
Before becoming raw NMR data, these signals undergo amplification and digitization. Raw NMR data, as depicted in Figure 1's middle section, vividly illustrate signal changes over time, capturing dynamic variations and thus are termed time domain NMR data. When multiple signals are present, they diminish over time, mix together, and become difficult to analyze. Therefore, these raw data must undergo a series of preprocessing steps to prepare them for examination in the frequency domain, where signals are separated into distinct peaks. Domain transformation is just one unavoidable preprocessing step among many.

In addition to domain transformation, preprocessing addresses various other issues inherent in raw NMR data. The presence of noise, baseline distortions, and other artifacts necessitates preprocessing to ensure an accurate interpretation of the data. Furthermore, for tasks such as peak-based molecule identification and quantification, it becomes imperative to enhance peak resolution and employ intelligent peak definition and deconvolution methods. Moreover, to facilitate meaningful data comparison across spectra, calibration, alignment, normalization, and transformation steps are often indispensable.

*Corresponding author: Aixiang Jiang, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Department of Pathology and Laboratory Medicine, University of British Columbia, and Centre for Lymphoid Cancer, British Columbia Cancer, Canada. Email: aijiang@bccrc.ca

© The Author(s) 2024. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

Figure 1
Illustration of the process of generating a proton signal in NMR spectroscopy



Many commercial and open-source software tools are available for NMR data preprocessing. Notable commercial software includes TopSpin, ACD/Labs, Mnova, Chenomx, and iNMR (see the Appendix). On the open-source front, there are options like NMRbox, NMRPipe, AlpsNMR, NMRphasing, nmrrr, PepsNMR, Rnmr1D, speaq, nmrespy, dnpLab, Protomix, peakipy, ssnmr, metabolabpy, nmrglue, spike-py, klassez, nmrpy, and pynmr (see the Appendix).

While a plethora of software options are available, this review article does not focus on delineating the functionalities of each tool. Instead, our examination delves into the algorithms employed for common NMR preprocessing steps.

To conduct this review, pertinent literature was identified using targeted keywords such as "NMR," "preprocess" or "pre-process," and the names of specific preprocessing steps. Searches were executed across databases including PubMed, IEEE, Google Scholar, Copilot, and other relevant platforms, with screening methods involving the assessment of titles, abstracts, and full-text articles.

We categorize preprocessing steps based on common practice, starting with the time domain, where NMR raw data originate, and then proceeding to the frequency domain, where NMR data analysis is conducted on. Within each domain, we arrange the steps by their typical order of application, providing a structured framework for analysis.

2. NMR Preprocessing Steps in the Time Domain

2.1. Direct current (DC) offset removal

The initial step in NMR data preprocessing involves converting raw NMR time domain files, referred to as free induction decay (FID) data, from a binary format to text. After reading in the FID, the first issue we need to address is the removal of the direct current (DC) offset, which is a constant voltage added to the NMR signal due to various factors like instrument imperfections or interference.

In Figure 2A, the signal, centered at zero with 10 cycles per second (10 Hz), is converted from the time domain to the frequency domain to distinguish signals (Figure 2B).

However, if a signal detector in an NMR spectrometer has a DC voltage offset, it shifts the signal's center away from zero in the time domain plot (Figure 2C), causing an unexpected non-signal line on the left in the frequency domain plot (Figure 2D). To tackle this, removing the DC offset in the time domain is necessary. Three methods are available:

1) Last data point method: Subtract the last data point's value from all data points.
2) Tail points method: Average the last quarter, 20%, or 10% of data points, and subtract this average from all data points (http://anorganik.uni-tuebingen.de/klaus/nmr/processing/index.php?p=dcoffset/dcoffset). This method is generally more reliable than the last point method [20].
3) Phase cycling method: Typically, only one detector is used to detect NMR signals. However, an additional detector positioned 180 degrees apart can be utilized. In this case, clean signals without DC offset can be obtained by subtracting the data of the additional detector from the original data. Note that signal amplitudes are doubled by this method. In this context, "phase" refers to the angular displacement of the NMR signal. To illustrate, considering the signal in Figure 2A, the extra signal detector begins recording the same signal when it reaches its first local minimum. While this idea may not be applicable to most 1D NMR data due to the absence of extra data, it is easily applied to MRI imaging sequences and extended to other phase cycling angles [21].

The most reliable approach to handling DC offset is phase cycling when an extra detector is available. In cases with sufficiently long FID recording times, estimating DC offset using the tail points can be considered. Unfortunately, no optimal solution exists for handling DC offset in other situations.

2.2. Eddy current (EC) correction

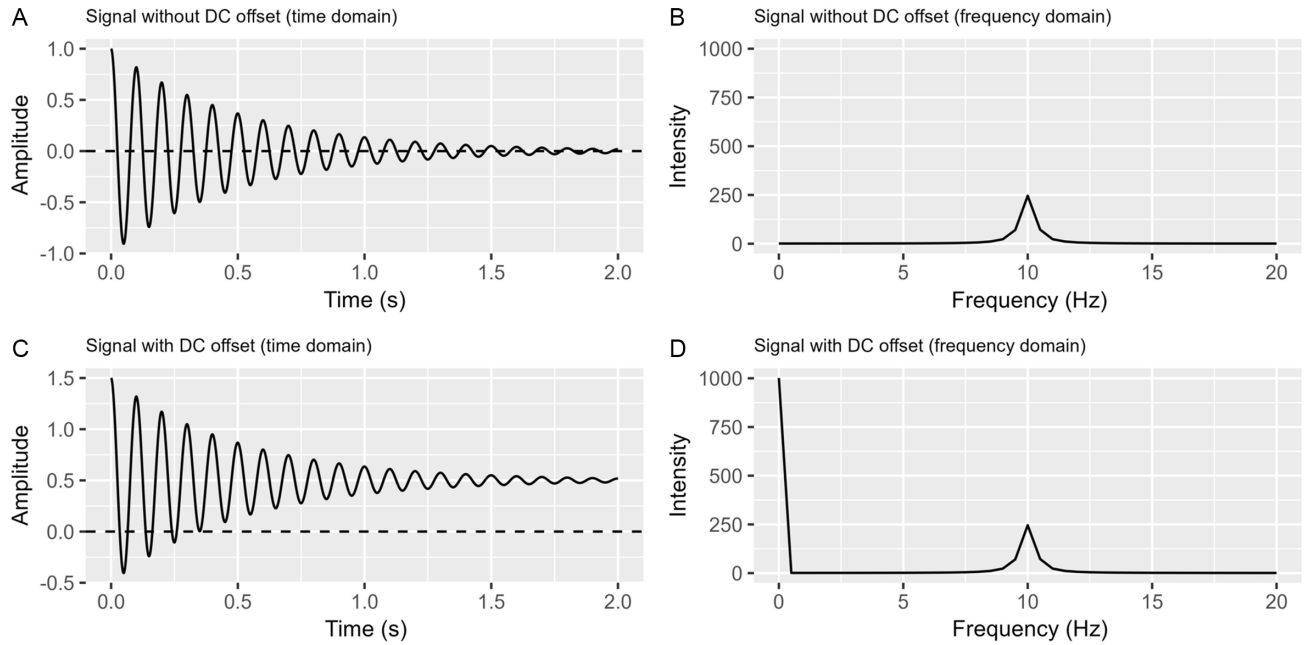
Eddy currents (EC) are induced by the interaction of changing magnetic fields with conductive elements in both the NMR sample and machine [22]. These currents create their own magnetic fields, which subsequently affect the designed magnetic field in the NMR system. As a result, these currents lead to variations in observed frequencies, fluctuations in signal amplitude, and phase distortions in acquired NMR signals.

In Figure 3A, the absence of EC results in a consistent cyclic signal in the time domain (Figure 3B), producing a single symmetric peak in the frequency domain (Figure 3C). The presence of EC (Figure 3D) leads to irregular time domain signals (Figure 3E) and multiple peaks, including negative ones, in the frequency domain (Figure 3F), causing significant signal distortion.

To correct EC effects, we discuss two NMR methods and one MRI method:

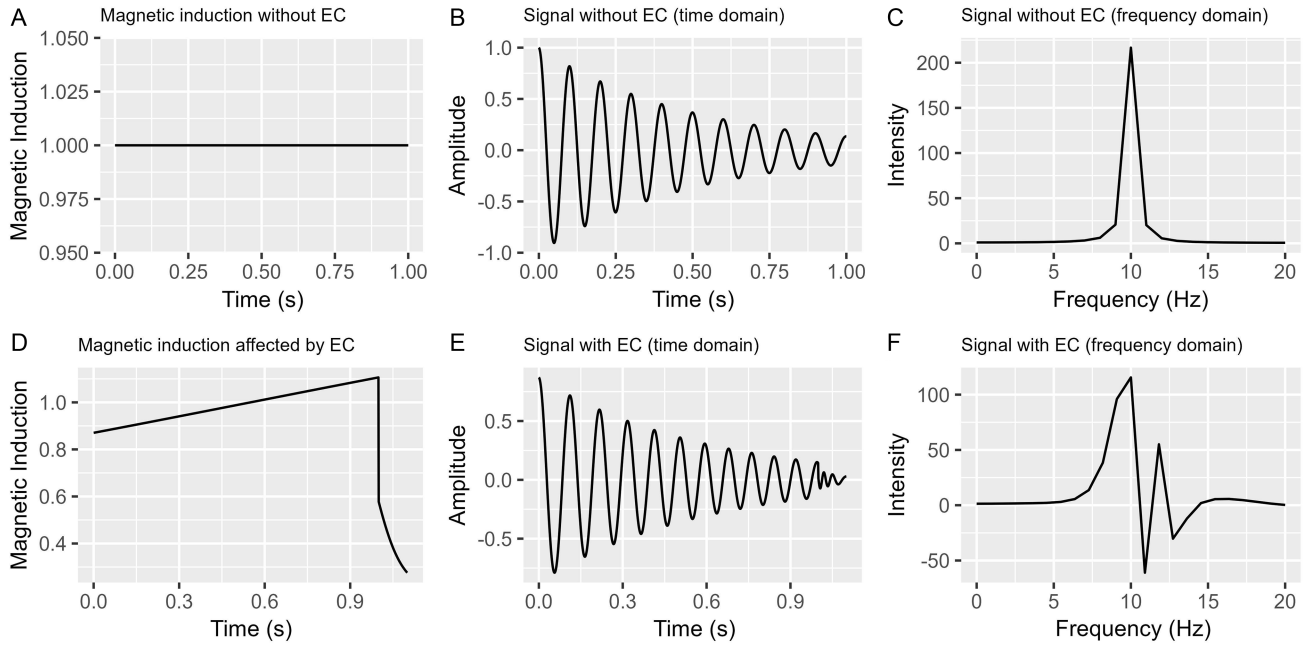
1) Phase correction with reference FID: Assuming that a reference-only FID is available and has the exact settings as an experimental FID without the reference, we can subtract the phase vector of the reference FID from that of the experimental FID, resulting in an EC-corrected phase vector. This corrected phase vector is then used to reconstruct a new FID file. The term "reference" here does not refer to a spike-in internal reference for signal quantification; instead, it pertains to a solvent, like water. In practice, utilizing a water-unsuppressed FID as a pseudo-reference-only FID and a water-suppressed FID as our experimental FID for EC correction is cost-effective because we just need to run the same sample twice with two different conditions, and water has significantly higher concentration, allowing us to disregard metabolite [23].
2) Phase error correction with opposite induction directions: Employing a two-step approach with positive and negative

Figure 2
Effect of direct current (DC) voltage on a signal.



(A) Time domain signal without DC offset.
(B) Frequency domain signal without DC offset.
(C) Time domain signal with DC offset.
(D) Frequency domain signal with DC offset, indicating a "glitch" (unexpected non-signal line)

Figure 3
Illustration depicting the eddy current effect in time and frequency domains.



(A) Magnetic induction without eddy currents.
(B) NMR signal without eddy currents (time domain).
(C) NMR signal without eddy currents (frequency domain).
(D) Magnetic induction affected by eddy currents.
(E) NMR signal with eddy currents (time domain). (F) NMR signal with eddy currents (frequency domain)

magnetic inductions effectively eliminates EC-induced phase errors, resulting in mitigated and even phase corrections [23, 24]. This is better than "Phase correction with reference FID" especially when a pseudo-reference-only FID is used.

3) EC-induced magnetic model: This model, more applicable to MRI, establishes the relationship between the EC-induced magnetic field and spatial coordinates. An iterative optimization process, facilitated by specialized software such as "eddy" (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy), determines model parameters to correct MRI data for EC effects. While low-order polynomials (up to the second order) are commonly used to model and correct EC-induced distortions, higher-order models (quadratic and cubic) might also find applications [23, 25–28].

While EC correction is recommended, it may not be applicable when additional data is unavailable. In such cases, alternative options can be employed to partially mitigate EC effects in the subsequent steps, including domain transformation, phase error correction, solvent filtering, and chemical shift calibration.

2.3. FID shift and linear prediction

The beginning of the FID sequence is particularly prone to distortion due to sudden radiofrequency shifts compared to the rest of the data. To mitigate this distortion, we may implement left shifts, moving points before time 0 beyond the FID, a method suitable for fully recorded FIDs with minor adjustments. Conversely, right shifts intentionally delay the FID [29], which effectively addresses significant distortions at the sequence start.

Whether using left or right shifts, data gaps inevitably occur. To address these gaps, linear prediction (LP) methods are employed to recover lost FID data caused by shifts. Backward LP focuses on data missing at the sequence start due to right shifts [29], while forward LP extends or fills missing segments at the sequence tail due to left shifts.

Furthermore, concern the LP formulas [30]:

$$F_n = \sum_{m=1}^{P} C_m F_{n+m} + \varepsilon_n \quad (1)$$

$$F_n = \sum_{m=1}^{P} C_m F_{n-m} + \varepsilon_n \quad (2)$$

Here, m represents the base point index, P is the total number of base points, F_n is the predicted point, F_{n+m} and F_{n-m} are the base points used for backward (1) and forward (2) LP, respectively, C_m stands for the coefficient of a base point, and ε_n is the random error associated with the predicted point. The prediction process involves an iterative optimization utilizing a loss function, such as squared differences between F_n and $\sum_{m=1}^{P} C_m F_{n+m}$.

Careful attention is necessary for FID shifts and backward LP due to potential data distortion [31]. A small left shift is generally safer than a larger right shift, and a forward LP is considered safer than a backward LP.

2.4. Weighting

Weighting involves multiplying the FID by nonlinear functions like exponential, Gaussian, half-Gaussian, or sine bell functions, to enhance sensitivity or resolution [32], commonly used in modern methodologies [10, 33–38]. However, not all weighting functions are appropriate for molecule quantitation [29].

Figure 4A–B illustrates a simulated FID in the time and frequency domains. Employing a decreasing exponential decay as a weighting function notably attenuates the FID's tail while preserving its initial segment, as demonstrated in Figure 4C. This enhances the signal-to-noise ratio (SNR) but may potentially broaden peaks and cause overlap in the frequency domain [17, 39]. When this process reduces the tail values to zero, it is also referred to as apodization [39]. Apodization can enhance visualization but is cautioned against when preparing data for spectral analysis. Applying apodization before such analysis may compromise the statistical assumptions tied to the fitting model [27].

Conversely, using an increasing exponential function enhances FID resolution (Figure 4E), resulting in a narrow peak in the frequency domain (Figure 4F). However, it amplifies noise in the FID tail, reducing SNR and potentially causing asymmetric peaks [40].

Utilizing weighting functions involves a delicate balance between sensitivity and resolution, where enhancement in one aspect may come at the expense of the other and could potentially introduce distortions, complicating data recovery. Applying these functions without a comprehensive understanding of the data or a specific sensitivity/resolution goal requires caution. Furthermore, maintaining consistent application of these functions throughout an experiment is vital to ensure comparability of the data.

Alternatively, for enhancing SNR without sacrificing resolution, employing singular-value decomposition-based approaches, such as Cadzow and principal component analysis (PCA), alongside a new wavelet transform routine, proves effective in efficiently enhancing SNR and robustly denoising 1D and 2D NMR spectra [41]. However, these methods should be applied after molecule identification and quantification, as they could potentially distort quantification results.

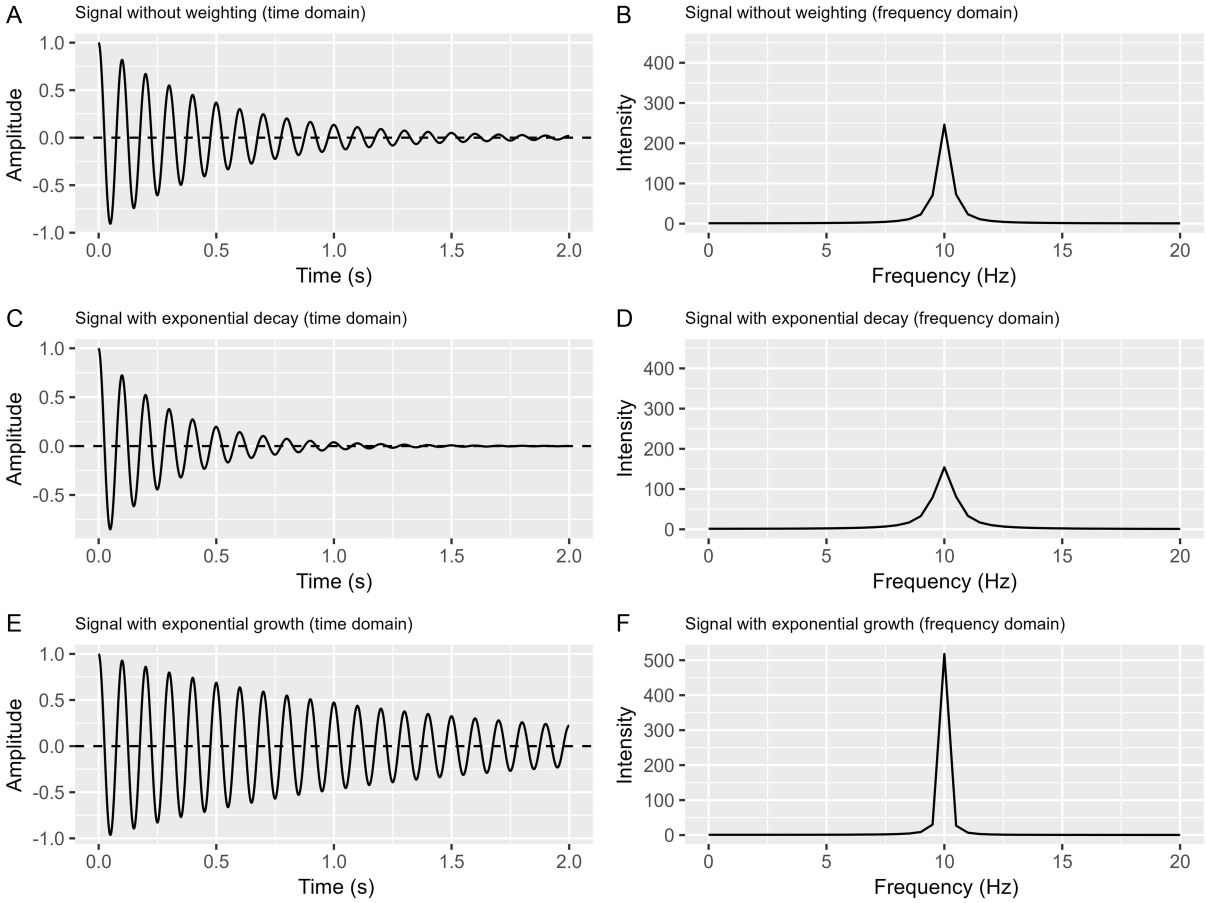
2.5. Zero filling

Zero filling involves adding zeros to the end of an FID, creating the illusion of higher digital resolution [20, 27]. For instance, doubling the spectral length and improving digital resolution are achieved by appending zeros equal to the number of experimental points, while additional zero filling aids in data interpretation through interpolation [29]. However, it's crucial to note that zero filling does not contribute to real signal data and may increase noise due to the introduction of zeros.

When employing zero filling, maintaining near-zero endpoints in the FID is crucial. Figure 5A–B illustrates a simulated FID in both time and frequency domains. In Figure 5C, utilizing only half of the signal points from Figure 5A, while preserving near-zero endpoints produces a frequency domain (Figure 5D) resembling the original peak (Figure 5B). Failure can lead to distorted peaks (Figure 5E–F). Techniques like forward LP and apodization decay functions [20] may aid in such scenarios, but their effectiveness varies.

Zero filling is generally safer with ample data and near-zero endpoints, but it's less beneficial with very few data points (Figure 5E–F). Prioritizing extended recording over zero filling is advisable. Consistency in zero filling across all FIDs within an experiment ensures data comparability without exception. Essentially, zero filling merely interpolates points in the frequency domain data without adding new information. Therefore, relying solely on zero filling is insufficient; extending signal recording time is vital.

Figure 4
Illustration depicting the effect of weighting functions. Only the real part is shown.



(A) Time domain plot of a simulated FID with a single peak.
(B) Frequency domain plot corresponding to A.
(C) Time domain plot of A times an exponential decay ($e^{-2.5(j-1)/N}$). Here, j is index of a given point, and N is the total number of data points in FID.
(D) Frequency domain plot corresponding to C.
(E) Time domain plot of A times an exponential growth ($e^{2.5(j-1)/N}$).
(F) Frequency domain plot corresponding to E

2.6. Domain transformation

Domain transformation is pivotal in converting FID time domain data into the frequency domain. The primary method used for this transformation is the discrete Fourier transform, which mathematically produces the frequency content of discrete signals through the following formula (3):

$$x_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N} \tag{3}$$

Here's a breakdown of the formula components:

- x_k represents the kth complex number in the frequency domain.
- x_n represents the nth complex number in the time domain.
- N is the total number of data points in the sequence.
- k is the frequency bin index, ranging from 0 to N–1.

This transformation allows FID signals to manifest as single peaks in the resulting spectrum [39].

There are several alternative methods for domain transformations. The linear model, although good for complementing the Fourier transform, is generally less accurate for independently analyzing multiple signal FIDs [42]. Bayesian methods rely on prior distributions [43, 44]. The wavelet transform is adept at handling uneven frequencies [45].

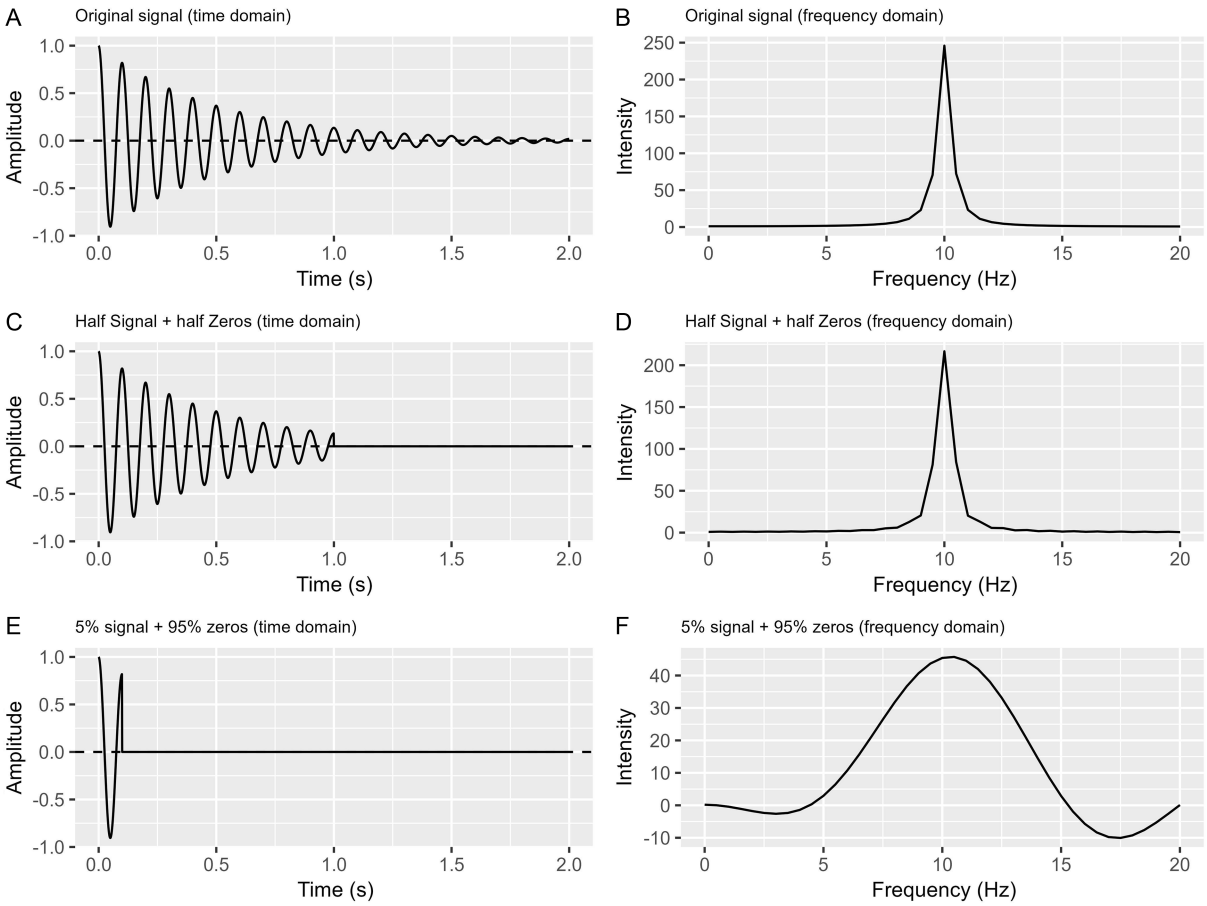
In standard scenarios without EC issues, the Fourier transformation is recommended for its reliability. However, when addressing EC-induced frequency alterations, the wavelet transform is preferable as it doesn't require additional data [46]. Post-wavelet transformation, it's important to note that the symmetry of frequency domain peaks might not be perfect, risking information loss if forcibly shaped into predetermined forms.

3. NMR Preprocessing Steps in the Frequency Domain

3.1. Phase error correction

Raw NMR signals in the time domain are complex numbers representing the nuclei's energy changes along two orthogonal directions. After being transformed into the frequency domain, they remain complex numbers, with the real part referred to as

Figure 5
Illustration of the impact on ending value, zero filling, and signal percentage. Only real part is shown.



(A) Original signal in the time domain.
(B) Frequency domain plot of the original signal.
(C) Half signal + half zeros in the time domain.
(D) Frequency domain plot of the half signal + half zeros.
(E) 5% signal + 95% zeros in the time domain.
(F) Frequency domain plot of the 5% signal + 95% zeros

absorption and the imaginary part as *dispersion*. Figure 6A shows a simulated absorption spectrum with three sharp and concentrated peaks. Correspondingly, Figure 6B displays the simulated dispersion spectrum. The phase, calculated as $Phase = \tan^{-1}\frac{Imaginary}{Real}$, indicates the relationship between absorption and dispersion. Its corresponding plot is shown in Figure 6C. Figure 6A–C represents ideal signals with no phase errors; thus, Figure 6D shows a phase error plot with all values at 0.

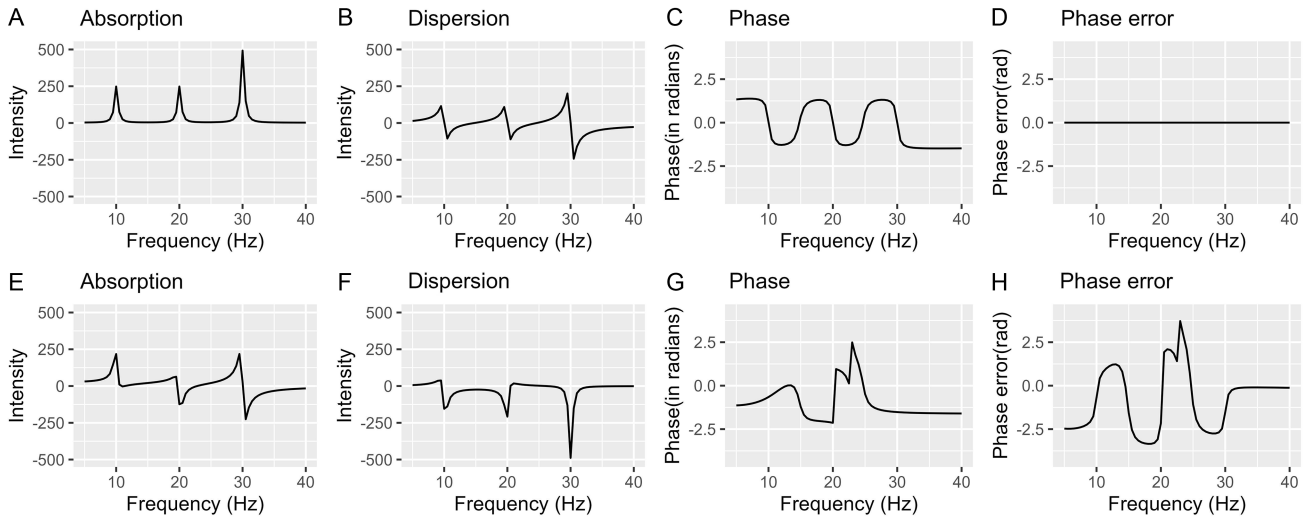
However, NMR data always contain phase errors, which can significantly alter the absorption, dispersion, phase, and phase error plots, as illustrated in Figure 6E–H. Consequently, a naïve data analysis based on peak locations and areas under curves in the absorption plot (Figure 6E), without considering phase errors, is unreliable because the apparent peaks differ from true peaks without phase errors (Figure 6A). Therefore, phase error correction must be performed before any data analysis.

Real signal components are shown in A and E, while imaginary signal components are shown in B and F. Phase plots are located in C and G, while phase error plots can be found in D and H.

Unfortunately, phase error correction is one of the most challenging preprocessing steps in the frequency domain. Current NMR phase error correction approaches mainly rely on a simple linear model applied to the entire spectrum [17]. This model searches for the intercept (zero-order parameter) and slope (first-order parameter) through an optimization process. Different algorithms employ various optimization functions [27, 47–62]. However, this simple linear model approach cannot effectively handle nonlinear phase errors such as shown in Figure 6H. Recent research continues to rely on manual phase error correction [27, 29, 36, 38, 63, 64]. Unfortunately, manual phase error correction heavily depends on individual experience, leading to inconsistencies and a lack of inter-user reliability.

To address nonlinear phase errors, researchers incorporated nonlinear terms into a linear model, surpassing the performance of a simple linear model [47, 52, 65, 66]. In contrast to standard simultaneous parameter correction, Jaroszewicz et al. [66] proposed an iterative order-by-order search for a linear model extended with quadratic terms. This involves optimizing the first-order parameter,

Figure 6
Illustration of NMR phase errors.



A frequency domain data example of three signals without phase error (A–D) and with errors (E–H). Real signal components are shown in A and E, while imaginary signal components are shown in B and F. Phase plots are located in C and G, while phase error plots can be found in D and H

followed by the zeroth order, and concluding with the second order in each iteration. The phase range progressively narrows between iterations. The method continues until reaching the maximum iteration limit or observing no significant changes, aligning with other optimization strategies. The authors stress the automatic nature of the approach, requiring no prior knowledge. However, this approach identifies phase values that maximize absorption and minimize dispersion, respectively, and so it might overlook solutions where both objectives are optimized simultaneously.

In addressing all types of phase errors, whether constant, linear, or nonlinear, we have recently developed a new R package, "NMRphasing" (https://cran.r-project.org/web/packages/NMRphasing/). One algorithm in "NMRphasing" starts with phase error-free data, such as magnitude and power spectra, which theoretically should not contain any phase errors. Subsequently, the algorithm derives the phase error-free absorption spectrum, as illustrated in Figure 6A. Alternatively, we propose multiple linear models to correct phase errors in different peak ranges. In addition, we introduced a novel optimization function aimed at minimizing the disparity between the absolute area under a curve and the net area under the same curve. This approach seeks to maximize absorption through net area while simultaneously minimizing dispersion via absolute area. A smaller absolute area of absorption implies less contamination of dispersion within the observed absorption, consequently reducing the net area of dispersion. This is desirable, as ideal dispersion should ideally exhibit zero net area.

Spatially varying phase errors can be effectively corrected using adaptive phase correction (APC) in MRI. Unlike traditional phase correction processes relying on regularization, APC utilizes MRI noise information for complex-valued image regularization, addressing noise bias and improving accuracy in diffusion MRI, especially in regions with diverse noise characteristics. The method involves applying a regularization operator and adjusting the phase based on noise variance estimates, resulting in a final image that accommodates noise characteristics in different regions [67]. However, phase-corrected images from this approach still contain phase errors and negative intensities. It is recommended to manually inspect these images and ensure their compatibility with subsequent processing steps.

3.2. Baseline correction

Baseline distortion refers to a non-flat and nonzero baseline, primarily caused by uncorrected DC (direct current) offsets and phase errors. Baseline correction involves estimating the baseline bias and subtracting it from the spectrum data. Various algorithms exist for estimating baseline bias; here are some examples:

1). Iterative polynomial fitting [10, 14, 17, 63, 64, 68]
2). Robust estimation procedure [14, 68]
3). Locally weighted scatter plot smoothing [14]
4). Asymmetric least squares regression with penalized least square approach [14, 68]
5). B-spline fixed or mixed model with or without penalization [14, 27]
6). Continuous wavelet transform [69]

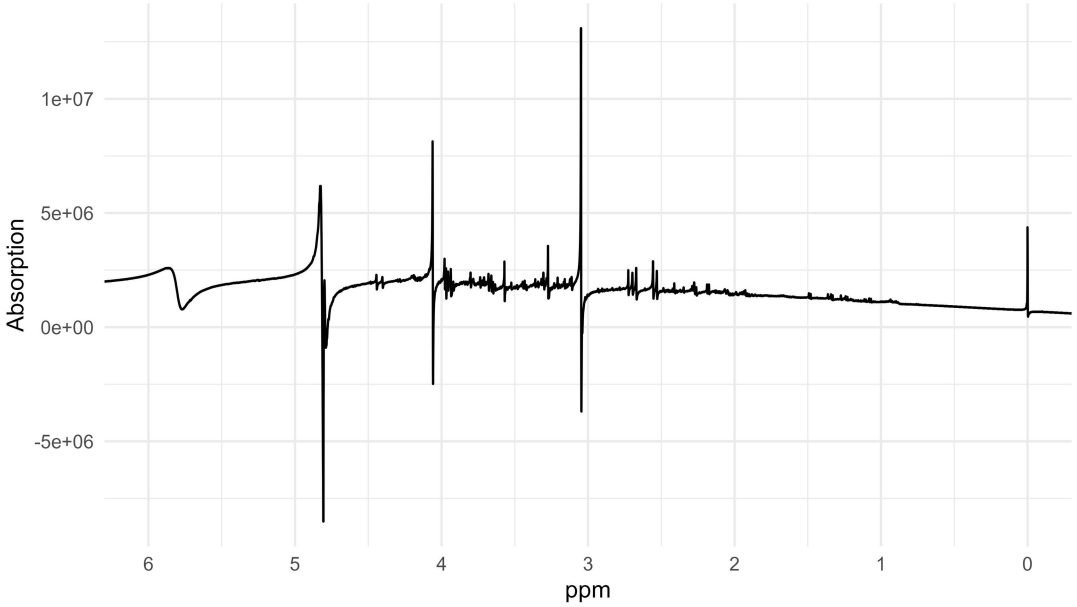
Baseline bias estimation in all these algorithms is based on regions without signals [17]. Of course, it might be challenging to distinguish noise and signal regions when no prior information about signal locations is available. One method is to classify individual points as either signal or noise points and subsequently employ linear interpolation between noise points to establish the baseline. After the baseline is constructed, it is subtracted from the corresponding spectrum. Most baseline correction methods are automated, although semiautomatic or manual baseline correction methods also exist [29, 70–73].

Regardless of the baseline correction method used, it is essential to be aware that baseline correction itself can introduce distortion and bias to the data, as it is intertwined with noise modeling [29].

3.3. Solvent filtering

As mentioned in Section 2.2 on EC correction, solvent filtering becomes a viable alternative when EC correction is not possible due to the unavailability of additional data, as intense solvent peaks often

Figure 7
A partial segment of a real-world absorption spectrum illustrating the distorted water peak around 4.7–5.0 ppm in a urine sample



capture most of the effects of EC, excluding the solvent signal range in the frequency domain and minimizing the impact of EC [23].

When the solvent is water, as depicted in Figure 7 around 4.7–5.0 ppm, the common practice is to run samples with water suppressed [74, 75].

If samples are run without solvent suppression and EC effects persist uncorrected in the time domain, and if phase error correction fails to rectify the distortion, solvent peaks can become severely distorted. To address this issue, common methods for solvent filtering include the following:

1). Subtract a solvent-only FID from the experimental FID of the sample and transforming the solvent-filtered FID into the frequency domain.
2). Create a pseudo-solvent-only FID by isolating data within the solvent peak range from the frequency spectrum, setting other data points to zero, and transforming it into the time domain. Subsequently, subtract this pseudo-solvent-only FID from the experimental FID of the sample and transform the resulting data into the frequency domain.
3). Use specialized filters targeting the solvent's frequency range to eliminate the solvent signal [23].
4). Integrate solvent peak removal with baseline correction in the frequency domain [23].
5). Zero out data points within the solvent peak range or set them to baseline values to effectively remove the solvent peaks [70–72, 76, 77].
6). Employ wavelet transformation to remove the solvent signal [78].

However, it is important to note that filtering solvent peaks may also inadvertently remove some true signals from their neighboring components.

3.4. Calibration and alignment

To ensure comparability of NMR spectra across different spectrometers, frequencies are expressed in parts per million (ppm) using the ratio of a signal's frequency to the spectrometer's frequency. Calibration, also known as global alignment, sets the internal reference signal's ppm to zero by shifting the entire spectrum [17, 63, 70, 72, 79]. On the other hand, (local) alignment is to adjust each peak across a group of spectra to the same ppm position [17, 68].

The following are example methods for alignment:

1). Fast Fourier transform cross-correlation [68, 80]
2). Correlation optimized warping [68, 81]
3). Peak alignment by beam search [82]
4). Fuzzy warping [83]
5). Hierarchical cluster-based peak alignment [82]
6). Local window peak alignment [68]
7). Selection of reference spectrum [68]
8). Recursive segment-wise peak alignment [68]
9). Spectral alignment via wavelet transform and clustering [69, 77]

More alignment methods can be found in the alignment review article [82].

Regardless of the method chosen, during the alignment process, the distance between two neighboring peaks might be increased or decreased.

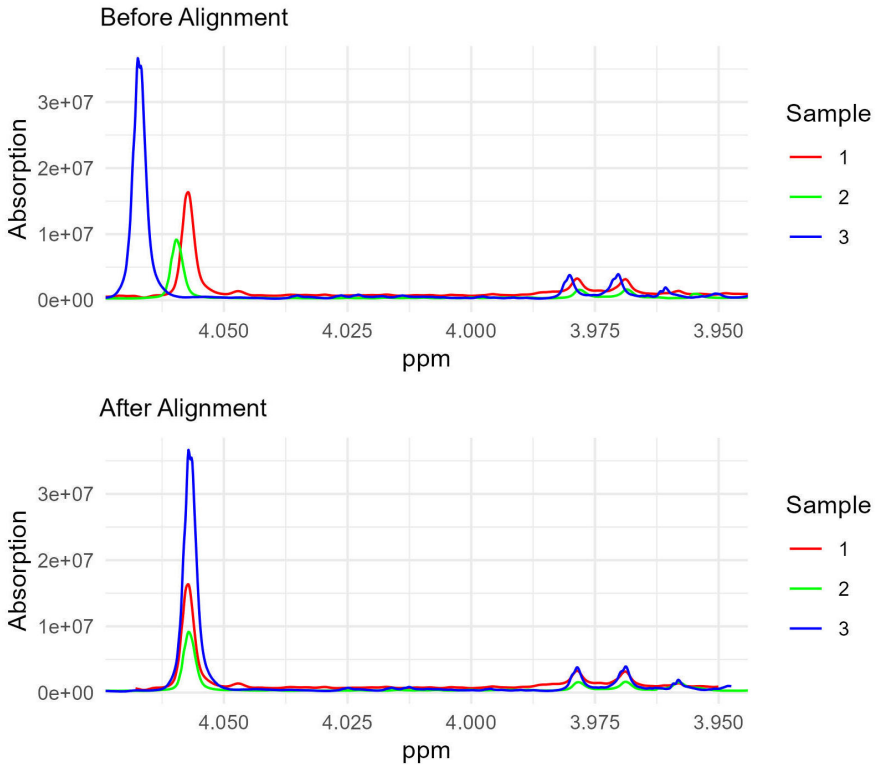
Figure 8 shows the effect of alignment. While alignment ensures that the same peaks are matched across different spectra, improving visual consistency, it can also alter the distances between peaks within a single spectrum. This is evident in peaks within sample 2 and sample 3, where the intra-sample peak distances have changed.

The decrease in intra-sample peak distance could affect peak areas and quantification [14]. Therefore, it has been suggested that quantification should be processed on unaligned spectra to avoid this potential issue [14]. Also, calibration can be applied without alignment; however, alignment should not be applied before calibration.

3.5. Reference deconvolution

In reference deconvolution, the internal reference signal undergoes a transformation into a Lorentzian line, defining an ideal peak shape. This process then extends to all signals, eliminating lineshape

Figure 8
Illustration of alignment in NMR spectra.



Top panel: Before alignment. Bottom panel: After alignment. Each color represents a partial segment of a real-world absorption spectrum in a urine sample

distortion across the spectrum. FIDDLE (Free Induction Decay Deconvolution for Lineshape Enhancement), a widely used reference deconvolution method [29, 84–86], begins with the generation of a pseudo-reference-only spectrum. The spectrum is then transformed into the time domain to obtain the reference-only FID. Through simulation, this FID is deconvolved to achieve an ideal FID with a Lorentzian lineshape. Adjustments using this ideal FID are made to the full FID, resulting in a corrected whole FID, which is then finalized by transforming it back into the frequency domain.

For 2D NMR data, a combined approach integrates reference deconvolution with peak alignment using a "reference spectrum" (also known as the "average spectrum") derived from PCA [87]. The calculation of the first principal component (PC1) for each peak represents it in the "average spectrum," aligning peaks across spectra with matching phase values to those in the PC1 "average spectrum." Despite not requiring a Lorentzian lineshape, the PCA-based method's alignment process may lead to discontinuous baselines and distorted overlapping peaks [88]. Some researchers adopt a hybrid approach, integrating FIDDLE and PCA, by replacing FIDDLE's lineshape with the average lineshape from PCA. While effective for groups of spectra in 2D NMR, this method assumes aligned peaks share the same shape and location, making it particularly suitable for DOSY (diffusion-ordered spectroscopy) data but not universally applicable.

Reference deconvolution primarily addresses lineshape distortion from phase errors [88], not corrected in the phase error correction step. Given the strong assumptions inherent in all reference methods, caution is advised against the indiscriminate application of reference deconvolution.

3.6. Binning, peak picking, and intelligent binning

Binning, or bucketing, divides a spectrum into fixed-width ranges [70, 76], while peak picking, also known as intelligent binning, identifies peaks [70, 89].

Fixed-width binning might cause signals to be split or combined, resulting in nonmeaningful bin summary data [14]. It also struggles with overlapping peaks, and comparability is hindered by alignment issues [39, 68].

Intelligent binning, employing artificial intelligence (AI) approaches, overcomes these challenges, generating more meaningful divided ranges [68, 89, 90]. Techniques such as wavelet transformations, dynamic algorithms, and Gaussian or exponential functions are used to detect peak edges [14, 69, 91]. AI binning allows small ppm adjustments across spectra and can be applied to each bin after fixed binning when complex computations are involved [69, 92].

Challenges in AI binning include peak screening, necessitating threshold definition considering factors like signal-to-noise ratio and variance. Prior knowledge and manual intervention may also be necessary for effective peak screening [90].

NMRNet, a deep learning approach for automated peak picking [93], identifies peaks by locating points with higher intensity in the spectrum, excluding those below the noise level. The key challenge is distinguishing true peaks from noise, treated as a binary classification problem. NMRNet addresses this by inputting retained peaks into a convolutional neural network, calculating probabilities for their significance. The final step refines the peak list through rule-based filtering. This

process involves normalizing resolution and intensity, aiding peak identification. However, normalized data isn't directly usable for quantification.

Another novel algorithm, DEEP Picker, focuses on peak picking and extends its functionality to include deconvolution. Developed by Li et al. [89], DEEP Picker utilizes a deep neural network-based approach, employing a sliding window and stacked convolutional layers for point-by-point spectrum prediction. The algorithm classifies each spectrum point into three categories (Class 2 peaks, Class 1 peaks, and Class 0 non-peaks) using a neural network architecture that includes seven 1D convolutional layers, a max-pooling layer, and a SoftMax activation function for classification. However, similar to other peak picking and intelligent binning methods, the determination of low peak amplitude cutoffs, which could vary from protein to protein and from sample to sample for the same protein, relies on prior knowledge [89].

There is no doubt that peak picking and intelligent binning are much better than fixed-width binning, and while their existing methods show promise and can be automated, human inspection is necessary to train more accurate models and allow room for the development of new methods in the future. Additionally, if the peak picking process involves normalization, these normalized data should not be used for further analysis especially quantification.

3.7. Peak fitting/deconvolution and compound identification

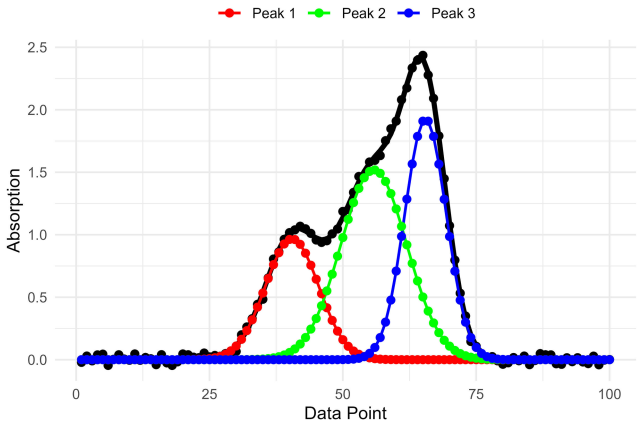
In this step, our aim is to identify molecules from data signals or "peaks" using peak fitting and deconvolution. Peak fitting precisely defines peak characteristics, while deconvolution untangles overlapping peaks, separating different molecule contributions. Figure 9 illustrates the challenging preprocessing step of deconvolution, which involves optimization using specific loss functions, such as the sum of squared differences [90].

Multiple deconvolution methods have been presented in recent years, including the following examples:

1). In the DEEP pipeline by Li et al. [94], a single convolutional layer with a linear activation function is employed to forecast peak characteristics, including position, amplitude, width, and the Lorentzian fraction of the peak shape.
2). Häckl et al. [71] created a user-friendly R package for fully automated deconvolution of overlapping signals using Lorentzian lineshapes. The process involves constructing individual Lorentz curves for each signal, requiring a peak selection procedure and parameter approximation method. The integral of the Lorentz curve is ultimately used as the area under the curve for singlets or multiplets.
3). Prostko et al. [60] developed a customized and automated deconvolution method for ssNMR mixture spectra, employing linear combination modeling by integrating reference spectra of pure solid-state components.
4). Schmid et al. [95] presented a robust deep learning-based deconvolution algorithm for 1D experimental NMR spectra, leveraging a neural network trained on synthetic spectra with customized preprocessing and labeling for accurate estimation of critical peak parameters.

The compound identification usually relies on libraries like the Human Metabolome Database and Biological Magnetic Resonance Data Bank. Lefort et al. [77] developed the R package ASICS, which includes a metabolite library comprising 190 spectra. The identification of metabolites is accomplished by fitting a mixture model to the library spectra, employing a sparse penalty, and quantifying the concentration of metabolites in a complex spectrum. Wang et al. [96] developed NMRQNet, which aims to establish a deep learning-based pipeline for the automatic identification and quantification of predominant metabolite candidates in human plasma samples.

Figure 9
An example of a partial simulated 1D NMR spectrum (in black) and its three deconvolution peaks (color-coded)



Challenges arise due to potential data-library incompatibility from different sources or handling methods. We must remain vigilant about these issues during compound identification [10, 97].

3.8. Integration and quantification

In the integration and quantification process, we employ summation within specific ranges, facilitating the quantification process. This involves determining the concentration of each molecule in the sample based on the area under the curve of the peaks [64].

Integration of signals is conceptually straightforward with raw intensities, although some researchers prefer to integrate absolute intensities [71]. The challenge lies in defining signal edges intelligently, a task initiated in the peak picking and peak fitting/deconvolution steps. A simple approach is to use a range of 24 times the signal width for integration, but caution is advised as this may inadvertently include unintended signals [98]. A more practical way is to set the integration range to be at least twice the full width at half maximum. However, using a too-narrow range may lead to challenges like overlapping peaks. Therefore, it is advisable to restrict integration to datasets featuring sparse, well-phased peaks and devoid of baseline or macromolecule interference [27].

Quantification relies on factors such as the area of a signal, the number of nuclei in the signal, the area of a reference signal, the number of nuclei in the reference, and notably, the reference's concentration in the specimen. In the absence of an internal reference concentration, alternative methods include external references or electronic references [98]. While internal references generally offer more accurate concentration estimations, care should be taken if they interact with other signals [91]. In cases where area determination is challenging, such as in 13C NMR spectra, height may be used instead.

When multiple signals contribute to the concentration estimation of a compound, the choice is between selecting the most stable and isolated peak or calculating the mean value from multiple signals. In instances of multiple technical replicates, concentration estimation should be based on the mean value across these spectra [64].

Although quantification is typically based on a reference signal, for comparative analysis, Canlet et al. [63] employed additional methods in the metabolite quantification process. These methods included determining concentrations from peak areas using a regression model, a calibration curve, calibration-range solutions, and a sum of Voigt pseudo-function shapes fitted through a combination of Gaussian and Lorentzian functions with optimization. Other researchers may also use peak fitting for quantification [29, 87, 88].

While these fit lines contain fewer or no random errors, they might deviate from observed spectrum data [34], leading to inaccurate peak areas and compound concentrations. Additionally, if the research goal is to identify significantly different peaks between two groups of spectra, using "error-free" numbers can potentially reduce or underestimate variance between groups and increase the false positive rate.

3.9. Normalization and transformation

This step aims to make data comparable or suitable for the assumptions needed in subsequent statistical analysis.

3.9.1. Normalization

Normalization is to make data comparable, which can be classified into spectrum-wise and location-wise normalization and can involve various approaches.

a. Spectrum-wise normalization

Spectrum-wise techniques, like dividing peak areas by total spectrum area [74, 77, 90, 99], assume equal total signal quantities, possibly impractical in diverse spectra. An alternative is normalizing using an internal reference area [14, 76], adaptable to binned NMR data.

Less common spectrum-wise techniques include distribution-based strategies like quantile normalization [37, 100, 101], histogram (matching) normalization [14], and spline normalization to align data distributions. Quantile normalization orders and transforms values across spectra to achieve uniform distributions. Histogram normalization scales data based on minimum and maximum values from a reference spectrum [102]. When a reference spectrum is unavailable, the average or median spectrum across a group of spectra can serve for histogram normalization. Spline normalization fits quantiles from experimental and reference spectra to a smooth cubic spline, which is then used to generate normalized data for the experimental spectrum [101, 103]. Similarly, the cubic spline can be replaced by LOWESS (locally weighted scatterplot smoothing) [104].

Among these techniques, reference-based normalization, relying on a spike-in internal reference with a known concentration, is widely regarded as the most robust choice.

b. Location-wise normalization

Location-wise normalization ensures the comparability of a variable across different locations. While methods in this section can be applied to spectrum-wise normalization (Section a), those in Section a are generally not applicable here. The simplest method is variable centering, which involves subtracting the mean or median across spectra for the same location and adding a constant [14].

Level scaling adjusts variables by dividing them by their mean at the same location across spectra, promoting alignment and facilitating comparative analysis [14, 17]. Unit variance scaling (auto-scaling) standardizes variables by dividing each by its standard deviation, ensuring all variables contribute equally to analysis regardless of their initial scale. Vast scaling enhances sensitivity to mean differences by multiplying unit variance-scaled data by their coefficients of variation, highlighting variations effectively [14, 105]. Pareto scaling mitigates the impact of large variances while preserving data structure, making it suitable for datasets with heterogeneous variance. Range scaling adjusts variables based on their range, facilitating comparisons across different scales by normalizing their values relative to their spread [96].

Standardization, a traditional normalization method, involves subtracting the mean and dividing by the standard deviation. However, direct standardization is not applicable to NMR data due to positivity concerns. A variation involves subtracting the mean, adding a constant, and then dividing by the standard deviation.

Vignoli et al. [106] compiled a list of 23 state-of-the-art normalization methods, recognizing the elusive consensus on optimal normalization due to the contingent nature of method choices based on available information and research goals.

The absence of consensus regarding the ideal normalization method emphasizes the necessity for ongoing research and evaluation in this field. Diverse methods may yield varying interpretations regarding the data's structure and variable significance and impact results [105]. In current practice, it is essential to ensure consistency by applying a specific normalization method consistently throughout an entire experiment to maintain data comparability. Additionally, regardless of the chosen normalization approach, there is a risk of amplifying the noise range, potentially compromising the integrity of the entire dataset if noise is misclassified as peaks during the peak picking step. Lastly, it is crucial to note that location-wise normalized data should not be utilized for quantification purposes, as it obscures quantity differences among peaks.

3.9.2. Transformation

Transformation is applied to each variable in NMR data to align the data with the assumptions required by a statistical method. The most commonly used transformation is the log transformation, enhancing normality and mitigating heteroscedasticity [14]. It's important to note that log transformation is unsuitable for nonpositive numbers, and its nonlinear nature may lead to noise amplification. The G-log transformation refers to a generalized log transformation or its variants [107–109]. While high values are logarithmically transformed similar to the regular log transformation, low values or noise undergo specialized transformation to avoid noise amplification issues. Implementing G-log requires prior knowledge of high- and low-value thresholds [17, 90].

The Box-Cox transformation aims to find the optimal power transformation for effective normalization, reducing non-normality effects and eliminating heteroscedasticity [14, 110].

Regardless of the transformation method used, it is crucial to note that while variable values can be transformed back to the original scale, reverting variances and 95% confidence intervals to the original scale poses challenges.

With the exception of internal reference-based area spectrum-wise normalization, all other methods in Section 3.9 are tailored for statistical analysis, not molecule quantification.

4. Conclusion

In conclusion, this comprehensive review provides a detailed exploration of NMR data preprocessing in both the time and frequency domains. In the time domain, it carefully covers key preprocessing steps, including DC offset removal through methods like phase cycling, EC correction primarily addressing phase errors, two directions of FID shift and LP, the impact of weighting functions, zero-filling ratio, and choices in domain transformation. Emphasizing the importance of each step for reliable data analysis, the review discusses potential distortions and provides guidelines for application.

Transitioning to the frequency domain, the article delves into the intricacies of NMR data preprocessing, spotlighting critical steps. Dealing with nonlinear phase errors can be challenging, but the "NMRphasing" R package offers potential solutions. Baseline correction methods and solvent filtering techniques are discussed with attention to potential distortions. The review also covers alignment methods and their impact on quantification. While reference deconvolution aims to address lineshape distortion, the assumptions behind it are often not practical. Additionally, the review discusses binning strategies and emerging AI approaches, recognizing the need for human intervention. Challenges in compound identification, integration, quantification, and a comprehensive overview of normalization and transformation techniques tailored for statistical analysis are addressed, underscoring the careful selection of methods to ensure accurate NMR data interpretation.

Among these preprocessing steps, nonlinear phase error correction, peak picking, intelligent binning, and peak deconvolution present notable challenges. While various methods exist, promising avenues for improvement are offered by optimization processes, particularly those aided by AI techniques and deep learning with neural networks. However, adapting neural networks to NMR data requires balancing complexity with practical application, which poses a significant challenge similar to other deep learning applications. Additionally, the size limitation of NMR datasets poses a formidable obstacle to effectively training deep learning models.

Strategies to overcome this size limitation include aggregating NMR spectra from various sources and implementing normalization methods across datasets to create large, comparable training datasets. However, as discussed earlier, normalized spectra cannot be used for quantification, adding another layer of challenge compared to deep learning in other fields such as natural language processing. A potential solution is to apply a traceable normalization method before deep learning on training datasets, maintaining consistency with new spectra but reverting to non-normalized spectra after intelligent binning and peak deconvolution.

Alternatively, incorporating various factors, such as source differences, into deep learning models may enhance performance and ease of application, albeit at the expense of increased complexity.

Ultimately, the pursuit of innovative approaches that strike a balance between complexity and applicability will drive advancements in NMR data preprocessing. These advancements have the potential to not only improve the accuracy and reliability of NMR data analysis but also facilitate broader utilization across diverse research domains.

Looking ahead, an exciting future involves developing fully automatic AI programs capable of generating comprehensive data, including lists of components and quantities, immediately after NMR analysis. Achieving this goal requires building a robust training database from diverse samples and developing tailored deep learning algorithms for NMR data. This approach aims to simplify data analysis across scientific disciplines and enhance real-time clinical applications such as MRI and fMRI. By focusing on practical usability, these advancements aim to support researchers in various fields.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

Aixiang Jiang is an Editorial Board Member for *Journal of Data Science and Intelligent Systems* and was not involved in the editorial review or the decision to publish this article. The author delcares that she has no conflicts of interest to this work.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author Contribution Statement

Aixiang Jiang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Barajas, R. F. J., Politi, L. S., Anzalone, N., Schöder, H., Fox, C. P., Boxerman, J. L., . . . , & Hu, L. S. (2021). Consensus recommendations for MRI and pet imaging of primary central nervous system lymphoma: Guideline statement from the international primary CNS lymphoma collaborative group (IPCG). *Neuro-Oncology*, *23*(7), 1056–1071. https://doi.org/10.1093/neuonc/noab020

[2] Bertholdo, D., Watcharakorn, A., & Castillo, M. (2013). Brain proton magnetic resonance spectroscopy: Introduction and overview. *Neuroimaging Clinics of North America*, *23*(3), 359–380. https://doi.org/10.1016/j.nic.2012.10.002

[3] Choi, K., Myoung, S., Seo, Y., & Ahn, S. (2021). Quantitative NMR as a versatile tool for the reference material preparation. *Magnetochemistry*, *7*(1), 15. https://doi.org/10.3390/magnetochemistry7010015

[4] Li, H., Zhang, Q., Duan, Q., Jin, J., Hu, F., Dang, J., & Zhang, M. (2021). Brainstem involvement in amyotrophic lateral sclerosis: A combined structural and diffusion tensor MRI analysis. *Frontiers in Neuroscience*, *15*, 675444. https://doi.org/10.3389/fnins.2021.675444

[5] Liu, S., Xiong, Y., Dai, E., Zhang, J., & Guo, H. (2021). Improving distortion correction for isotropic high-resolution 3D diffusion MRI by optimizing Jacobian modulation. *Magnetic Resonance in Medicine*, *86*(5), 2780–2794. https://doi.org/10.1002/mrm.28884

[6] Silva, M. S. (2017). Recent advances in multinuclear NMR spectroscopy for chiral recognition of organic compounds. *Molecules*, *22*(2), 247. https://doi.org/10.3390/molecules22020247

[7] Sorger, B., & Goebel, R. (2020). Real-time fMRI for brain-computer interfacing. *Handbook of Clinical Neurology*, *168*, 289–302. https://doi.org/10.1016/B978-0-444-63934-9.00021-4

[8] Wei, Y., Yang, C., Jiang, H., Li, Q., Che, F., Wan, S., ..., & Song, B. (2022). Multi-nuclear magnetic resonance spectroscopy: State of the art and future directions. *Insights into Imaging*, *13*(1), 135. https://doi.org/10.1186/s13244-022-01262-z

[9] Blonder, N., & Delaglio, F. (2021). The NMR spectral measurement database: A system for organizing and accessing NMR spectra of therapeutic proteins. *Journal of Research of the National Institute of Standards and Technology*, *126*, 126035. https://doi.org/10.6028/jres.126.035

[10] Cuperlovic-Culf, M., Cormier, K., Touaibia, M., Reyjal, J., Robichaud, S., Belbraouet, M., & Turcotte, S. (2016). ^{1}H NMR metabolomics analysis of renal cell carcinoma cells: Effect of VHL inactivation on metabolism. *International Journal of Cancer*, *138*(10), 2439–2449. https://doi.org/10.1002/ijc.29947

[11] Ferry-Dumazet, H., Gil, L., Deborde, C., Moing, A., Bernillon, S., Rolin, D., ..., & Jacob, D. (2011). MeRy-B: A web knowledgebase for the storage, visualization, analysis and annotation of plant NMR metabolomic profiles. *BMC Plant Biology*, *11*(1), 1–12. https://doi.org/10.1186/1471-2229-11-104

[12] Patra, A., & Bera, M. (2014). Spectroscopic investigation of new water soluble MnII2 and MgII2 complexes for the substrate binding models of xylose/glucose isomerases. *Carbohydrate Research*, *384*, 87–98. https://doi.org/10.1016/j.carres.2013.12.002

[13] Pintér, G., Hohmann, K. F., Grün, J. T., Wirmer-Bartoschek, J., Glaubitz, C., Fürtig, B., & Schwalbe, H. (2021). Real-time nuclear magnetic resonance spectroscopy in the study of biomolecular kinetics and dynamics. *Magnetic Resonance*, *2*(1), 291–320. https://doi.org/10.5194/mr-2-291-2021

[14] Smolinska, A., Blanchet, L., Buydens, L. M. C., & Wijmenga, S. S. (2012). NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta*, *750*, 82–97. https://doi.org/10.1016/j.aca.2012.05.049

[15] Bonneau, E., & Legault, P. (2014). NMR localization of divalent cations at the active site of the Neurospora VS ribozyme provides insights into RNA-metal-ion interactions. *Biochemistry*, *53*(3), 579–590. https://doi.org/10.1021/bi401484a

[16] Brzezinska, J., Gdaniec, Z., Popenda, L., & Markiewicz, W. T. (2014). Polyaminooligonucleotide: NMR structure of duplex DNA containing a nucleoside with spermine residue, N-[4,9,13-triazatridecan-1-yl]-2′-deoxycytidine. *Biochimica et Biophysica Acta (BBA) – General Subjects*, *1840*(3), 1163–1170. https://doi.org/10.1016/j.bbagen.2013.12.008

[17] Ebbels, T. M. D., & De Iorio, M. (2011). Statistical data analysis in metabolomics. In Michael P. H. Stumpf, David J. Balding, Mark Girolami (Eds.), *Handbook of Statistical Systems Biology*, 163–180. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119970606.ch8

[18] Lu, G. J., & Opella, S. J. (2014). Resonance assignments of a membrane protein in phospholipid bilayers by combining multiple strategies of oriented sample solid-state NMR. *Journal of Biomolecular NMR*, *58*(1), 69–81. https://doi.org/10.1007/s10858-013-9806-y

[19] Smith, E. A. (2013). Advanced techniques in pediatric abdominopelvic oncologic magnetic resonance imaging. *Magnetic Resonance Imaging Clinics*, *21*(4), 829–841. https://doi.org/10.1016/j.mric.2013.06.002

[20] Rule, G. S., & Hitchens, T. K. (2006). *NMR spectroscopy*. Netherlands: Springer.

[21] Han, P. K., Park, H., & Park, S. H. (2017). DC artifact correction for arbitrary phase-cycling sequence. *Magnetic Resonance Imaging*, *38*, 21–26. https://doi.org/10.1016/j.mri.2016.12.015

[22] Ahn, C. B., & Cho, Z. H. (1991). Analysis of the eddy-current induced artifacts and the temporal compensation in nuclear magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *10*(1), 47–52. https://doi.org/10.1109/42.75610

[23] Jiru, F. (2008). Introduction to post-processing techniques. *European Journal of Radiology*, *67*(2), 202–217. https://doi.org/10.1016/j.ejrad.2008.03.005

[24] Bieri, O., Markl, M., & Scheffler, K. (2005). Analysis and compensation of eddy currents in balanced SSFP. *An Official Journal of the International Society for Magnetic Resonance in Medicine*, *54*(1), 129–137. https://doi.org/10.1002/mrm.20527

[25] Andersson, J. L. R., & Sotiropoulos, S. N. (2015). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078. https://doi.org/10.1016/j.neuroimage.2015.10.019

[26] Klose, U. (1990). *In vivo* proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine*, *14*(1), 26–30. https://doi.org/10.1002/mrm.1910140104

[27] Near, J., Harris, A. D., Juchem, C., Kreis, R., Marjańska, M., Öz, G., ..., & Gasparovic, C. (2021). Pre-processing, analysis and quantification in single-voxel magnetic resonance spectroscopy: Experts' consensus recommendations. *NMR in Biomedicine*, *34*(5), e4257. https://doi.org/10.1002/nbm.4257

[28] Tax, C. M. W., Bastiani, M., Veraart, J., Garyfallidis, E., & Okan Irfanoglu, M. (2022). What's new and what's next in diffusion MRI pre-processing. *NeuroImage*, *249*, 118830. https://doi.org/10.1016/j.neuroimage.2021.118830

[29] Ben-Tal, Y., Boaler, P. J., Dale, H. J. A., Dooley, R. E., Fohn, N. A., Gao, Y., ..., & Lloyd-Jones, G. C. (2022). Mechanistic analysis by NMR spectroscopy: A users guide. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *129*, 28–106. https://doi.org/10.1016/j.pnmrs.2022.01.001

[30] Feng, H., Cai, S., Chen, Z., Lin, M., & Feng, J. (2008). Application of the forward linear prediction on high-resolution NMR spectra in inhomogeneous fields. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *71*(3), 1027–1031. https://doi.org/10.1016/j.saa.2008.02.041

[31] Bigler, P. (2008). *NMR spectroscopy: Processing strategies*. USA: John Wiley & Sons.

[32] Atta-ur-Rahman, & Choudhary, M. I. (1996). The basics of modern NMR spectroscopy. In Atta-ur-Rahman & M. I. Choudhary (Eds.), *Solving problems with NMR spectroscopy* (pp. 1–89). Academic Press. https://doi.org/10.1016/B978-012066320-0/50003-3

[33] Beckmann, N. (2006). *In Vivo* magnetic resonance techniques and drug discovery. *Brazilian Journal of Physics*, *36*, 16–22.

[34] Betson, T. R., Augusti, A., & Schleucher, J. (2006). Quantification of deuterium isotopomers of tree-ring cellulose using nuclear magnetic resonance. *Analytical Chemistry*, *78*(24), 8406–8411. https://doi.org/10.1021/ac061050a

[35] Croitor-Sava, A., Beck, V., Sandaite, I., Van Huffel, S., Dresselaers, T., Claus, F., ..., & Deprest, J. (2015). High-resolution ^{1}H NMR spectroscopy discriminates amniotic fluid of fetuses with congenital diaphragmatic

hernia from healthy controls. *Journal of Proteome Research*, *14*(11), 4502–4510. https://doi.org/10.1021/acs.jproteome.5b00131

[36] del Campo, G., Zuriarrain, J., Zuriarrain, A., & Berregi, I. (2016). Quantitative determination of carboxylic acids, amino acids, carbohydrates, ethanol and hydroxymethylfurfural in honey by ^{1}H NMR. *Food Chemistry*, *196*, 1031–1039. https://doi.org/10.1016/j.foodchem.2015.10.036

[37] Luck, M., Le Moyec, L., Barrey, E., Triba, M., Bouchemal, N., Savarin, P., & Robert, C. (2015). Energetics of endurance exercise in young horses determined by nuclear magnetic resonance metabolomics. *Frontiers in Physiology*, *6*, 198. https://doi.org/10.3389/fphys.2015.00198

[38] Motegi, H., Tsuboi, Y., Saga, A., Kagami, T., Inoue, M., Toki, H., . . . , & Kikuchi, J. (2015). Identification of reliable components in multivariate curve resolution-alternating least squares (MCR-ALS): A data-driven approach across metabolic processes. *Scientific Reports*, *5*, 15710–15710. https://doi.org/10.1038/srep15710

[39] Ebbels, T. M. D., Lindon, J. C., & Coen, M. (2011). Processing and modeling of nuclear magnetic resonance (NMR) metabolic profiles. In J. M. Walker (Ed.), *Methods in molecular biology* (pp. 365–388). Springer. https://doi.org/10.1007/978-1-61737-985-7_21

[40] Van Horn, W. D., Beel, A. J., Kang, C., & Sanders, C. R. (2010). The impact of window functions on NMR-based paramagnetic relaxation enhancement measurements in membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1798*(2), 140–149. https://doi.org/10.1016/j.bbamem.2009.08.022

[41] Altenhof, A. R., Mason, H., & Schurko, R. W. (2023). Desperate: A Python library for processing and denoising NMR spectra. *Journal of Magnetic Resonance*, *346*, 107320. https://doi.org/10.1016/j.jmr.2022.107320

[42] Koehl, P. (1999). Linear prediction spectral analysis of NMR data. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *34*(3–4), 257–299. https://doi.org/10.1016/s0079-6565(99)00002-3

[43] Bretthorst, G. L. (1992). Bayesian analysis. V. Amplitude estimation for multiple well-separated sinusoids. *JMRES Journal of Magnetic Resonance (1969)*, *98*(3), 501–523.

[44] Krishnamurthy, K. (2013). CRAFT (complete reduction to amplitude frequency table)—Robust and time-efficient Bayesian approach for quantitative mixture analysis by NMR. *Magnetic Resonance in Chemistry*, *51*(12), 821–829. https://doi.org/10.1002/mrc.4022

[45] Zhang, S., Zheng, C., Lanza, I. R., Nair, K. S., Raftery, D., & Vitek, O. (2009). Interdependence of signal processing and analysis of urine ^{1}H NMR spectra for metabolic profiling. *Analytical Chemistry*, *81*(15), 6080–6088.

[46] Barache, D., Antoine, J. P., & Dereppe, J. M. (1997). The continuous wavelet transform, an analysis tool for NMR spectroscopy. *YJMRE Journal of Magnetic Resonance*, *128*(1), 1–11.

[47] Balacco, G. (1994). A new criterion for automatic phase correction of high-resolution NMR-spectra which does not require isolated or symmetrical lines. *Journal of Magnetic Resonance, Series A*, *110*(1), 19–25. https://doi.org/10.1006/jmra.1994.1175

[48] Bao, Q., Feng, J., Chen, L., Chen, F., Liu, Z., Jiang, B., & Liu, C. (2013). A robust automatic phase correction method for signal dense spectra. *Journal of Magnetic Resonance*, *234*, 82–89. https://doi.org/10.1016/j.jmr.2013.06.012

[49] Brown, D. E., Campbell, T. W., & Moore, R. N. (1989). Automated phase correction of FT NMR spectra by baseline optimization. *Journal of Magnetic Resonance (1969)*, *85*(1), 15–23. https://doi.org/10.1016/0022-2364(89)90315-6

[50] Chen, L., Weng, Z., Goh, L., & Garland, M. (2002). An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization. *Journal of Magnetic Resonance*, *158*(1), 164–168. https://doi.org/10.1016/S1090-7807(02)00069-1

[51] de Brouwer, H. (2009). Evaluation of algorithms for automated phase correction of NMR spectra. *Journal of Magnetic Resonance*, *201*(2), 230–238. https://doi.org/10.1016/j.jmr.2009.09.017

[52] Džakula, Ž. (2000). Phase angle measurement from peak areas (PAMPAS). *Journal of Magnetic Resonance*, *146*(1), 20–32. https://doi.org/10.1006/jmre.2000.2123

[53] Ernst, R. R. (1969). Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy. *Journal of Magnetic Resonance (1969)*, *1*(1), 7–26. https://doi.org/10.1016/0022-2364(69)90003-1

[54] Hardy, E. H., Hoferer, J., Mertens, D., & Kasper, G. (2009). Automated phase correction via maximization of the real signal. *Magnetic Resonance Imaging*, *27*(3), 393–400. https://doi.org/10.1016/j.mri.2008.07.009

[55] Herring, F. G., & Phillips, P. S. (1984). Automatic phase correction in magnetic resonance using DISPA. *Journal of Magnetic Resonance (1969)*, *59*(3), 489–496. https://doi.org/10.1016/0022-2364(84)90083-0

[56] Heuer, A. (1991). A new algorithm for automatic phase correction by symmetrizing lines. *Journal of Magnetic Resonance (1969)*, *91*(2), 241–253. https://doi.org/10.1016/0022-2364(91)90189-Z

[57] Hoffman, R. E., Delaglio, F., & Levy, G. C. (1992). Phase correction of two-dimensional NMR spectra using DISPA. *Journal of Magnetic Resonance (1969)*, *98*(2), 231–237. https://doi.org/10.1016/0022-2364(92)90129-U

[58] Larry Bretthorst, G. (2008). Automatic phasing of MR images. Part I: Linearly varying phase. *Journal of Magnetic Resonance*, *191*(2), 184–192. https://doi.org/10.1016/j.jmr.2007.12.010

[59] Montigny, F., Elbayed, K., Brondeau, J., & Canet, D. (1990). Automatic phase correction of Fourier-transform NMR data and estimation of peak area by fitting to a Lorentzian shape. *Analytical Chemistry*, *62*(8), 864–867. https://doi.org/10.1021/ac00207a019

[60] Prostko, P., Pikkemaat, J., Selter, P., Lukaschek, M., Wechselberger, R., Khamiakova, T., & Valkenborg, D. (2022). R Shiny App for the automated deconvolution of NMR spectra to quantify the solid-state forms of pharmaceutical mixtures. *Metabolites*, *12*(12), 1248. https://doi.org/10.3390/metabo12121248

[61] Sotak, C. H., Dumoulin, C. L., & Newsham, M. D. (1984). Automatic phase correction of Fourier transform NMR spectra based on the dispersion versus absorption (DISPA) lineshape analysis. *Journal of Magnetic Resonance (1969)*, *57*(3), 453–462. https://doi.org/10.1016/0022-2364(84)90260-9

[62] Wachter, E. A., Sidky, E. Y., & Farra, T. C. (1989). Calculation of phase-correction constants using the DISPA phase-angle estimation technique. *Journal of Magnetic Resonance (1969)*, *82*(2), 352–359.

[63] Canlet, C., Deborde, C., Cahoreau, E., Da Costa, G., Gautier, R., Jacob, D., . . . , & Giraudeau, P. (2023). NMR metabolite

quantification of a synthetic urine sample: An inter-laboratory comparison of processing workflows. *Metabolomics*, *19*(7), 65. https://doi.org/10.1007/s11306-023-02028-4

[64] Corol, D. I., Harflett, C., Beale, M. H., & Ward, J. L. (2014). An efficient high throughput metabotyping platform for screening of biomass willows. *Metabolites*, *4*(4), 946–976. https://doi.org/10.3390/metabo4040946

[65] Gan, Z., & Hung, I. (2022). Second-order phase correction of NMR spectra acquired using linear frequency-sweeps. *Magnetic Resonance Letters*, *2*(1), 1–8. https://doi.org/10.1016/j.mrl.2021.100026

[66] Jaroszewicz, M. J., Altenhof, A. R., Schurko, R. W., & Frydman, L. (2023). An automated multi-order phase correction routine for processing ultra-wideline NMR spectra. *Journal of Magnetic Resonance*, *354*, 107528. https://doi.org/10.1016/j.jmr.2023.107528

[67] Pizzolato, M., Gilbert, G., Thiran, J. P., Descoteaux, M., & Deriche, R. (2020). Adaptive phase correction of diffusion-weighted images. *NeuroImage*, *206*, 116274. https://doi.org/10.1016/j.neuroimage.2019.116274

[68] Sun, J., & Xia, Y. (2023). Pretreating and normalizing metabolomics data for statistical analysis. *Genes & Diseases, 11*(3), 100979. https://doi.org/10.1016/j.gendis.2023.04.018

[69] Vu, T. N., Valkenborg, D., Smets, K., Verwaest, K. A., Dommisse, R., Lemière, F., . . . , & Laukens, K. (2011). An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics*, *12*(1), 405. https://doi.org/10.1186/1471-2105-12-405

[70] Chai, X., Liu, C., Fan, X., Huang, T., Zhang, X., Jiang, B., & Liu, M. (2023). Combination of peak-picking and binning for NMR-based untargeted metabonomics study. *Journal of Magnetic Resonance*, *351*, 107429. https://doi.org/10.1016/j.jmr.2023.107429

[71] Häckl, M., Tauber, P., Schweda, F., Zacharias, H. U., Altenbuchinger, M., Oefner, P. J., & Gronwald, W. (2021). An R-package for the deconvolution and integration of 1D NMR data: MetaboDecon1D. *Metabolites*, *11*(7), 452. https://doi.org/10.3390/metabo11070452

[72] Wang, X., Mickiewicz, B., Thompson, G. C., Joffe, A. R., Blackwood, J., Vogel, H. J., & Kopciuk, K. A. (2022). Comparison of two automated targeted metabolomics programs to manual profiling by an experienced spectroscopist for ^{1}H-NMR spectra. *Metabolites*, *12*(3), 227. https://doi.org/10.3390/metabo12030227

[73] Zailer, E., & Diehl, B. W. K. (2016). Alternative determination of blood alcohol concentration by ^{1}H NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, *119*, 59–64. https://doi.org/10.1016/j.jpba.2015.11.030

[74] Deborde, C., Fontaine, J. X., Jacob, D., Botana, A., Nicaise, V., Richard-Forget, F., . . . , & Molinié, R. (2019). Optimizing 1D ^{1}H-NMR profiling of plant samples for high throughput analysis: Extract preparation, standardization, automation and spectra processing. *Metabolomics*, *15*(3), 28. https://doi.org/10.1007/s11306-019-1488-3

[75] Giraudeau, P., Silvestre, V., & Akoka, S. (2015). Optimizing water suppression for quantitative NMR-based metabolomics: A tutorial review. *Metabolomics*, *11*(5), 1041–1055. https://doi.org/10.1007/s11306-015-0794-7

[76] Altenbuchinger, M., Berndt, H., Kosch, R., Lang, I., Dönitz, J., Oefner, P. J., . . . & Zacharias, H. U. (2022). Bucket fuser: Statistical signal extraction for 1D ^{1}H NMR metabolomic data. *Metabolites*, *12*(9), 812. https://doi.org/10.3390/metabo12090812

[77] Lefort, G., Liaubet, L., Canlet, C., Tardivel, P., Père, M. C., Quesnel, H., . . . , & Servien, R. (2019). ASICS: An R package for a whole analysis workflow of 1D ^{1}H NMR spectra. *Bioinformatics*, *35*(21), 4356–4363. https://doi.org/10.1093/bioinformatics/btz248

[78] Günther, U. L., Ludwig, C., & Rüterjans, H. (2002). WAVEWAT—Improved solvent suppression in NMR spectra employing wavelet transforms. *Journal of Magnetic Resonance*, *156*(1), 19–25. http://doi.org/10.1006/jmre.2002.2534

[79] Ribay, V., Dey, A., Charrier, B., Praud, C., Mandral, J., Dumez, J. N., . . . , & Giraudeau, P. (2023). Hyperpolarized 13C NMR spectroscopy of urine samples at natural abundance by quantitative dissolution dynamic nuclear polarization. *Angewandte Chemie International Edition*, *62*(27), e202302110. https://doi.org/10.1002/anie.202302110

[80] Savorani, F., Tomasi, G., & Engelsen, S. B. (2010). Icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, *202*(2), 190–202. https://doi.org/10.1016/j.jmr.2009.11.012

[81] Pravdova, V., Walczak, B., & Massart, D. L. (2002). A comparison of two algorithms for warping of analytical signals. *Analytica Chimica Acta*, *456*(1), 77–92. https://doi.org/10.1016/S0003-2670(02)00008-9

[82] Vu, T. N., & Laukens, K. (2013). Getting your peaks in line: A review of alignment methods for NMR spectral data. *Metabolites*, *3*(2), 259–276. https://doi.org/10.3390/metabo3020259

[83] Wu, W., Daszykowski, M., Walczak, B., Sweatman, B. C., Connor, S. C., Haselden, J. N., . . . , & Lutz, M. W. (2006). Peak alignment of urine NMR spectra using fuzzy warping. *Journal of Chemical Information and Modeling*, *46*(2), 863–875. https://doi.org/10.1021/ci050316w

[84] Gibbs, A., Morris, G. A., Swanson, A. G., & Cowburn, D. (1993). Suppression of t1 noise in 2D NMR spectroscopy by reference deconvolution. *Journal of Magnetic Resonance, Series A*, *101*(3), 351–356. https://doi.org/10.1006/jmra.1993.1058

[85] Metz, K., Lam, M., & Webb, A. (2000). Reference deconvolution: A simple and effective method for resolution enhancement in nuclear magnetic resonance spectroscopy. *Concepts in Magnetic Resonance*, *12*, 21–42.

[86] Morris, G. A., Barjat, H., & Home, T. J. (1997). Reference deconvolution methods. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *31*(2), 197–257. https://doi.org/10.1016/S0079-6565(97)00011-3

[87] Witjes, H., Melssen, W. J., van der Graaf, M., Heerschap, A., & Buydens, L. M. C. (2000). Automatic correction for phase shifts, frequency shifts, and lineshape distortions across a series of single resonance lines in large spectral data sets. *Journal of Magnetic Resonance*, *144*(1), 35–44. https://doi.org/10.1006/jmre.2000.2021

[88] Huo, R., Wehrens, R., & Buydens, L. M. C. (2004). Improved DOSY NMR data processing by data enhancement and combination of multivariate curve resolution with non-linear least square fitting. *Journal of Magnetic Resonance*, *169*(2), 257–269. https://doi.org/10.1016/j.jmr.2004.04.019

[89] Li, D. W., Hansen, A. L., Bruschweiler-Li, L., Yuan, C., & Brüschweiler, R. (2022). Fundamental and practical aspects of machine learning for the peak picking of biomolecular NMR spectra. *Journal of Biomolecular NMR*, *76*(3), 49–57. https://doi.org/10.1007/s10858-022-00393-1

[90] Morris, G. A. (2017). NMR data processing. *Encyclopedia of Spectroscopy and Spectrometry*, 125–133. https://doi.org/10.1016/B978-0-12-409547-2.05103-9

[91] Puchades-Carrasco, L., Palomino-Schätzlein, M., Pérez-Rambla, C., & Pineda-Lucena, A. (2016). Bioinformatics tools for the analysis of NMR metabolomics studies focused on the identification of clinically relevant biomarkers. *Briefings in Bioinformatics*, *17*(3), 541–552. https://doi.org/10.1093/bib/bbv077

[92] De Meyer, T., Sinnaeve, D., Van Gasse, B., Tsiporkova, E., Rietzschel, E. R., De Buyzere, M. L., . . . , & Van Criekinge, W. (2008). NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Analytical Chemistry*, *80*(10), 3783–3790. https://doi.org/10.1021/ac7025964

[93] Klukowski, P., Augoff, M., Zieba, M., Drwal, M., Gonczarek, A., & Walczak, M. J. (2018). NMRNet: A deep learning approach to automated peak picking of protein NMR spectra. *Bioinformatics*, *34*(15), 2590–2597. https://doi.org/10.1093/bioinformatics/bty134

[94] Li, D. W., Hansen, A. L., Yuan, C., Bruschweiler-Li, L., & Brüschweiler, R. (2021). DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra. *Nature Communications*, *12*(1), 5229. https://doi.org/10.1038/s41467-021-25496-5

[95] Schmid, N., Bruderer, S., Paruzzo, F., Fischetti, G., Toscano, G., Graf, D., . . . , & Wilhelm, D. (2023). Deconvolution of 1D NMR spectra: A deep learning-based approach. *Journal of Magnetic Resonance*, *347*, 107357. https://doi.org/10.1016/j.jmr.2022.107357

[96] Wang, W., Ma, L., Maletic-Savatic, M., & Liu, Z. (2023). NMRQNet: A deep learning approach for automatic identification and quantification of metabolites using nuclear magnetic resonance (NMR) in human plasma samples. *bioRxiv: 2023.03.01.530642*. https://doi.org/10.1101/2023.03.01.530642

[97] Hao, J., Astle, W., De Iorio, M., & Ebbels, T. M. D. (2012). BATMAN—An R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics*, *28*(15), 2088–2090. https://doi.org/10.1093/bioinformatics/bts308

[98] Argyropoulos, D., & Avizonis, D. (2010). Electronic referencing in quantitative NMR. *eMagRes*. https://doi.org/10.1002/9780470034590.emrstm1168

[99] Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry*, *78*(7), 2262–2267. https://doi.org/10.1021/ac0519312

[100] Huang, Y., Su, B. N., Ye, Q., Palaniswamy, V. A., Bolgar, M. S., & Raglione, T. V. (2014). Improving the efficiency of quantitative ^{1}H NMR: An innovative external standard-internal reference approach. *Journal of Pharmaceutical and Biomedical Analysis*, *88*, 1–6. https://doi.org/10.1016/j.jpba.2013.07.043

[101] Kohl, S. M., Klein, M. S., Hochrein, J., Oefner, P. J., Spang, R., & Gronwald, W. (2012). State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics: Official Journal of the Metabolomic Society*, *8*, 146–160. https://doi.org/10.1007/s11306-011-0350-z

[102] Sun, X., Shi, L., Luo, Y., Yang, W., Li, H., Liang, P., . . . , & Wang, D. (2015). Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. *BioMedical Engineering OnLine*, *14*(1), 73. https://doi.org/10.1186/s12938-015-0064-y

[103] Workman, C., Jensen, L. J., Jarmer, H., Berka, R., Gautier, L., Nielser, H. B., . . . , & Knudsen, S. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology*, *3*(9), 1–16. https://doi.org/10.1186/gb-2002-3-9-research0048

[104] Berger, J. A., Hautaniemi, S., Järvinen, A. K., Edgren, H., Mitra, S. K., & Astola, J. (2004). Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC Bioinformatics*, *5*(1), 194. https://doi.org/10.1186/1471-2105-5-194

[105] van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, *7*(1), 142. https://doi.org/10.1186/1471-2164-7-142

[106] Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., . . . , & Luchinat, C. (2019). High-throughput metabolomics by 1D NMR. *Angewandte Chemie International Edition*, *58*(4), 968–994. https://doi.org/10.1002/anie.201804736

[107] Durbin, B. P., Hardin, J. S., Hawkins, D. M., & Rocke, D. M. (2002). A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*, *18*, S105–S110. https://doi.org/10.1093/bioinformatics/18.suppl_1.S105

[108] Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., & Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, *18*, S96–S104. https://doi.org/10.1093/bioinformatics/18.suppl_1.S96

[109] Rocke, D., Lee, G. C., Tillinghast, J., Durbin-Johnson, B., & Wu, S. (2012). *LMGene software for data transformation and identification of differentially expressed genes in gene expression arrays*. Retrieved from: http://www.bioconductor.org/packages//2.7/bioc/html/LMGene.html

[110] Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, *15*(1), 12. https://doi.org/10.7275/qbpc-gk17

How to Cite: Jiang, A. (2024). Insights into Nuclear Magnetic Resonance Data Preprocessing: A Comprehensive Review. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS42022556

APPENDIX

Websites for NMR data preprocessing software mentioned in the text

Software	Website
TopSpin	https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html
ACD/Labs	https://www.acdlabs.com/
Mnova	https://mestrelab.com/software/mnova/
Chenomx	https://www.chenomx.com/
iNMR	https://www.inmr.net/
NMRbox	https://nmrbox.nmrhub.org/
NMRPipe	https://www.ibbr.umd.edu/nmrpipe/
AlpsNMR	https://bioconductor.org/packages/release/bioc/html/AlpsNMR.html
NMRphasing	https://cran.r-project.org/web/packages/NMRphasing/
nmrrr	https://cran.r-project.org/web/packages/nmrrr/index.html
PepsNMR	https://bioconductor.org/packages/release/bioc/html/PepsNMR.html
Rnmr1D	https://cran.r-project.org/web/packages/Rnmr1D/index.html
speaq	https://cran.r-project.org/web/packages/speaq/index.html
nmrespy	https://pypi.org/project/nmrespy/
dnpLab	https://pypi.org/project/dnplab/
Protomix	https://pypi.org/project/Protomix/
peakipy	https://pypi.org/project/peakipy/
ssnmr	https://pypi.org/project/ssnmr/
metabolabpy	https://pypi.org/project/metabolabpy/
nmrglue	https://pypi.org/project/nmrglue/
spike-py	https://pypi.org/project/spike-py/
klassez	https://pypi.org/project/klassez/
nmrpy	https://pypi.org/project/nmrpy/
pynmr	https://pypi.org/project/pynmr/