*Article*

# Unsupervised Analysis of Small Molecule Mixtures by Wavelet-Based Super-Resolved NMR

**Aritro Sinha Roy** [1] 0000-0001-5900-1997 **and Madhur Srivastava** [2,3]* 0000-0002-2095-0412

[1] Department of Chemistry and Chemical Biology, Cornell University; as836@cornell.edu
[2] Department of Chemistry and Chemical Biology, Cornell University; ms2736@cornell.edu
[3] National Biomedical Center for Advanced ESR Technology, Cornell University; ms2736@cornell.edu
* Correspondence: ms2736@cornell.edu

**Abstract:** Resolving small molecule mixtures by nuclear magnetic resonance (NMR) spectroscopy has been of great interest for a long time for its precision, reproducibility and efficiency. However, spectral analyses for such mixtures are often highly challenging due to overlapping resonance lines and limited chemical shift windows. The existing experimental and theoretical methods to produce shift NMR spectra in dealing with the problem have limited applicability owing to sensitivity issues, inconsistency and / or requirement of prior knowledge. Recently, we have resolved the problem by decoupling multiplet structures in NMR spectra by the wavelet packet transform (WPT) technique. In this work, we developed a scheme for deploying the method in generating highly resolved WPT NMR spectra and predicting the composition of the corresponding molecular mixtures from their $^{1}$H NMR spectra in an automated fashion. The four-step spectral analysis scheme consists of calculating WPT spectrum, peak matching with a WPT shift NMR library, followed by two optimization steps in producing the predicted molecular composition of a mixture. The robustness of the method was tested on an augmented dataset of 1000 molecular mixtures, each containing 3 to 7 molecules. The method successfully predicted the constituent molecules with a median true positive rate of 1.0 against the varying compositions, while a median false positive rate of 0.04 was obtained. The approach can be scaled easily for much larger datasets.

**Keywords:** NMR; shift spectra; wavelet packet transform; automated small molecule mixture analysis

## 1. Introduction

Identification of the components of small molecule mixtures is a crucial and challenging step in the research and developments activities in pharmaceutical drug discovery [1–3], metabolomics [4–6], natural product synthesis [7–9], food quality control [10,11] and environmental sciences [12,13]. Different types of nuclear magnetic resonance (NMR) spectroscopic methods, high-performance liquid chromatography (HPLC) and mass spectrometry (MS) are widely used across the associated industries for the purpose. The main advantages of NMR over the other techniques are that (1) its results are highly reproducible, (2) it requires very little sample preparation effort and (3) it is a nondestructive method [14–16]. However, its relatively poor resolution and sensitivity often made NMR an essential but non-exhaustive analytic tool [5,9]. While recent developments for sensitivity improvement has largely been successful [17–19], limited progress has been made towards achieving the desired resolution. This is primarily due to the limited range of chemical shift windows ($\sim$ 10 ppm) and overlapping resonance lines of small molecules. It is possible to enhance resolution in homonuclear decoupled $^{1}$H NMR spectroscopy by producing pure shift spectra [20–24]. The technique failed to gain wide applicability owing to its experimental complexity and poor sensitivity [20,25]. In order to resolve the problem theoretically, the maximum entropy method has been used in converting NMR to pure shift spectra by deconvolution [26,27]. One of its major drawbacks is the requirement of prior knowledge about the scalar coupling patterns and the coupling constants, which is reasonable for $^{13}$C NMR, but makes it unsuitable for $^{1}$H NMR spectroscopy. Apart from those, a series of spectral analysis tools have been developed, which include peak matching strategies [28–30], spectral editing [31,32], similarity measure [33,34] and deep learning-based tools

 doi:10.20944/preprints202212.0278.v1



[35–37], for identifying small molecule mixture constituents from the corresponding NMR spectra. However, those applications can be seldom generalized, often suffer from low reliability and / or require extremely large and specifically designed training datasets. As a result, none of the methods are suitable for high-throughput analysis of small molecule mixtures using $^1$H NMR as the primary tool.

In a recent work, we showed that wavelet packet transform (WPT) can work as a multi-resolution signal processing tool in transforming an $^1$H NMR to a pure-shift spectrum [38]. Successive decomposition of a spectrum by WPT yields pairs of approximation and detail components at each level, which contain some of the low- and high-frequency spectral features from the chemical shift domain, respectively. The approximation component produced at the final level of decomposition of an $^1$H NMR spectrum produces only singlet structures, while the multiplet structures are transferred to the various detail components. We illustrated that the former can be used to calculate a simple pure-shift spectrum and the robustness of the WPT-based NMR spectral analysis method against significant level of noise had been established [38].

In this work, we developed an automated method for predicting molecular compositions from the corresponding $^1$H NMR spectra. The problem can be divided into two, (1) predicting the number of molecules in a mixture and (2) predicting their identities. For the purpose, we created an extensive database of 1000 augmented NMR spectra of molecular mixtures, each containing 3 to 7 spectra of the constituent molecules. A library of WPT shift NMR was created from 500 MHz NMR spectra of 74 molecules. The mixed NMR spectra were analyzed in an automated fashion by implementing a four-step algorithm. The algorithm in its first two steps, calculates a WPT shift spectrum from an NMR spectrum and obtains a potential molecular composition by matching the peaks in the shift spectrum with those of the spectral library. Next, the composition is optimized by employing a gradient descent method to minimize the mean squared error in predicting the WPT shift spectrum of the mixture. The top 15 entries from the potential composition are forwarded to the next step, where another gradient descent-based minimization in predicting the the WPT spectrum of the mixture produces the final list of molecules,

In analyzing the performance of the method, we used the true positive and false positive rates, which represented the number of accurate and false predictions with respect to the actual compositions of a molecular mixture, respectively. After the first optimization step, we obtained an average true positive rate of 1.0, while the average false positive rate was very high (0.3). This *elimination* step removed the molecules (choices) with zero or very low probability to be present in the composition from the probable list. Among the remaining choices, the top entries by their calculated probability of existence described the true compositions for all the cases in the augmented dataset. In fact, we observed that for mixtures with 3 to 4 molecules, a true positive rate of 1.0 could have been obtained considering only the top 6 to 8 entries, respectively. Therefore, selecting the optimal number of entries from the potential list of molecules without a priori information required a second optimization. In this *identification* step, the top 15 entries obtained at the end of the elimination step was optimized by another gradient descent method, which resulted into a median true positive rate of 1.0, while reducing the false positive rate to 0.04 for the analysis.

## 2. Materials and Methods

### *2.1. Spectral Library and Augmented Dataset Creation*

We built a spectral library with the NMR spectra of 74 small molecules. Both experimental and predicted NMR spectra were used in the library based on data availability from a peer-reviewed publication [35] and the Human Metabolome Database [39]. The corresponding WPT spectral library for the molecules was computed using Daubechies-9 (Db9) wavelet and a full reduction of all multiplets to singlets in a spectrum was used as the criterion to select the optimal decomposition level. We mixed the NMR spectra of 20 molecules from the library to create an augmented dataset of 1000 spectra, shown in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2022 doi:10.20944/preprints202212.0278.v1



Figure 1. Only 20 molecules were chosen in creating the augmented spectra for two reasons, (1) setting a known list of true negatives and (2) making the analysis realistic, because the mixtures in most applications usually contain structurally related molecules. Each mixed spectrum was calculated by mixing 3 to 7 randomly selected molecules' NMR spectra in varying proportions from 0.15 to 0.4. In creating a mixed spectrum, the component spectra were added in such a way that the number of data points in the mixed spectrum equaled the mean length of the individual spectra [38].

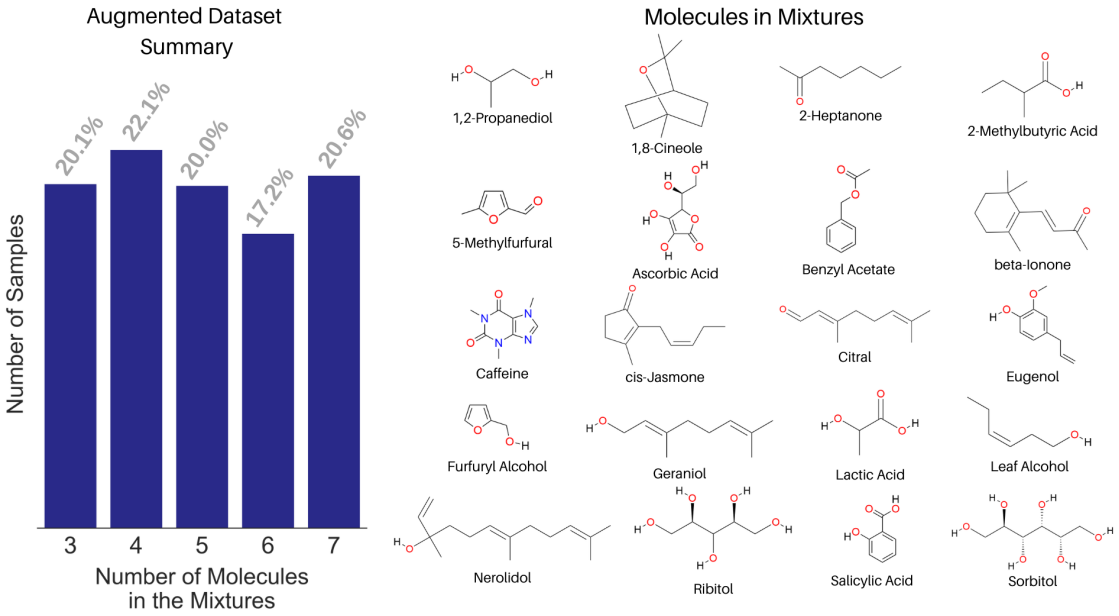


**Figure 1.** Summary of the augmented NMR spectral dataset with fraction of samples against the number of constituent molecules in the mixtures (left) and the structure of the 20 molecules used in creating the augmented dataset (right).

### *2.2. Spectral Analysis Algorithm*

The algorithm used can be seen as a four-step process, (1) conversion of a NMR spectrum to its WPT and WPT shift versions, (2) matching peaks with the WPT shift NMR library and producing a sorted list (L-I) of potential components, (3) optimizing L-I to L-II by applying a linear gradient descent algorithm, and (4) optimizing the top 15 entries of L-II to produce the final prediction of the molecular composition of a mixture. The scheme is summarized in Figure 2. Both the optimization steps used linear gradient descent algorithms, but the targets ($Y$) were taken to be WPT shift spectra for step-(3) and the WPT spectra for step-(4). WPT shift spectra are much simplified versions of the corresponding WPT spectra, where only the peak positions and peak heights from the latter are used [38]. The density matrices ($X$), whose columns correspond to the potential molecules in L-I or L-II, were constructed from the intersections of the chemical shift values from $Y$ and the WPT shift / WPT spectral intensities for the individual molecules. The cost function, $J$ [40], and the gradient descent minimization are given by

$$\begin{aligned} J(\Theta) &= \sum (Y - X \cdot \Theta)^2 / n \\ \Theta_{i+1} &= \Theta_i - \alpha \, \nabla_\Theta J(\Theta_i) \end{aligned} \tag{1}$$

where $n$ is the dimension of $Y$ and $\Theta$ contains the probabilities for a set of molecules to be present in a mixture, $\nabla_\Theta J(\Theta)$ represents the gradient of the cost function and $\alpha$ is the learning rate. The value of $\alpha$ was chosen to be 0.1 in step-(3) and for each iteration in step-(4), $\alpha$ was selected randomly from a uniform distribution in the range of 0.01 and 0.1. The steps in the algorithm are summarized as follows

 doi:10.20944/preprints202212.0278.v1



1. Calculate WPT and WPT shift spectrum from an NMR spectrum
2. Match the WPT shift spectrum with the WPT shift spectral library
   (a) $p$ = Count number of match for each molecule in the library
   (b) Probability for a molecule to be in the mixture = p / The number of peaks in the WPT shift spectrum of the molecule
   (c) Continue for all the molecules in the library and short-list the ones with non-zero probabilities into the list, L-I
3. Optimize the short-listed molecules by a gradient descent method
   (a) Define the WPT shift NMR spectrum of a molecular mixture as the target variable, $Y_1$
   (b) Create a design matrix, $X_1$, from the intersection of the chemical shift values from $Y_1$ and the intensities of the spectra for the molecules in L-I
   (c) Minimize $\sum (Y_1 - X_1 \cdot \Theta)^2 / n_1$, where $n_1$ is the dimension of $Y_1$ and $\Theta$ is probabilities associated with the molecules in L-I, using a gradient descent method with a learning rate, $\alpha$ = 0.1
   (d) An optimized list of molecules, L-II, associated with non-zero probabilities is obtained
4. Top 15 entries from L-II are used as the input to another optimization step
   (a) Define the WPT NMR spectrum of a molecular mixture as the target variable, $Y_2$
   (b) Create a design matrix, $X_2$, from the intersection of the chemical shift values from $Y_2$ and the intensities of the spectra for the molecules in L-II
   (c) Minimize $\sum (Y_2 - X_2 \cdot \Theta)^2 / n_2$ using a gradient descent method with the learning rate chosen randomly from a uniform distribution between 0.01 and 0.1
   (d) An optimized list of molecules associated with probabilities greater than 0.1 is obtained

We used true positive and false positive rates as the metrics in evaluating the performance of our spectral analysis method, defined as follows

$$\text{True positive rate} = \frac{\text{True assignments}}{\text{Actual composition}}$$

$$\text{False positive rate} = \frac{\text{False assignments}}{\text{Spectral library - True assignments}}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2022 doi:10.20944/preprints202212.0278.v1



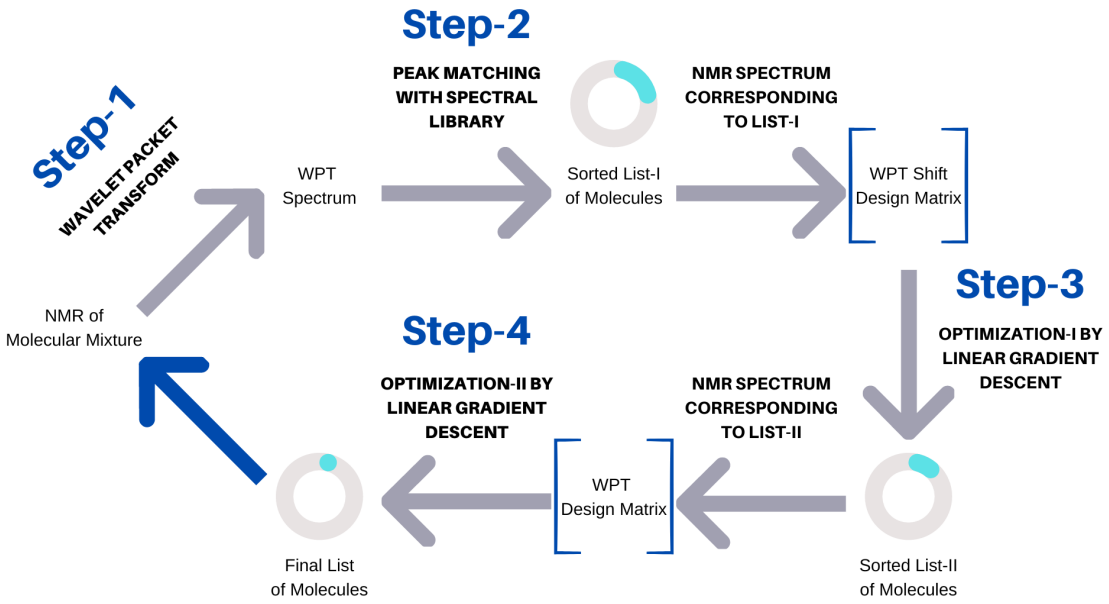


**Figure 2.** Schematic representation of the spectral analysis algorithm.

### 3. Results and Discussion

As a benchmark, we analyzed the dataset of mixed spectra by matching them with the NMR spectra of individual molecules, which is the most commonly used strategy at present [28,29]. From the summary of the analysis shown in Figure 3, it can be seen that the average true positive rate is $\sim 0.7$ for the entire dataset as well as for the mixtures with different number of constituent molecules in them. Both the subplots in the figure show large variations in the true positive rate, which demonstrates the high uncertainty involved in the analysis. The false positive rate for all the cases were equal to 0. It should also be noted that this kind of direct matching may not be feasible for much larger libraries than the one with 74 molecules used in this work.

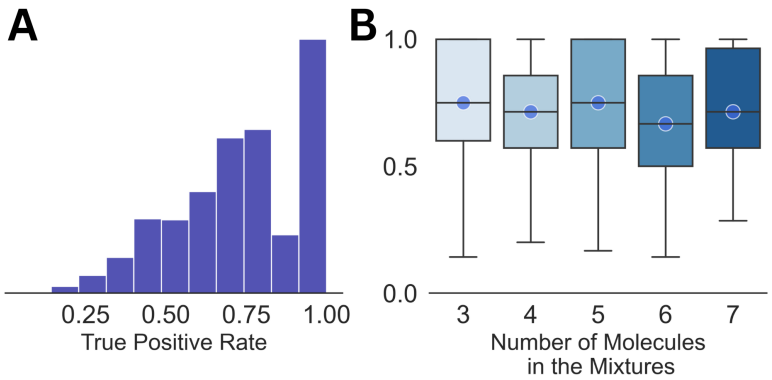


**Figure 3.** Distribution of the true positive rates for the entire dataset (A) and against the size of the mixture (B) are shown. The circles in (B) emphasizes the median for each of the distributions.

The results obtained at step-(3) of our scheme (Figure 2) are summarized in Figure 4. At this stage of our analysis, a median true positive rate of 1.0 was obtained for the entire spectral dataset. This observation demonstrates the robustness of the WPT shift representation of a NMR spectrum and its ability to enhance resolution [38]. However, the impressive true positive rate was associated with a very high false positive rate across all the cases, with a median value of 0.3. Both the average false positive rate, and its variations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2022 doi:10.20944/preprints202212.0278.v1



increased with the size of the mixtures or the increasing complexity of the corresponding spectra.

Peaking into the individual analyses and the corresponding mixture compositions, we noticed that while $\sim$ 30 molecules were present in each prediction on an average, leaving a few outliers, the top 6 to 15 entries by their probability of existence contained all the components of the mixtures. Hence, in the final step of our spectral analysis scheme, we employed another optimization, which used the top 15 entries from a predicted molecular composition at step-(3). This step resulted into a massive reduction in false positive rate from 0.3 at step-(3) to 0.04, shown in Figure 5, while the true positive rate remained mostly unaffected except for the case with molecular mixtures comprising of 7 molecules. Even for the latter case, we obtained a median true positive rate of 1.0 with a standard deviation of 0.08.

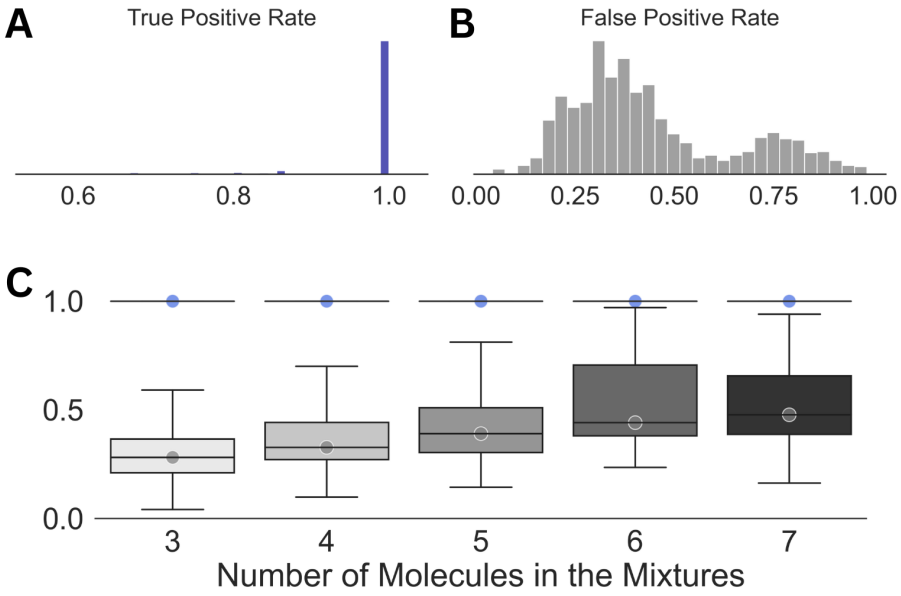


**Figure 4.** Summary of the results obtained at step-(3) of the analysis. Distributions of the true positive (A) and the false positive rates (B) for the entire dataset along with those against the size of the mixtures (C) are shown. The circles in (C) emphasizes the median true positive rate (blue) and false positive rate (gray) for each of the distributions. For all the cases, a true positive rate of 1.0 was achieved (standard deviation = 0).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2022 doi:10.20944/preprints202212.0278.v1



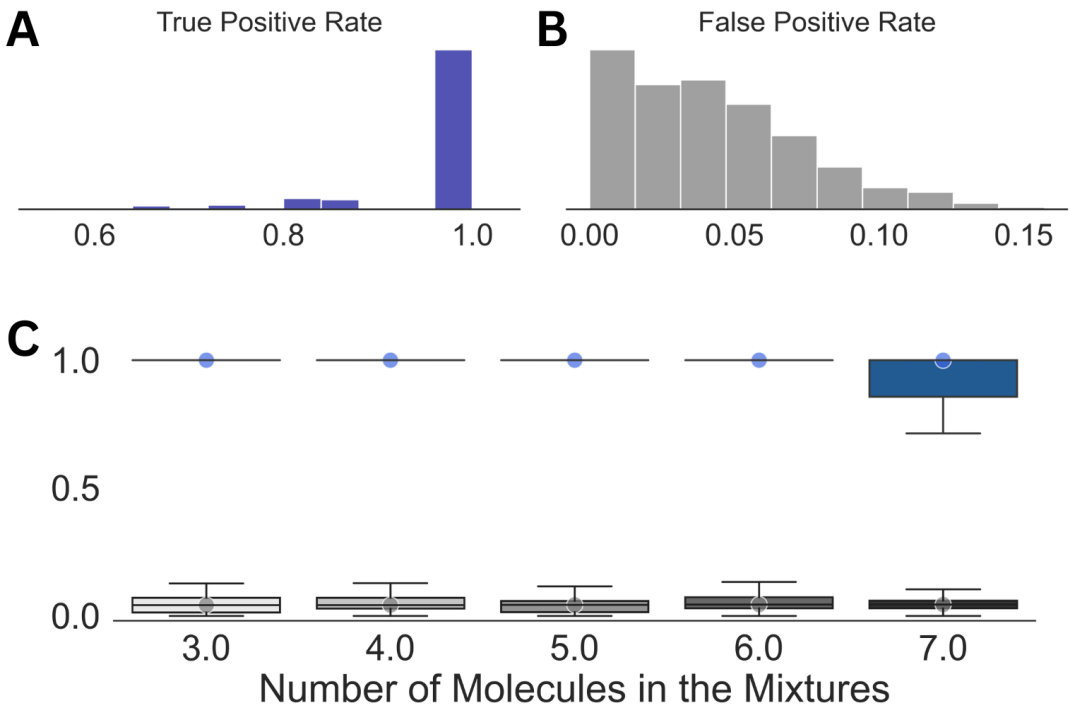


**Figure 5.** Summary of the results obtained after step-(4) in the analysis. Distributions of the true positive (A) and false positive rates (B) for the entire dataset along with those against the size of the mixtures (C) are shown. The circles in (C) emphasizes the median true positive rate (blue) and false positive rate (gray) for each of the distributions.

For visualization purpose, we plotted the predicted NMR spectra from the component analysis for a set of four representative cases and compared those with the corresponding mixed NMR spectra, shown in Figure 6. The descriptions for the representative mixtures are given in Table 1. In Figure 6 (A), a simple visual inspection could remove the false entries; astaxanthin, indolelactive acid and L-fucose. The probable cause for their inclusion in the final prediction was partial overlap between the molecular and mixed NMR spectra. In contrast, the top 4 molecules of the prediction in Figure 6 (B) correspond to the composition of the molecular mixture-23. Two of the three false positives, nicotinuric acid and linalyl acetate, could be discarded by visual inspection and the probability of the third one, sulcatone, is less than half of that of the 1,8-cineole. Figure 6 (C) illustrates a similar analysis for a mixture containing 5 molecules, predicted by the top 5 molecules in the analysis. An easy elimination of the false positives, nicotine and catechin, by visual inspection is achievable in this case as well. In the last example, Figure 6 (D), the top 7 molecules in the prediction capture all the 6 molecules in the corresponding mixture. The false positive, shikimic acid, shows up in this list because of its high degree of overlap with the mixed spectrum. However, the missing peak in the mixed spectrum between 7 and 8 ppm could be used to remove it from the predicted composition. The rest of the false positives, nicotine and sucrose, can be eliminated by visual comparison of the actual and the predicted spectra. Our method's performance summary is given in Table 2.

 doi:10.20944/preprints202212.0278.v1



**Table 1.** Representative set of molecular mixtures and the corresponding prediction summary.

| Mixture No. | Number of Molecules | Molecules (Proportions %) | true positive rate | false positive rate |
|---|---|---|---|---|
| 5 | 3 | Caffeine (39), Ribitol (33), cis-Jasmone (28) | 1.0 | 0.04 |
| 23 | 4 | Nerolidol (35), 1,8-Cineole (22), Leaf alcohol (22), Furfuryl alcohol (21) | 1.0 | 0.04 |
| 35 | 5 | Sorbitol (28), Eugenol (26), Ribitol (18), Ascorbic acid (15), Salicylic acid (13) | 1.0 | 0.03 |
| 20 | 6 | Ribitol (20), Eugenol (19), cis-Jasmone (18), 5-Methylfurfural (17), Ascorbic acid (15), 1,8-Cineole (12) | 1.0 | 0.04 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2022 doi:10.20944/preprints202212.0278.v1



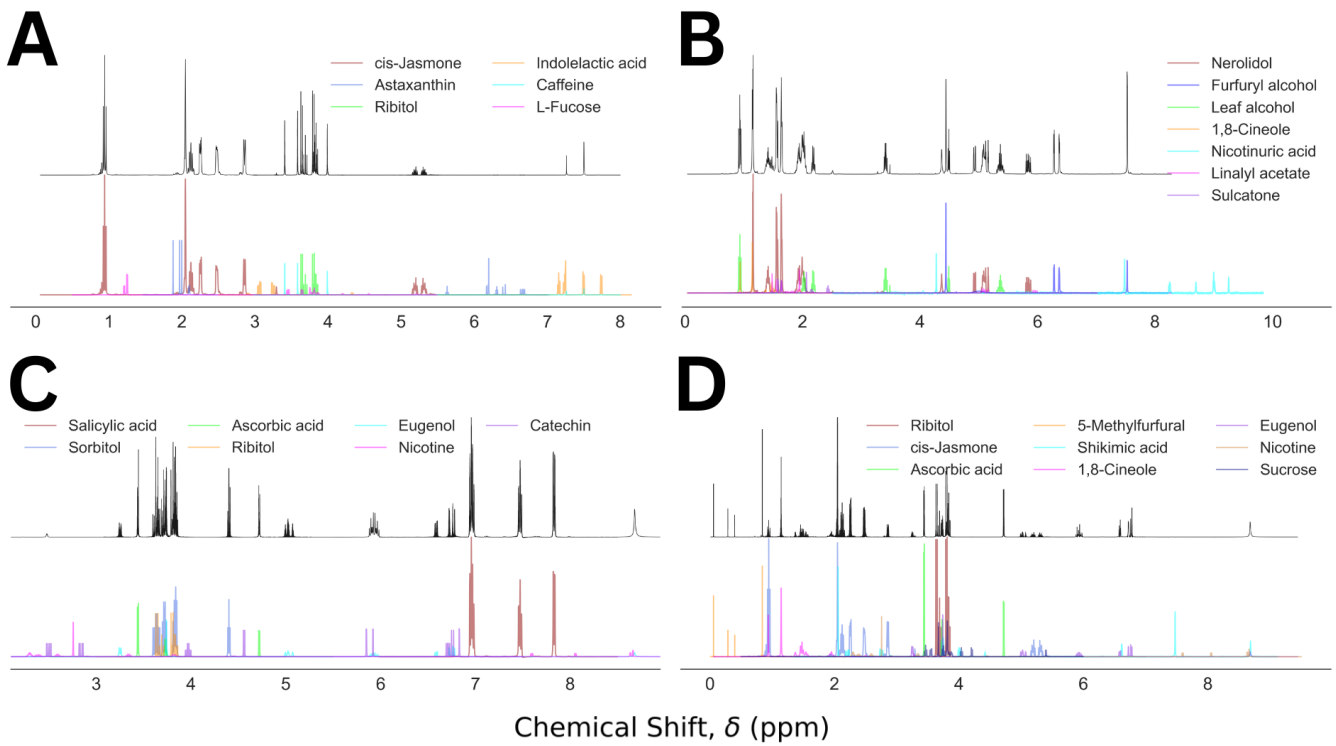


**Figure 6.** Mixed NMR spectra (black) and the predicted components (color coded) for mixture number 5 (A), 23 (B), 35 (C) and 20 (D), containing 3, 4, 5 and 6 molecules, respectively.

**Table 2.** Summary of the automated molecular mixture analyzer's performance for the augmented NMR dataset.

| Parameters | true positive rate | false positive rate |
|---|---|---|
| **Mean** | 0.97 | 0.05 |
| **Median** | 1.0 | 0.04 |
| **Standard Deviation** | 0.09 | 0.03 |

## 4. Conclusions

Composition analysis of small molecule mixtures is essential across a wide range of biological and organic research activities. While $^{1}$H NMR spectroscopy is a very powerful and effective technique in identifying small molecules, NMR spectra of molecular mixtures are often poorly resolved due to spectral overlapping and the presence of multiplet structures. In this work, we presented an automated spectral analysis algorithm, which enhances spectral resolution by the application of the wavelet packet transform and predicts the associated molecular composition in a probabilistic manner. An augmented dataset of 1000 NMR spectra, corresponding to molecular mixtures containing 3 to 7 molecules, was used to test the efficiency of our method. We obtained a median true positive rate of 1.0 for all the mixtures with zero variation for the mixtures containing up to 6 molecules; the true positive rate for mixtures with 7 molecules had a median and standard deviation of 1.0 and 0.08, respectively. A reasonably low false positive rate of 0.04 was achieved for the dataset. In addition, we demonstrated that the precision of the analysis could be further improved by visual inspection of the actual and predicted NMR spectrum of a molecular mixture, which can be automated as well. We believe that this method can enable high-throughput

 doi:10.20944/preprints202212.0278.v1



analysis of small molecule mixture compositions using $^1$H NMR as the primary or only spectroscopic tool.

**Author Contributions:** Conceptualization, A.S.R; methodology, A.S.R; software, A.S.R.; formal analysis, A.S.R.; resources, M.S.; data curation, A.S.R.; writing—original draft preparation, A.S.R.; writing—review and editing, A.S.R. and M.S.; visualization, A.S.R.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NSF grant number 2044599.

**Data Availability Statement:** The data used in this paper can be accessed via https://github.com/Signal-Science-Lab/Unsupervised_Molecular_Mixture_Analysis.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| WPT | Wavelet packet transform |
| DWT | Discrete wavelet transform |
| CWT | Continuous Wavelet transform |
| NMR | Nuclear magnetic resonance |

## References

1. Pellecchia, M.; Sem, D.S.; Wüthrich, K. NMR in drug discovery. *Nature Reviews Drug Discovery* **2002**, *1*, 211–219.
2. Shi, L.; Zhang, N. Applications of solution NMR in drug discovery. *Molecules* **2021**, *26*, 576.
3. Softley, C.A.; Bostock, M.J.; Popowicz, G.M.; Sattler, M. Paramagnetic NMR in drug discovery. *Journal of Biomolecular NMR* **2020**, *74*, 287–309.
4. Emwas, A.H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Gowda, G.N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; et al. NMR spectroscopy for metabolomics research. *Metabolites* **2019**, *9*, 123.
5. Markley, J.L.; Brüschweiler, R.; Edison, A.S.; Eghbalnia, H.R.; Powers, R.; Raftery, D.; Wishart, D.S. The future of NMR-based metabolomics. *Current opinion in biotechnology* **2017**, *43*, 34–40.
6. Wishart, D.S. NMR metabolomics: A look ahead. *Journal of Magnetic Resonance* **2019**, *306*, 155–161.
7. Pauli, G.F.; Jaki, B.U.; Lankin, D.C. Quantitative 1H NMR: Development and potential of a method for natural products analysis. *Journal of Natural Products* **2005**, *68*, 133–149.
8. Breton, R.C.; Reynolds, W.F. Using NMR to identify and characterize natural products. *Natural Product Reports* **2013**, *30*, 501–524.
9. Robinette, S.L.; Brüschweiler, R.; Schroeder, F.C.; Edison, A.S. NMR in metabolomics and natural products research: Two sides of the same coin. *Accounts of Chemical Research* **2012**, *45*, 288–297.
10. Capitani, D.; Sobolev, A.P.; Di Tullio, V.; Mannina, L.; Proietti, N. Portable NMR in food analysis. *Chemical and Biological Technologies in Agriculture* **2017**, *4*, 1–14.
11. Martínez-Yusta, A.; Goicoechea, E.; Guillén, M.D. A review of thermo-oxidative degradation of food lipids studied by 1H NMR spectroscopy: Influence of degradative conditions and food lipid nature. *Comprehensive Reviews in Food Science and Food Safety* **2014**, *13*, 838–859.
12. Whitfield Åslund, M.L.; McShane, H.; Simpson, M.J.; Simpson, A.J.; Whalen, J.K.; Hendershot, W.H.; Sunahara, G.I. Earthworm sublethal responses to titanium dioxide nanomaterial in soil detected by 1H NMR metabolomics. *Environmental Science & Technology* **2012**, *46*, 1111–1118.
13. Cardoza, L.; Korir, A.; Otto, W.; Wurrey, C.; Larive, C. Applications of NMR spectroscopy in environmental science. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2004**, *45*, 209–238.
14. Pauli, G.F.; Godecke, T.; Jaki, B.U.; Lankin, D.C. Quantitative 1H NMR. Development and potential of an analytical method: An update. *Journal of Natural Products* **2012**, *75*, 834–851.
15. Caligiani, A.; Acquotti, D.; Palla, G.; Bocchi, V. Identification and quantification of the main organic components of vinegars by high resolution 1H NMR spectroscopy. *Analytica Chimica Acta* **2007**, *585*, 110–119.
16. Barison, A.; Pereira da Silva, C.W.; Campos, F.R.; Simonelli, F.; Lenz, C.A.; Ferreira, A.G. A simple methodology for the determination of fatty acid composition in edible oils through 1H NMR spectroscopy. *Magnetic Resonance in Chemistry* **2010**, *48*, 642–650.
17. Lee, J.H.; Okuno, Y.; Cavagnero, S. Sensitivity enhancement in solution NMR: Emerging ideas and new frontiers. *Journal of Magnetic Resonance* **2014**, *241*, 18–31.

 doi:10.20944/preprints202212.0278.v1




18. Mompeán, M.; Sánchez-Donoso, R.M.; De La Hoz, A.; Saggiomo, V.; Velders, A.H.; Gomez, M. Pushing nuclear magnetic resonance sensitivity limits with microfluidics and photo-chemically induced dynamic nuclear polarization. *Nature Communications* **2018**, *9*, 1–8.
19. Kovacs, H.; Moskau, D.; Spraul, M. Cryogenically cooled probes—A leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005**, *46*, 131–155.
20. Zangger, K. Pure shift NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2015**, *86*, 1–20.
21. Foroozandeh, M.; Morris, G.A.; Nilsson, M. PSYCHE pure shift NMR spectroscopy. *Chemistry–A European Journal* **2018**, *24*, 13988–14000.
22. Aguilar, J.A.; Nilsson, M.; Morris, G.A. Simple proton spectra from complex spin systems: Pure shift NMR spectroscopy using BIRD. *Angewandte Chemie* **2011**, *123*, 9890–9891.
23. Lupulescu, A.; Olsen, G.L.; Frydman, L. Toward single-shot pure-shift solution 1H NMR by trains of BIRD-based homonuclear decoupling. *Journal of Magnetic Resonance* **2012**, *218*, 141–146.
24. Castañar, L.; Parella, T. Broadband 1H homodecoupled NMR experiments: Recent developments, methods and applications. *Magnetic Resonance in Chemistry* **2015**, *53*, 399–426.
25. Giraudeau, P. Challenges and perspectives in quantitative NMR. *Magnetic Resonance in Chemistry* **2017**, *55*, 61–69.
26. Delsuc, M.A.; Levy, G.C. The application of maximum entropy processing to the deconvolution of coupling patterns in NMR. *Journal of Magnetic Resonance (1969)* **1988**, *76*, 306–315.
27. Shimba, N.; Stern, A.S.; Craik, C.S.; Hoch, J.C.; Dötsch, V. Elimination of 13C$\alpha$ splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. *Journal of the American Chemical Society* **2003**, *125*, 2382–2383.
28. Cui, Q.; Lewis, I.A.; Hegeman, A.D.; Anderson, M.E.; Li, J.; Schulte, C.F.; Westler, W.M.; Eghbalnia, H.R.; Sussman, M.R.; Markley, J.L. Metabolite identification via the madison metabolomics consortium database. *Nature Biotechnology* **2008**, *26*, 162–164.
29. Steinbeck, C.; Krause, S.; Kuhn, S. NMRShiftDB–Constructing a free chemical information system with open-source components. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 1733–1739.
30. Tulpan, D.; Léger, S.; Belliveau, L.; Culf, A.; Čuperlović-Culf, M. MetaboHunter: An automatic approach for identification of metabolites from 1H-NMR spectra of complex mixtures. *BMC Bioinformatics* **2011**, *12*, 1–22.
31. Vu, T.N.; Laukens, K. Getting your peaks in line: A review of alignment methods for NMR spectral data. *Metabolites* **2013**, *3*, 259–276.
32. Lepre, C.A. Library design for NMR-based screening. *Drug Discovery Today* **2001**, *6*, 133–140.
33. dos Santos Ribeiro, H.S.; Dagnino, D.; Schripsema, J. Rapid and accurate verification of drug identity, purity and quality by 1H-NMR using similarity calculations and differential NMR. *Journal of Pharmaceutical and Biomedical Analysis* **2021**, *199*, 114040.
34. Wei, S.; Zhang, J.; Liu, L.; Ye, T.; Gowda, G.N.; Tayyari, F.; Raftery, D. Ratio analysis nuclear magnetic resonance spectroscopy for selective metabolite identification in complex samples. *Analytical chemistry* **2011**, *83*, 7616–7623.
35. Wei, W.; Liao, Y.; Wang, Y.; Wang, S.; Du, W.; Lu, H.; Kong, B.; Yang, H.; Zhang, Z. Deep Learning-Based Method for Compound Identification in NMR Spectra of Mixtures. *Molecules* **2022**, *27*, 3653.
36. Pomyen, Y.; Wanichthanarak, K.; Poungsombat, P.; Fahrmann, J.; Grapov, D.; Khoomrung, S. Deep metabolome: Applications of deep learning in metabolomics. *Computational and Structural Biotechnology Journal* **2020**, *18*, 2818–2825.
37. Corsaro, C.; Vasi, S.; Neri, F.; Mezzasalma, A.M.; Neri, G.; Fazio, E. NMR in Metabolomics: From Conventional Statistics to Machine Learning and Neural Network Approaches. *Applied Sciences* **2022**, *12*, 2824.
38. Sinha Roy, A.; Srivastava, M. Analysis of Small-Molecule Mixtures by Super-Resolved 1H NMR Spectroscopy. *The Journal of Physical Chemistry A* **2022**.
39. Wishart, D.S.; Knox, C.; Guo, A.C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D.D.; Psychogios, N.; Dong, E.; Bouatra, S.; et al. HMDB: A knowledgebase for the human metabolome. *Nucleic Acids Research* **2009**, *37*, D603–D610.
40. Ray, S. A quick review of machine learning algorithms. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, 2019, pp. 35–39.