TARGETED FREQUENCY DOMAIN CORRECTION OF LINESHAPE DISTORTIONS INTRODUCED BY MAGNETIC RESONANCE INHOMOGENEITY IN QUANTITATIVE NMR SPECTROSCOPY

by

Theodore Street

Submitted in partial fulfilment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
August 2022

© Copyright by Theodore Street, 2022

For all those whose hard work this builds upon.

Table of Contents

Abstract . . . . . **vi**

Acknowledgements . . . . . **vii**

Chapter 1 Introduction **1**

Chapter 2 Objectives and Approach **3**

Chapter 3 Literature Review **6**

3.1 Background 6

3.2 FID Inputs 10
3.2.1 FID to FID 10
3.2.2 FID to Parameters 11
3.2.3 FID to Frequency Domain 15

3.3 Frequency Inputs 16
3.3.1 Frequency to Frequency 16
3.3.2 Frequency to Parameters 19
3.3.3 Machine Learning Techniques 22
3.3.4 Frequency to Area 23

3.4 Discussion 26
3.4.1 Local vs Global Methods 26
3.4.2 Prior Knowledge 26
3.4.3 Time vs Frequency Domain 28
3.4.4 Automation 29
3.4.5 Stand alone vs Integrated 29
3.4.6 Identification of Potential Areas for Improvement 30

Chapter 4 Theoretical Elements of Magnetic Resonance Inhomogeneity **32**

4.1 Representation of Peaks 32

4.2 Magnetic Resonance Inhomogeneity 33

Chapter 5 Proposed algorithm **37**

5.1 Implementation 37

5.1.1 Singlet Software Implementation 37
5.1.2 Multiple Peaks Software Implementation 38

5.2 Testing Methodology 41
5.2.1 Simulated Data 41
5.2.2 Experimental Data - Singlet with Inserted Shim Error 44
5.2.3 Experimental Data - Multiple Compound Spectrum 47
5.2.4 Reference Deconvolution Comparison 49

Chapter 6 Results & Discussion 50

6.1 Simulated Data 50

6.2 Experimental Data - Singlet with Inserted Shim Error 52

6.3 Robustness Results 54

6.4 Experimental Data - Multiple Compound spectrum 60

6.5 Reference Deconvolution Testing 63

Chapter 7 Conclusion 68

Appendix A Initial Parameters for Multiple Compound Spectrum . 71

Appendix B Full Results for Multiplet Compound Spectrum 73

Appendix C Singlet Algorithm - Complete R Package 74

Appendix D Multiplet Algorithm - Complete R Package 83

Bibliography 85

List of Tables

Table 5.1 Initial spectral parameters for simulated NMR multiplet data. . 42

Table 5.2 Magnetic resonance inhomogeneity parameters for simulated NMR multiplet data. 44

Table 5.3 Shim error types and values introduced in singlet experimental data. 46

Table 5.4 Initial peak parameters for singlet with inserted shim error. . . 46

Table 5.5 Initial algorithm parameters for singlet with inserted shim error. 46

Table 5.6 Initial robustness testing parameters for singlet with inserted shim error. 47

Table 6.1 Area accuracy results for simulated NMR data comparison. . . 52

Table 6.2 Area accuracy results for experimental NMR singlet data comparison. 54

Table 6.3 Truncated area accuracy results for experimental NMR multiplet data comparison. 63

Table 6.4 Area accuracy results for reference deconvolution and proposed algorithm on NMR multiplet data and a comparison of the difference in area accuracy results. 66

Table A.1 Initial peak parameters for multiplet spectrum with lineshape inhomogeneity. 71

Table A.2 Initial algorithm parameters for multiplet spectrum with lineshape inhomogeneity. 72

Table B.1 Area accuracy results for experimental NMR multiplet data comparison. 73

List of Figures

Figure 3.1 NMR data visualised as an FID in the time domain (top) and a set of peaks in the frequency domain (bottom). 7

Figure 3.2 NMR data processing steps visualised as a pipeline of analysis techniques. 9

Figure 4.1 Representation of a physical sample suspended in NMR magnetic coil (left) and a demonstration of lineshape distortions produced by magnetic field inhomogeneity and its impact on a molecule's chemical shift (right). 35

Figure 4.2 Examples of common NMR shim errors in the frequency (left) and time (right) domains. 36

Figure 5.1 Overview of the implementation process for correction of lineshape distortion produced by magnetic resonance inhomogeneity that is present across a single NMR peak. 38

Figure 5.2 Theoretical frequency domain representation of a single peak containing lineshape distortions due to magnetic resonance inhomogeneity as the summation of multiple ideal Lorentzian peak areas with offset chemical shifts. 39

Figure 5.3 Overview of the implementation process for correction of lineshape distortion produced by magnetic resonance inhomogeneity that is present across multiple NMR peaks. 40

Figure 5.4 Frequency domain spectrum consisting of simulated Lorentzian NMR data with lineshape distortion due to magnetic resonance inhomogeneity added. 43

Figure 5.5 Experimental data consisting of the NMR spectrum of alanine (3.8, 1.5 ppm), glycine (3.5 ppm), and ethanol (1.3, 3.7 ppm) in water in the frequency domain. 48

Figure 5.6 Experimental data consisting of the NMR spectrum of alanine (3.8, 1.5 ppm), glycine (3.5 ppm), ethanol (1.3, 3.7 ppm), and sucrose (4.8 ppm) in water in the frequency domain. 48

Figure 6.1 Qualitative comparison of a simulated NMR spectrum with introduced lineshape distortions and the results of the algorithm fitting that spectrum. 51

Figure 6.2 Qualitative comparison of NMR singlet spectrum with introduced lineshape distortions and the results of the algorithm fitting that spectrum. 53

Figure 6.3 Qualitative comparison of NMR sidebands and the results of the algorithm fitting those sidebands. 55

Figure 6.4 Area accuracies and lineshape fit robustness testing results of varying FWHM initial guess parameter. 57

Figure 6.5 Area accuracies and lineshape fit robustness testing results of varying peak position (F0) initial guess parameter. 58

Figure 6.6 Area accuracies and lineshape fit robustness testing results of varying peak height/intensity initial guess parameter. 59

Figure 6.7 Area accuracies and lineshape fit robustness testing results of varying upper/lower bounds initial guess parameter. 59

Figure 6.8 Qualitative comparison of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum. 60

Figure 6.9 Qualitative comparison displaying the lower intensity values of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum. 61

Figure 6.10 Qualitative comparison displaying the ethanol peaks portion of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum. 62

Figure 6.11 Qualitative comparison displaying the glycine peaks portion of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum. 62

Figure 6.12 NMR lineshape of a spectrum containing alanine, glycine, ethanol, and sucrose demonstrating the presence of phase error in the sample. 64

Figure 6.13 Qualitative comparison of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the reference deconvolution package MetaboDecon1D fitting that spectrum. 64

Figure 6.14 Qualitative comparison displaying amino acid peaks from an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the reference deconvolution package MetaboDecon1D fitting that spectrum. 65

Abstract

A review has been performed of the NMR software landscape that determines ideal qNMR software to be local, frequency-domain based and as automated as feasible. Within this landscape, reference deconvolution corrects lineshape distortions using a time-domain local/global technique and thus is a technique that has been singled out for improvement. A novel algorithm was created based on theoretical knowledge of magnetic resonance inhomogeneity as a summation of scaled, shifted, central peaks. This algorithm produced sufficient accuracy when applied to simulated (0.067-0.232%), and experimentally obtained data (0.025-18.04%), is robust to reasonable changes in initial peak parameters, and outperforms reference deconvolution by 0.85-25.7%. Overall, this algorithm is an effective solution for the correction of lineshape distortions produced by magnetic resonance inhomogeneity that operates entirely in the frequency domain and implements a global consensus fit, allowing for distinct improvement over reference deconvolution.

Acknowledgements

First and foremost, I'd like to thank my supervisor, Dr. Stanislav Sokolenko for all the guidance and expertise he has provided over the past two years. Likewise, I am grateful to Dr. Jan Haelsigg and Dr. Rob Beiko whose mentorship on my supervisory committee has been invaluable. I am also immensely grateful for the assistance in running samples of Mike Lumsden, Michelle Combe, and Ruis MacDonald, and for the support of Leila Rezaei. Without you all this thesis would not have been possible, and I am forever grateful to each of you. Without Julie O'Grady and Paula Colicchio the past two years would have been unnavigable and I am grateful for their administrative support. I am further grateful to Dr. Sue Budge for her advice as the department graduate coordinator. Finally, I'd like to thank my friends and family both near and far who have supported me steadfastly throughout my degree. I couldn't have made it without you all.

Chapter 1

Introduction

Nuclear Magnetic Resonance (NMR) is an important analytical tool that is widely used to determine the presence of certain organic compounds (characterisation), and calculation of relative concentrations (quantification) of those compounds within solutions or mixtures [1, 2]. As a tool for characterisation, NMR is used in organic chemistry [3, 4], structural biology [5–7], drug discovery [8], and recently with larger, more complex datasets in metabolomic profiling [9–14]. There are also other wide ranging applications of NMR detailed in literature [15–17]. Quantitative NMR (qNMR) is especially important in biological analysis as it allows for the determination of concentrations of compounds within biological systems, and can allow for studies of how the concentration of these compounds change over time [18, 19], however there are also applications for qNMR in pharmaceuticals [20], law enforcement [21], and organic chemistry [22]. Each of these fields require a precise and accurate determination of peak area and as such need to have processing software with the capacity to deal with a variety of errors that impact the lineshape and peak area.

This breadth of applications of NMR spectroscopy coupled with differences in complexity of datasets requires software that is effective, efficient, and robust in its ability to characterise the compounds present and quantify the concentration of those compounds. Software approaches vary, and some techniques are better suited to analyse certain datasets than others. It is important to consider these fundamental differences when creating NMR data processing software. For example, samples with many atoms, such as biological or metabolomic mixtures, often experience spectrally overlapping peaks, decreasing the ability of NMR to characterise and quantify the atoms present. NMR also has limited applicability to non-organic compounds, as the atoms elucidated by this technique are limited to those with specific spin properties (1H, 13C, 2H, 3He, 15N, 19F, etc). Likewise, it is important to understand that each

NMR experiment requires relatively large amounts of sample compared to GCMS or other analytical techniques [23], thus increasing the cost of each sample run. Further, NMR machines are expensive and thus it is desirable that each sample is run as few times as possible. It should be noted, however, that low-cost low field NMR spectrometers exist, although these suffer from greater magnetic field inhomogeneity which causes a greater degree of lineshape error [24]. NMR spectrometers also require skilled technicians to operate and resolve errors and thus it is desirable to limit the amount of time each sample takes to run and analyse. These constraints suggest that software should be automated, fast, accurate and capable of quantifying areas without a large number of samples run.

A class of errors that NMR technicians are expected to understand and account for are shim errors. These errors cause distortion in the lineshape peaks and are caused by inhomogeneity in the NMR magnetic field. Fitting NMR data with this peak distortion is especially challenging in crowded NMR spectra as overlapping peaks can become indistinguishable. Traditionally, a set of small magnetic coils termed shim coils are added to the NMR spectrometer, and the technician varies the strength of these coils prior to the experimental run in order to minimise the field inhomogeneity. Extensive effort has been made to design and optimise these coils [25, 26], and software exists that can correct for a certain amount of this class of error [27], however these approaches currently still require a technician determining the presence of shim error and making manual adjustments prior to the NMR experiment being run.

Chapter 2

Objectives and Approach

NMR spectroscopy data processing techniques employ a variety of approaches to refine and process experimental data. These approaches include techniques applied to time domain data, transformation techniques to convert time domain data to frequency domain data, and processing approaches that focus solely on frequency domain data. Each approach has advantages and disadvantages. In this thesis a detailed literature review of these different software techniques is presented and discussed, and each technique is catalogued according to a pipeline approach of NMR data processing. This review serves to highlight the strengths and deficiencies of each approach in order to identify weaknesses or gaps in current NMR data processing tools. In addition, it creates a conceptual framework for researchers to classify and compare their techniques relative to similar approaches at each step in the data processing pipeline.

An analysis of the results of this literature review revealed several potential areas of NMR data processing where improvements could be made. It was determined that a significant bottleneck in data processing is the correction of a class of lineshape errors introduced by magnetic resonance inhomogeneity, including what is commonly termed "shim error". This step in the pipeline has been singled out and a novel technique developed in order to correct for this class of errors. The method developed and applied herein implements a correction technique that accounts specifically for the presence of lineshape distortions due to magnetic resonance inhomogeneity in NMR spectra and creates a model lineshape to represent this error as a summation of scaled, shifted central peaks. These peaks are then optimised to determine correction factors for a single peak. Further, this algorithm expands the technique to multiple peaks by implementing a consensus fit across multiple NMR peaks in the same spectrum and determines a common set of shim correction factors that can be applied to each peak

and optimised across all peaks. This algorithm is implemented in the R programming language.

Entirely automated methods for the correction of lineshape distortions due to magnetic resonance inhomogeneity have previously been proposed and implemented as part of a "catch-all" approach termed reference deconvolution [27, 28]. This approach determines a reference signal in the NMR spectrum, preferably a well-resolved singlet, and then compares the experimental peak to the ideal theoretical peak. The difference in lineshape is then applied to the other peaks in the spectrum. A key benefit of reference deconvolution is that it corrects for a wide variety of hardware errors that are applied equally to each peak, such as shifts in peak position, radio-frequency pulse phase error, and shim error [29]. This is a well established technique, however it is not targeted at magnetic resonance inhomogeneities specifically, and does not take into account any knowledge of the theoretical basis of this class of error. Further, reference deconvolution uses one well-resolved peak to calculate errors and applies these across every peak. Thus, any lineshape modification that corrects for an error that is not actually uniform across every peak will nevertheless be applied to each peak. These undesirable properties have been corrected for in the proposed technique.

The overall goal of this thesis is to further the field of NMR spectroscopy data processing by researching current approaches in the field, assigning those techniques to a logical framework, and improving a weak link in this pipeline by developing a targeted algorithm to correct for lineshape distortions caused by magnetic resonance inhomogeneity in the frequency domain using a global consensus fit method. This novel technique has been validated using both simulated and experimental data and the results are presented and discussed. Within this thesis is presented the following:

1. A literature review of qNMR spectroscopy data processing- focusing on the processes applied to the spectrum at each stage and classifying each process in a pipeline of steps

2. A detailed review of magnetic resonance errors and the theory underpinning the proposed algorithm

3. A broad overview of the proposed method and the algorithm, and the testing methodology used to validate the algorithm

4. The results of algorithm testing and a discussion of its effectiveness

Chapter 3

Literature Review

3.1 Background

As detailed previously, NMR spectroscopy can either be used to qualitatively determine whether certain compounds are present in a sample, or quantitatively determine the concentration of specific compounds in a sample. qNMR is more challenging as it requires a greater level of accuracy in the NMR spectrum in order to achieve accurate results. Within NMR spectroscopy, quantification refers to obtaining an accurate calculation of the concentration of each component within the organic molecule studied. There are two primary domains that NMR data may be used in; the time or frequency domain. Raw data from the NMR spectrometer is an exponentially decaying sinusoid in the time domain called a free induction decay (FID) which may then be transformed into a spectral plot of the different frequencies present in the FID (the frequency domain). These two typical domains are shown in Fig. 3.1. In the frequency domain, relative areas of each peak can be compared to one another to determine the relative concentrations of each element, or a reference compound with known concentration can be included in the sample [30]. The area of each peak can then be compared to this reference peak and used for absolute concentration quantification. Quantification can also occur in the time domain by determining the parameters and type of lineshape and then calculating the area based on knowledge of the shape of the lineshape. An overview of possible processing steps in qNMR analysis follows:

A. FID collection: An NMR spectrometer creates a discrete, complex valued plot of sinusoidal intensity decaying as a function of time (FID). This data is input to the data processing software program

B. The FID input from the spectrometer is manipulated in a way to enhance resolution and return an FID output. These techniques include zero-filling, apodisation,

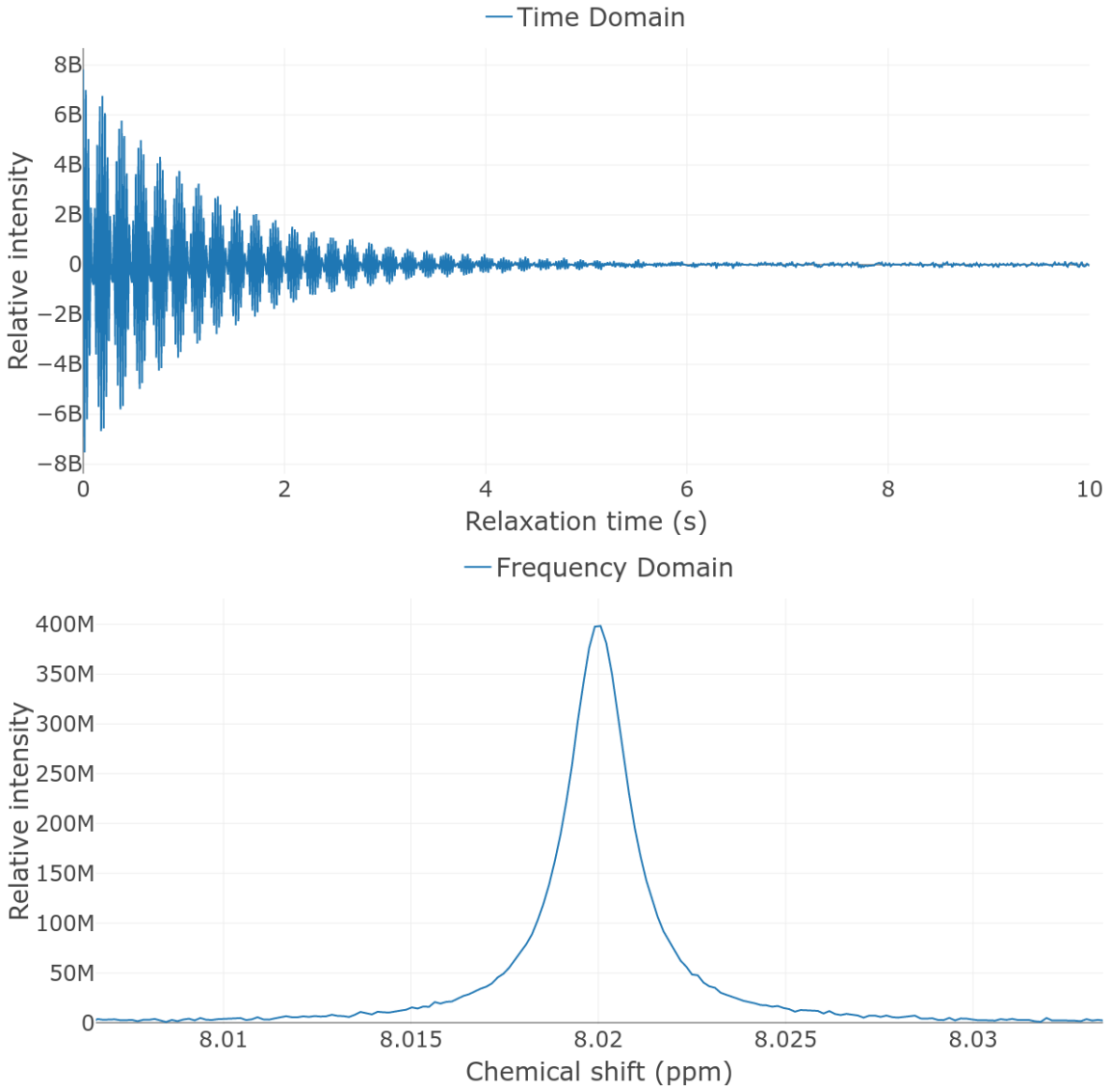

Figure 3.1: NMR data visualised as an FID in the time domain (top) and a set of peaks in the frequency domain (bottom).

etc and are discussed in section 3.2.1. This step is commonly referred to as pre-processing

C. The FID can at this stage be mathematically decomposed using linear algebra based methods like SVD or LP and a set of lineshape parameters (peak height, width etc) returned. The area of the peak, and thus the relative concentration is then easily calculated using these peak parameters. These techniques are discussed in detail in section 3.2.2

D. Alternatively, the FID may be decomposed using the Fourier Transform, returning a dataset corresponding to the varying sinusoids with their respective intensities. See section 3.2.3 for more details

E. The frequency domain input can be manipulated in a similar way to the FID inputs in B. This generally consists of creating a model lineshape, baseline, shim, or phase correction, and returns a modified set of intensities. See section 3.3.1 for more details

F. Similarly to the FID, it is possible to use SVD, statistical methods, or machine learning to determine the peak parameters directly from the frequency plot. These approaches are discussed in section 3.3.2

G. Alternatively, it is possible to directly compute the area of the frequency inputs using simple integration or binning, as is shown in section 3.3.4

Each of these steps may be conceptually considered as a process performed on a type of data input in order to produce a modified data output. For example, given the raw FID from the NMR spectrometer, it is possible to manipulate the FID data to produce a new FID dataset with improved characteristics. The NMR analyst chooses whether to include this process or not, and then which technique to apply to the data (zero-filling, apodisation, etc). Once the analyst has chosen which processes to include, data flows through this 'pipeline' in order to produce an accurate area. This approach is shown graphically in Fig. 3.2. It should also be noted that since each step in this process represents a possible data processing step, there are thus a variety of ways one can determine the area of the lineshape peak. For example, one can work

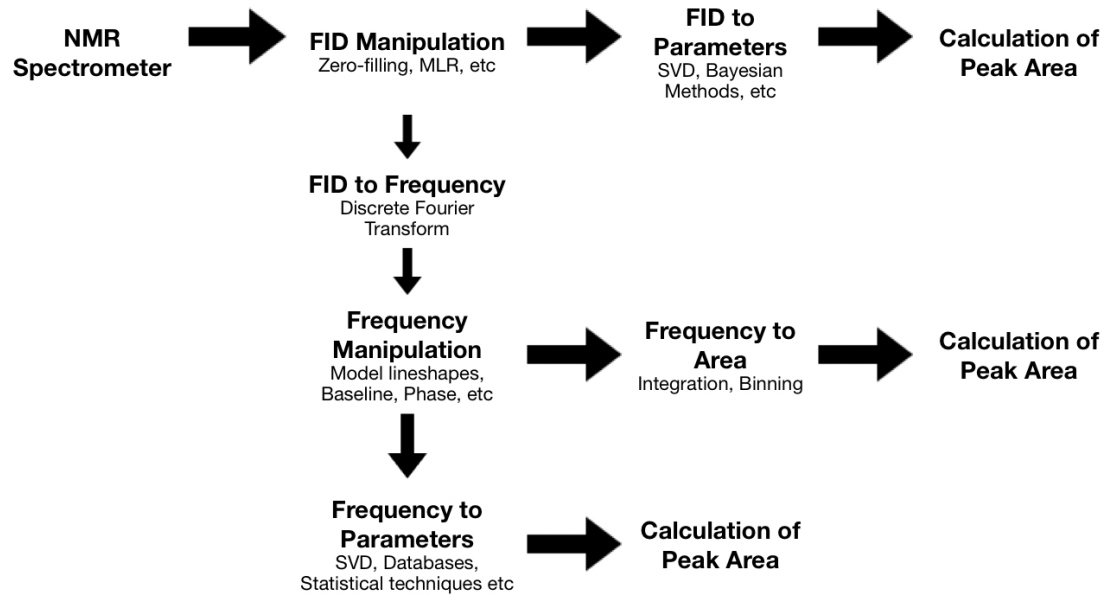

Figure 3.2: NMR data processing steps visualised as a pipeline of analysis techniques.

solely with FIDs (time domain) and perform only steps B and C. Or the analyst may wish to work almost exclusively with frequency domain datasets and go from A to D to E or G. These steps are thus building blocks that represent a variety of different pipelines that can be used to analyse NMR data.

When considering different NMR data processing approaches, it should be noted that this review is a sampling of common techniques and is not a comprehensive review of every technique or method. Those interested in particular approaches should conduct further research using the references herein. Further, there are a number of excellent reviews that focus on some of these techniques and packages, such as those by Vanhamme [31], Mierisova [32], and Poullet [33]. Those reviews focus on techniques and approaches, and there are also many reviews describing and comparing complete packages that implement these techniques, such as those by Izquierdo-Garcia [34], and Spicer [35]. These reviews are more suited to those researchers looking for comprehensive packages already created rather than those who wish to implement their own package consisting of a desired pipeline of techniques.

3.2 FID Inputs

This section focuses on techniques that may be applied to FID data. There are three broad groups of techniques. The first focuses on applying techniques to FID data and producing another set of improved FID datapoints. The vast majority of techniques used to resolve or improve the FID are based on the principle of linear prediction. This class of algorithms was introduced to NMR by Barkhuijsen et al in 1985 [36]. The second group of techniques transforms FID data into a set of peak parameters using algebraic decomposition or statistical methods, and the third produces a set of intensity datapoints using the FID data, generally through application of the Fourier Transform (FT).

3.2.1 FID to FID

Zero-Filling

A common data processing technique to improve the digital resolution of FIDs is zero-filling. It is desirable to collect as long of an FID as possible to ensure a high resolution dataset, however, since the signal decays over time any increase in the amount of time that the FID is recorded will rapidly decrease the signal-to-noise ratio and result in a poor quality spectrum. This challenge can be overcome by appending zeroes to the end of the FID. This approach is non-ideal as it introduces discontinuity when the lineshape data abruptly changes from non-zero to zero value. These discontinuities introduce error in calculating areas in future steps and are termed leakage errors [37]. An effective approach to mitigate this class of errors is to multiply the FID with an apodisation or weighting function that emphasises the initial data points and decays to zero. For example, an exponentially decaying function, or a linear function that begins at a non-zero intensity value and gradually decays to a zero value across the length of the FID. Zero filling using these approaches is an effective way to improve resolution of NMR spectra, and by minimising the latter parts of the FID serves to also increase the signal to noise ratio of the data.

Maximum Likelihood Reconstruction

An alternative to zero-filling is to replace the latter part of the FID with an extrapolation of data points based on the trend of decaying signal observed in the earlier parts of the FID. This can improve the signal to noise ratio, and avoids the loss of data that zero-filling the latter part of the signal includes. The downsides to this approach are that estimating the latter part of the FID can be challenging and must be performed as accurately as possible to ensure an improvement in resolution. This approach is often implemented using Burg maximum entropy, or maximum entropy extrapolation and has been applied in various packages [37, 38]. It is important to note that nomenclature is not consistent, and there are many maximum entropy methods that instead use model lineshapes to determine the peak parameters [39]. See 3.2.2 for a more in depth discussion of these methods.

In a broader sense, rather than just appending values to the end of the FID that fit the earlier data points it is possible to add values to any sparsely sampled point within the FID. Maximum likelihood reconstruction is a technique to add points in between any sampled points in order to improve the resolution of FIDs. This technique estimates where a point (or multiple points) should be placed in between already sampled points, and is used to more accurately and easily quantify the spectral peaks. It is based on the principle that sparsely sampled time domain data produces a low resolution spectrum, and that the resolution can be improved by inserting calculated data points into the sparse regions. This is especially important in 2D NMR and can greatly improve multidimensional NMR data analysis. Maximum likelihood reconstruction is also easier to implement than reconstructing the latter part of the FID as simple linear or low order interpolation models can be more accurately used to calculate the points.

3.2.2 FID to Parameters

Model Functions

An important technique in time domain NMR data processing is to convert the FID into a set of mathematical parameters based on an understanding of the FID as

a decaying exponential function. A common approach to accomplish this is to create a model lineshape of an exponentially decaying sinusoid and minimise the squared difference between the data and the model lineshape. This is typically accomplished using nonlinear least squares (NLLS) optimisation, however other techniques to minimise the difference and obtain the maximum likelihood estimate have been explored. As detailed rigorously in Vanhamme's review [31] and the references therein, these include minimising variable projection functions, using global procedures such as simulated annealing or genetic algorithms, and expectation-maximisation (EM) methods. These model lineshape techniques are similar to the maximum likelihood reconstruction techniques described in 3.2.1 and differ mainly in application. MLR is used to improve the FID resolution whereas creating model lineshapes is intended to determine the parameters of the FID and thus quantify the peak area. Further, creating a model lineshape requires estimating the mathematical function behind the data and reconstituting the entire lineshape, whereas MLR only interpolates sample points within the FID or extrapolates the FID to future points.

The first method for reconstructing time domain data was linear prediction (LP) [36]. A large number of reconstruction methods build on this fundamental approach and include those techniques by Chen, Stern, Mobli, Qu, and Hyberts [38–46]. Reconstruction techniques specifically applied to multidimensional NMR spectra also include those by Orekhov and Bostock [47, 48]. Often in multidimensional NMR, 1D spectra are obtained for preliminary characterisation before performing the multidimensional scan. Packages that use these guide-FIDs to improve the accuracy of their model lineshape reconstruction include those by Orekhov, Bodart, and Ueda [47, 49, 50].

To improve model lineshape linear prediction and total least squares techniques. Zhu et al. [51] also introduced an iterative quadratic maximum likelihood method. This method is more accurate than LP or TLS methods, however it requires the computation of a root of a polynomial. Similarly to LP this may generate numerical issues. Another approach to improve MLR is the FMLR (fast maximum likelihood reconstruction) algorithm. This is an iterative spectral deconvolution approach for 1D and 2D NMR spectra for the purpose of accurate signal quantification [52]. The

algorithm is a complete package for peak determination based on fitting a model function and is a hybrid time-frequency approach. It demonstrates greater accuracy than peak height analysis and peak integral analysis with greatly reduced operator intervention. FMLR considers well-dispersed, partially dispersed, or overlapped peaks, for treatment of overlapping peaks using regions of interest for the analysis of 1D and 2D NMR spectra.

A major limitation of model lineshapes and MLR is that each technique assumes that the obtained data is of high enough sensitivity that predictive models of reconstruction will be beneficial [53]. If the data points obtained are too sparse, or the signal has a very low signal to noise ratio (SNR) then these techniques will not benefit the analysis and quantification. Additionally, in the time domain, peaks are harder to distinguish from the background noise and thus require techniques to accommodate these background signals when creating and optimising the model lineshape [54].

Singular Value Decomposition

A class of statistical techniques have been proposed that assume the experimental NMR spectrum is formed from multiple exponentially decaying sinusoids. These methods use estimation techniques to determine the parameters of these sinusoids by solving singular value decomposition problems. An introduction to SVD theory as applied to NMR is given by James [37]. These algorithms include linear prediction SVD [36], or Hankel matrix decomposition SVD [55], and provide a fast and accurate method for processing a single spectrum. These methods are ideal for single, non overlapping peaks, and become less reliable when there is a low SNR, or multiple peaks overlap. Limitations of SVD methods include cases where the signal intensity is significantly large (rapidly decaying signals). SVD methods generally fit the original data to Lorentzian models and thus are inferior to methods that have the flexibility to fit the data to Gaussian/Voigt lineshapes in addition to Lorentzian.

There are a variety of different SVD techniques, including early efforts, such as that of Millhauser [56], which are effective at obtaining necessary peak parameters, however due to the lack of prior knowledge can result in infeasible results. The AMARES

algorithm [57] was developed to address this issue by extending the VARPRO [58] peak-fitting method to allow for a greater level of prior knowledge. AQSES [59] improves on the early SVD approaches by using a combination of time domain fitting and splines to model the baseline and uses the variable projection method to estimate the amplitudes of the metabolite basis set, thus reducing the number of model parameters. Similarly, a variant of the package Totally Automatic Robust Quantitation in NMR (TARQUIN) eliminates baseline errors and models the resultant signal with a parameterised basis set using least squares projection to estimate signal amplitudes [60].

Bayesian Methods

Bayesian methods are a prior knowledge based approach that uses statistical methods to determine the probability distribution of the parameters, assuming that the parameters are random variables [61, 62]. In the time domain, Bayesian or other prior knowledge based approaches can improve methods based on linear prediction or state-space theory. Early attempts were only capable of using Lorentzian peaks (an NMR signal with intensity conforming to a Lorentzian distribution), however methods have been introduced, such as that by Belkic et al. [63] that is based on the Padé transform. This method can extract the correct number of resonances from the time domain signal, however it requires a similar level of prior knowledge as SVD based methods. The Padé approximation groups single components even when they may have contributions from more than one biochemical source. Additionally, it struggles with resonance amplitudes with a low SNR [64].

A significant limitation of methods based on linear prediction is that they require an estimation of the number of signals, and thus are less suited for automated signal processing. A different approach includes defining a model lineshape that can be improved using statistical methods such as Maximum Likelihood (ML) or Bayesian analysis as in section 3.2.2 and either minimising the likelihood that the signal is incorrect (ML), or maximising the likelihood that the model is correct (Bayesian). These two approaches have been described in detail in the review by Mobli [42] and

these methods have been applied in various packages, for example in the ChiFit program of Chylla and Markley [65]. An approach that incorporates Bayesian methods in FID parameter estimation using model lineshapes is given by Matviychuk et al [66]. Their approach produces resulting parameter estimates that are robust to both sample noise and field inhomogeneity.

3.2.3 FID to Frequency Domain

FID to frequency is a processing step that transforms the exponentially decaying sinusoidal FID into a frequency domain function. The Fourier Transform is used almost exclusively to accomplish this in NMR and decomposes the FID into a set of points that displays the intensity of each sinusoid as a function of its frequency. It is the most common software approach in NMR as it allows for users to visualise the atoms present as a spectra of lines. This is easier to visualise than a set of decaying sinusoids superimposed on one another. As the FID is a set of discrete points, the Fourier Transform is implemented using a fast Discrete Fourier Transform (DFT) algorithm. Due to subtle mathematical differences between the DFT and FT the frequency domain plot will need pre-processing (such as normalising the height of the spectrum) in order to determine peak parameters or area [37].

Challenges are encountered by the DFT when the FID sample spacing is irregular. A large amount of peak data is included in the earlier parts of the FID, and thus it is preferable to increase the number of samples taken in the early stages of the signal decay, then lower the frequency of sampling as the signal decays further. This is known as non-uniform sampling. The main disadvantage of non-uniform sampling is that it decreases the speed of the FFT. Matsuki's technique [67] incorporates MLR in a unique way in order to improve the speed of the FFT when using non-uniformly sampled data. This is accomplished by using MLR to predict the values in the latter portion of the FID and then Fourier Transforming the spectrum. This is applied in the Spectroscopy with Integration of Frequency and Time domain (SIFT) package. A similar technique has been applied to multidimensional NMR by Mayzel [68].

3.3 Frequency Inputs

The second class of processes discussed in this review deals with inputs in the frequency domain after transformation of the FID. Similarly to the time domain it is possible to manipulate the data in the frequency domain in order to obtain an improved frequency domain plot. For example by creating a model lineshape or through baseline or phase error correction. It is also possible to obtain peak parameters from the frequency data through SVD, machine learning, or other statistical techniques. Finally, it is possible to directly calculate the area from the frequency data by integrating or binning the lineshape over the desired peaks.

3.3.1 Frequency to Frequency

Creating Model Lineshapes

It is possible to create another frequency domain spectrum using a model lineshape that fits the experimentally observed data. Creating model lineshapes using frequency inputs is similar to the time domain methods using FID inputs as discussed in section 3.2.2. This is a fundamental technique in NMR data processing that many packages use in addition to the other techniques described herein (i.e the packages created by Hao, Sokolenko, and Heinecke [69–71]). It is entirely possible, and common, for example, to create a model lineshape that matches the frequency domain experimental data and then integrate this function to determine the concentrations in lieu of integrating the experimental data directly. It is also possible to use the Bayesian statistics of section 3.3.2 to incorporate prior knowledge and improve the lineshape model before calculating the peak area. Model lineshapes are commonly used in conjunction with SVD methods and PCA, and these techniques are necessary in order to determine the parameters of the model lineshape.

An ideal NMR signal has a Lorentzian lineshape. Due to a variety of errors and noise, the real portion of the frequency domain spectrum can be viewed as a sum of either Lorentzian, Gaussian, or a mixture (Voigt) of these two lineshapes with specific signal dependent parameters, or factors that account for the shape and non-ideality of the peak. Thus, a model lineshape can be created that is based on one of these

mathematical forms and then fit to the experimental NMR data. Creating these lineshapes requires the estimation and optimisation of parameters such as lineshape width, height, and how close the peak is to an ideal Lorentzian. Many techniques focus on solving these parameters to create the model lineshape. After parameter estimation, the fit is then compared to the experimental data, often with a least squares approach [72]. A common tool to minimise the non linear sum of squares using local optimisation is by using the Levenberg-Marquardt algorithm [73, 74]. The process is then iterated until the difference between the model lineshape and experimental data is within acceptable bounds and the concentrations are calculated.

Creating a model lineshape is often a suitable approach because it makes the techniques less susceptible to noise and often improves the quantification results [57, 59, 75], however, in the frequency domain, methods based on model functions have the additional wrinkle that for a non Lorentzian FID a simple, exact analytical expression of Voigt and Gaussian lineshapes in the frequency domain are often not available, although numerical approximations abound [76, 77]. Voigt lineshapes have been approximated, and the model function has been fit using the Levenberg-Marquardt algorithm in papers by Gillies and Marshall [77, 78]. The complexity of these model functions in the frequency domain mean they are more computationally intensive, and thus the less parameters required, the better the results will be [79].

When creating the spectral lineshape, it is important that the algorithm used is capable of efficiently determining how many components should be included. Whether that is choosing which metabolite profiles to use, as is the case in AQSES or QUEST, or how many peaks to allow, as in VARPRO or AMARES. The use of a variable projection method (as in VARPRO and AQSES, but not AMARES or QUEST) has been beneficial when using model functions [80] as it encounters fewer problems when experimental peak amplitudes are very small [59]. Thus, variable projection algorithms are preferred particularly when there is uncertainty about which components are present in the data. Iterative quantification methods that are not based on variable projection (such as AMARES) are thus less beneficial for use with complex NMR signals.

The number of components in a spectrum impacts the computational efficiency of every quantification approach, however some methods are particularly sensitive to the number of components. In order to mitigate this weakness, techniques have been proposed that create model lineshapes using maximum likelihood principles, which divide the problem into independent optimisations equivalent to the number of components in the signal [65, 81, 82]. These techniques are very similar to the time domain parameter estimation methods described in section 3.2.2.

A continuous wavelet transform is detailed in research by Serrai [83] that iteratively extracts each resonance from the experimental data. This approach accommodates viewing NMR spectral peaks as either the Lorentzian and Gaussian lineshapes and picks the lineshape that gives the best fit. The choice of lineshape can impact the metabolite peak areas, and thus Voigt lineshapes are preferred when creating model functions as they are a combination of Lorentzian and Gaussian peaks and thus are able to more flexibly model the experimental data accurately [84]. Algorithms such as KNOB-TLS [85] provide parameter estimates which can then be used as starting values in other techniques. Metropolis Frequency-Selective (MeFreS) is a technique based on rank minimisation of a Hankel matrix [86]. This algorithm does not do any preprocessing, however the signal model function is directly fitted. This techniques is compared to AMARESW and VARPRO in the package by Romano [86].

Other Frequency Domain Manipulation Techniques

Other techniques which manipulate the frequency domain data and return a modified frequency spectrum are only discussed briefly herein. For a more detailed review of frequency domain methods the reader is directed to the review article by Mierisova and Ala-Korpela [32]. Briefly, these techniques include correcting any discontinuities or distortions in the baseline of the frequency spectrum. These distortions can be introduced due to phase errors in recording the FID, or the presence of noise artefacts. Baseline errors can be corrected by fitting the baseline either as a linear or low order polynomial function and normalising the spectrum using this fit. Further errors include lineshape distortion due to magnetic resonance inhomogeneity which are introduced due to magnetic field inhomogeneity in the NMR magnet. These will

be discussed in more detail in chapter 4. A final class of errors often accounted for include resolving the presence of large solvent peaks, which can block out other peaks and are often cancelled using baseline fitting and interpolation across the peak in question.

3.3.2 Frequency to Parameters

Frequency Domain SVD

SVD techniques in the frequency domain are similar to those in the time domain. LCModel(TM) [74] incorporates metabolite basis sets into the fitting model. It is effective at short echo time data and works in the frequency domain using a linear combination of signals combined with a smoothing spline to account for baseline signals. It requires appropriate phasing and that the spectrum is in absorption mode. Methods have also been proposed that use complex SVD to analyse spectral data despite the existence of phase errors in the NMR spectrum [87, 88]. A recently developed SVD tool incorporating both automated (SVD) and prior knowledge components used with lineshape fitting is described in the package by Bazgir [89], and SVD techniques as applied to solid state NMR are described in the article by Andrade [90].

Five advanced SVD techniques that are applied to specific spectral regions in the frequency domain are described in the review by Sandgren [91]. These techniques include filter diagonalisation [92, 93], selective frequency SELF-MODE [94], a data filtering and decimation approach FIDO, the ARMA-modelling based technique SB-HOYWSVD [95], and SELF-SVD [96]. Each of these techniques has advantages and disadvantages that are described fully in Sandgren's review [91]. SB-HOYWSVD is the technique that requires the most input parameters, SELF-SVD is generally the fastest method whereas filter diagonalisation gives better estimates. Generally, these frequency selective local approaches are less computationally intensive than time domain SVD methods. Limitations of these SVD methods specific to the frequency domain are often due to pre-processing. For example, baseline errors are often an issue, as the lack of prior knowledge about the peak itself can lead to inclusion or omission of area. This is especially true when due to the broadness of the peak, the signal is more appropriately modelled as a Gaussian rather than a Lorentzian

lineshape [97].

Using Prior Knowledge and Databases

Bayesian statistics are an effective approach that is commonly used to estimate model parameters [61, 62, 66]. Bayesian techniques incorporate prior knowledge into the model, and allow for the estimation of uncertainty of results. Bayesian methods are often used in software that focuses on peak classification [98, 99] although not exclusively [100]. Prior knowledge often takes the form of providing the software with a reference peak and knowledge of the metabolite in question from databases or from simulated metabolite spectra. Likewise, many interactive software approaches use a metabolite profile as a basis. Not all packages require a metabolite basis profile. For example, AMARES (and the updated version AMARESW) [101] is a local optimisation procedure based on the Levenberg-Marquardt algorithm. It is a robust and flexible software package that allows for the use of prior knowledge and can fit echo signals. Packages that use metabolite basis sets/databases are AQSES [59], and QUEST [102]. This basis can be built from simulated spectra (NMR-SCOPE [103], GAMMA [104]), or from in vitro spectra. For example, GAMMA is used in Young's technique [73] to simulate the spectrum and determine information for use in parametric spectral analysis.

As briefly discussed previously, AQSES is a software package that minimises the sum of squares of the difference between the acquired signal and a linear combination of templates. This method overcomes baseline and other distortions, however the templates must be pre-generated and thus requires prior information about the expected lineshape. Further, it is impractical for long FIDs due to its computational load. Other iterative model function fitting methods are also this computationally expensive [59]. Recent focus on Bayesian techniques has been intense, and numerous packages have appeared, including CRAFT [105], packages in completely automated analysis (BQuant) [106], applied to complex 1D spectra (BATMAN) [69], and to 2D NMR [71].

Statistical Methods

Statistical analysis can be performed on NMR datasets in order to identify the components of the substance being analysed and quantify the amplitude or area differences of the peaks. Statistical approaches can be either targeted (local), analysing only specific compounds, or untargeted (global), which evaluates all the features of the spectrum depending on the level of prior knowledge [107]. Likewise, lineshape data can be analysed using multivariate or univariate analysis. Multivariate methods are generally used to visualise biological data, identify clusters, and build predictive models, and univariate methods are used to identify metabolites and pathways that are altered or correlated with specific biological conditions [108, 109]. The first step in statistical analysis is generally automated, and is used to summarise, explore, and discover clusters or trends in the data. This helps visualise the data and discover possible outliers.

Principal component analysis (PCA) is a statistical tool that allows for the determination of variables within a dataset and distinguishes them from noise. It is an untargeted method that is generally only used in the frequency domain. It does not depend on previous information, and is independent of the shape of the peak [110]. This method is advantageous when dealing with large sets of unknown lineshapes, or spectra that include a reference peak, however is not recommended for single peaks or for sets of spectra that have been obtained under different conditions as these variations may be statistically significant. Modern PCA NMR methods incorporate phase, chemical shift, and utilise both the real and complex parts of the frequency spectrum [88]. Further, the use of a similar technique, independent component analysis, has been explored [111]. Updated PCA methods, such as Group-Wise Principal Component Analysis (GPCA) [112] have potential applications to improve automated metabolomic algorithms.

Statistical techniques that require previous knowledge include partial least squares (PLS) regression, partial least squares-discriminant analysis (PLS-DA) and orthogonal partial least squares (OPLS). These techniques are beneficial, however can lead to overfitting [35]. To avoid this, PLS-DA models are usually cross-validated using

internal or external data sets [109]. Another similar class of statistical approaches is data clustering. This technique subdivides samples into clusters of samples that are similar in metabolic features. This includes methods such as K-Means (KM) [113–115], Spectral Clustering (SC) [116], KODOMA [117, 118], and Self-Organising Maps (SOM) [119, 120]. An additional statistical technique that is exclusive to NMR is statistical correlation spectroscopy (STOCSY). STOCSY takes advantage of the multi-collinearity of the frequency variables in a set of 1D NMR spectra to generate a pseudo-two-dimensional NMR spectrum that displays the correlation among the intensities of the various peaks across the whole sample. It is particularly suitable for the identification of metabolites in complex mixtures, such as urine [35].

3.3.3 Machine Learning Techniques

Machine learning is a subset of artificial intelligence that is premised on pattern recognition and the theory that computers can learn without being specifically programmed to perform certain tasks. It depends on training a model using large data sets so the model learns complex patterns and relationships and is able to predict similar relationships given another previously unseen dataset. Machine learning has great applicability in NMR peak prediction and has been used to specify groups of organic compounds starting with alkanes, alkenes and benzenes and later generalised for many other classes of organic molecule. Early pioneers of using machine learning in NMR were Reilly and Kowalski in 1971, and later, the machine learning mainstay, the k-nearest neighbours algorithm was also introduced to NMR [121]. Other approaches such as Support Vector Machines (SVM) [122], Random Forests (RF) [123], and principal least regression [124] have also been implemented. An excellent mini-review of recent ML tools in NMR has been written by Cobas [125].

A major challenge in NMR signal processing with machine learning methods is that this technique is strongly dependent on having a suitable training set, and obtaining a large number of NMR spectra is expensive. This can result in poor or unbalanced training sets. A common approach to mitigate this is K-fold cross validation. K-fold cross validation optimises the accuracy of training using this data set by dividing the training set into K number of subsets (folds). The first fold is used as a validation

set to test the classifier while the remaining K-1 sets are used to train the model. Next, the second fold is used as a validation set and the remaining folds are used to train. This is repeated K times and ensures the classifier has seen the entire training set and validated the model across all folds. K is typically 10-30. Another approach to mitigate the lack of suitable data for a training set is to use simulated data. Qu et al created 40,000 simulated FIDs using a Poisson-gap method [126], and Hansen created a simulated dataset with approximately 8,000,000 FIDs [127]. An alternative method is bootstrapping, which generates new samples by drawing spectra from the original sample with replacement [125].

A major benefit of machine learning methods is that they are much faster than traditional database searching, in many cases orders of magnitude faster. Further, databases in traditional search algorithms must be free of errors and the spectral lineshape of the NMR data must match the database fairly precisely in order to generate accurate results. Machine learning methods only require a suitable training set and are better capable of generalising over the bounds of that training set, allowing them more robustness with spectral data [125]. A critical need in NMR machine learning approaches is the ability to disentangle overlapping peaks. This is important because integration or binning will not produce accurate results if multiple NMR peaks overlap. Techniques to disentangle the overlapping peaks are also important in order to isolate and identify the spectrum of interest [125]. Recent approaches expanding on machine learning include those using Deep Learning (DL) such as using densely connected convolution neural networks [126]. These have been used by NMR manufacturer Bruker, and by third party developers in peak picking applications, [128] among other applications [129, 130].

3.3.4 Frequency to Area

Binning

Binning is a local, automated, frequency domain technique that is common in chemometrics. It allows for simple, rapid calculation of peak area, and can overcome chemical shift errors, however it provides a lower resolution than other methods and is also susceptible to bin placement errors. As a technique, binning can be performed

either without prior knowledge, or by incorporating knowledge of the spectrum into the bin boundaries. Binning is accomplished by dividing the frequency spectrum into small spectral regions (bins) and forming a new spectrum using maximum intensity or the area under the curve as the intensity for each bin. This corrects small chemical shifts in the peaks, reduces the noise of the spectrum, and reduces the number of data points. This helps decrease the computation requirements of the algorithm, however it reduces resolution as any information on intensity points within each bin is lost and wrapped up in an aggregate intensity of all the points. Binning has previously been used as a validation technique [131], and in conjunction with curve-fitting [132], direct peak fitting, and peak alignment. Bins must only contain one real peak and not split peaks in order to not lose information or create artefacts. Improper placement of the bins can remove information or produce peaks with large errors. If the bins are too wide, they may cover more than one peak, and thus each peak area will have large errors. Another issue that arises is that if a peak has a large chemical shift, an area of spectral intensity may be split into multiple bins, and thus added to the area of a neighbouring peak volume. This can make statistical analysis less precise [133].

Binning techniques are divided into two categories of algorithm based on how the bin size is determined. Rigid, or uniform binning is a traditional method that involves dividing the spectrum into hundreds of bins of equal size that do not overlap each other. This technique has been applied frequently in literature and has been shown to be a suitable and effective method [134–139]. Generally, the spectral width of each bin is between 0.01 to 0.05 ppm [140], with the most common size being 0.04 ppm, however bins as small as 0.001 ppm have been used [141]. Rigid binning requires careful user guidance in order to ensure that peaks are not overlapping, bin widths are not smaller than the chemical shift, and that the peaks are not too close to the bin boundaries. Different approaches exist to determine the size of the bins. The most common approach is equidistant binning, where the spectrum is divided into equal width bins with size dependant on the peak parameters. A number of packages using equidistant binning are reviewed in the paper by Izquierdo-Garcia [34]. A less common rigid boundary method is Gaussian binning, which still uses a single bin size, however the bin size is determined using a Gaussian weighting algorithm [142].

The second class of binning algorithms is based on flexible boundaries. Flexible boundaries allow for a greater level of control over the size of the bins, allowing for boundaries to be adjusted to include an entire spectral peak. This approach allows algorithms to dynamically determine the size and location of each bin. These methods include adaptive intelligent binning [143] which recursively identifies bin edges in existing bins, dynamic adaptive binning [144] using undecimated wavelet transforms to smooth a composite spectrum and forming bins from this composite spectrum, [145], and an optimised bucketing algorithm [140].

Integration

Integration of the spectral peak is an automated local technique performed in the frequency domain. It is the simplest and is the most common method of determining peak area, however it is prone to errors from peaks that have a low SNR, it can be challenging to determine the integration bounds, and integration is highly sensitive to baseline errors [146]. Integration was the first quantification technique, and can be performed manually although has not been since the development of reliable computer algorithms [147]. Integration can either be performed on a model lineshape function that approximates the experimental data, or directly on the signal [148, 149]. Fitting the experimental data prior to integration of the model function is more accurate [150–153]. A benefit of this technique is that no assumptions have to be made concerning the lineshape, and thus no prior information is required. It is important to disentangle overlapping peaks prior to integrating, as multiple contributions will lead to inaccuracies in area. Baseline signals that have not been addressed, and low SNR will also stymie this method to a greater degree than in binning. Appropriate phasing is also necessary.

Defining the integral bounds is the most important part of this process, as some part of the tail of the peak to be integrated will be neglected, and thus the area under the peak in every case will be underestimated. These inaccuracies will depend on the chosen integral bounds, and can range up to 40% [147]. Because of the impact on the accuracy of the concentration estimation, bounds should be as large as possible, Griffiths recommends 25 times the spectral width for less than 1% error [154]. The

peak must also contain an adequate number of points for an accurate integral. If the experimental data does not provide this, then zero filling the FID can be performed [154, 155].

3.4 Discussion

3.4.1 Local vs Global Methods

The lineshape can be created and optimised either by locally focusing on a spectral region, generally limited to one peak, or globally by modelling the entire spectrum. Global optimisation is generally performed in the time domain, and has been used in approaches by Weber and Metzger [156, 157], however it is disadvantaged by its poor computational efficiency. The downside to global methods are that there is some risk of the solution converging to a local minimum, especially in multidimensional datasets [156]. This requires starting values for each parameter to be closer to the global optimum than is required in well positioned local techniques. Due to the poor computational efficiency of global techniques, most quantitative methods are based on local optimisation [58, 158]. Efforts to improve the efficiency of global methods include incorporating heuristics and knowledge of molecules present in the spectrum [156, 157]. Local techniques are generally applied in the frequency domain, as the peaks and a frequency subsection of the spectrum to analyse can be selected easily. In the literature, these local areas are sometimes termed Regions of Interest (ROIs). Focusing on these areas alone allows for faster computation because only a subset of data is considered, however the package must employ a type of picking algorithm to determine which areas contain peaks and which contain noise. If the spectral regions are not picked satisfactorily they can return local minima. These local techniques include SVD, integration, and binning.

3.4.2 Prior Knowledge

Many packages require the user to manually set initial guess parameters for the location of peaks or type of molecule present. The algorithm then iteratively optimises the lineshape based on these guesses. Alternatively, the package may search

a database in order to find a metabolite basis set that broadly matches the experimental data and then optimise based on those parameters [57, 59, 102, 159]. These prior knowledge based approaches to lineshape quantification can decrease the computational demands of the program as the initial guesses are closer to the solution lineshape. Using prior knowledge without incorporating a metabolite basis set can save computational time, and is ideal for well separated peaks. Unfortunately, if peaks are too close together, or there are unwanted peaks in the same frequency region, and have large amplitudes, this method breaks down [64, 86]. This method can also work in the short-echo time spectrum [160]. These issues, and the large degree of required user interaction make this method less preferable than using a metabolite database since more knowledge about the expected peaks, and any overlapping peaks can be implicitly included in the model. This also makes it easy to identify the nuisance peaks and distortions introduced in the experimental conditions.

Using a metabolite basis set is preferable as it allows for the separation of overlapping peaks as long as the metabolite profiles are not overlapping themselves. This prior knowledge brings greater accuracy [161]. When adding a metabolite basis set, it is important to think about the acquisition parameters, such as the NMR magnetic field, temperature, repetition time, pH, etc. The database metabolite profiles may differ from the experimentally obtained signals because of any of these conditions. The influence of this on the signal quantification, using QUEST, has been analysed in Cudalbu's research [162]. Prior knowledge also allows for decreased computational resources, however requires user interaction for each lineshape. This is time consuming and requires experienced operators to accurately determine the initial parameters. Cavassila estimates the benefit to accuracy that the inclusion of prior knowledge, while cumbersome for the user, allows [163].

Software approaches exist that do not rely on applying prior knowledge to the dataset. These methods are generally termed "black box" and are usually based on statistical analysis: SVD, PCA, or least squares. Black box is an ambiguous term, and in the literature there are methods that claim to be Black box while incorporating limited amounts of prior knowledge [85, 164]. Black box and machine learning methods are generally not as accurate as prior knowledge based methods [165] although

this depends on the level of skill of the individual, or the ability of the database to provide accurate prior knowledge. Black box methods also struggle more with large overlapping peaks as there is no way for the software to intuitively disentangle the overlapping peaks.

3.4.3 Time vs Frequency Domain

The model lineshape can be created and optimised either in the time domain [31] or the frequency domain [32], as they are mathematically equivalent when the Fourier Transform is applied. This is true in theory, however practical considerations mean working in each domain is often performed differently [166]. Time domain model lineshape fitting methods have been studied extensively [59, 102, 159] and in many ways working in the time domain is desirable as there is no need to Fourier Transform the data prior to processing, and all calculations are carried out in the same domain as that in which the signal was obtained. These methods, however, are still not widely adopted and software approaches that use the time domain are mostly used for automated qualitative analysis and characterisation of compounds [167]. These packages are thus generally only paired with lower power bench-top based NMR equipment designed specifically for characterisation over quantification. The data processing step in these approaches is simpler and quicker than for high resolution NMR.

Working in the frequency domain is most commonly used for quantification as well as peak characterisation, as this allows for manipulation and visualisation of peaks, and thus is easier for the user to characterise the compounds. This also allows for the focus of the quantitative analysis (if any is performed) to be on a smaller subset of data. Frequency selective approaches must take into account overlapping or highly distorted peaks, along with baseline and phase errors, and the processing load that accompanies those corrections. Frequency based techniques must also use the Discrete Fourier Transform (DFT) to convert the experimental data into the frequency domain, and thus may in certain cases require more computational resources than time domain approaches, although that is highly situation and package specific, as time domain techniques break down frequencies into sinusoids in a similar way to the FFT, and

thus are not inherently faster in every case.

3.4.4 Automation

Due to the increasing demand for high throughput and the large number of NMR datasets used in 2D and 3D NMR experiments, fully automated packages that require no user feedback have been explored. These approaches include using the Markov chain Monte Carlo (MCMC) algorithm to automate metabolite quantification from NMR spectral data [69], and an approach by Hiltunen et al. to quantify metabolites in real time using an artificial neural network [168]. This method, when compared to established fitting algorithms had a similar level of accuracy, and is a promising step, as it required less user interaction than traditional fitting algorithms. Fully automated software is often more computationally intensive than those that require user interaction, and can be less accurate than manual fitting [169]. Despite this limitation, automated NMR lineshape fitting has clear benefits, and the approaches that currently exist are accurate enough for the majority of purposes.

3.4.5 Stand alone vs Integrated

Most NMR tools (i.e TopSpin, BAYESIL, AQuA, ASICS) [170] are packaged as integrated platforms, designed to perform all the steps in NMR data processing. This allows for close integration between processing steps, requires less work for the user to export and import large datasets to different packages, and reduces unnecessary duplication of steps. A slight disadvantage is that without extensive work (when the source code is available at all) a user is unable to choose the best features of each package to use for their datasets. For example, NMRSpec [171] is an open source integrated software package written in MATLAB that focuses on peak preprocessing and lineshape integration. If users wish to include statistical methods on datasets using this package it would require finding or writing MATLAB compatible code, parsing through the existing code to understand how the data is manipulated by the core NMRSpec package, and then integrating the statistical analysis package into the original package. This is arduous and requires knowledge of programming principles and NMR data processing. A significant downside is that stand alone processes must be adapted to account for the variety of manufacturer specific data formats for FID

or frequency spectra. This lack of consistency greatly limits the feasibility of each step in the NMR data processing pipeline having a separate implementation package, and thus the vast majority of packages are an integrated platform of processing steps.

3.4.6 Identification of Potential Areas for Improvement

As discussed in section 3.4.1 qNMR tools generally use local optimisation techniques due to the poor computational efficiency of global methods. Due to the challenges of isolating individual signals in the time domain, and the ease of doing so in the frequency domain, local techniques are also generally performed in the frequency domain. Further, most techniques require prior knowledge in the form of initial peak parameters or comparison datasets, however it is desirable to have automated techniques due to the complexity and time required for each sample. These general observations can then be applied to specific segments of the pipeline in order to determine current techniques' weaknesses and strengths in meeting those goals.

Reference deconvolution is a technique with the goal of correcting for lineshape distortions due to magnetic resonance inhomogeneity. A reference peak is isolated in the frequency domain and then Fourier Transformed in order to determine the FID of that specific peak. An ideal Lorentzian peak with the sample peak parameters (chemical shift, FWHM, and intensity) is then generated and compared to this FID. Theoretically, the ideal Lorentzian peak multiplied by the error FID introduced due to the magnetic resonance inhomogeneity will produce the experimental FID. Thus, an estimate for this error lineshape is generated and the entire FID comprising each peak has this error FID removed through division. This is an effective technique and corrects for a variety of errors, however there are significant weaknesses in using reference deconvolution for qNMR. The first is that it is a time domain technique, and thus computational resources are expended in transforming the frequency domain spectrum most qNMR software use into the correct domain and back again. Further, this technique determines the lineshape error using a local segment of the frequency domain and applies this correction to every peak. This concoction of local and global techniques relies on the assumption that every error present in the reference peak is also present in every other peak. This assumption is not always valid and a truly local

or global technique is preferable. These deficiencies make reference deconvolution and the correction of lineshape distortions caused by magnetic resonance inhomogeneities a prime point within the pipeline to attempt to improve. A clearly improved algorithm should be entirely in the frequency domain to match other qNMR algorithms, and should either be completely local or global in order to avoid introducing local errors to the entire spectrum.

Chapter 4

Theoretical Elements of Magnetic Resonance Inhomogeneity

4.1 Representation of Peaks

NMR peaks are represented mathematically in the time domain as a decaying sum of sinusoidal frequencies, given by Eq. 4.1, where S_0 is the maximum peak intensity, Ω is the peak resonant frequency (chemical shift, or location on the frequency axis in ppm), and R is the peak decay rate, which is related to the peak Full Width at Half Maximum (FWHM).

$$S(t) = S_0 e^{i\Omega t} e^{-Rt} \tag{4.1}$$

This time domain data can be transformed into the frequency domain using the Fourier Transform to produce a typical NMR spectrum with compounds represented as peaks. The location of specific peaks depends on their atom's unique magnetic resonance and its proximity to other atoms in the mixture analysed. The area of a specific peak corresponds to the relative amount of that atom present in the sample analysed [31, 32]. The frequency domain peak is represented mathematically either as a Lorentzian, Gaussian, or Voigt peak as described by Eqs, 4.2, 4.3, and 4.4 respectively, where $erfc$ is the complementary error function and R_g is the Gaussian portion of the combined decay rate.

$$S(\nu) = S_0 \frac{1 + iz}{1 + z^2} \qquad z = \frac{-2\pi(\nu - \nu_0)}{R} \tag{4.2}$$

$$S(\nu) = S_0 e^{-z^2} erfc(-iz) \qquad z = \frac{-2\pi(\nu - \nu_0)}{R\sqrt{2}} \tag{4.3}$$

$$S(\nu) = S_0 e^{-z^2} erfc(-iz) \qquad z = \frac{-2\pi(\nu - \nu_0 + iR)}{iR_g\sqrt{2}} \tag{4.4}$$

It is possible to determine peak parameters using these mathematical models in both the time and frequency domain and calculate their respective areas; thus determining the concentration of compounds.

NMR peak quantification is traditionally performed in the frequency domain and once the lineshape has been created, can be accomplished by either integrating between two defined points on the frequency lineshape, or by determining an average value of the lineshape across a series of chemical shifts and then summing the areas that correspond to the lineshape peak (binning). Integration can be incredibly accurate, with a measurement uncertainty as low as 0.15% for samples >2.5mg in recent studies [172], although Malz's landmark study indicates that the measurement uncertainty can be much larger at >1.5% [173]. In order to decrease this uncertainty as much as possible, the signal is usually processed to remove a series of common lineshape errors before integration.

Multiple peaks in a NMR spectrum are simply additional sinusoids summed in the time domain, and separate frequency peaks in the frequency domain. These additional peaks are defined by unique Lorentzian/Gaussian/Voigt functions with different peak parameters. If peaks are close enough to one another then in the frequency domain the intensity values of those two peaks are summed together in the experimental lineshape. Thus the area of each peak can be summed individually if each lineshape's peak parameters have been determined. Similarly to single peaks, errors can be present in samples containing multiple peaks. Certain types of errors are spatially limited within the sample and thus only impact specific peaks, however, assuming the NMR sample is well mixed, most errors (for example magnetic resonance inhomogeneity errors) are uniformly applied to every peak.

4.2 Magnetic Resonance Inhomogeneity

Within the NMR spectrometer is a superconducting magnet oriented such that the sample within the NMR bore chamber is held within a constant magnetic field. This is critical for the correct operation of the spectrometer, as the magnetic resonance of the atoms (the Larmor frequency) of each NMR signal is proportional to the

strength of the magnetic field. This relationship is described in Eq. 4.5 where γ is the gyromagnetic ratio of the atom studied, and B_0 is the magnetic field strength (in Tesla)

$$\nu_0 = \frac{-\gamma B_0}{2\pi} \tag{4.5}$$

It is important for this magnetic field to be as uniform as possible. Keeler details the impact of minute changes to the magnetic field for proton NMR. For example, a variation of $2.0 * 10^{-10}$ in the magnetic field of a 500 MHz (11.75 T) spectrometer will cause a 0.1 Hz variation in the Larmor frequency. Commonly, NMR spectral lines are narrow and on the order of 1 Hz, so this is a significant distortion in the lineshape. This degree of homogeneity is impossible with a single magnet, and modern spectrometers include a series of small coils along each spatial axis of the bore chamber to allow for local variations in the magnetic field. Along with these three shim coils (X, Y, and Z) there is also a shim coil that varies quadratically across the Z axis, which is the direction of the magnetic field B (Z2). There are also additional shim coils that correlate with hydrogen atomic orbitals and correct for other specific inhomogeneities in the lineshape [174]. Any inhomogeneity remaining in the magnetic field after the sample is run will cause distortions in the time and frequency domain. These lineshape distortions caused by magnetic resonance inhomogeneity are often termed shim errors.

A graphical example of shim error is shown in Fig. 4.1. On the left is shown the NMR sample tube oriented within the magnetic field B. The field varies spatially across the Z axis of the sample such that a lower magnetic field strength is present at one end of the sample and a higher strength is present at the other end. The change in magnetic field strength spatially across the sample will cause a variance in the resonant frequencies such that, while large groups of molecules resonate at the central frequency $\nu_0 = 8.02$ ppm, these magnetic field imperfections will also cause a distribution of frequencies resonating across the possible distribution of values. This phenomenon is present to a small degree in every sample (hence the Lorentzian distribution commonly observed), however, if the magnetic field fluctuates non-linearly or has significant variance it can cause line-broadening as more and more molecules are resonating at a different frequency to those in the central peak. Lineshape distortions caused by

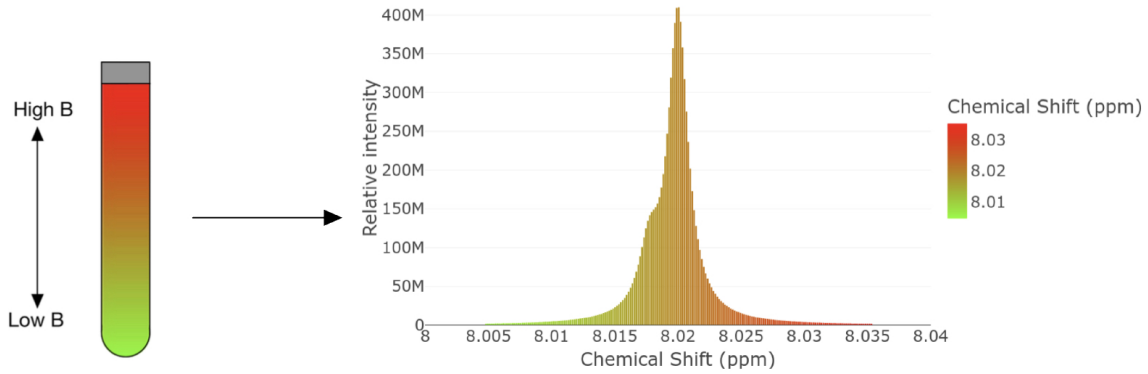

Figure 4.1: Representation of a physical sample suspended in NMR magnetic coil (left) and a demonstration of lineshape distortions produced by magnetic field inhomogeneity and its impact on a molecule's chemical shift (right).

magnetic resonance inhomogeneity vary depending on the degree and type of magnetic field inhomogeneity. For example, there is often distortion of the main peak with the impression of another small peak present at an offset to the central peak, however sometimes the shim error is as subtle as a small degree of line-broadening. Similarly, in the time domain shim error is often distinguishable from the "beats" that are formed as the signal decays, however may also only be distinguished by an FID that decays quickly. Some representative shim errors in the time and frequency domain have been modelled and are shown in Fig. 4.2. Further details of the different types of shim error are given by Chmurny [175].

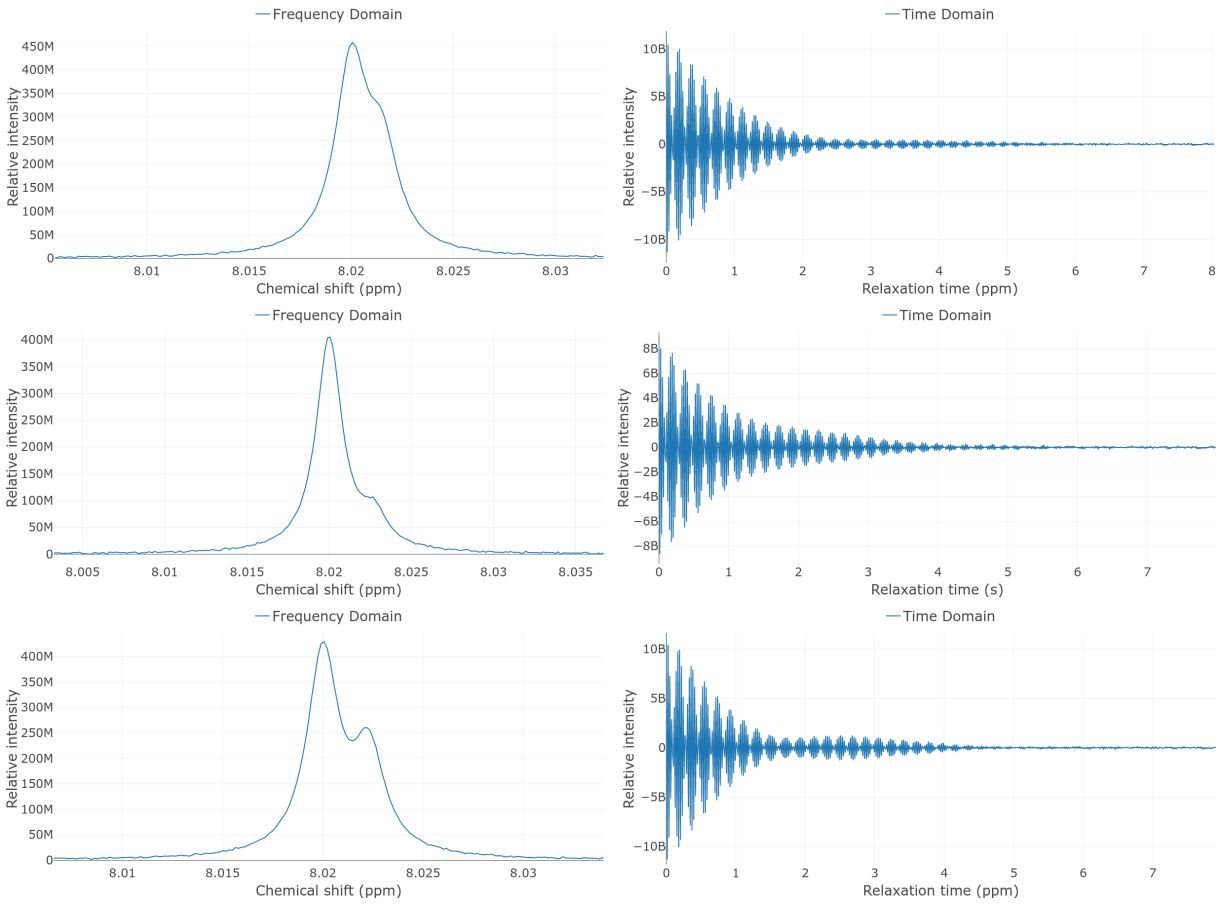

Figure 4.2: Examples of common NMR shim errors in the frequency (left) and time (right) domains.

Chapter 5

Proposed algorithm

5.1 Implementation

The proposed algorithm has been programmed in the R language. A brief overview of the important algorithm processes and parameters is given below, and the full algorithm can be found in Appendix C (singlet) and D (multiplet). The complete algorithm codebase and testing files have been uploaded to Dalhousie's institutional repository (Dalspace) where they are currently publicly available.

5.1.1 Singlet Software Implementation

For a single peak containing lineshape distortions due to magnetic resonance inhomogeneity, the experimental NMR peak is represented as a summation of scaled, shifted central peaks and the intensity of these scaled peaks is optimised to determine the ideal initial parameters and shim factors. This technique is described mathematically by Eq. 5.1 where n represents each peak used to describe the overall lineshape (i.e there is one central peak with height S_0 at position ν_0, the first error peak with height S_1 at position ν_1 and so on).

$$S(\nu) = \sum_{n=0}^{n} S_n \frac{1 + iz_n}{1 + z_n^2} \qquad\qquad z_n = \frac{-2\pi(\nu - \nu_n)}{R} \tag{5.1}$$

When implemented, the error peaks have intensities that are fractions of the central peak (i.e $S_1 = f_1 * S_0$ however this is merely for ease of use and has no impact on the theoretical considerations. A flowchart of the error correction process is shown in Fig. 5.1.

Within the algorithm, a set of user-defined parameters representing the peak location, type of peak (Lorentzian or Gaussian), range of points to be considered, and

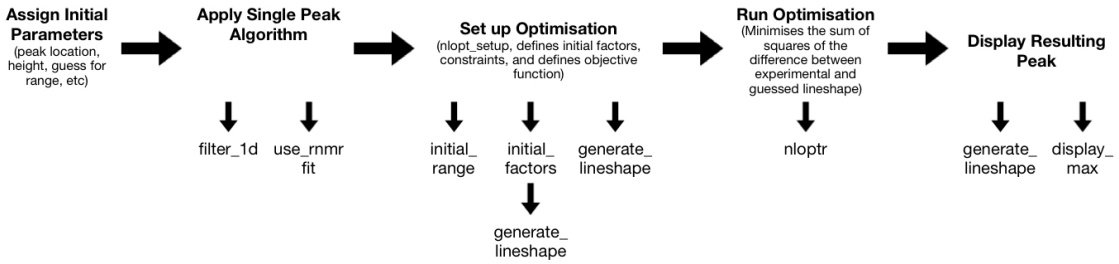

Figure 5.1: Overview of the implementation process for correction of lineshape distortion produced by magnetic resonance inhomogeneity that is present across a single NMR peak.

certain constraints is passed to the algorithm. The algorithm then opens the experimental lineshape and locates the peak using these parameters. A guess is created for initial shim error factors over the range specified by the user by creating a Lorentzian or Gaussian peak guess for the central peak lineshape, then multiplying this guess by an initial shim error factor and placing this peak at the location in the range of frequency values. The algorithm increases the error factors incrementally until the error peaks summed with the guess peaks approximates the experimental lineshape and gives a suitable initial guess for the magnetic resonance inhomogeneity factors. This process is shown graphically in Fig. 5.2 where Shim Peak 1 and Shim Peak 2 represent error peaks that are scaled copies of the central Lorentzian guess peak that have been added at locations offset from the central peak. These peaks when summed equal the experimental lineshape.

After these initial parameters have been calculated, the algorithm runs NLOPT COBYLA (a derivative free non-linear optimisation algorithm) to optimise the shim error factors and minimise the difference between the guess lineshape and experimental lineshape. The area of the central peak can then be summed with the areas of the generated shim error peaks to obtain the total peak area and thus the relative concentration of each compound in the spectrum.

5.1.2 Multiple Peaks Software Implementation

For multiple peaks generated in a single NMR experiment, an initial set of shim correction factors are generated using an initial reference peak that the user specifies

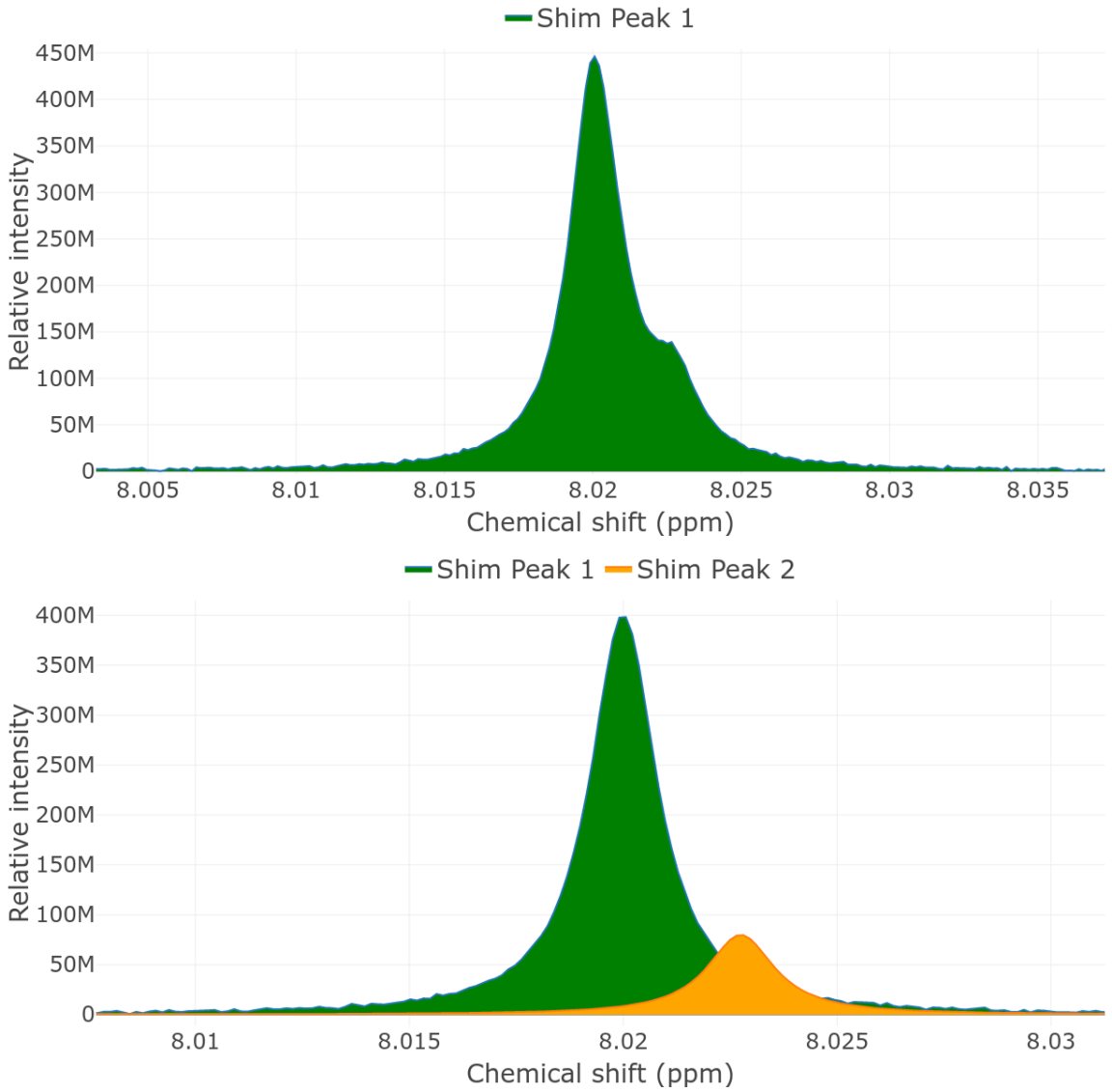

Figure 5.2: Theoretical frequency domain representation of a single peak containing lineshape distortions due to magnetic resonance inhomogeneity as the summation of multiple ideal Lorentzian peak areas with offset chemical shifts.

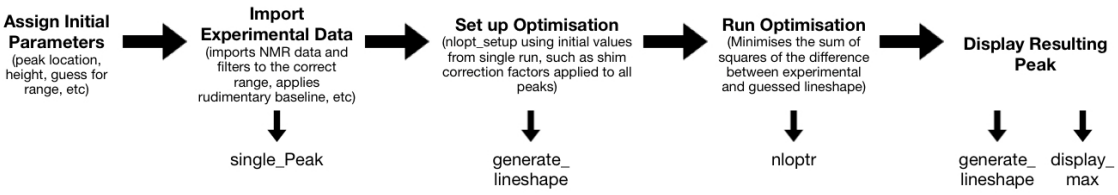

Figure 5.3: Overview of the implementation process for correction of lineshape distortion produced by magnetic resonance inhomogeneity that is present across multiple NMR peaks.

within the spectrum. This peak should be a singlet that is not overlapped by any other peaks. A common set of shim correction factors can then be applied to each peak and optimised across all peaks in the dataset simultaneously. This is accomplished according to the flowchart in Fig. 5.3. First, the user chooses a well-resolved peak and determines the initial parameters such as peak location and initial range to pass to the algorithm. The singlet algorithm detailed previously is then run on this peak and an initial set of magnetic resonance correction parameters is determined for that peak. These correction parameters are then applied to every peak in the spectrum and NLOPT COBYLA is run again to optimise this one set of factors across all peaks.

Initial parameters required by the singlet and multiplet algorithms that should be provided by the user follow. Representative values for many of these parameters as applied to a singlet are described in section **??**. Parameters that are not dependent on the specifics of the peak have default values that generally do not need changing between each spectrum.

A. **Decay rate (R):** This is a guess for the broadness of the central peak and can be estimated using the FWHM. Each peak must be specified separately

B. **Peak location (F_0):** The location of the central peak in ppm. This can be estimated as the location of the peak maximum. Each peak must be specified separately

C. **Peak intensity (A):** The maximum height of the central peak. Each peak must be specified separately

D. **Range of points (upr/lwr limit):** How many points from the central peak the algorithm should consider on each side of the central point

E. **Multiplet scaling factor:** Magnetic resonance inhomogeneity often broadens the lineshape peak- lowering the maximum intensity. This parameter allows NLOPT to accommodate for any height variations imposed by the magnetic resonance inhomogeneity. This parameter is only required if multiple NMR peaks are analysed. A default value of 0.75 was sufficient for each test performed in this work

F. **# of NLOPT evaluations:** The number of times NLOPT should execute in its optimisation stage. Separate values are provided for the reference peak and the multiplet peak evaluation step. Values of 10,000 for each parameter were sufficient for each test performed in this work

G. **Factor constraints:** i.e, if this factor is equal to five, NLOPT cannot guess magnetic resonance inhomogeneity factors that are greater than five times the initial guess. A value of 5 was sufficient for each test performed in this work

H. **NLOPT tolerance:** NLOPT should stop when this relative tolerance parameter is met or exceeded. Separate values are provided for the reference peak and the multiplet evaluation step. A value of $1.0 * 10^{-10}$ was sufficient for each test performed in this work

5.2 Testing Methodology

5.2.1 Simulated Data

As an initial validation step, the algorithm's accuracy on a simulated dataset containing a known degree of frequency error was tested. Using simulated data to validate this algorithm is important for two reasons. First, it is an excellent test of whether the algorithm does what it purports to do; fit lineshape distortions due to simulated magnetic resonance inhomogeneity. This allows for the elimination of any algorithm implementation errors on a known dataset without yet testing that the theoretical

assumptions are valid. Second, in order to compare the area of the algorithm's fit and the experimental lineshape, a numerical integration method has been chosen with the integration limits set as the upper and lower bounds of the peak considered by the algorithm. As detailed in section 3.3.4 this introduces errors in the estimate for the area. Comparing the actual area of each peak with the numerically integrated area will allow for an estimate of the measurement error introduced by using numerical integration. This testing was accomplished by generating a spectrum as three Lorentzian peaks using similar peak parameters to real NMR spectra on a 500 MHz spectrometer. Once the three initial peaks had been generated, a simulated lineshape distortion due to magnetic resonance inhomogeneity was applied to each peak by taking the central peak and multiplying it by three scaling factors. Three scaling factors were chosen in order to show the presence of error on both the left and right side of the peak as well as close to the central peak itself. These error peaks were then added to each pure Lorentzian peak at points offset to the main peak and random error was added to the entire peak according to a normal distribution with a mean of 0, standard deviation of 1, and an intensity multiplier of $1.0 * 10^{-6}$. The resulting combination of peaks is shown in Fig. 5.4 and the scaling factors and error peak location are detailed in Table 5.1. The chosen peaks had FWHM of 1Hz (0.002 ppm). Three runs of this experiment were completed with different levels of simulated error. Each dataset consisted of 3934 points and was simulated based on a 500 MHz spectrometer. In addition to the Lorentzian peaks, a simulated dataset was tested using Gaussian peaks in order to ensure the algorithm could handle each type of lineshape.

Table 5.1: Initial spectral parameters for simulated NMR multiplet data.

	FWHM (ppm)	**Relative Intensity**	**F0 (ppm)**
Peak 1	0.002	4	8.00
Peak 2	0.002	2	7.99
Peak 3	0.002	1	8.01

Once the simulated peaks had been generated, the algorithm was applied and a guess obtained for the shim error factors and the area of each peak. The ideal area of each peak can then be determined by summing the areas of each individual Lorentzian peak. The ideal peak area can then be compared to the area of the lineshape produced

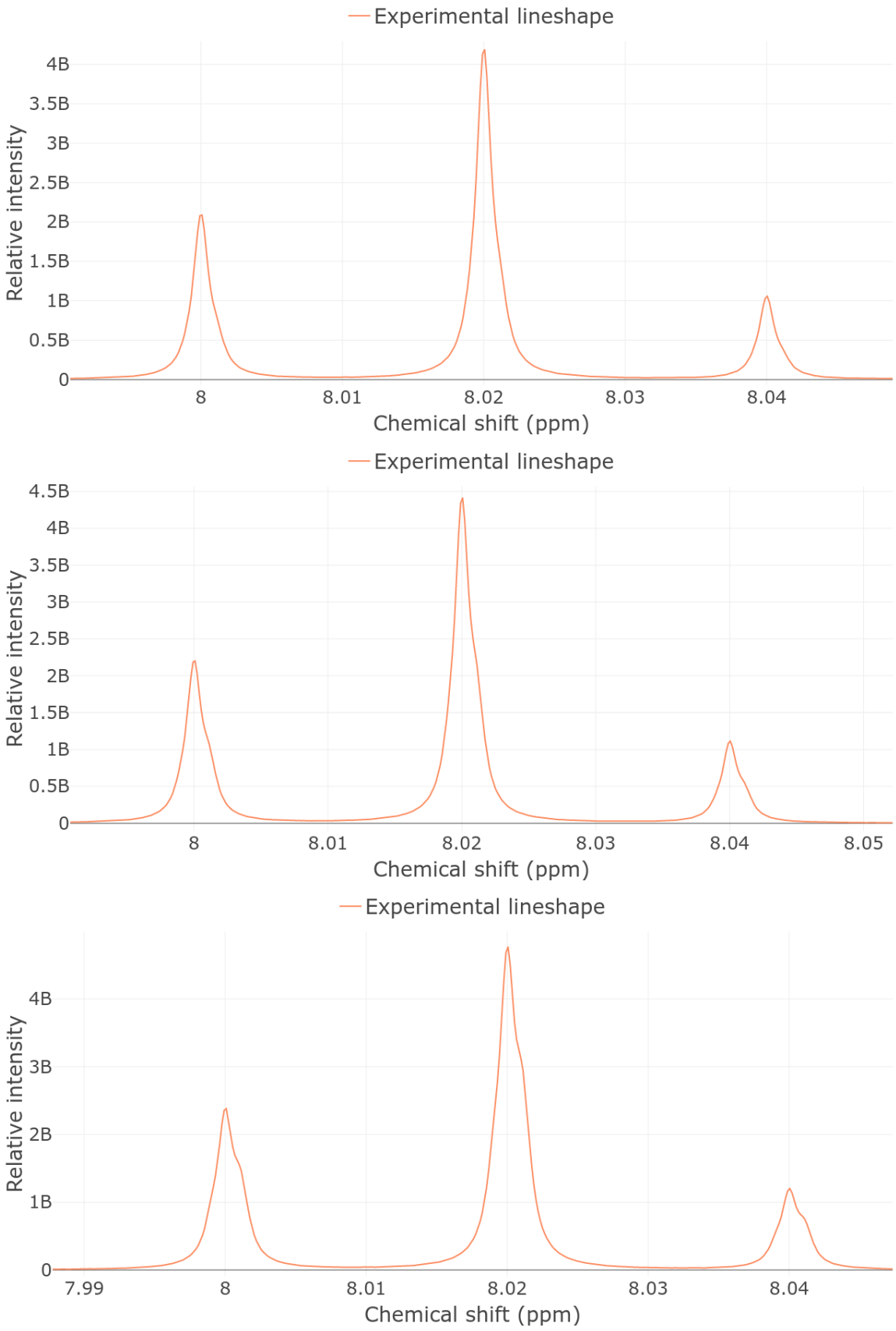

Figure 5.4: Frequency domain spectrum consisting of simulated Lorentzian NMR data with lineshape distortion due to magnetic resonance inhomogeneity added.

Table 5.2: Magnetic resonance inhomogeneity parameters for simulated NMR multiplet data.

No. of pts from F0	**Error factor set 1 (Large)**	**Error factor set 2 (Med.)**	**Error factor set 3 (Small)**
-8	0.3	0.2	0.1
-6	0.2	0.1	0.05
0	1.0	1.0	1.0
6	0.2	0.1	0.05

by the algorithm. The lineshape areas were each calculated as previously described using numerical integration and the limits of integration set as the upper and lower number of points defined in the initial parameters. The percent area as defined in Eq. 5.2 was then calculated where f_{ex} and f_{fit} are locations on the fit and experimental lineshape, and w is the width in ppm between two points on the lineshape. Three sets of peaks were chosen in order to represent small, medium, and large amounts of lineshape distortion due to magnetic resonance inhomogeneity. After comparing the empirically calculated total lineshape area with the experimental numerical area, a comparison of the experimental numerical area with the fitted numerical area has been performed. This test allows for the determination of how much difference in areas is introduced when creating the model lineshape and thus the degree of error the algorithm introduces when fitting the experimental lineshape.

$$\text{Percent Difference Area} = \frac{|A_{ex} - A_{fit}|}{|A_{ex}|} \tag{5.2}$$

$$A_{ex} = \sum (f_{ex})(w) + \frac{f_{ex} - f_{ex-1}}{2} w + ... \tag{5.3}$$

$$A_{fit} = \sum (f_{fit})(w) + \frac{f_{fit} - f_{fit-1}}{2} w + ... \tag{5.4}$$

5.2.2 Experimental Data - Singlet with Inserted Shim Error

Experimental NMR data consisting of 1% chloroform in deuterated acetone (Acetone-d6) has been obtained and varying types of shim error has been manually applied. The algorithm's fit has been compared both qualitatively and quantitatively and the sensitivity of results to changes in the initial parameters has been determined by

applying robustness testing. The purpose of testing this algorithm on a real dataset containing a single large central peak is threefold: First, it is important to validate the algorithm's effectiveness on a real dataset. Second, it is desirable to test the robustness of initial parameters. This requires many runs, and thus it is more efficient to use less peaks and thus less computational resources. Third, there are a class of errors included in complex multi-peak spectra (including overlapping peaks). These are not types of lineshape error that this algorithm purports to solve and thus is not a representative test. Including multiple peaks may skew the accuracy of the robustness results.

Data has been obtained on a Bruker Avance 500 MHz NMR spectrometer and used for algorithm testing. This NMR plot consists of a large central peak at 8.02 ppm and two small C13 satellite peaks on either side of this central peak. These peaks are of no interest in this analysis, and are small and far enough from the central peak that they are not likely to impact the results of this test. Eleven runs were obtained, each with a different shim error deliberately introduced as detailed in Table 5.3. Sample 1 was shimmed by a trained NMR technician, and the rest of the samples were obtained by changing the specific shim value by the magnitude indicated. Only Z1 through Z5 were considered as higher order shims do not need to be optimised before every run. The accuracy of the area calculation between experimental and fit lineshape was determined in order to test the algorithm's ability to create an accurate model lineshape. Algorithm robustness to initial parameters using one of the above samples was also determined by varying the algorithm initial parameters. For the initial test of area accuracy, each of these spectra were given reasonable initial guesses for each peak parameter and the algorithm was run. The initial parameters are shown in Table 5.5.

The plots were each checked visually to ensure results qualitatively fit the experimental data, then the percent difference between the experimental and fit lineshapes was determined according to Eq. 5.2. A series of tests were then conducted on sample 2 to determine the algorithm's robustness to initial guesses. This was accomplished by varying each initial parameter and determining the accuracy of the algorithm in determining the peak area. This was also measured by comparing the percent relative

Table 5.3: Shim error types and values introduced in singlet experimental data.

Sample Number	**Shim Type**	**Magnitude**
1	NA	NA
2	Z	+20
3	Z	-20
4	Z2	+20
5	Z2	-20
6	Z3	+50
7	Z3	-50
8	Z4	+100
9	Z4	-100
10	Z5	+200
11	Z5	-200

Table 5.4: Initial peak parameters for singlet with inserted shim error.

	R	**A**	**F0 (ppm)**	**Multiplet Factor**
Central Peak	0.0006	393067000	8.022349	0.9
Satellite 1	0.0006	2864836	7.804503	0.9
Satellite 2	0.0006	2806424	8.23455	0.9

Table 5.5: Initial algorithm parameters for singlet with inserted shim error.

	Parameter Value
NLOPT	LN COBYLA
Xtolrel	1E-10
Maxeval	10,000
Upper limit	60
Lower limit	60
Fconstraint	5

difference in peak area. Each of the user defined initial parameters outlined in Table 5.6 was tested and the results plotted as a function of the input guesses. Each range was chosen to represent plausible cases of user inputted data. A reasonably observant analyst should be able to obtain more accurate guesses than this range allows, so some margin of error has been included, however completely unreasonable guesses have not been considered.

Table 5.6: Initial robustness testing parameters for singlet with inserted shim error.

	Lower	**Upper**	**Increment**	**Runs**
R	0.0004	0.0154	0.0003	50
F0	8.02	8.03	0.0005	20
Fconstraint	1	21	1	20
Maxeval	100	2600	100	25
upr/lwr limits	5	105	2	50
Intensity (A)	10000001	460000001	10000000	45

5.2.3 Experimental Data - Multiple Compound Spectrum

A larger set of experimental data consisting of a mixture of common metabolomic components was then prepared and NMR spectra obtained. Testing the algorithm on a set of data with multiple peaks is important as a key benefit of this algorithm is that it implements a global consensus fit across all peaks. This consensus fit needs to be tested to ensure the algorithm supports analysis of multiple peaks. The test was performed on two sets of compounds consisting of alanine, glycine, and ethanol dissolved in water, and alanine, glycine, ethanol, and sucrose dissolved in water. A representative NMR plot of Alanine, glycine, and ethanol in water is shown in Fig. 5.5. The two large groupings of peaks at 1.3 and 3.7 ppm represent the ethanol, with the amino acids centred around 3.8 ppm, 3.5 ppm, and 1.5 ppm. Fig. 5.6 shows the sample containing alanine, glycine, ethanol, and sucrose. Sucrose is the large peak located at 4.8 ppm. Reasonable estimates for peak initial parameters were provided to the algorithm and can be found in Appendix A.

There were 30+ discernible peaks in these mixtures with their intensities ranging from less than 100,000 to sucrose's maximum of 500 million. These vastly different scales represent the different concentrations of each element in the mixture. After testing the algorithm on all of the peaks to determine the qualitative accuracy of the fit, the percent area difference was calculated for each peak in order to determine the quantitative accuracy. Each mixture sample was tested to ensure the algorithm could accurately determine the concentrations of these compounds. These specific amino acid/ethanol/sucrose mixture were chosen as complex mixtures are often biological in nature, and thus amino acids with multiplets that are close to one another are

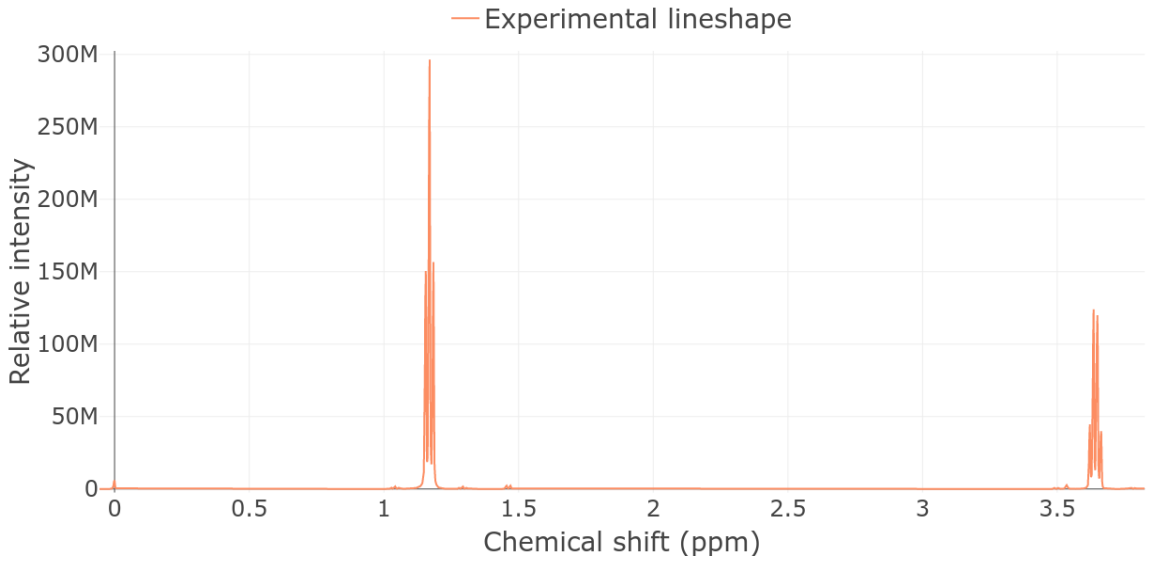

Figure 5.5: Experimental data consisting of the NMR spectrum of alanine (3.8, 1.5 ppm), glycine (3.5 ppm), and ethanol (1.3, 3.7 ppm) in water in the frequency domain.

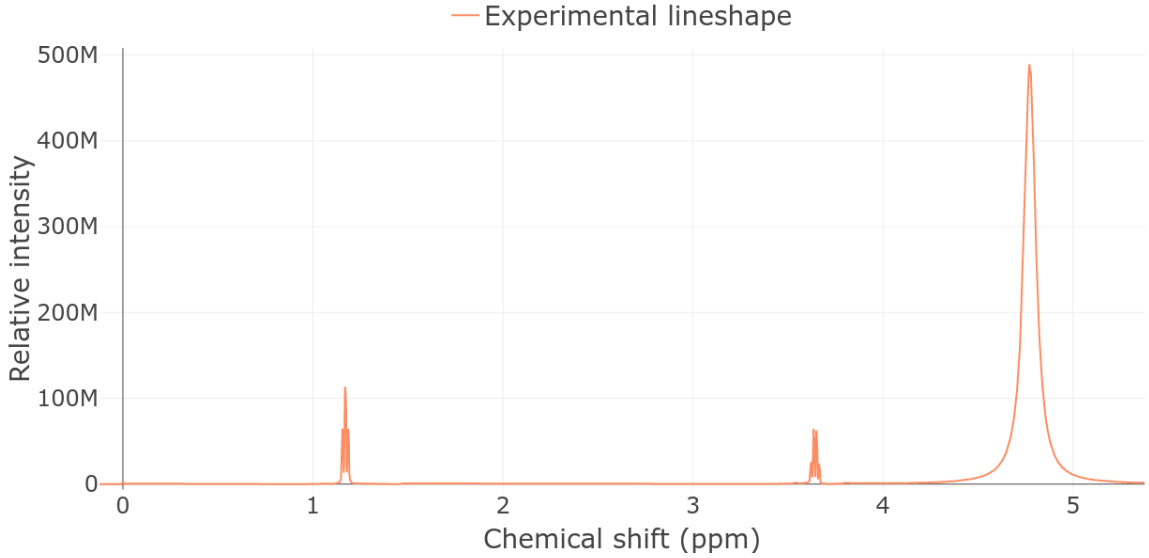

Figure 5.6: Experimental data consisting of the NMR spectrum of alanine (3.8, 1.5 ppm), glycine (3.5 ppm), ethanol (1.3, 3.7 ppm), and sucrose (4.8 ppm) in water in the frequency domain.

excellent test candidates. Further, ethanol and sucrose are common organic molecules found in biological mixtures.

5.2.4 Reference Deconvolution Comparison

As a final test, the accuracy of this algorithm has been compared to the Lorentzian area calculated using reference deconvolution. This is an important test as a key benefit of this algorithm is that it creates a global consensus fit for error across all peaks in the spectrum. Reference deconvolution only applies the error obtained from one peak to all other peaks and thus this may include extraneous error in each peak's fit. Thus, it should be determined whether the global consensus fit improves the accuracy of the areas obtained. The open source reference deconvolution software package MetaboDecon1D [176] was applied to the sample containing alanine, glycine, and ethanol in water from the mixtures in section 5.2.3. Default package parameters were used, and no software pre-processing of the mixture data was performed before applying the package. This package fits spectra as a series of Lorentzian peaks convoluted with a single shim error parameter determined from a reference peak. The resulting peaks were then integrated using the same numerical method as for the mixture data described previously, and the percent area differences compared with the previously obtained mixture values in section 5.2.3. The reference deconvolution package creates a single ideal Lorentzian lineshape that contains the area for each peak, whereas the proposed algorithm creates a sum of a series of peaks. As such, the reference deconvolution lineshape will be slightly broader, and in order to ensure a similar area comparison between peaks, the integration bounds are 30 points on either side of the peak as opposed to 15 used in section 5.2.3. This should not introduce any extra error into the reference deconvolution area calculation as there is no noise in the ideal Lorentzian lineshape produced by MetaboDecon1D and thus increasing the bounds is acceptable.

Chapter 6

Results & Discussion

6.1 Simulated Data

Qualitatively from Fig. 6.1 it is clear that the proposed algorithm gives an excellent fit for each of the simulated lineshape distortions. Further, it is clear that 10 points on either side of the peak, giving a range of 8.0214 to 8.0185 ppm are sufficient to capture the majority of the peak's area. A comparison of the percent error between the numerically integrated area and the actual lineshape area of 1.6-4.5% reinforces this notion and confirms that numerical integration gives an adequate approximation of the empirically determined lineshape area. Three samples are insufficient to determine statistical significance, however it is reasonable to conclude that any difference in algorithm approaches less than the mean of these values (<3%) could be due to errors introduced by using numerical integration rather than the difference between the algorithm's performance itself. This will become important only when comparing the algorithm to reference deconvolution in section 6.5. It should be noted that the measurement error was on the upper end of the range on the largest inserted error peak. This is likely due to the broadening effect of the magnetic resonance inhomogeneity on the lineshape, although 4.5% error is still a reasonable value when compared to the 10%+ error detailed in section 3.3.4, however, it is still clearly higher than the 1.5% error indicated by Malz [173]

Quantitatively, calculation of the percent area differences between the simulated experimental and the fit data yields impressive results that are shown in Table 6.1. These validate the qualitative observations that this algorithm creates an excellent fit for the simulated data and is able to accurately incorporate the simulated error.

It should be noted that when the algorithm was tested using simulated Gaussian peaks, the accuracy of the algorithm appeared to depend on the degree of frequency

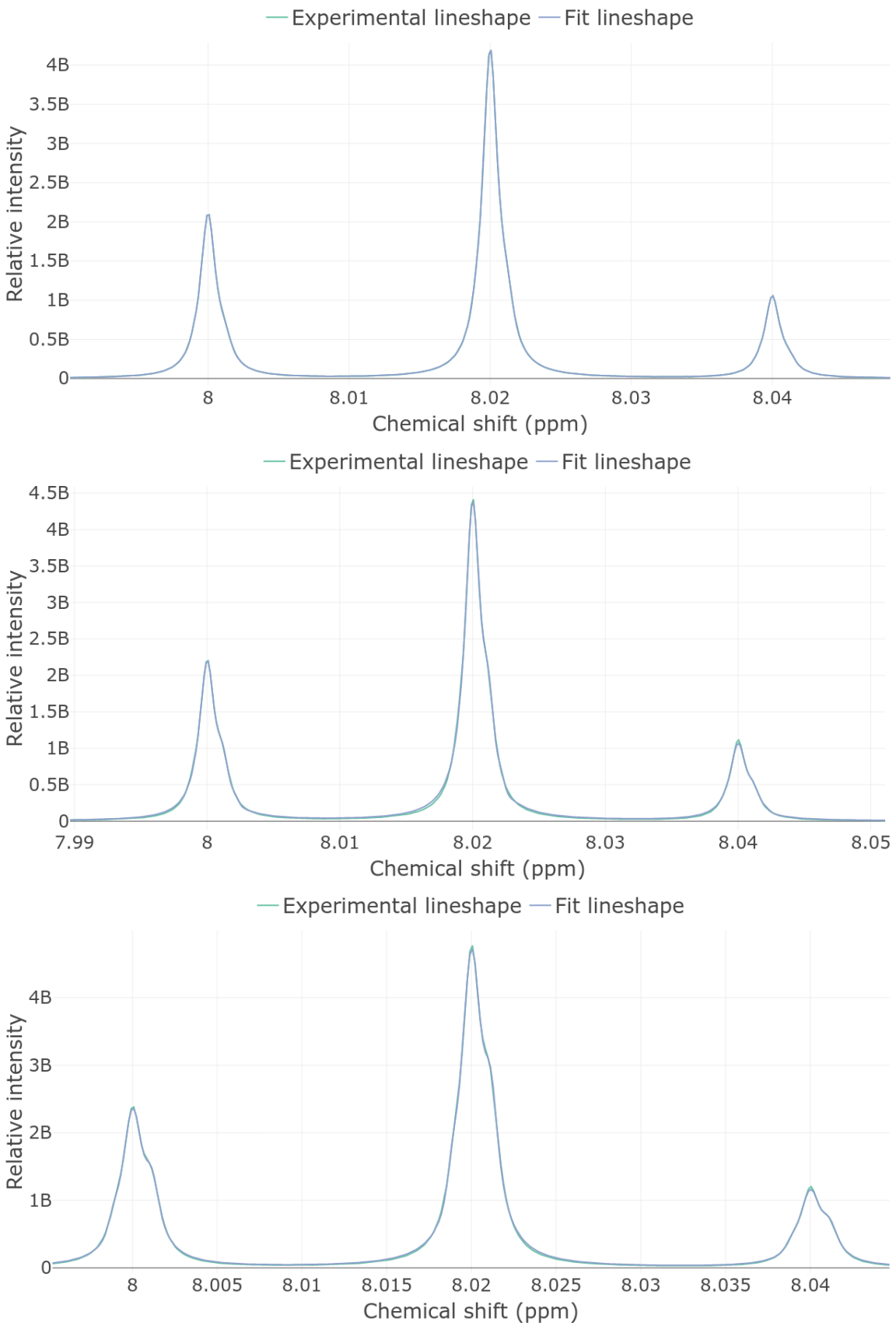

Figure 6.1: Qualitative comparison of a simulated NMR spectrum with introduced lineshape distortions and the results of the algorithm fitting that spectrum.

Table 6.1: Area accuracy results for simulated NMR data comparison.

	% Diff Area
Small	0.097309
Med	0.255327
Large	0.231969

distortion error present. The algorithm had greater accuracy when large extraneous peaks added to the central peak (0.178%, 0.12%, and 0.007% difference area for small, medium, and large error presence) and has approximately 30 times lower accuracy when exposed to the dataset with the lowest degree of error. It is important, however, to keep these values in perspective. 0.178% difference in lineshape areas is still incredibly accurate, despite being 30 times greater than the large error dataset at 0.007% error. This trend of decreasing accuracy with smaller errors was not present in the Lorentzian samples, and is likely due to the broadness of the Gaussian peaks.

It is also important to insert a caveat. This test validates the algorithm only on a dataset that was created using the assumptions inherent in the algorithm itself, and thus the only conclusion that can be drawn from this test is that the algorithm accomplishes its designed purpose. It is able to accurately determine the area of a peak that has been summed with other error peaks. In order to validate the algorithm's inherent assumptions, a set of experimental data with real frequency distortion error is required.

6.2 Experimental Data - Singlet with Inserted Shim Error

The first task when using a real experimental dataset with varying types of shim error is to ensure the algorithm can fit each type of error. The results of this test are shown in Fig. 6.2. It is clear that this algorithm is capable of correcting for Z1 through Z5 error and fitting the experimental lineshape with very accurate results. Quantifying these results, Table 6.2 gives the percent area difference and the lineshape difference for the large central peak in each spectrum. Again, it is clear that the algorithm is able to accurately fit an experimental lineshape with varying degrees of shim error as the percent difference between the model lineshape and experimental

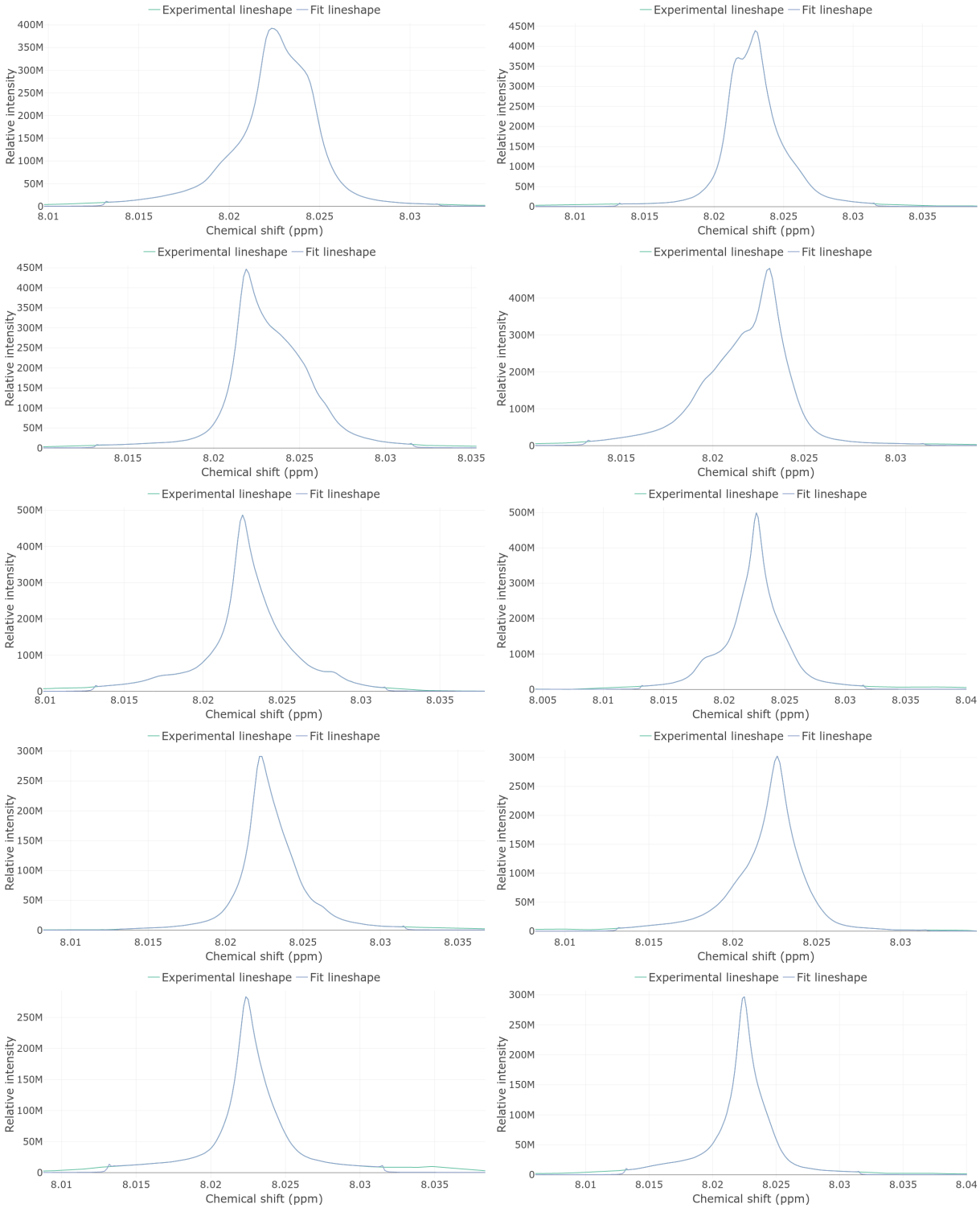

Figure 6.2: Qualitative comparison of NMR singlet spectrum with introduced lineshape distortions and the results of the algorithm fitting that spectrum.

Table 6.2: Area accuracy results for experimental NMR singlet data comparison.

Sample	**% Diff Area**
2	0.044721
3	0.065938
4	0.051199
5	0.100722
6	0.050139
7	0.068681
8	0.059250
9	0.041514
10	0.157828
11	0.097364

lineshape in all cases is below 0.16%. Qualitatively and quantitatively, the accuracy of this algorithm for determining the area of a large central peak in a compound with a large central peak and two small satellite peaks is high. Each of these datasets was the same fundamental compound and so the same initial parameters were used. This indicates that the type of shim error does not materially affect the initial parameters, and that this algorithm is capable of broad application to various types of shim error.

While the central peak area is accurate when compared to the experimental data, the same is not true for the small satellite peaks. These peaks, shown in Fig. 6.3 are of no interest in this analysis, and the large differences in the two lineshapes is almost certainly due to the large baseline distortion and low signal to noise ratio (SNR) present. It should be noted that despite the large errors present when considering these small peaks, this algorithm is still able to determine the presence of lineshape distortions due to lineshape inhomogeneity error and the consensus fit is largely able to capture the inhomogeneity. Peaks of this size with this degree of baseline error/SNR should not be considered in quantitative analysis.

6.3 Robustness Results

It was found that varying the factor constraint input from 1 to 20 had negligible effect on the area accuracy (0.072% to 0.026% area difference). Further, there was no trend in the data and as such the plot of these results have been omitted with

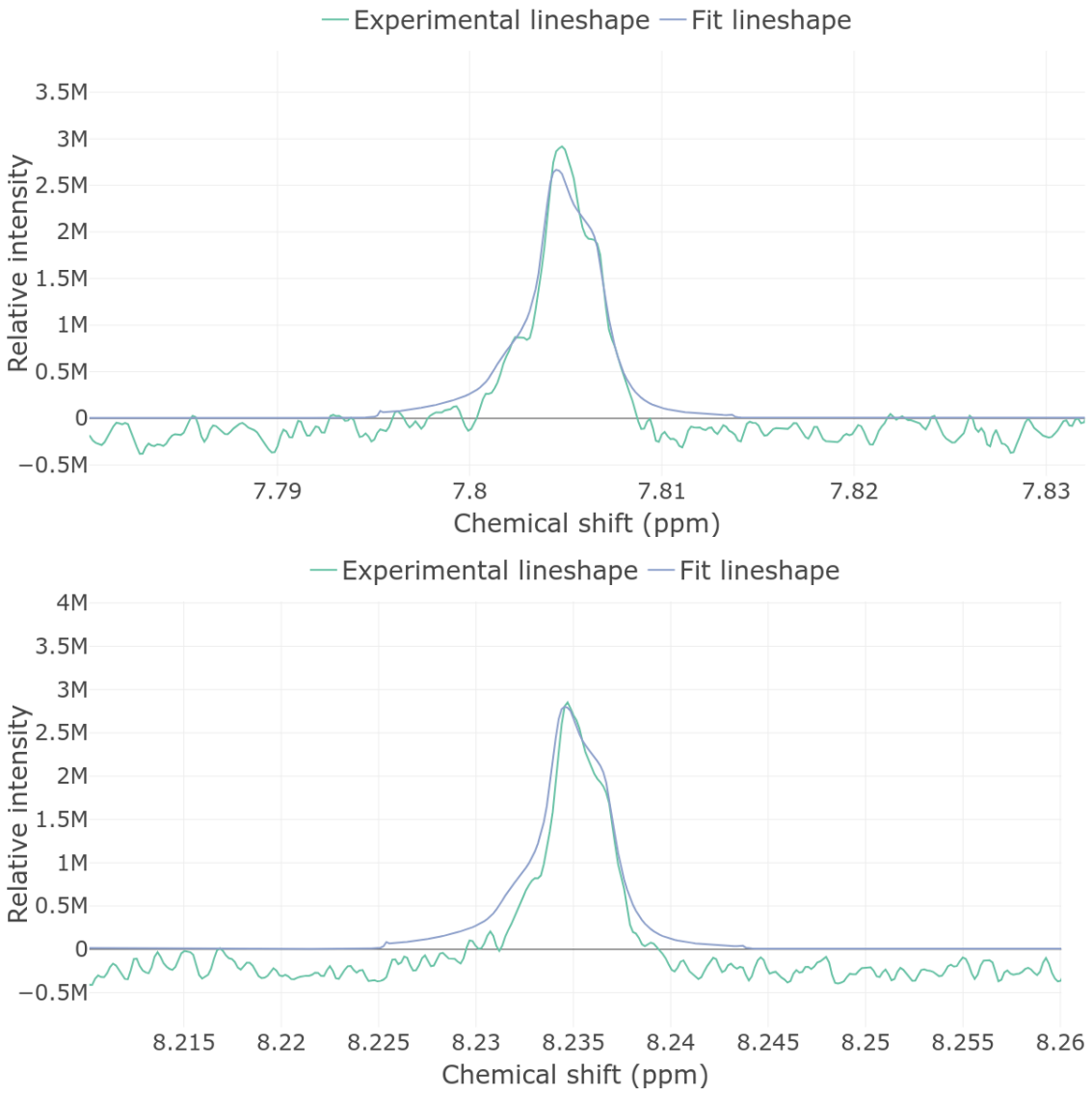

Figure 6.3: Qualitative comparison of NMR sidebands and the results of the algorithm fitting those sidebands.

the sole remark that varying this factor does not appear to impact results. This is intuitive, as this parameter is a simple constraint on error factors going out of bounds. A reasonable value of between 5-10 should suffice for most cases. Greater loosening of this factor may be required if the initial guesses are vastly different from the results, and a smaller factor may decrease the number of runs, or increase the accuracy of the algorithm as there are less values NLOPT will check. Further omitted is the plot of area accuracy versus the number of NLOPT evaluations initial parameter. As is expected, increasing the number of evaluations gives greater accuracy along an exponentially decaying curve. After approximately 10,000 samples, the accuracy does not improve enough to warrant the increased time required.

Fig. 6.4 shows the impact on area accuracy when varying the decay rate/FWHM guess. It is clear that at FWHM values below 0.0015 ppm (0.75 Hz on a 500MHz spectrometer) the accuracy is acceptable, then decreases slightly before a discernible improvement up to 0.0038 ppm (1.88 Hz) and increases again to unsuitable levels. It is likely that below 0.75 Hz, the algorithm is underestimating the central peak width and making up for that by including extra correction peaks. In terms of calculating area this is not a large flaw, however it does serve to increase the number of error peaks, and thus the computation time required to calculate. Further, when implementing consensus fit for multiple peaks, if FWHM is too small (and thus each peak is described by numerous error peaks) then it will become more challenging to develop a consensus fit as the reference peak is likely to have been overfit and had errors that are unique to that peak corrected for rather than solely magnetic resonance inhomogeneity errors. From visual inspection of the spectrum, the ideal FWHM is approximately 1.5-2.5 Hz which correlates well with the dip observed between 1.4 and 1.88 Hz, although there is very little room for error in estimation of FWHM. A more detailed analysis of these results indicate that if the algorithm is given values of FWHM that are too large, upon optimisation a single large Lorentzian peak is formed rather than a series of peaks representing the different regions of the error peak. Thus it is preferable to slightly underestimate this factor as opposed to overestimating it, however a reasonable guess for FWHM can often be determined with sufficient accuracy.

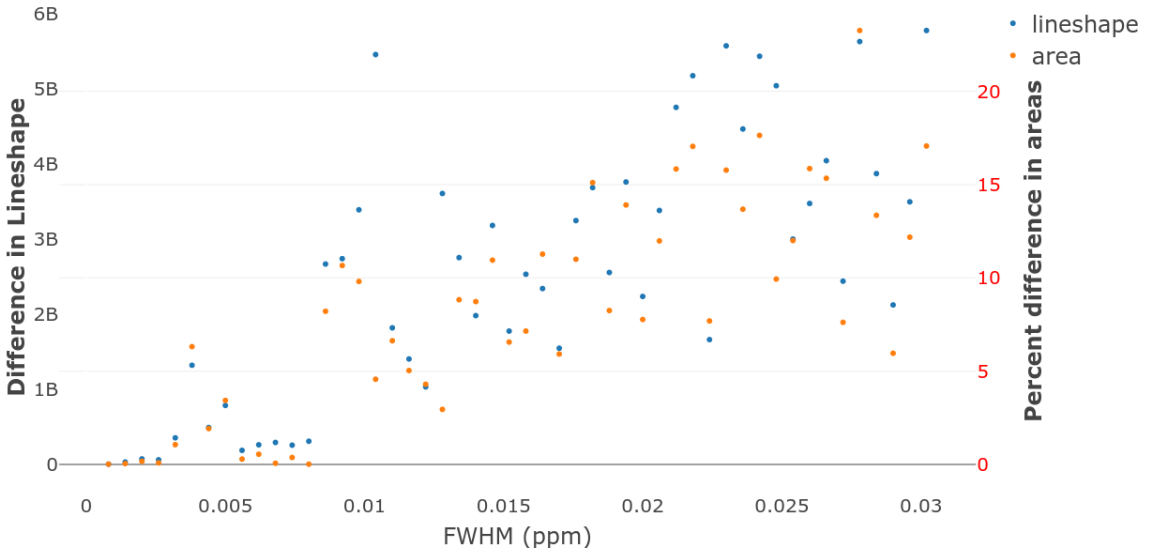

Figure 6.4: Area accuracies and lineshape fit robustness testing results of varying FWHM initial guess parameter.

It is important to note that alongside the values for the percent difference in areas calculated from the algorithm, the difference in lineshapes is also presented. This represents the difference between the sum of each intensity value for the fit and experimental lineshape across the upper and lower integral bounds. This evaluation metric has been included as another tool to assess the degree of difference between the two lineshapes because in certain edge cases when testing, it was found that the percent difference in areas metric failed as a measure of fit. For example, when testing the robustness of the specified range of considered points (Fig. 6.7), the algorithm would optimise noise lineshapes and thus the area difference would be small, however the total lineshape difference reveals increases in the absolute difference in lineshapes, thus making this a less optimal solution overall. In most cases, however, there is very little variation between the trends observed using both measures.

Varying the peak position initial guess gives Fig. 6.5, and shows that the algorithm is sensitive to improper initial guesses for the initial position. Further, higher initial guesses appear to have a greater impact on the accuracy of the algorithm, although that region is overrepresented in this test and may not hold true uniformly. Regardless, it is clear that the algorithm requires a guess between +/- 0.001 ppm of the actual lineshape position. This is a simple parameter to accurately determine

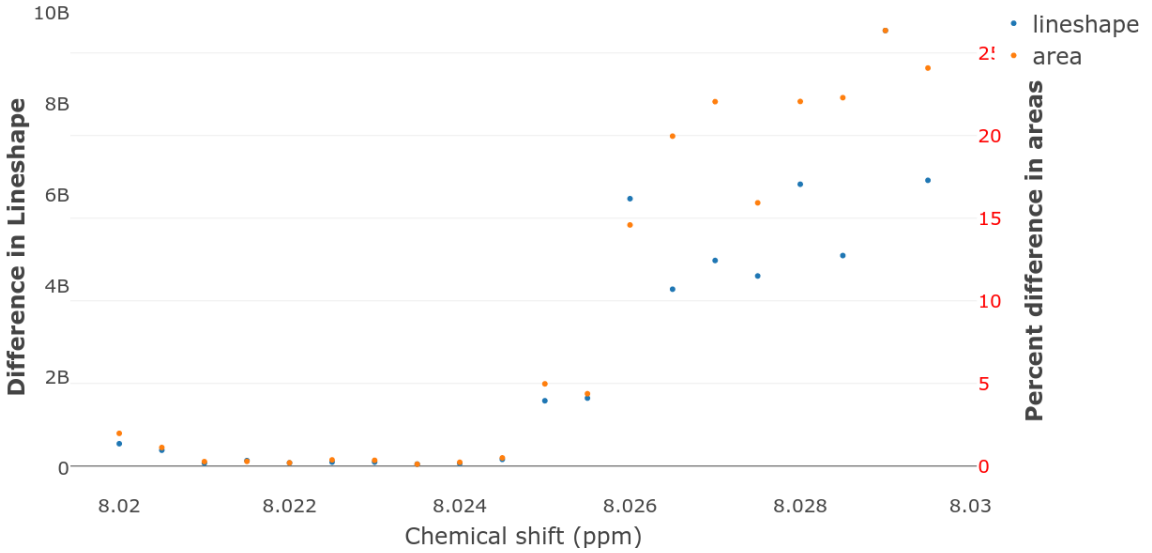

Figure 6.5: Area accuracies and lineshape fit robustness testing results of varying peak position (F0) initial guess parameter.

in many cases and this level of tolerance should be attainable in the vast majority of spectra. Rounding out the peak parameters, varying peak intensity initial guess yielded increasing accuracy the closer the guess was to the actual peak height. As shown in Fig. 6.6 it is clear that any guess between 65% and 130% of the actual peak height gives suitable accuracies and that this parameter is the least sensitive to variations of the three initial peak parameters tested.

Varying the range of points considered from either side of the peak maximum gives Fig. 6.7. It is clear that closer to the central peak, the accuracies are unacceptable. This makes sense as the bounds need to be wide enough to account for the majority of the peaks area. A suitable number of points appears to be 60. This corresponds to 0.009 ppm or 4.5 Hz on either side of the central peak. It is interesting to note that the accuracy decreases when more points are added. This is likely due to the inclusion of more of the baseline and thus more error which is incorrectly fitted as lineshape distortion due to magnetic resonance inhomogeneity.

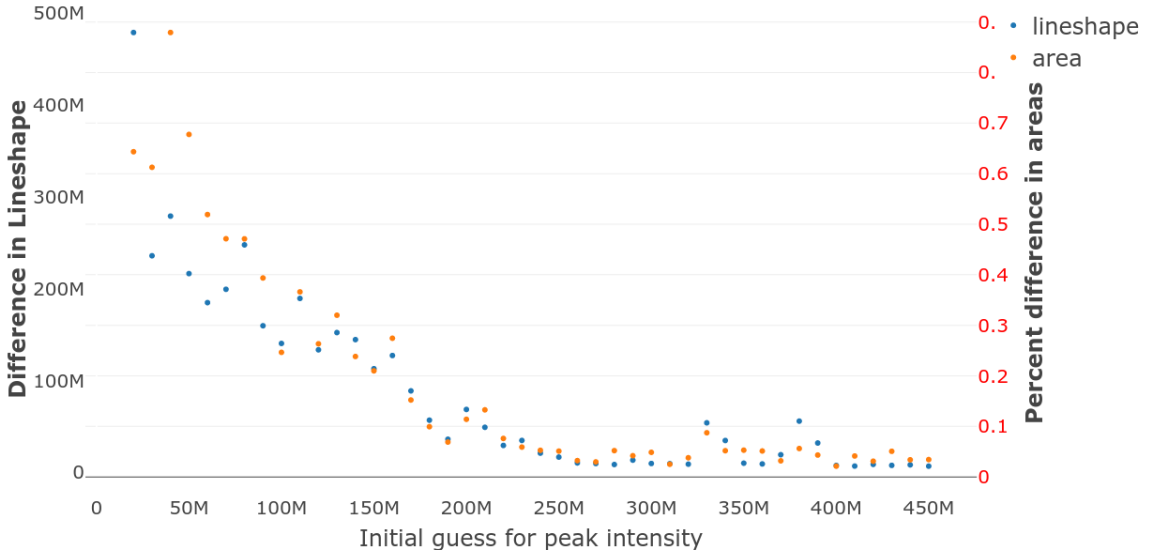

Figure 6.6: Area accuracies and lineshape fit robustness testing results of varying peak height/intensity initial guess parameter.

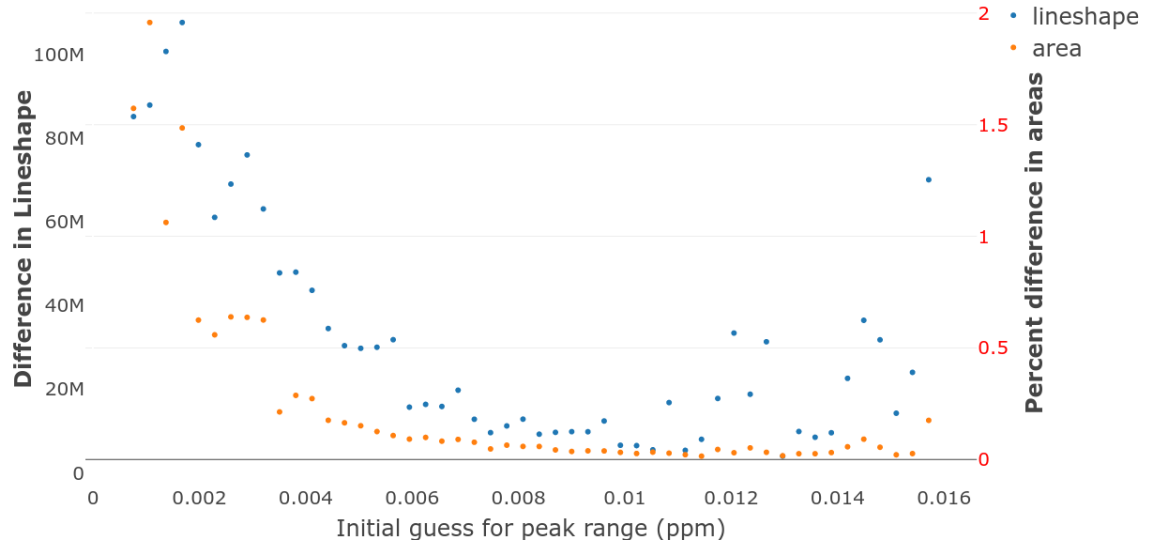

Figure 6.7: Area accuracies and lineshape fit robustness testing results of varying upper/lower bounds initial guess parameter.

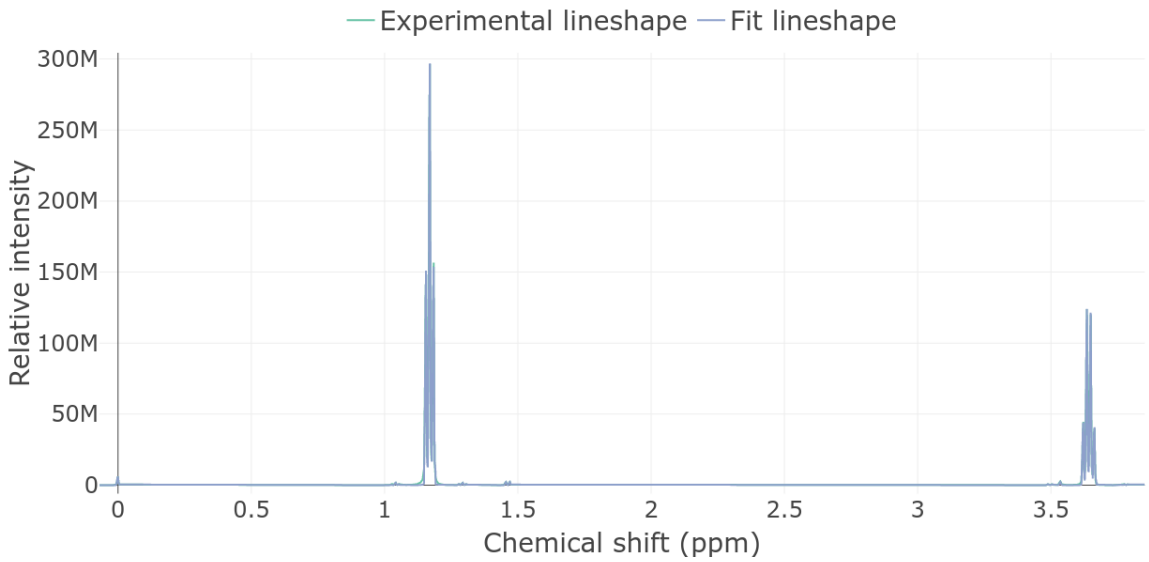

Figure 6.8: Qualitative comparison of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum.

6.4 Experimental Data - Multiple Compound spectrum

The algorithm has been tested on a spectrum containing alanine, glycine, and ethanol in water, and the plot showing the spectrum fitted with the algorithm is given in Fig. 6.8. It is clear from this plot that the algorithm was able to determine the location of the ethanol peaks, however due to the different scales, it is only possible to determine the qualitative fit of the amino acid peaks with a subset of this spectrum shown in Fig. 6.9. Focusing on these smaller peaks validates the algorithms ability to qualitatively fit these smaller peaks.

A subsection of this plot focusing on the ethanol grouping around 1.17 ppm is shown in Fig. 6.10. The ethanol peak has clearly been fit adequately, and the percent area difference of 8.27-8.85% confirms this. The other ethanol grouping centred on 3.64 ppm had percent area differences of 6.14-18.05% as shown in Table 6.3. These errors are large considering the size of the peaks, and the understanding that smaller peaks will naturally have larger percent area differences, however in combination with the qualitative fit these errors are acceptable. Focusing on the smaller peaks reveals greater accuracy (0.02-7.86%) with the notable exception of the glycine peak at 3.545

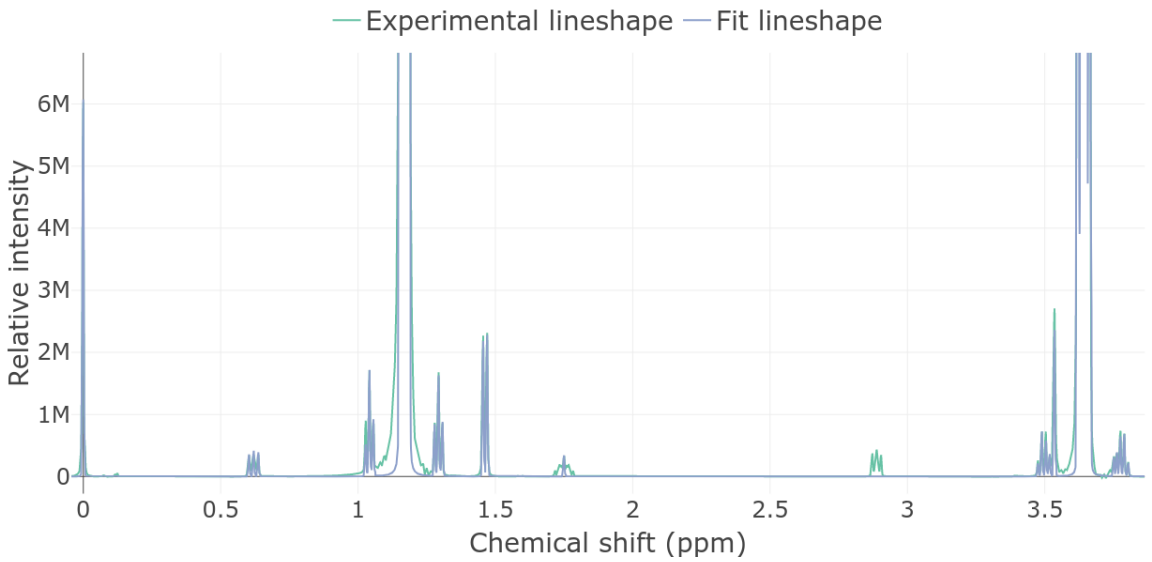

Figure 6.9: Qualitative comparison displaying the lower intensity values of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum.

ppm which had an error of 32.23%. This unusually large error appears to be specific to glycine as Fig. 6.11 clearly shows the shape of the glycine peak is different to all other peaks. It is possible another peak is very close to the glycine peak and thus is obfuscating these results. Note that in this run, all 30 observed peaks were considered, however only the important amino acid/organic molecule peaks are displayed in Table 6.3. For full results see Appendix B.

It is important to stress that in all likelihood, the area errors observed when applying this algorithm are likely due to the presence of other errors such as phase or baseline error. This algorithm does not correct for other kinds of errors in the peak. As an example of this, when applied to the sample containing alanine, glycine, ethanol, and sucrose, the algorithm failed to converge when tested as there was intense phase error present in the sample (Fig. 6.12). This algorithm is designed to target lineshape distortion due to magnetic resonance inhomogeneity and be used as a complementary technique to other lineshape improvement methods, and as such

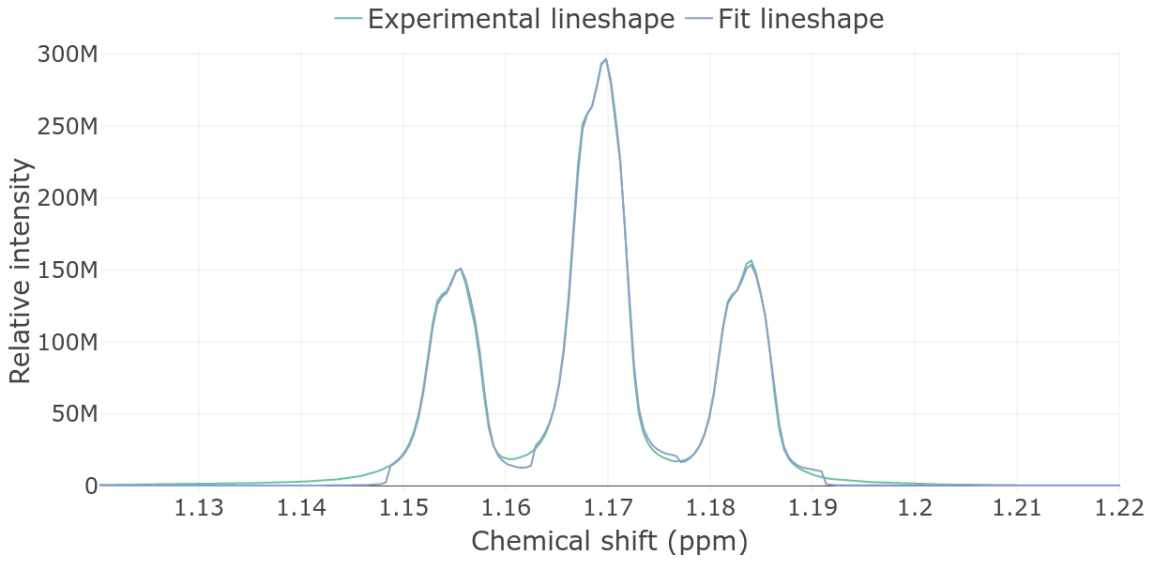

Figure 6.10: Qualitative comparison displaying the ethanol peaks portion of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum.

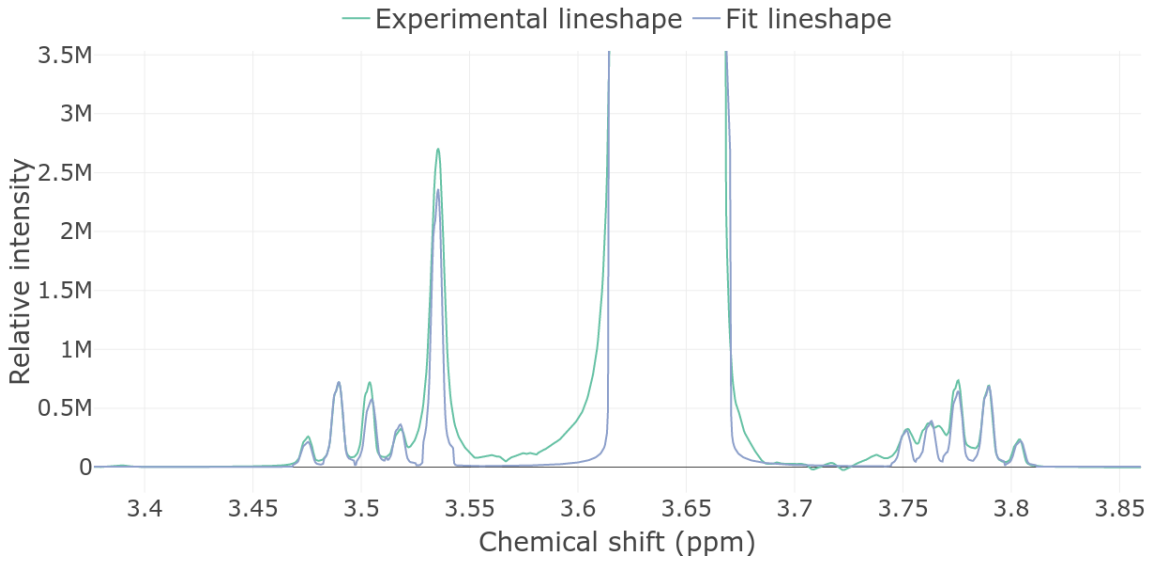

Figure 6.11: Qualitative comparison displaying the glycine peaks portion of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the algorithm fitting that spectrum.

Table 6.3: Truncated area accuracy results for experimental NMR multiplet data comparison.

Peak	Chemical Shift (ppm)	% Diff Area
1 (Alanine)	3.8	7.862609
2 (Alanine)	3.79	0.618946
3 (Alanine)	3.78	2.232451
5 (Alanine)	3.75	1.730346
6 (Glycine)	3.54	32.232686
11 (Ethanol)	3.66	6.278110
12 (Ethanol)	3.62	18.047230
13 (Ethanol)	3.65	6.143892
14 (Ethanol)	3.63	16.101096
16 (Alanine)	1.47	0.579300
17 (Alanine)	1.46	0.024619
24 (Ethanol)	1.18	8.634065
25 (Ethanol)	1.17	8.270839
26 (Ethanol)	1.16	8.850864

is unsuitable when applied as a freestanding algorithm on spectra containing other types of error.

6.5 Reference Deconvolution Testing

A plot of the results of applying MetaboDecon1D to Sample A is shown in Fig. 6.13. Qualitatively, it appears the reference deconvolution package is able to find and fit each peak. A subset of peaks showing the ethanol cluster at 1.17 ppm and the alanine peaks at 3.5 - 3.8 ppm is shown in Fig. 6.14. Table 6.4 details the percent difference in areas obtained using reference deconvolution, and compares them to the area errors obtained using the proposed algorithm.

The differences between area when calculated using reference deconvolution is higher than when calculated using the proposed algorithm for the majority of peaks. Using the proposed algorithm yields improvements from 25.7% for the glycine peak at 3.53 ppm to a low of 0.86% for the ethanol peak at 1.18 ppm (Table 6.4). The majority of these improvements were in the high single digits to low double digits. The notable exceptions to this observed improvement was the 8.64% decrease in accuracy obtained when applied to the ethanol peak at 1.16 ppm and 2.24% decrease on the

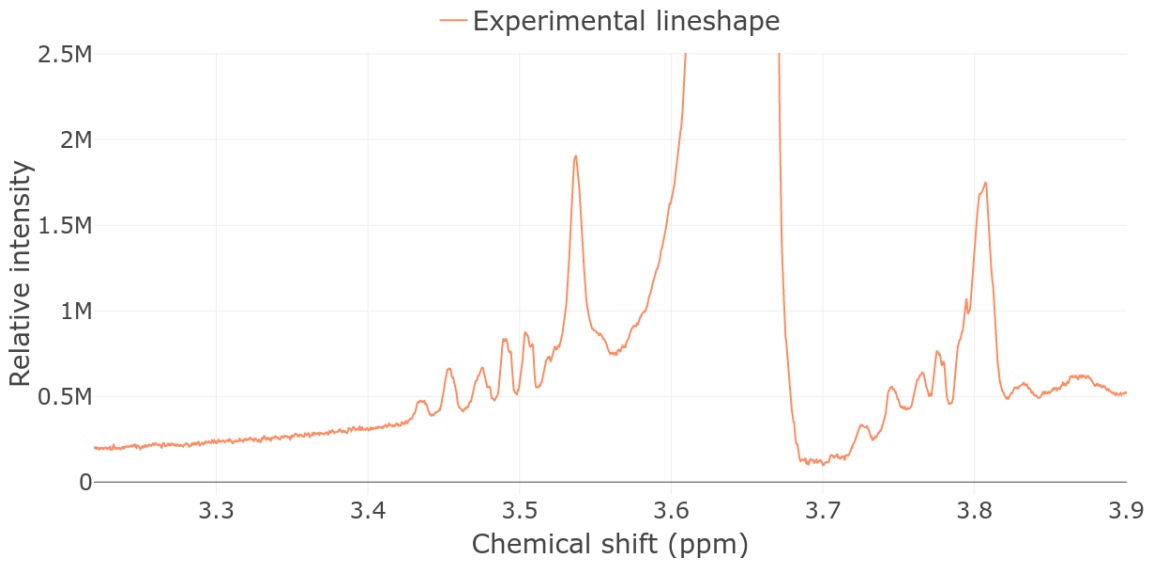

Figure 6.12: NMR lineshape of a spectrum containing alanine, glycine, ethanol, and sucrose demonstrating the presence of phase error in the sample.

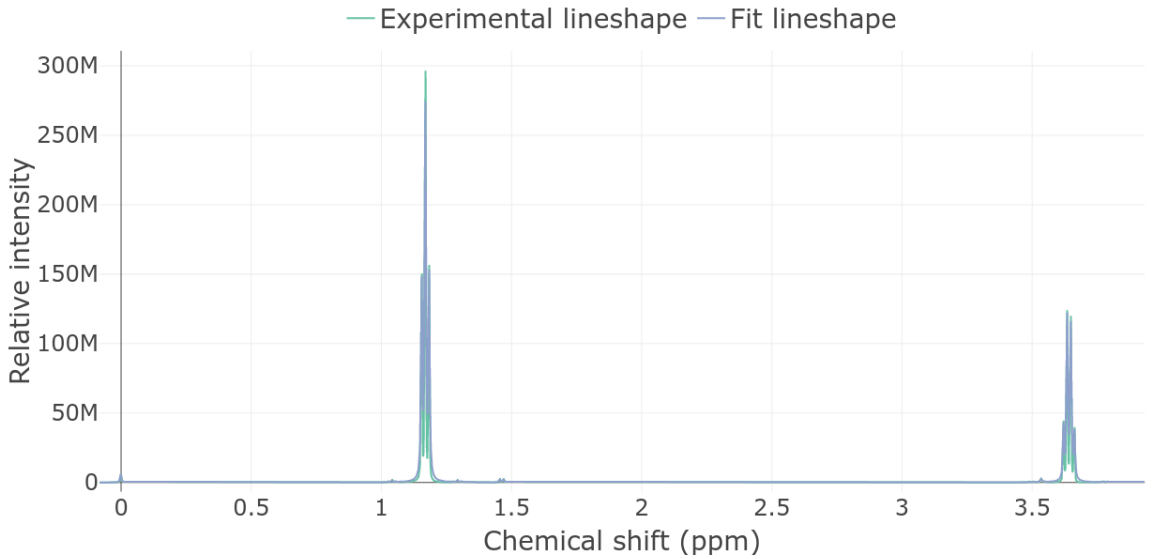

Figure 6.13: Qualitative comparison of an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the reference deconvolution package MetaboDecon1D fitting that spectrum.

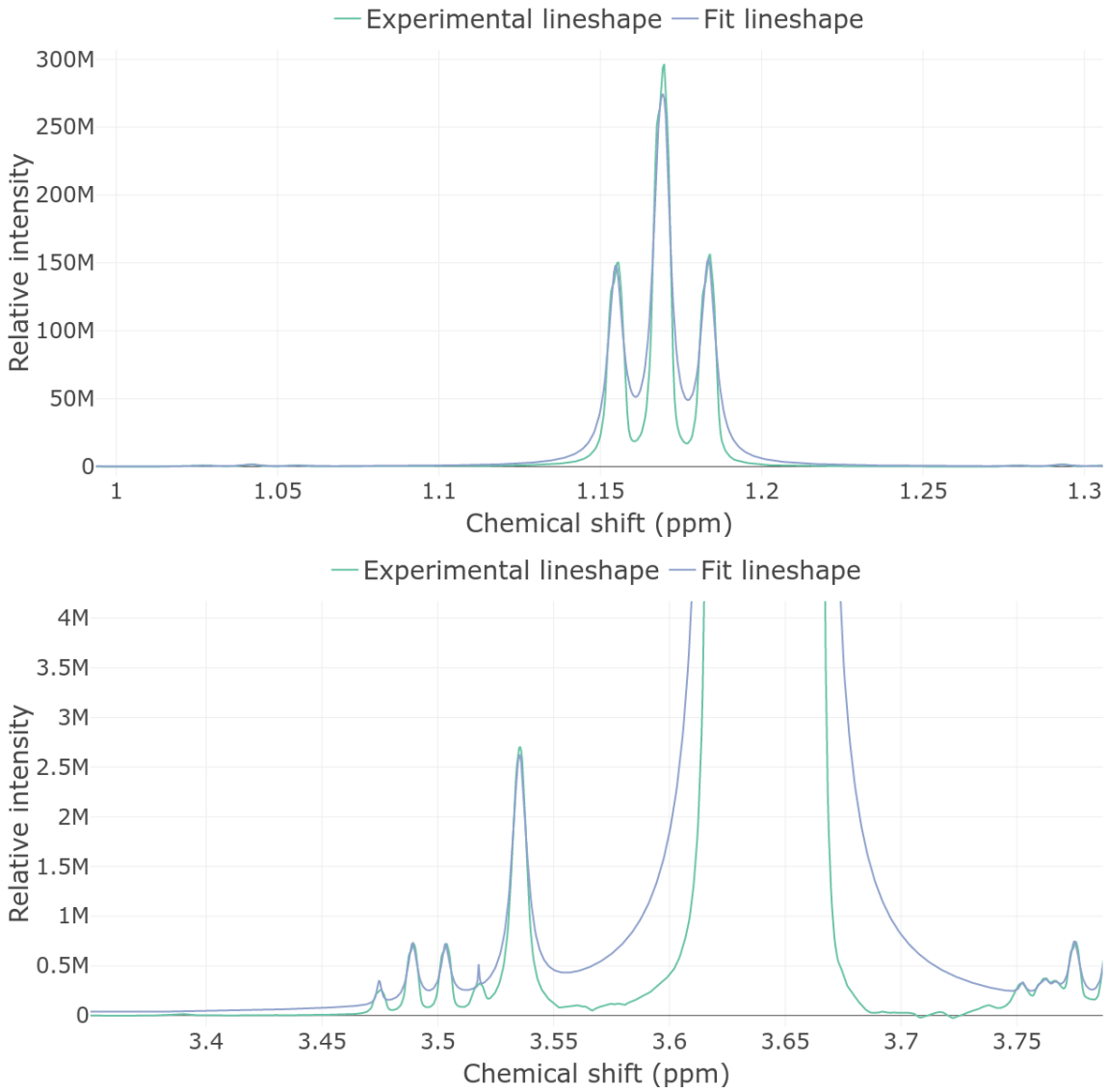

Figure 6.14: Qualitative comparison displaying amino acid peaks from an experimental multiple peak NMR spectrum containing alanine, glycine, and ethanol in water with lineshape distortions due to magnetic resonance inhomogeneity and the results of the reference deconvolution package MetaboDecon1D fitting that spectrum.

Table 6.4: Area accuracy results for reference deconvolution and proposed algorithm on NMR multiplet data and a comparison of the difference in area accuracy results.

Peak	**Ref. Deconvolution % Diff Area**	**Prop. Algorithm % Diff Area**	**Difference**
1 (Alanine @ 3.8 ppm)	31.501443	7.862609	23.638834
2 (Alanine @ 3.79 ppm)	15.932152	0.618946	15.313206
3 (Alanine @ 3.77 ppm)	10.544190	2.232451	8.311739
6 (Alanine @ 3.75 ppm)	5.472757	1.730346	3.742411
7 (Glycine @ 3.55 ppm)	57.931141	32.232686	25.698455
11 (Ethanol @ 3.66 ppm)	15.686288	6.278110	9.408178
12 (Ethanol @ 3.62 ppm)	19.511574	18.047230	1.464344
13 (Ethanol @ 3.65 ppm)	15.035421	6.143892	8.891529
14 (Ethanol @ 3.63 ppm)	13.859835	16.101096	-2.241261
16 (Alanine @ 1.47 ppm)	10.934718	0.579300	10.355418
17 (Alanine @ 1.45 ppm)	13.462582	0.024619	13.437963
25 (Ethanol @ 1.18 ppm)	9.489721	8.634065	0.855636
26 (Ethanol @ 1.17 ppm)	15.830813	8.270839	7.595997
27 (Ethanol @ 1.16 ppm)	0.208868	8.850864	-8.641996

ethanol peak at 3.63 ppm. All ethanol peaks have lower improvement in accuracy than by applying reference deconvolution. This is likely due to the broadness of the peaks impacting both the accuracy of the algorithm, and the increased measurement uncertainty observed when analysing broader peaks. This algorithm shows a difference of greater than +/-3% over reference deconvolution in 11 of 14 samples, which suggests that the differences between the results obtained by each algorithm are due to the algorithm's inherent capabilities and not due to measurement error introduced by using numerical integration. Overall then, these results demonstrate that this algorithm outperforms reference deconvolution for the vast majority of peaks analysed and thus accomplishes its goals.

Another important observation is that this compares the worst possible case for both algorithms. First, no parameters have been optimised and tweaked to ensure either algorithm has excellent initial guesses. Second, only rudimentary phase and baseline correction has been applied to the experimental dataset, and more work could be done to clean up artefacts in the spectrum in order to improve the algorithms effectiveness. The importance of correcting these errors before focusing on

lineshape distortions due to magnetic field inhomogeneity cannot be understated. Simply stated- the presence of phase error, baseline error, and overlapping peaks are not representative of lineshape distortions due to magnetic field inhomogeneity and *will* skew results unfavourably. In general, however, this test demonstrates that for the amino acids studied, and for most of the ethanol peaks, the proposed algorithm demonstrates improved accuracy over reference deconvolution in creating a model fit for the experimental lineshape.

Chapter 7

Conclusion

In this text, a literature review of available NMR spectroscopy data processing techniques has been performed. This review focuses on techniques and categorises the approaches according to a data pipeline of processes applied to different types of data. The review describes techniques in the time and frequency domain, model lineshape based, statistically based, stand alone vs integrated, amongst other parameters. This review highlights the need for extensive data processing of NMR spectra in order to minimise experimental run time (and associated costs) and to correct for types of errors commonly present, including phase error, baseline error, and lineshape distortions produced by magnetic resonance inhomogeneity. Magnetic resonance inhomogeneity is singled out as a potential opportunity for algorithm improvements due to the traditional technique (reference deconvolution) having the undesirable characteristics of being time domain based, and a mixture of local/global methods that can propagate local errors across every peak. In chapter 4 the fundamentals of magnetic resonance inhomogeneity are set forth and important concepts are described. An algorithm is proposed that identifies and corrects for the presence of lineshape distortion due to magnetic resonance inhomogeneity by representing the experimental peak in the frequency domain as a summation of a central Lorentzian peak with multiple error peaks. The error peaks are generated as scaled copies of the central peak and a non-linear optimisation algorithm is applied to minimise the sum of squares between the generated summation of peaks and the experimental peak. This conceptualisation is generalised across all peaks in the spectrum and the algorithm seeks a set of consensus global factors that fit each peak in the spectrum; thus creating a global, frequency-domain based algorithm that overcomes reference deconvolution's primary weaknesses.

This approach is effective in determining lineshape distortion due to magnetic resonance inhomogeneity in simulated and experimental error. When applied to simulated data, the measurement error introduced by using numerical integration was between 1.6 and 4.5%, and when the fit lineshape was compared to the simulated lineshape the percent difference in areas was 0.067 to 0.232%. These are excellent results and qualitative analysis of the peaks confirms that this algorithm fits simulated data well. Likewise, when applied to a singlet with inserted shim error, the algorithm was able to effectively fit shim error from Z1 to Z5. When quantitatively tested on each type of shim error, the percent area difference between the fit and experimental peak ranged from 0.042 to 0.16%. This validates the applicability of the proposed algorithm across a variety of different types of shim error. Further, it was found that when applied to the peak with Z1 shim error, the algorithm is robust to an acceptable range of initial guesses. The least tolerant initial guess was the peak initial chemical shift parameter which required a value of +/- 0.001 ppm from the actual peak lineshape central position. This parameter is, however, simple to accurately guess. It was found that the factor constraint variable had no practical impact on the fit (0.072-0.026% area difference) and that a value of 5-10 is suitable for this parameter in most cases. Further, increasing the maximum number of NLOPT evaluations is only beneficial until 10,000 iterations. For the FWHM initial guess it is important to choose a guess close to the observed FWHM or slightly lower (1.4-1.88 Hz for an actual FWHM of 1.5-2.5 Hz) and that guesses that are too large will detrimentally affect the algorithm. Similarly, the guess for the peak height must be appropriate to within 65 and 130% of the actual peak height. Finally, a choice for the number of points the algorithm considers must take into account as much lineshape area as possible without including too much noisy baseline. This is simple to estimate, and for this compound was 60 points or 4.5 Hz from either side of the central peak. Overall this is a robust algorithm that is able to accommodate a variety of reasonable initial guesses.

When this algorithm was tested on a complex metabolomic spectrum containing alanine, glycine, and ethanol in water it demonstrated an area accuracy of 0.025 to 18.04% with the notable outlier of glycine (32.23%) likely corresponding to a heavily overlapped peak. These values have been qualitatively shown to be accurate and

are satisfactory considering the lack of proper peak pre-processing. When this technique was compared with reference deconvolution it was found that the proposed algorithm generally outperformed reference deconvolution by 0.85% to 25.7% with the notable exception of two ethanol peaks which the algorithm fit less accurately than reference deconvolution by 8.6 and 2.2%. This technique appears to outperform reference deconvolution when applied to smaller peaks, and performs at a similar or slightly decreased level when compared to large ethanol peaks. This may be due to the measurement error introduced by analysing large broad peaks, however the outperformance of the proposed algorithm was greater than the estimate for measurement error in 11 of 14 peaks which suggests measurement error is not a cause of the algorithm's outperformance compared to reference deconvolution. The proposed algorithm failed to converge when tested on a sample containing the previous amino acids, ethanol, and sucrose due to the presence of large phase error.

The technique outlined and applied herein should be incorporated into a software package as part of a pre-processing step for NMR spectra. It accurately and efficiently identifies and corrects for the presence of lineshape distortions produced by magnetic resonance inhomogeneity and can be targeted in the frequency domain. Further testing of this algorithm within a package of pre-processing steps should be performed to ensure that the algorithm performs well with other steps added. The method should also be investigated further to possibly incorporate prior knowledge (for example a guess for the shape of the magnetic field lineshape) that might serve to improve the algorithm. A rigorous robustness study should be performed on complex metabolomic data and the results analysed to ensure this algorithm is suitable when installed in a package of correction tools and applied to complex spectra.

Appendix A

Initial Parameters for Multiple Compound Spectrum

Table A.1: Initial peak parameters for multiplet spectrum with lineshape inhomogeneity.

	R	**A**	**F0 (ppm)**	**Multiplet Factor**
Central Peak	.0005	6025735	-.000214	0.75
2	.0005	219267	3.804445	0.75
3	.0005	691724	3.789785	0.75
4	.0005	736521	3.775584	0.75
5	.0005	373554	3.76184	0.75
6	.0005	323486	3.751762	0.75
7	.0005	2705812	3.535531	0.75
8	.0005	323536	3.518122	0.75
9	.0005	724331	3.50392	0.75
10	.0005	723300	3.489719	0.75
11	.0005	259627	3.475517	0.75
12	.0005	39846370	3.663345	0.75
13	.0005	44599630	3.621198	0.75
14	.0005	120040000	3.649143	0.75
15	.0005	123982000	3.634942	0.75
16	.0005	334634	1.750249	0.75
17	.0005	2315054	1.47034	0.75
18	.0005	2268221	1.45568	0.75
19	.0005	862064	1.307709	0.75
20	.0005	1679096	1.293507	0.75
21	.0005	865020	1.279305	0.75
22	.0005	912225	1.056202	0.75
23	.0005	1712931	1.042001	0.75
24	.0005	901254	1.028257	0.75
25	.0005	156634800	1.184017	0.75
26	.0005	296451300	1.169815	0.75
27	.0005	150543700	1.155614	0.75
28	.0005	351144	0.6205334	0.75
29	.0005	349886	0.6374837	0.75
30	.0005	343131	0.6035831	0.75

Table A.2: Initial algorithm parameters for multiplet spectrum with lineshape inhomogeneity.

	Parameter Value
NLOPT	LN COBYLA
Xtolrel	1E-10
Maxeval	10,000
Upper limit	15
Lower limit	15
Fconstraint	5

Appendix B

Full Results for Multiplet Compound Spectrum

Table B.1: Area accuracy results for experimental NMR multiplet data comparison.

Peak	**Chemical Shift (ppm)**	**% Diff Area**
1 (Alanine)	3.8	7.862609
2 (Alanine)	3.79	0.618946
3 (Alanine)	3.78	2.232451
4	3.76	1.618198
5 (Alanine)	3.75	1.730346
6 (Glycine)	3.54	32.232686
7	3.52	1.766067
8	3.5	20.165006
9	3.49	11.330895
10	3.48	15.706083
11 (Ethanol)	3.66	6.278110
12 (Ethanol)	3.62	18.047230
13 (Ethanol)	3.65	6.143892
14 (Ethanol)	3.63	16.101096
15	1.75	0.070520
16 (Alanine)	1.47	0.579300
17 (Alanine)	1.46	0.024619
18	1.31	37.209537
19	1.3	0.333908
20	1.28	20.654302
21	1.06	6.538963
22	1.04	25.264942
23	1.02	26.308041
24 (Ethanol)	1.18	8.634065
25 (Ethanol)	1.17	8.270839
26 (Ethanol)	1.16	8.850864
27	.62	8.545717
28	.64	74.967863
29	.6	4.427567
30	0	82.466954

Appendix C

Singlet Algorithm - Complete R Package

```
#=============================================================
# MISC Functions & Required Libraries
#=============================================================
# Required Libraries
    library(rnmrfit)
    library(plotly)
    library('nloptr')
    library(RcppFaddeeva)

# generates a lorentzian peak along f using the decay rate
  (R), height (A), and position (F0)

lorentz <- function(R, A, F0, f){
     # This function generates a set of intensity points along
       the frequency axis (f) that correspond to a Lorentzian
       peak of height (A), decay rate (R), and centred at
       peak position (F0)

     z <- -2*pi*(f - F0) / R
     ya <- A * complex(r = 1, i = z)/(1+z^2)

     return(ya)
}

# generates a gaussian peak along f using the decay rate (R),
  height (A), and position (F0)

gauss <- function(R, A, F0, f){
     # This function generates a set of intensity points along
       the frequency axis (f) that correspond to a Gaussian
       peak of height (A), decay rate (R), and centred at peak
       position (F0)
```

```
    z <- -2*pi*(f - F0) / (R * sqrt(2))
    ya <- A * Faddeeva_w(z)

    return(ya)
}

# opens nmr data using rnmrfit

use_rnmrfit <- function(nmrdata1, peaks){
    # generating scaffold and fitting
    scaffold <- nmrscaffold_1d(peaks, nmrdata1)
    fit <- nmrfit_1d(scaffold)

    d <- processed(nmrdata(fit))
    d <- d[order(d$direct.shift), ]
    baseline <- calc_baseline(fit, d$direct.shift)
    all.fits <- calc_lineshape(fit, d$direct.shift, FALSE)

    sum.fits <- aggregate(all.fits$intensity,
                by = list(all.fits$direct.shift), sum)

    colnames(sum.fits) <- c('direct.shift', 'intensity')
    sum.fits$intensity <- sum.fits$intensity + baseline

    sum.fits$direct.shift <- rev(sum.fits$direct.shift)
    sum.fits$intensity <- rev(sum.fits$intensity)

    nmrdata1@processed$intensity <-
        nmrdata1@processed$intensity - baseline1

    return(list('axis'= sum.fits$direct.shift,
                'fit'= sum.fits$intensity,
                'exp'= nmrdata1@processed$intensity))
    print("Read file done")
}
```

```
# generates the shim corrected lineshape given factors and an
  initial peak

generate_lineshape <- function(R, A, F0, f_axis, shim_factors,
     shim_range){

     mg <- Re(lorentz(R, A, F0, f_axis))
     lineshape <- 0

     # scaling the central peak
     for(i in 1:(length(shim_factors))){
         temp_line <- mg * shim_factors[i]

         # shifting all elements in the central peak to the
           right by i points
         if(shim_range[i] == 0){
            temp_line2 <- temp_line

         }else{
            temp_line2 <- c(tail(temp_line, shim_range[i]),
                            head(temp_line, -shim_range[i]))
         }

         # recombining the lineshapes
         lineshape <- lineshape + temp_line2
     }

     return(lineshape)
}

# what points and how far out from the peak maximum should the
  algorithm go to

initial_range <- function(par, upr_limit, lwr_limit){
     print("Beginning initial range")

     lwr_points <- seq(1:lwr_limit)
```

```
    lwr_points <- -rev(lwr_points)
    upr_points <- seq(1:upr_limit)

    # What point on the freqency axis is the range going to
    lwr_stop <- -max(abs(lwr_points))
    upr_stop <- max(upr_points)

    total_range <- c(lwr_points,0, upr_points)

    range <- list("shim_range" = total_range,
                  "lwr_stop" = lwr_stop,
                  "upr_stop" = upr_stop)
    print("Range done")
    return(range)
}

# The initial shim correction factors to feed to nlopt

initial_factors <- function(par, shim_range){
    print("Beginning initial factors")

    # determining the peak location
    peak_location <- which.min(abs(f_axis - par[3]))

    n_factor <- 0
    f_low <- 0
    f_high <- 0
    m_guess <- Re(lorentz(par[1], par[2], par[3], f_axis))
    factor_diff <- 0
    temp_diff <- rep(0, length(shim_range))
    for(i in 1:10){
        if(i == 1){

        }else{
           for(i in 1:length(n_factor)){
               if(n_factor[i] < 0){
               n_factor[i] <- 1 - n_factor[i]
```

```
            }
        }

        m_guess <- generate_lineshape(par[1], par[2],
        par[3], f_axis, n_factor, shim_range)
    }

    d_guess <- Re(f_ex) - m_guess
    comp_points <- shim_range + peak_location
    d_points <- d_guess[comp_points]

    factor_diff <- d_points / max(m_guess)

    for(j in 1:(length(factor_diff))){
        if(j >= 1 && j <= 30){
           factor_diff[j] <- 0.25*d_points[j] /
                             max(m_guess)
        }else if(j > 30 && j < 70){
           factor_diff[j] <- 0.5*d_points[j] /
                             max(m_guess)
        }else if(j > 70){
           factor_diff[j] <- 0.25*d_points[j] /
                             max(m_guess)
        }
    }

    n_factor <- n_factor + factor_diff
}

# making sure no factor is above 1 or below 0
for(i in 1:length(n_factor)){
    if(n_factor[i] > 1){
       n_factor[i] <- .99
    }else if(n_factor[i] < -1){
       n_factor[i] <- -.99
    }
}
```

```
    print("Factors done")
    return(n_factor)
}

# Creates the NLOPT objective function

nlopt_setup <- function(par){

    range <- initial_range(par, upr_limit, lwr_limit)

    # extracting important range parameters
    s_range <<- range$shim_range
    lwr_r_max <- range$lwr_stop
    upr_r_max <- range$upr_stop

    # the factors are then retrieved
    factors <<- initial_factors(par, s_range)

    # setting a constraint that the factors are no more than
      x% of initial guess
    factors_constrained <- ifelse(factors >=0,
                                  f_constraint * factors,
                                  factors/f_constraint)

    for(i in 1:length(factors_constrained)){
        if(factors_constrained[i] > 1){
           factors_constrained[i] <- 1
        }else if(factors_constrained[i] < -1){
           factors_constrained[i] <- -1
        }
    }

    # The initial guess
    par0 <<- c(par, factors)
    lbnew <<- c(lb, rep(-1,length(factors)))
    ubnew <<- c(ub, factors_constrained)
```

```
     # the objective function for NLOPT
     function( parnew ){

          shim_guess <- generate_lineshape(parnew[1],
          parnew[2], parnew[3], f_axis, parnew[4:(3 +
          length(factors))], s_range)

     return(list("objective" = sum( (Re(f_ex)
                 Re(shim_guess))^2),
                 "gradient" = rep(0, length(parnew))))
     }
}

#=============================================================
# Importing Experimental Data
#=============================================================

start_time <- Sys.time()

# retrieving the data and creating the fit
data1 <- use_rnmrfit(nmrdata1, peaks)

# the x axis, rnmr's fit of the experimental data, and the
  experimental data itself

f_axis <<- data1$axis
f_ex <<- data1$exp
f_rnmr_fit <- data1$fit

#=============================================================
# NLOPT fitting
#=============================================================

# creating the evaluation function and setting up initial
  range/factors
```

```
eval_f <- nlopt_setup(par)

opts <- list( "algorithm" = algorithm,
              "xtol_rel" = xtol_rel,
              "maxeval" = maxeval)

pre_opt_time <- Sys.time()
res <- nloptr(  x0=par0,
                eval_f=eval_f,
                lb=lbnew,
                ub=ubnew,
                opts=opts)
opt_time <- Sys.time()

pre_opt_diff <- difftime(pre_opt_time, start_time,
                         units = "secs")
opt_diff <- difftime(opt_time, pre_opt_time, units = "secs")

#================================================================
# Displaying Results
#================================================================

# displaying the final fit
f_fit <<- generate_lineshape(res$solution[1], res$solution[2],
                             res$solution[3], f_axis,
                             res$solution[4
                             (3+length(factors))], s_range)

# the initial guess
#f_fit <<- generate_lineshape(par0[1], par0[2], par0[3], f_axis,
   par0[4:(3+length(factors))], s_range)


# displaying the experimental data, the fit, (Toggle TRUE when
  qualitative results are desired)
if(FALSE){
     p <- plot_ly(x = f_axis,
```

```
                    y = Re(f_ex),
                    color = 'green',
                    name = 'f_ex',
                    type = 'scatter',
                    mode = 'lines')%>%
                    layout(showlegend = TRUE,
                           legend = list(orientation = "h",
                           xanchor = "center",
                           x = 0.5, y = 7))
     p <- p %>% add_trace(x = f_axis,
                          y = Re(f_fit),
                          color = 'red',
                          name = 'f_fit',
                          type = 'scatter',
                          mode = 'lines')
    p
}
```

Appendix D

Multiplet Algorithm - Complete R Package

```
# Required Libraries
     library(rnmrfit)
     library(plotly)
     library('nloptr')
     library(RcppFaddeeva)

#=============================================================
# Applying shim correction to reference peak
#=============================================================

source("FILEPATH", encoding = "UTF-8", echo = TRUE)

factors_final <- res$solution[4:(3+length(factors))]

peak_par_multiple <- c(peak_par_multiple, factors_final)
lb_multiple <- c(lb_multiple, rep(-1, length(factors)))
ub_multiple <- c(ub_multiple, rep(2, length(factors)))
#=============================================================
# Applying correction to other peaks
#=============================================================

# the objective function for NLOPT
eval_g <- function( peak_par_m ){
     shim_guess_m <- 0
     j <- 0

     for(i in 1:(length_par_multiple)){
         shim_guess_m <-
           shim_guess_m + generate_lineshape(peak_par_m[1+j],
           peak_par_m[3*length_par_multiple + i] *
                 peak_par_m[2+j],
           peak_par_m[3+j], f_axis,
```

```
            peak_par_m[(3*length_par_multiple +
                length_par_multiple + 1):(3*length_par_multiple
                +length_par_multiple + length(factors_final))],
            s_range)

        j <- j + 3
    }


    return(list( "objective" = sum( (Re(f_ex)
                Re(shim_guess_m))^2),
                "gradient" = rep(0,length(peak_par_multiple))
                ))
}

opts <- list( "algorithm" = m_algorithm,
              "xtol_rel" = m_xtol_rel,
              "maxeval" = m_maxeval)

res <- nloptr(  x0=peak_par_multiple,
                eval_f=eval_g,
                lb=lb_multiple,
                ub=ub_multiple,
                opts=opts)

#=============================================================
# Displaying Results
#=============================================================

# displaying the experimental data, the fit,

final_lineshape <- 0
j <- 0

for(i in 1:length_par_multiple){
    final_lineshape <- final_lineshape +
    generate_lineshape(res$solution[1+j],
```

```
                        res$solution[3*length_par_multiple + i]
                        * res$solution[2+j], res$solution[3+j],
                        f_axis,
                        res$solution[(3*length_par_multiple +
                        length_par_multiple + 1)
                        (3*length_par_multiple +
                        length_par_multiple +
                        length(factors_final))],
                        s_range)
    j <- j + 3
}

# Toggle if qualitative results are desired
if(FALSE){
     p <- plot_ly(x = f_axis,
                  y = Re(f_ex),
                  color = 'green',
                  name = 'Experimental lineshape',
                  type = 'scatter',
                  mode = 'lines')%>%
                  layout(showlegend = TRUE,
                         xaxis = list(title = 'Chemical shift
                                      (ppm)'),
                         yaxis = list(title = 'Relative
                                      intensity'),
                         legend = list(orientation = "h",
                         xanchor = "center",
                         x = 0.5, y = 7))
     p <- p %>% add_trace(x = f_axis,
                          y = Re(final_lineshape),
                          color = 'red',
                          name = 'Fit lineshape',
                          type = 'scatter',
                          mode = 'lines')
     p
}
```

Bibliography

[1] E. Skordi, I. D. Wilson, J. C. Lindon, and J. K. Nicholson, "Characterization and quantification of metabolites of racemic ketoprofen excreted in urine following oral administration to man by ¡sup¿1¡/sup¿ h-nmr spectroscopy-directly coupled hplc-ms and hplc-nmrand circular dichroism," *Xenobiotica*, vol. 34, pp. 1075–1089, 11-12 Jan. 2004, ISSN: 0049-8254. DOI: 10.1080/00498250412331281098.

[2] S. Ok, "Characterization and quantification of microstructures of a fluorinated terpolymer by both homonuclear and heteronuclear two-dimensional nmr spectroscopy," *Magnetic Resonance in Chemistry*, vol. 53, pp. 130–134, 2 Feb. 2015, ISSN: 07491581. DOI: 10.1002/mrc.4156.

[3] N. E. Jacobsen, *NMR spectroscopy explained : simplified theoryapplications and examples for organic chemistry and structural biology*. HobokenN.J. : Wiley, 2007.

[4] E. Breitmaier, *Carbon-13 NMR spectroscopy : high-resolution methods and applications in organic chemistry and biochemistry*, 3rd comple, W. Voelter and E. Breitmaier, Eds. New York : VCH Publishers, 1987.

[5] L. Mureddu and G. W. Vuister, "Simple high-resolution nmr spectroscopy as a tool in molecular biology," *FEBS Journal*, vol. 286, pp. 2035–2042, 11 2019, ISSN: 17424658. DOI: 10.1111/febs.14771.

[6] J. L. Miller, *Solid-state nmr resolves protein structures - no deuteration required*, Oct. 2016. DOI: 10.1063/PT.3.3318.

[7] D. L. Rabenstein and A. A. Isab, "Determination of the intracellular ph of intact erythrocytes by 1h nmr spectroscopy," *Analytical Biochemistry*, vol. 121, pp. 423–432, 2 1982, ISSN: 0003-2697. DOI: https://doi.org/10.1016/0003-2697(82)90502-4. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0003269782905024.

[8] G. Hanna and C. Lau-Cam, "A stability-indicating proton nuclear magnetic resonance spectroscopic method for the analysis of propantheline bromide in pharmaceutical samples," *Die Pharmazie*, vol. 56, pp. 700–703, Oct. 2001.

[9] T. W.-M. Fan, R. M. Higashi, A. N. Lane, and O. Jardetzky, "Combined use of 1h-nmr and gc-ms for metabolite monitoring and in vivo 1h-nmr assignments," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 882, pp. 154–167, 2 1986, ISSN: 0304-4165. DOI: https://doi.org/10.1016/0304-4165(86)90150-9. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0304416586901509.

[10] W. T. Evanochko, T. T. Sakai, T. C. Ng, N. Krishna, H. D. Kim, R. B. Zeidler, V. K. Ghanta, R. Brockman, L. M. Schiffer, P. Braunschweiger, and J. D. Glickson, "Nmr study of in vivo rif-1 tumors: Analysis of perchloric acid extracts and identification of 1h31p and 13c resonances," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 805, pp. 104–116, 1 1984, ISSN: 0167-4889. DOI: https://doi.org/10.1016/0167-4889(84)90042-9. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167488984900429.

[11] J. R. Bales, D. P. Higham, I. Howe, J. K. Nicholson, and P. J. Sadler, "Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine.," *Clinical chemistry*, vol. 30, pp. 426–432, 3 1984, ISSN: 00099147. DOI: 10.1093/clinchem/30.3.426.

[12] J. K. Nicholson, M. J. Buckingham, and P. J. Sadler, "High resolution 1h n.m.r. studies of vertebrate blood and plasma," *Biochemical Journal*, vol. 211, pp. 605–615, 3 1983, ISSN: 02646021. DOI: 10.1042/bj2110605.

[13] T. W. Fan, P. K. Lorkiewicz, K. Sellers, H. N. Moseley, R. M. Higashi, and A. N. Lane, "Stable isotope-resolved metabolomics and applications for drug development," *Pharmacology and Therapeutics*, vol. 133, pp. 366–391, 3 2012, ISSN: 01637258. DOI: 10.1016/j.pharmthera.2011.12.007. [Online]. Available: http://dx.doi.org/10.1016/j.pharmthera.2011.12.007.

[14] T. W. Fan and A. N. Lane, "Applications of nmr spectroscopy to systems biochemistry," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 92-93, pp. 18–53, 2016, ISSN: 00796565. DOI: 10.1016/j.pnmrs.2016.01.005. [Online]. Available: http://dx.doi.org/10.1016/j.pnmrs.2016.01.005.

[15] A. Moreno and C. Arús, "Quantitative and qualitative characterization of 1h nmr spectra of colon tumorsnormal mucosa and their perchloric acid extracts: Decreased levels of myo-inositol in tumours can be detected in intact biopsies," *NMR in Biomedicine*, vol. 9, pp. 33–45, 1 1996, ISSN: 09523480. DOI: 10.1002/(SICI)1099-1492(199602)9:1<33::AID-NBM391>3.0.CO;2-G.

[16] E. Servens, D. Löwik, A. Bosman, L. Nelissen, and P. J. Lemstra, "Synthesis and characterization of poly[(2,6-dimethyl-1,4-phenylene oxide)-block-isoprene] diblock copolymers," *Macromolecular Chemistry and Physics*, vol. 198, pp. 379–389, 2 1997, ISSN: 10221352. DOI: 10.1002/macp.1997.021980213.

[17] K. Jankoa and J. Kops, "1h-nmr investigation of quantitative functionalization of poly(ethylene glycol)s," *Journal of Applied Polymer Science*, vol. 54, pp. 1027–1032, 8 Nov. 1994, ISSN: 0021-8995. DOI: https://doi.org/10.1002/app.1994.070540804. [Online]. Available: https://doi.org/10.1002/app.1994.070540804.

[18] E. Luchinat and L. Banci, "A unique tool for cellular structural biology: In-cell nmr," *Journal of Biological Chemistry*, vol. 291, pp. 3776–3784, 8 Feb. 2016, ISSN: 00219258. DOI: 10.1074/jbc.R115.643247.

[19] A. Leitner, G. Dorn, and F. H.-T. Allain, "Combining mass spectrometry (ms) and nuclear magnetic resonance (nmr) spectroscopy for integrative structural biology of protein–rna complexes," *Cold Spring Harbor Perspectives in Biology*, vol. 11, a032359, 7 Jul. 2019, ISSN: 1943-0264. DOI: 10.1101/cshperspect.a032359.

[20] U. Holzgrabe, I. Wawer, and B. Diehl, *NMR Spectroscopy in Pharmaceutical Analysis*. Elsevier, 2008, ISBN: 9780444531735. DOI: 10.1016/B978-0-444-53173-5.X0001-7.

[21] S. Balayssac, E. Retailleau, G. Bertrand, M.-P. Escot, R. Martino, M. Malet-Martino, and V. Gilard, "Characterization of heroin samples by 1h nmr and 2d dosy 1h nmr.," *Forensic science international*, vol. 234, pp. 29–38, Jan. 2014, ISSN: 1872-6283. DOI: 10.1016/j.forsciint.2013.10.025.

[22] R. W. Morrison and M. Zhang, "Applications of benchtop nmr in the organic chemistry instructional laboratory," *Magnetic Resonance in Chemistry*, Jun. 2020, ISSN: 07491581. DOI: 10.1002/mrc.5031.

[23] J. C. Chatham, B. Bouchard, and C. D. Rosiers, "A comparison between nmr and gcms 13c-isotopomer analysis in cardiac metabolism," *Molecular and Cellular Biochemistry*, vol. 249, pp. 105–112, 1/2 2003, ISSN: 03008177. DOI: 10.1023/A:1024786622400.

[24] N. Bryden, M. Antonacci, M. Kelley, and R. T. Branca, "An open-sourcelow-cost nmr spectrometer operating in the mt field regime.," *Journal of magnetic resonance (San DiegoCalif. : 1997)*, vol. 332, p. 107 076, Nov. 2021, ISSN: 1096-0856. DOI: 10.1016/j.jmr.2021.107076.

[25] Y. Wang, Q. Wang, J. Liu, J. Cheng, and F. Liu, "Insert magnet and shim coils design for a 27t nuclear magnetic resonance spectrometer with hybrid high and low temperature superconductors," *Superconductor Science & Technology*, vol. 33, p. 64 004, 6 2020.

[26] Z. Ni, L. Li, G. Hu, C. Wen, X. Hu, F. Liu, and Q. Wang, "Design of superconducting shim coils for a 400 mhz nmr using nonlinear optimization algorithm," *IEEE Transactions on Applied Superconductivity*, vol. 22, pp. 4 900 505–4 900 505, 3 Jun. 2012, ISSN: 1051-8223. DOI: 10.1109/TASC.2011.2176699.

[27] S. Reutter and A. Privalov, "Compensation of magnetic field instabilities in field cycling nmr by reference deconvolution," *Applied Magnetic Resonance*, vol. 44, pp. 55–63, 1-2 Feb. 2013, ISSN: 0937-9347. DOI: 10.1007/s00723-012-0396-8.

[28] G. Hajjar, N. Merchak, C. Daniel, T. Rizk, S. Akoka, and J. Bejjani, "Improved lipid mixtures profiling by 1h nmr using reference lineshape adjustment and deconvolution techniques," *Talanta*, vol. 208, p. 120 475, Feb. 2020, ISSN: 00399140. DOI: 10.1016/j.talanta.2019.120475.

[29] P. Ebrahimi, M. Nilsson, G. A. Morris, H. M. Jensen, and S. B. Engelsen, "Cleaning up nmr spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra," *Journal of Chemometrics*, vol. 28, pp. 656–662, 8 Aug. 2014, ISSN: 08869383. DOI: 10.1002/cem.2607.

[30] G. Maniara, K. Rajamoorthi, S. Rajan, and G. W. Stockton, "Method performance and validation for quantitative analysis by 1h and 31p nmr spectroscopy. applications to analytical standards and agricultural chemicals," *Analytical Chemistry*, vol. 70, pp. 4921–4928, 23 1998, ISSN: 00032700. DOI: 10.1021/ac980573i.

[31] L. Vanhamme, T. Sundin, P. V. Hecke, and S. V. Huffel, "Mr spectroscopy quantitation: A review of time-domain methods," *NMR in Biomedicine*, vol. 14, pp. 233–246, 4 2001, ISSN: 09523480. DOI: 10.1002/nbm.695.

[32] Š. Mierisová and M. Ala-Korpela, "Mr spectroscopy quantitation: A review of frequency domain methods," *NMR in Biomedicine*, vol. 14, pp. 247–259, 4 2001, ISSN: 09523480. DOI: 10.1002/nbm.697.

[33] J. B. Poullet, D. M. Sima, and S. V. Huffel, "Mrs signal quantitation: A review of time- and frequency-domain methods," *Journal of Magnetic Resonance*, vol. 195, pp. 134–144, 2 Dec. 2008, ISSN: 10907807. DOI: 10.1016/j.jmr.2008.09.005.

[34] J. L. Izquierdo-García, P. Villa, A. Kyriazis, L. D. Puerto-Nevado, S. Pérez-Rial, I. Rodriguez, N. Hernandez, and J. Ruiz-Cabello, "Descriptive review of current nmr-based metabolomic data analysis packages," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 59, pp. 263–270, 3 2011, ISSN: 00796565. DOI: 10.1016/j.pnmrs.2011.02.001.

[35] R. Spicer, R. M. Salek, P. Moreno, D. Cañueto, and C. Steinbeck, "Navigating freely-available software tools for metabolomics analysis," *Metabolomics*, vol. 13, pp. 1–16, 9 2017, ISSN: 15733890. DOI: 10.1007/s11306-017-1242-7.

[36] H. Barkhuijsen, R. D. Beer, W. M. M. J. Bovee, J. H. N. Creyghton, and D. V. Ormondt, "Application of linear prediction and singular value decomposition (lpsvd) to determine nmr frequencies and intensities from the fid," *Magnetic Resonance in Medicine*, vol. 2, pp. 86–89, 1 Feb. 1985, ISSN: 0740-3194. DOI: https://doi.org/10.1002/mrm.1910020111. [Online]. Available: https://doi.org/10.1002/mrm.1910020111.

[37] T. L. James, ; S. Macura, B. T. Farmer, and L. R. Brown, "[12] modern spectrum analysis in nuclear magnetic resonance: Alternatives to the fourier transform," 1984, p. 493.

[38] S. G. Hyberts, G. J. Heffron, N. G. Tarragona, K. Solanky, K. A. Edmonds, H. Luithardt, J. Fejzo, M. Chorev, H. Aktas, K. Colson, K. H. Falchuk, J. A. Halperin, and G. Wagner, "Ultrahigh-resolution 1h-13c hsqc spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction," *Journal of the American Chemical Society*, vol. 129, pp. 5108–5116, 16 Apr. 2007, ISSN: 00027863. DOI: 10.1021/ja068541x.

[39] A. S. Stern, K. B. Li, and J. C. Hoch, "Modern spectrum analysis in multidimensional nmr spectroscopy: Comparison of linear-prediction extrapolation and maximum-entropy reconstruction," *Journal of the American Chemical Society*, vol. 124, pp. 1982–1993, 9 Mar. 2002, ISSN: 00027863. DOI: 10.1021/ja011669o.

[40] J. Chen, A. J. Shaka, and V. A. Mandelshtam, "Rrt: The regularized resolvent transform for high-resolution spectral estimation," *Journal of Magnetic Resonance*, vol. 147, pp. 129–137, 1 2000, ISSN: 10907807. DOI: 10.1006/jmre.2000.2176.

[41] M. Mobli and J. C. Hoch, *Maximum entropy spectral reconstruction of nonuniformly sampled data*, Nov. 2008. DOI: 10.1002/cmr.a.20126.

[42] ——, *Nonuniform sampling and non-fourier signal processing methods in multidimensional nmr*, 2014. DOI: 10.1016/j.pnmrs.2014.09.002.

[43] X. Qu, M. Mayzel, J. F. Cai, Z. Chen, and V. Orekhov, *Accelerated nmr spectroscopy with low-rank reconstruction*, Jan. 2015. DOI: 10.1002/anie.201409291.

[44] S. G. Hyberts, D. P. Frueh, H. Arthanari, and G. Wagner, "Fm reconstruction of non-uniformly sampled protein nmr data at higher dimensions and optimization by distillation," *Journal of Biomolecular NMR*, vol. 45, pp. 283–294, 3 2009, ISSN: 09252738. DOI: 10.1007/s10858-009-9368-1.

[45] S. G. Hyberts, K. Takeuchi, and G. Wagner, "Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein nmr data," *Journal of the American Chemical Society*, vol. 132, pp. 2145–2147, 7 Mar. 2010, ISSN: 00027863. DOI: 10.1021/ja908004w.

[46] S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, and G. Wagner, "Application of iterative soft thresholding for fast reconstruction of nmr data non-uniformly sampled with multidimensional poisson gap scheduling," *Journal of Biomolecular NMR*, vol. 52, pp. 315–327, 4 Apr. 2012, ISSN: 09252738. DOI: 10.1007/s10858-012-9611-z.

[47] V. Y. Orekhov and V. A. Jaravine, *Analysis of non-uniformly sampled spectra with multi-dimensional decomposition*, 2011. DOI: 10.1016/j.pnmrs.2011.02.002.

[48] M. J. Bostock, D. J. Holland, and D. Nietlispach, "Compressed sensing reconstruction of undersampled 3d noesy spectra: Application to large membrane proteins," *Journal of Biomolecular NMR*, vol. 54, pp. 15–32, 1 Sep. 2012, ISSN: 09252738. DOI: 10.1007/s10858-012-9643-4.

[49] P. R. Bodart, J. P. Amoureux, and F. Taulelle, "Anafor: Application of a restricted linear least squares procedure to nmr data processing," *Solid State Nuclear Magnetic Resonance*, vol. 21, pp. 1–20, 1-2 Feb. 2002, ISSN: 09262040. DOI: 10.1006/snmr.2001.0040.

[50] T. Ueda, C. Yoshiura, M. Matsumoto, Y. Kofuku, J. Okude, K. Kondo, Y. Shiraishi, K. Takeuchi, and I. Shimada, "Development of a method for reconstruction of crowded nmr spectra from undersampled time-domain data," *Journal of Biomolecular NMR*, vol. 62, pp. 31–41, 1 May 2015, ISSN: 15735001. DOI: 10.1007/s10858-015-9908-9.

[51] G. Zhu, W. Y. Choy, and B. C. Sanctuary, "Spectral parameter estimation by an iterative quadratic maximum likelihood method," *Journal of Magnetic Resonance*, vol. 135, pp. 37–43, 1 1998, ISSN: 10907807. DOI: 10.1006/jmre.1998.1539.

[52] R. A. Chylla, K. Hu, J. J. Ellinger, and J. L. Markley, "Deconvolution of two-dimensional nmr spectra by fast maximum likelihood reconstruction: Application to quantitative metabolomics," *Analytical Chemistry*, vol. 83, pp. 4871–4880, 12 Jun. 2011, ISSN: 00032700. DOI: 10.1021/ac200536b.

[53] Y. Shrot and L. Frydman, "Compressed sensing and the reconstruction of ultrafast 2d nmr data: Principles and biomolecular applications," *Journal of Magnetic Resonance*, vol. 209, pp. 352–358, 2 Apr. 2011, ISSN: 10907807. DOI: 10.1016/j.jmr.2011.01.017.

[54] H. Ratiney, Y. Coenradie, S. Cavassila, D. V. Ormondt, and D. Graveron-Demilly, "Time-domain quantitation of 1h short echo-time signals: Background accommodation," *Magnetic Resonance Materials in PhysicsBiology and Medicine*, vol. 16, pp. 284–296, 6 2004, ISSN: 09685243. DOI: 10.1007/s10334-004-0037-9.

[55] S. Y. Kung, K. S. Arun, and D. V. B. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *Journal of the Optical Society of America*, vol. 73, pp. 1799–1811, 12 1983. DOI: 10.1364/JOSA.73.001799. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=josa-73-12-1799.

[56] G. L. Millhauser, A. A. Carter, D. J. Schneider, J. H. Freed, and R. E. Oswald, "Rapid singular value decomposition for time-domain analysis of magnetic resonance signals by use of the lanczos algorithm," *Journal of Magnetic Resonance (1969)*, vol. 82, pp. 150–155, 1 1989, ISSN: 0022-2364. DOI: https://doi.org/10.1016/0022-2364(89)90175-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022236489901753.

[57] L. Vanhamme, A. V. D. Boogaart, and S. V. Huffel, "Improved method for accurate and efficient quantification of mrs data with use of prior knowledge," 1997, pp. 35–43.

[58] J. W. van der Veen, R. de Beer, P. R. Luyten, and D. van Ormondt, "Accurate quantification of in vivo 31p nmr signals using the variable projection method and prior knowledge.," *Magnetic resonance in medicine*, vol. 6, pp. 92–8, 1 Jan. 1988, ISSN: 0740-3194. DOI: 10.1002/mrm.1910060111.

[59] J. B. Poullet, D. M. Sima, A. W. Simonetti, B. D. Neuter, L. Vanhamme, P. Lemmerling, and S. V. Huffel, "An automated quantification of short echo time mrs spectra in an open source software environment: Aqses," *NMR in Biomedicine*, vol. 20, pp. 493–504, 5 2007, ISSN: 09523480. DOI: 10.1002/nbm.1112.

[60] M. Wilson, G. Reynolds, R. A. Kauppinen, T. N. Arvanitis, and A. C. Peet, "A constrained least-squares approach to the automated quantitation of in vivo 1h magnetic resonance spectroscopy data.," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 65, pp. 1–12, 1 2011, ISSN: 15222594. DOI: 10.1002/mrm.22579.

[61] A. G. Wilson, Y. Wu, D. J. Holland, S. Nowozin, M. D. Mantle, L. F. Gladden, and A. Blake, "Bayesian inference for nmr spectroscopy with applications to chemical quantification," Feb. 2014. [Online]. Available: http://arxiv.org/abs/1402.3580.

[62] D. V. Rubtsov and J. L. Griffin, "Time-domain bayesian detection and estimation of noisy damped sinusoidal signals applied to nmr spectroscopy," *Journal of Magnetic Resonance*, vol. 188, pp. 367–379, 2 Oct. 2007, ISSN: 10907807. DOI: 10.1016/j.jmr.2007.08.008.

[63] D. Belkić and K. Belkić, "The fast padé transform in magnetic resonance spectroscopy for potential improvements in early cancer diagnostics," *Physics in Medicine and Biology*, vol. 50, pp. 4385–4408, 18 2005, ISSN: 00319155. DOI: 10.1088/0031-9155/50/18/010.

[64] D. C. Williamson, H. Hawesa, N. A. Thacker, and S. R. Williams, "Robust quantification of short echo time 1h magnetic resonance spectra using the padé approximant," *Magnetic Resonance in Medicine*, vol. 55, pp. 762–771, 4 2006, ISSN: 15222594. DOI: 10.1002/mrm.20842.

[65] R. A. Chylla and J. L. Markley, "Theory and application of the maximum likelihood principle to nmr parameter estimation of multidimensional nmr data," *Journal of Biomolecular NMR*, vol. 5, pp. 245–258, 3 1995, ISSN: 09252738. DOI: 10.1007/BF00211752.

[66] Y. Matviychuk, E. von Harbou, and D. J. Holland, "An experimental validation of a bayesian model for quantification in nmr spectroscopy," *Journal of Magnetic Resonance*, vol. 285, pp. 86–100, Dec. 2017, ISSN: 10960856. DOI: 10.1016/j.jmr.2017.10.009.

[67] Y. Matsuki, M. T. Eddy, and J. Herzfeld, "Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra," *Journal of the American Chemical Society*, vol. 131, pp. 4648–4656, 13 Apr. 2009, ISSN: 00027863. DOI: 10.1021/ja807893k.

[68] M. Mayzel, J. Rosenlöw, L. Isaksson, and V. Y. Orekhov, "Time-resolved multidimensional nmr with non-uniform sampling," *Journal of Biomolecular NMR*, vol. 58, pp. 129–139, 2 Feb. 2014, ISSN: 09252738. DOI: 10.1007/s10858-013-9811-1.

[69] J. Hao, W. Astle, M. D. Iorio, and T. Ebbels, "Batman-an r package for the automated quantification of me-tabolites from nmr spectra using a bayesian model." [Online]. Available: http://www1.imperial.ac.uk/medicine/people/t.ebbels/.

[70] S. Sokolenko, T. Jézéquel, G. Hajjar, J. Farjon, S. Akoka, and P. Giraudeau, "Robust 1d nmr lineshape fitting using real and imaginary data in the frequency domain," *Journal of Magnetic Resonance*, vol. 298, pp. 91–100, Jan. 2019, ISSN: 10960856. DOI: 10.1016/j.jmr.2018.11.004.

[71] A. Heinecke, L. Ye, M. D. Iorio, and T. Ebbels, "Bayesian deconvolution and quantification of metabolites from J -resolved nmr spectroscopy," *Bayesian Analysis*, May 2020, ISSN: 1936-0975. DOI: 10.1214/20-ba1208.

[72] S. Vanhuffel, H. Chen, C. Decanniere, and P. Vanhecke, "Algorithm for time-domain nmr data fitting based on total least squares," *Journal of magnetic resonance. Series A*, vol. 110, pp. 228–237, 2 1994, ISSN: 1064-1858. DOI: 10.1006/jmra.1994.1209.

[73] K. Young, B. J. Soher, and A. A. Maudsley, "Automated spectral analysis ii: Application of wavelet shrinkage for characterization of non-parameterized signals," *Magnetic Resonance in Medicine*, vol. 40, pp. 816–821, 6 1998, ISSN: 07403194. DOI: 10.1002/mrm.1910400606.

[74] "Estimation of metabolite concentrations from localized in vivo proton nmr spectra," *Magnetic Resonance in Medicine*, vol. 30, pp. 672–679, 6 Dec. 1993, ISSN: 0740-3194. DOI: https://doi.org/10.1002/mrm.1910300604. [Online]. Available: https://doi.org/10.1002/mrm.1910300604.

[75] S. W. Provencher, "Automatic quantitation of localized in vivo 1h spectra with lcmodel," *NMR in Biomedicine*, vol. 14, pp. 260–264, 4 2001, ISSN: 09523480. DOI: 10.1002/nbm.698.

[76] J. P. Grivet, "Accurate numerical approximation to the gauss-lorentz lineshape," *Journal of Magnetic Resonance*, vol. 125, pp. 102–106, 1 1997, ISSN: 10907807. DOI: 10.1006/jmre.1996.1076.

[77] I. Marshall, J. Higinbotham, S. Bruce, and A. Freise, "Use of voigt lineshape for quantification of in vivo 1h spectra," *Magnetic Resonance in Medicine*, vol. 37, pp. 651–657, 5 1997, ISSN: 07403194. DOI: 10.1002/mrm.1910370504.

[78] P. Gillies, I. Marshall, M. Asplund, P. Winkler, and J. Higinbotham, "Quantification of mrs data in the frequency domain using a wavelet filteran approximated voigt lineshape model and prior knowledge," *NMR in Biomedicine*, vol. 19, pp. 617–626, 5 2006, ISSN: 09523480. DOI: 10.1002/nbm.1060.

[79] G. L. Bretthorst, W. C. Hutton, J. R. Garbow, and J. J. Ackerman, "Exponential parameter estimation (in nmr) using bayesian probability theory," *Concepts in Magnetic Resonance Part A: Bridging Education and Research*, vol. 27, pp. 55–63, 2 2005, ISSN: 15466086. DOI: 10.1002/cmr.a.20043.

[80] A. Knijn, R. D. Beer, and D. V. Ormondt, "Frequency-selective quantification in the time domain," *Journal of Magnetic Resonance (1969)*, vol. 97, pp. 444–450, 2 1992, ISSN: 0022-2364. DOI: https://doi.org/10.1016/0022-2364(92)90330-A. [Online]. Available: https://www.sciencedirect.com/science/article/pii/002223649290330A.

[81] M. I. Miller, T. J. Schaewe, C. S. Bosch, and J. J. Ackerman, *Model-based maximum-likelihood estimation for phase- and frequency-encoded magnetic-resonance-imaging data*, 1995. DOI: 10.1006/jmrb.1995.1081.

[82] S. C. Chen, T. J. Schaewe, R. S. Teichman, M. I. Miller, S. N. Nadel, and A. S. Greene, "Parallel algorithms for maximum-likelihood nuclear magnetic resonance spectroscopy," *Journal of Magnetic Resonance*, pp. 16–23, 1992.

[83] H. Serrai, L. Senhadji, D. B. Clayton, C. Zuo, and R. E. Lenkinski, "Water modeled signal removal and data quantification in localized mr spectroscopy using a time-scale postacquistion method," *Journal of Magnetic Resonance*, vol. 149, pp. 45–51, 1 2001, ISSN: 10907807. DOI: 10.1006/jmre.2001.2292.

[84] I. Marshall, S. D. Bruce, J. Higinbotham, A. MacLullich, J. M. Wardlaw, K. J. Ferguson, and J. Seckl, "Choice of spectroscopic lineshape model affects metabolite peak areas and area ratios," *Magnetic Resonance in Medicine*, vol. 44, pp. 646–649, 4 2000, ISSN: 07403194. DOI: 10.1002/1522-2594(200010)44:4<646::AID-MRM20>3.0.CO;2-0.

[85] T. Laudadio, Y. Selén, L. Vanhamme, P. Stoica, P. V. Hecke, and S. V. Huffel, "Subspace-based mrs data quantitation of multiplets using prior knowledge," *Journal of Magnetic Resonance*, vol. 168, pp. 53–65, 1 2004, ISSN: 10907807. DOI: 10.1016/j.jmr.2004.01.015.

[86] R. Romano, A. Motta, S. Camassa, C. Pagano, M. T. Santini, and P. L. Indovina, "A new time-domain frequency-selective quantification algorithm," *Journal of Magnetic Resonance*, vol. 155, pp. 226–235, 2 2002, ISSN: 10907807. DOI: 10.1006/jmre.2002.2521.

[87] H. Witjes, W. J. Melssen, H. J. I. '. Zandt, M. V. D. Graaf, A. Heerschap, and L. M. Buydens, "Automatic correction for phase shiftsfrequency shiftsand lineshape distortions across a series of single resonance lines in large spectral data sets," *Journal of Magnetic Resonance*, vol. 144, pp. 35–44, 1 2000, ISSN: 10907807. DOI: 10.1006/jmre.2000.2021.

[88] R. Stoyanova and T. R. Brown, "Nmr spectral quantitation by principal component analysis: Iii. a generalized procedure for determination of lineshape variations," *Journal of Magnetic Resonance*, vol. 154, pp. 163–175, 2 2002, ISSN: 10907807. DOI: 10.1006/jmre.2001.2486.

[89] O. Bazgir, E. Walden, B. Nutter, and S. Mitra, "A novel data-driven magnetic resonance spectroscopy signal analysis framework to quantify metabolite concentration," *Algorithms*, vol. 13, 5 May 2020, ISSN: 19994893. DOI: 10.3390/A13050120.

[90] F. D. Andrade, L. A. Forato, R. B. Filho, and L. A. Colnago, "Quantification of protein secondary structure by 13c solid-state nmr," *Analytical and Bioanalytical Chemistry*, vol. 408, pp. 3875–3879, 14 May 2016, ISSN: 16182650. DOI: 10.1007/s00216-016-9484-1.

[91] N. Sandgren, Y. Selén, P. Stoica, and J. Li, "Parametric methods for frequency-selective mr spectroscopy - a review," *Journal of Magnetic Resonance*, vol. 168, pp. 259–272, 2 2004, ISSN: 10907807. DOI: 10.1016/j.jmr.2004.03.011.

[92] V. A. Mandelshtam, H. S. Taylor, and A. J. Shaka, "Application of the filter diagonalization method to one- and two-dimensional nmr spectra," *Journal of Magnetic Resonance*, vol. 133, pp. 304–312, 2 1998, ISSN: 10907807. DOI: 10.1006/jmre.1998.1476.

[93] V. A. Mandelshtam, "The multidimensional filter diagonalization method: I. theory and numerical implementation," *Journal of Magnetic Resonance*, vol. 144, pp. 343–356, 2 2000, ISSN: 10907807. DOI: 10.1006/jmre.2000.2023.

[94] M. Cedervall, P. Stoica, and R. Moses, "Mode-type algorithm for estimating dampedundamped or explosive modes," *Conference Record - Asilomar Conference on SignalsSystems and Computers*, vol. 2, pp. 1101–1105, 3 1995, ISSN: 10586393. DOI: 10.1109/ACSSC.1995.540870.

[95] M. Tomczak and E. H. Djermoune, "A subband arma modeling approach to high-resolution nmr spectroscopy," *Journal of Magnetic Resonance*, vol. 158, pp. 86–98, 1-2 2002, ISSN: 10907807. DOI: 10.1016/S1090-7807(02)00064-2.

[96] P. Stoica, N. Sandgren, Y. Selén, L. Vanhamme, and S. V. Huffel, "Frequency-domain method based on the singular value decomposition for frequency-selective nmr spectroscopy," *Journal of Magnetic Resonance*, vol. 165, pp. 80–88, 1 2003, ISSN: 10907807. DOI: 10.1016/S1090-7807(03)00188-5.

[97] U. Seeger, U. Klose, I. Mader, W. Grodd, and T. Nägele, "Parameterized evaluation of macromolecules and lipids in proton mr spectroscopy of brain diseases," *Magnetic Resonance in Medicine*, vol. 49, pp. 19–28, 1 2003, ISSN: 07403194. DOI: 10.1002/mrm.10332.

[98] S. B. Kim, Z. Wang, and B. Hiremath, "A bayesian approach for the alignment of high-resolution nmr spectra," *Annals of Operations Research*, vol. 174, pp. 19–32, 1 2010, ISSN: 02545330. DOI: 10.1007/s10479-008-0332-3.

[99] Y. Cheng, X. Gao, and F. Liang, "Bayesian peak picking for nmr spectra," *GenomicsProteomics and Bioinformatics*, vol. 12, pp. 39–47, 1 2014, ISSN: 16720229. DOI: 10.1016/j.gpb.2013.07.003. [Online]. Available: http://dx.doi.org/10.1016/j.gpb.2013.07.003.

[100] J. Hao, M. Liebeke, W. Astle, M. D. Iorio, J. G. Bundy, and T. M. D. Ebbels, "Bayesian deconvolution and quantification of metabolites in complex 1d nmr spectra using batman," *Nature Protocols*, vol. 9, pp. 1416–1427, 6 2014, ISSN: 17502799. DOI: 10.1038/nprot.2014.090.

[101] L. Vanhamme, T. Sundin, P. V. Hecke, S. V. Huffel, and R. Pintelon, "Frequency-selective quantification of biomedical magnetic resonance spectroscopy data," *Journal of Magnetic Resonance*, vol. 143, pp. 1–16, 1 2000, ISSN: 10907807. DOI: 10.1006/jmre.1999.1960.

[102] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly, "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR in Biomedicine*, vol. 18, pp. 1–13, 1 2005, ISSN: 09523480. DOI: 10.1002/nbm.895.

[103] D. Graveron-Demilly, A. Diop, A. Briguet, and B. Fenet, *Product-operator algebra for strongly coupled spin systems*, 1993. DOI: 10.1006/jmra.1993.1038.

[104] S. A. Smith, T. O. Levante, B. H. Meier, and R. R. Ernst, *Computer simulations in magnetic resonance. an object-oriented programming approach*, 1994. DOI: 10.1006/jmra.1994.1008.

[105] K. Krishnamurthy, "Craft (complete reduction to amplitude frequency table) - robust and time-efficient bayesian approach for quantitative mixture analysis by nmr," *Magnetic Resonance in Chemistry*, vol. 51, pp. 821–829, 12 Dec. 2013, ISSN: 07491581. DOI: 10.1002/mrc.4022.

[106] C. Zheng, S. Zhang, S. Ragg, D. Raftery, and O. Vitek, "Identification and quantification of metabolites in 1h nmr spectra by bayesian model selection," *Bioinformatics*, vol. 27, pp. 1637–1644, 12 Jun. 2011, ISSN: 13674803. DOI: 10.1093/bioinformatics/btr118.

[107] A. Vignoli, V. Ghini, G. Meoni, C. Licari, P. G. Takis, L. Tenori, P. Turano, and C. Luchinat, "Hochdurchsatz-metabolomik mit 1d-nmr," *Angewandte Chemie*, vol. 131, pp. 980–1007, 4 Jan. 2019, ISSN: 0044-8249. DOI: 10.1002/ange.201804736.

[108] S. Kostidis, R. D. Addie, H. Morreau, O. A. Mayboroda, and M. Giera, *Quantitative nmr analysis of intra- and extracellular metabolism of mammalian cells: A tutorial*, Aug. 2017. DOI: 10.1016/j.aca.2017.05.011.

[109] D. Djukovic, D. Raftery, and N. Gowda, *Mass spectrometry and nmr spectroscopy based quantitative metabolomics*, Jan. 2019. DOI: 10.1016/B978-0-12-818607-7.00016-5.

[110] R. Stoyanova, A. C. Kuesel, and T. R. Brown, *Application of principal-component analysis for nmr spectral quantitation*, 1995. DOI: 10.1006/jmra.1995.1177.

[111] C. Ladroue, F. A. Howe, J. R. Griffiths, and A. R. Tate, "Independent component analysis for automated decomposition of in vivo magnetic resonance spectra," *Magnetic Resonance in Medicine*, vol. 50, pp. 697–703, 4 2003, ISSN: 07403194. DOI: 10.1002/mrm.10595.

[112] J. Camacho, R. A. Rodríguez-Gómez, and E. Saccenti, "Group-wise principal component analysis for exploratory data analysis," *Journal of Computational and Graphical Statistics*, vol. 26, pp. 501–512, 3 2017, ISSN: 15372715. DOI: 10.1080/10618600.2016.1265527.

[113] J. A. Hartigan and M. A. Wong, "Algorithm as 136 a k-means clustering algorithm," *Journal of the Royal Statistical Society Series B Methodological*, vol. 28, pp. 100–108, 1 2012.

[114] A. Srinivasan, C. J. Galbán, T. D. Johnson, T. L. Chenevert, B. D. Ross, and S. K. Mukherji, "Utility of the k-means clustering algorithm in differentiating apparent diffusion coefficient values of benign and malignant neck pathologies," *American Journal of Neuroradiology*, vol. 31, pp. 736–740, 4 2010, ISSN: 01956108. DOI: 10.3174/ajnr.A1901.

[115] M. Čuperlović-Culf, N. Belacel, A. S. Culf, I. C. Chute, R. J. Ouellette, I. W. Burton, T. K. Karakach, and J. A. Walter, "Nmrmetabolic analysis of samples using fuzzy k-means clustering," *Magnetic Resonance in Chemistry*, vol. 47, SUPPL. 1 2009, ISSN: 07491581. DOI: 10.1002/mrc.2502.

[116] P. Tiwari, M. Rosen, and A. Madabhushi, "A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (mrs)," *Medical Physics*, vol. 36, pp. 3927–3939, 9 2009, ISSN: 00942405. DOI: 10.1118/1.3180955.

[117] S. Cacciatore, C. Luchinat, and L. Tenori, "Knowledge discovery by accuracy maximization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 5117–5122, 14 2014, ISSN: 10916490. DOI: 10.1073/pnas.1220873111.

[118] S. Cacciatore, L. Tenori, C. Luchinat, P. R. Bennett, and D. A. MacIntyre, "Kodama: An r package for knowledge discovery and data mining," *Bioinformatics*, vol. 33, pp. 621–623, 4 2017, ISSN: 14602059. DOI: 10.1093/bioinformatics/btw705.

[119] T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52–65, 2013, ISSN: 08936080. DOI: 10.1016/j.neunet.2012.09.018. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2012.09.018.

[120] H. Zheng, J. Ji, L. Zhao, M. Chen, A. Shi, L. Pan, Y. Huang, H. Zhang, B. Dong, and H. Gao, "Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps," *Oncotarget*, vol. 7, pp. 59 189–59 198, 37 2016, ISSN: 19492553. DOI: 10.18632/oncotarget.10830.

[121] O. Beckonert, M. E. Bollard, T. M. Ebbels, H. C. Keun, H. Antti, E. Holmes, J. C. Lindon, and J. K. Nicholson, "Nmr-based metabonomic toxicity classification: Hierarchical cluster analysis and k-nearest-neighbour approaches," *Analytica Chimica Acta*, vol. 490, pp. 3–15, 1-2 2003, ISSN: 00032670. DOI: 10.1016/S0003-2670(03)00060-6.

[122] S. Kuhn, B. Egert, S. Neumann, and C. Steinbeck, "Building blocks for automated elucidation of metabolites: Machine learning methods for nmr prediction," *BMC Bioinformatics*, vol. 9, pp. 1–19, 2008, ISSN: 14712105. DOI: 10.1186/1471-2105-9-400.

[123] R. Laatikainen, T. Hassinen, J. Lehtivarjo, M. Tiainen, J. Jungman, T. Tynkkynen, S. P. Korhonen, M. Niemitz, P. Poutiainen, O. Jääskeläinen, T. Väisänen, J. Weisell, P. Soininen, P. Laatikainen, H. Martonen, and K. Tuppurainen, "Comprehensive strategy for proton chemical shift prediction: Linear prediction with nonlinear corrections," *Journal of Chemical Information and Modeling*, vol. 54, pp. 419–430, 2 2014, ISSN: 15499596. DOI: 10.1021/ci400648s.

[124] K. A. Blinov, Y. D. Smurnyy, T. S. Churanova, M. E. Elyashberg, and A. J. Williams, "Development of a fast and accurate method of 13c nmr chemical shift prediction," *Chemometrics and Intelligent Laboratory Systems*, vol. 97, pp. 91–97, 1 2009, ISSN: 01697439. DOI: 10.1016/j.chemolab.2009.01.010. [Online]. Available: http://dx.doi.org/10.1016/j.chemolab.2009.01.010.

[125] C. Cobas, "Nmr signal processingpredictionand structure verification with machine learning techniques," *Magnetic Resonance in Chemistry*, vol. 58, pp. 512–519, 6 2020, ISSN: 1097458X. DOI: 10.1002/mrc.4989.

[126] X. Qu, Y. Huang, H. Lu, T. Qiu, D. Guo, T. Agback, V. Orekhov, and Z. Chen, "Accelerated nuclear magnetic resonance spectroscopy with deep learning," *Angewandte Chemie - International Edition*, vol. 59, pp. 10 297–10 300, 26 2020, ISSN: 15213773. DOI: 10.1002/anie.201908162.

[127] D. F. Hansen, "Using deep neural networks to reconstruct non-uniformly sampled nmr spectra," *Journal of Biomolecular NMR*, vol. 73, pp. 577–585, 10-11 2019, ISSN: 15735001. DOI: 10.1007/s10858-019-00265-1. [Online]. Available: https://doi.org/10.1007/s10858-019-00265-1.

[128] P. Klukowski, M. Augoff, M. ZieRba, M. Drwal, A. Gonczarek, and M. J. Walczak, "Nmrnet: A deep learning approach to automated peak picking of protein nmr spectra," *Bioinformatics*, vol. 34, pp. 2590–2597, 15 2018, ISSN: 14602059. DOI: 10.1093/bioinformatics/bty134.

[129] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015, ISSN: 18792782. DOI: 10.1016/j.neunet.2014.09.003. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2014.09.003.

[130] F. M. Alakwaa, K. Chaudhary, and L. X. Garmire, "Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data," *Journal of Proteome Research*, vol. 17, pp. 337–347, 1 2018, ISSN: 15353907. DOI: 10.1021/acs.jproteome.7b00595.

[131] A. M. Weljie, J. Newton, P. Mercier, E. Carlson, and C. M. Slupsky, "Targeted profiling: Quantitative analysis of 1h nmr metabolomics data.," *Analytical chemistry*, vol. 78, pp. 4430–42, 13 Jul. 2006, ISSN: 0003-2700. DOI: 10.1021/ac060209g. [Online]. Available: https://pubs.acs.org/doi/10.1021/ac060209g%20http://www.ncbi.nlm.nih.gov/pubmed/16808451.

[132] M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes, *Nmr-based metabonomic approaches for evaluating physiological influences on biofluid composition*, May 2005. DOI: 10.1002/nbm.935.

[133] K. H. Liland, *Multivariate methods in metabolomics - from pre-processing to dimension reduction and statistical analysis*, Jun. 2011. DOI: 10.1016/j.trac.2011.02.007.

[134] B. M. Beckwith-Hall, E. Holmes, J. C. Lindon, J. Gounarides, A. Vickers, M. Shapiro, and J. K. Nicholson, "Nmr-based metabonomic studies on the biochemical effects of commonly used drug carrier vehicles in the rat," 2002. DOI: 10.1021/tx020020. [Online]. Available: https://pubs.acs.org/sharingguidelines.

[135] S. C. Connor, R. A. Gray, M. P. Hodson, N. M. Clayton, J. N. Haselden, I. P. Chessell, and C. Bountra, "An nmr-based metabolic profiling study of inflammatory pain using the rat fca model," *Metabolomics*, vol. 3, pp. 29–39, 1 Mar. 2007, ISSN: 15733882. DOI: 10.1007/s11306-006-0039-x.

[136] J. L. Griffin, H. J. Williams, E. Sang, and J. K. Nicholson, "Abnormal lipid profile of dystrophic cardiac tissue as demonstrated by one-and two-dimensional magic-angle spinning 1 h nmr spectroscopy," 2001.

[137] Y. Wang, E. Holmes, J. K. Nicholson, O. Cloarec, J. Chollet, M. Tanner, B. H. Singer, and J. Utzinger, "Metabonomic investigations in mice infected with schistosoma mansoni: An approach for biomarker identification," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 12 676–12 681, 34 Aug. 2004, ISSN: 00278424. DOI: 10.1073/pnas.0404878101.

[138] T. L. Whitehead, B. Monzavi-Karbassi, and T. Kieber-Emmons, "1h-nmr metabonomics analysis of sera differentiates between mammary tumor-bearing mice and healthy controls," *Metabolomics*, vol. 1, pp. 269–278, 3 Jul. 2005, ISSN: 15733882. DOI: 10.1007/s11306-005-0006-y.

[139] D. G. Robertson, M. D. Reily, R. E. Sigler, D. F. Wells, D. A. Paterson, and T. K. Braden, "Metabonomics: Evaluation of nuclear magnetic resonance (nmr) and pattern recognition technology for rapid in vivo screening of liver and kidney toxicants," 2000.

[140] S. A. Sousa, A. Magalhães, and M. M. C. Ferreira, "Optimized bucketing for nmr spectra: Three case studies," *Chemometrics and Intelligent Laboratory Systems*, vol. 122, pp. 93–102, 2013, ISSN: 01697439. DOI: 10.1016/j.chemolab.2013.01.006. [Online]. Available: http://dx.doi.org/10.1016/j.chemolab.2013.01.006.

[141] J. E. Lee, G. S. Hwang, F. V. D. Berg, C. H. Lee, and Y. S. Hong, "Evidence of vintage effects on grape wines using 1h nmr-based metabolomic study," *Analytica Chimica Acta*, vol. 648, pp. 71–76, 1 Aug. 2009, ISSN: 00032670. DOI: 10.1016/j.aca.2009.06.039.

[142] P. E. Anderson, N. V. Reo, N. J. DelRaso, T. E. Doom, and M. L. Raymer, "Gaussian binning: A new kernel-based method for processing nmr spectroscopic data for metabolomics," *Metabolomics*, vol. 4, pp. 261–272, 3 2008, ISSN: 15733882. DOI: 10.1007/s11306-008-0117-3.

[143] T. D. Meyer, D. Sinnaeve, B. V. Gasse, E. Tsiporkova, E. R. Rietzschel, M. L. D. Buyzere, T. C. Gillebert, S. Bekaert, J. C. Martins, and W. V. Criekinge, "Nmr-based characterization of metabolic alterations in hypertension using an adaptiveintelligent binning algorithm," *Analytical Chemistry*, vol. 80, pp. 3783–3790, 10 2008, ISSN: 00032700. DOI: 10.1021/ac7025964.

[144] P. E. Anderson, D. A. Mahle, T. E. Doom, N. V. Reo, N. J. DelRaso, and M. L. Raymer, "Dynamic adaptive binning: An improved quantification technique for nmr spectroscopic data," *Metabolomics*, vol. 7, pp. 179–190, 2 2011, ISSN: 15733882. DOI: 10.1007/s11306-010-0242-7.

[145] R. A. Davis, A. J. Charlton, J. Godward, S. A. Jones, M. Harrison, and J. C. Wilson, "Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, pp. 144–154, 1 2007, ISSN: 01697439. DOI: 10.1016/j.chemolab.2006.08.014.

[146] S. K. Bharti and R. Roy, *Quantitative 1h nmr spectroscopy*, May 2012. DOI: 10.1016/j.trac.2012.02.007.

[147] R. A. Meyer, M. J. Fisher, S. J. Nelson, and T. R. Brown, "Evaluation of manual methods for integration of in vivo phosphorus nmr spectra," *NMR in Biomedicine*, vol. 1, pp. 131–135, 3 Jun. 1988, ISSN: 0952-3480. DOI: https://doi.org/10.1002/nbm.1940010306. [Online]. Available: https://doi.org/10.1002/nbm.1940010306.

[148] M. Geyer, K. P. Neidig, and H. R. Kalbitzer, *Automated peak integration in multidimensional nmr spectra by an optimized iterative segmentation procedure*, 1995. DOI: 10.1006/jmrb.1995.1143.

[149] C. Rischel, "Fundamentals of peak integration," *Journal of Magnetic ResonanceSeries A*, vol. 116, pp. 255–258, 2 1995, ISSN: 10641858. DOI: 10.1006/jmra.1995.0016.

[150] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "Nmrpipe: A multidimensional spectral processing system based on unix pipes," *Journal of Biomolecular NMR*, vol. 6, pp. 277–293, 3 1995, ISSN: 09252738. DOI: 10.1007/BF00197809.

[151] R. Koradi, M. Billeter, M. Engeli, P. Güntert, and K. Wüthrich, "Automated peak picking and peak integration in macromolecular nmr spectra using autopsy," *Journal of Magnetic Resonance*, vol. 135, pp. 288–297, 2 1998, ISSN: 10907807. DOI: 10.1006/jmre.1998.1570.

[152] D. M. Korzhnev, I. V. Ibraghimov, M. Billeter, and V. Y. Orekhov, "Munin: Application of three-way decomposition to the analysis of heteronuclear nmr relaxation data," *Journal of Biomolecular NMR*, vol. 21, pp. 263–268, 3 2001, ISSN: 09252738. DOI: 10.1023/A:1012982830367.

[153] R. Romano, D. Paris, F. Acernese, F. Barone, and A. Motta, "Fractional volume integration in two-dimensional nmr spectra: Cakea monte carlo approach," *Journal of Magnetic Resonance*, vol. 192, pp. 294–301, 2 2008, ISSN: 10907807. DOI: 10.1016/j.jmr.2008.03.010.

[154] L. Griffiths and A. M. Irving, "Assay by nuclear magnetic resonance spectroscopy: Quantification limits," *Analyst*, vol. 123, pp. 1061–1068, 5 1998, ISSN: 00032654. DOI: 10.1039/a800625c.

[155] J. C. Lindon and A. G. Ferrige, "Digitisation and data processing in fourier transform nmr," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 14, pp. 27–66, 1 1980, ISSN: 00796565. DOI: 10.1016/0079-6565(80)80002-1.

[156] O. M. Weber, C. O. Duc, D. Meier, and P. Boesiger, "Heuristic optimization algorithms applied to the quantification of spectroscopic data," *Magnetic Resonance in Medicine*, vol. 39, pp. 723–730, 5 1998, ISSN: 07403194. DOI: 10.1002/mrm.1910390509.

[157] G. J. Metzger, M. Patel, and X. Hu, "Application of genetic algorithms to spectral quantification," *Journal of Magnetic Resonance - Series B*, vol. 110, pp. 316–320, 3 1996, ISSN: 10641866. DOI: 10.1006/jmrb.1996.0050.

[158] R. E. Gabr, R. Ouwerkerk, and P. A. Bottomley, "Quantifying in vivo mr spectra with circles," *Journal of Magnetic Resonance*, vol. 179, pp. 152–163, 1 2006, ISSN: 10907807. DOI: 10.1016/j.jmr.2005.11.004.

[159] C. Elster, F. Schubert, A. Link, M. Walzel, F. Seifert, and H. Rinneberg, "Quantitative magnetic resonance spectroscopy: Semi-parametric modeling and determination of uncertainties," *Magnetic Resonance in Medicine*, vol. 53, pp. 1288–1296, 6 2005, ISSN: 07403194. DOI: 10.1002/mrm.20500.

[160] M. Kanowski, J. Kaufmann, J. Braun, J. Bernarding, and C. Tempelmann, "Quantitation of simulated short echo time 1h human brain spectra by lcmodel and amares," *Magnetic Resonance in Medicine*, vol. 51, pp. 904–912, 5 2004, ISSN: 07403194. DOI: 10.1002/mrm.20063.

[161] S. Cavassila, S. Deval, C. Huegen, D. V. Ormondt, and D. Graveron-Demilly, "Cramer-rao bound expressions for parametric estimation of overlapping peaks: Influence of prior knowledge," *Journal of Magnetic Resonance*, vol. 143, pp. 311–320, 2 2000, ISSN: 10907807. DOI: 10.1006/jmre.1999.2002.

[162] C. Cudalbu, S. Cavassila, H. Rabeson, D. van Ormondt, and D. Graveron-Demilly, "Influence of measured and simulated basis sets on metabolite concentration estimates," *NMR in Biomedicine*, vol. 21, pp. 627–636, 6 2008, ISSN: 09523480. DOI: 10.1002/nbm.1234.

[163] S. CAVASSILA, S. DEVAL, C. HUEGEN, D. V. ORMONDT, and D. GRAVERON-DEMILLY, "The beneficial influence of prior knowledge on the quantitation of in vivo magnetic resonance spectroscopy signals," *Invest Radiol*, vol. 34, pp. 242–246, 3 1999, ISSN: 0020-9996. DOI: 10.1097/00004424-199903000-00015.

[164] H. Chen, S. V. Huffel, D. V. Ormondt, and R. D. Beer, "Parameter estimation with prior knowledge of known signal poles for the quantification of nmr spectroscopy data in the time domain," *Journal of Magnetic Resonance - Series A*, vol. 119, pp. 225–234, 2 1996, ISSN: 10641858. DOI: 10.1006/jmra.1996.0077.

[165] H. Chen, S. Van Huffel, A. van den Boom, and P. van den Bosch, "Subspace-based parameter estimation of exponentially damped sinusoids using prior knowledge of frequency and phase," *Signal Processing*, vol. 59, no. 1, pp. 129–136, 1997, ISSN: 0165-1684. DOI: https://doi.org/10.1016/S0165-1684(97)00085-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168497000856.

[166] M. V. D. Graaf and A. Heerschap, "Common processing of in vivo mr spectra," *NMR in Biomedicine*, vol. 14, pp. 224–232, 4 2001, ISSN: 09523480. DOI: 10.1002/nbm.707.

[167] S. U. Schumacher, B. Rothenhäusler, A. Willmann, J. Thun, R. Moog, and M. Kuentz, "Time domain nmr as a new process monitoring method for characterization of pharmaceutical hydrates," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 137, pp. 96–103, 2017, ISSN: 1873264X. DOI: 10.1016/j.jpba.2017.01.017. [Online]. Available: http://dx.doi.org/10.1016/j.jpba.2017.01.017.

[168] Y. Hiltunen, J. Kaartinen, J. Pulkkinen, A. M. Häkkinen, N. Lundbom, and R. A. Kauppinen, "Quantification of human brain metabolites from in vivo 1h nmr magnitude spectra using automated artificial neural network analysis," *Journal of Magnetic Resonance*, vol. 154, pp. 1–5, 1 2002, ISSN: 10907807. DOI: 10.1006/jmre.2001.2457.

[169] H. E. Röhnisch, J. Eriksson, E. Müllner, P. Agback, C. Sandström, and A. A. Moazzami, "Aqua: An automated quantification algorithm for high-throughput nmr-based metabolomics and its application in human plasma," *Analytical Chemistry*, vol. 90, pp. 2095–2102, 3 Feb. 2018, ISSN: 15206882. DOI: 10.1021/acs.analchem.7b04324.

[170] G. Lefort, L. Liaubet, C. Canlet, P. Tardivel, M. C. Père, H. Quesnel, A. Paris, N. Iannuccelli, N. Vialaneix, and R. Servien, "Asics: An r package for a whole analysis workflow of 1d 1h nmr spectra," *Bioinformatics (OxfordEngland)*, vol. 35, pp. 4356–4363, 21 2019, ISSN: 13674811. DOI: 10.1093/bioinformatics/btz248.

[171] Y. Liu, J. Cheng, H. Liu, Y. Deng, J. Wang, and F. Xu, "Nmrspec: An integrated software package for processing and analyzing one dimensional nuclear magnetic resonance spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 162, pp. 142–148, October 2016 2017, ISSN: 18733239. DOI: 10.1016/j.chemolab.2017.01.005. [Online]. Available: http://dx.doi.org/10.1016/j.chemolab.2017.01.005.

[172] T. Schoenberger, "Determination of standard sample purity using the high-precision 1h-nmr process," *Analytical and Bioanalytical Chemistry*, vol. 403, pp. 247–254, 1 Apr. 2012, ISSN: 1618-2642. DOI: 10.1007/s00216-012-5777-1.

[173] F. Malz and H. Jancke, "Validation of quantitative nmr," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 38, pp. 813–823, 5 Aug. 2005, ISSN: 07317085. DOI: 10.1016/j.jpba.2005.01.043.

[174] J. Keeler, *Understanding NMR Spectroscopy*, 2nd ed. John Wiley and Sons, 2010.

[175] G. N. Chmurny and D. I. Hoult, "The ancient and honourable art of shimming," *Concepts in Magnetic Resonance*, vol. 2, pp. 131–149, 3 Jul. 1990, ISSN: 10437347. DOI: 10.1002/cmr.1820020303.

[176] M. Häckl, P. Tauber, F. Schweda, H. U. Zacharias, M. Altenbuchinger, P. J. Oefner, and W. Gronwald, "An r-package for the deconvolution and integration of 1d nmr data: Metabodecon1d," *Metabolites*, vol. 11, p. 452, 7 Jul. 2021, ISSN: 2218-1989. DOI: 10.3390/metabo11070452.