Curve Resolution by Sample Subtraction

Abstract

Blind Source Separation and Curve Resolution methods have been used increasingly in Spectroscopy and Process Analytical Techniques to extract, from spectral samples, estimates of the spectral signatures of the underlying components in a material. Spectral distortion, caused by sensor de-calibration or by external influences, is an important factor for which current component extraction techniques are unable to cope with. In this paper, a technique is proposed that is able to estimate the spectral components from a distorted spectral data set. The capabilities of the proposed approach are illustrated using simulated and laboratory data.

1 Introduction

Analysis of spectroscopic data has been increasingly used in the Pharmaceutical Industry, specifically in the area of Process Analytical Technology (PAT) [1, 2]. Szostak et al.[3], for example, demonstrated how Raman spectral measurements could be used to identify the compounds in a pharmaceutical tablet. A further use of this approach used spectral measurements to identify counterfeit medicines [4]. Another important motivation for spectral signature extraction is their use in the estimation of component concentrations [5].

There is a wide array of techniques commonly used for component extraction. Many of these techniques fall under the umbrella of Blind Source Separation (BSS) methods [6] or Self-Modelling Curve Resolution methods (SMCR) [7]. Of the many BSS/SMCR techniques available, for the analysis of spectral data, Principal Component Analysis (PCA), Alternating Least Squares (ALS), Non-Negative Matrix Factorization (NNMF), and Independent Component Analysis (ICA) have proved popular.

Principal Component Analysis (PCA) is able to reduce the dimensionality of a problem by identifying directions of greatest variation in the data, with the imposed constraint that each of the directions are orthogonal. Each of these directions can be considered a 'component' of the data, and the small-variance directions that are identified are typically considered to be 'noise' [8]. However, because only limited constraints are applied with this algorithm, the components that it identifies may have little physical meaning [9].

Non-negative Matrix Factorisation (NMF) assumes that all components that are identified have no negative parts, which can make their physical representations easier to render [9]. However, it has been found that in certain circumstances it is difficult for the algorithm to converge to a global optimum, or even to converge at all [10].

Alternating Least Squares (ALS) extracts the components from a data set by exploiting the assumption of a linear combination, however it is required that the number of components to be retrieved is known *a-priori*. Although this can be attempted in a variety of ways [7], the presence of noise and other factors can make this procedure non-trivial [7]. However, with ALS this issue can be circumvented by using an increasing amount of components until convergence is reached [11].

Independent Component Analysis (ICA) is an alternative approach to extracting meaningful components from a data set, assuming that all sources are independent of one another [12]. It provides an objective function that, when optimized, identifies components that have real physical meaning [13]. This technique has been used extensively in many different areas of science and engineering and its popularity is increasing [14, 15].

Much of the research conducted into using these BSS and SMCR methods to analyze spectral measurements has concentrated on their use in an off-line capacity. However, the ability to analyze spectral data in real-time offers important benefits to industrial automation. In particular, the ability to analyze spectral measurements in real-time provides the possibility of using such measurements in a feedback control system or for real-time release of product. Unfortunately, analyzing spectra in real-time introduces a number of complex challenges, such as coping with measurements that may be subject to distortion from variations of the parameters in the surrounding environment (e.g. temperature, pressure, humidity, etc.).

Spectral instrumentation requires precise and complex calibration methods, if it is to obtain spectra that can be used in spectral analysis [16–20]. Continuous calibration of spectral sensors is necessary as spectral data acquisition methods, such as Raman, Mass and Infrared, are sensitive to external influences. Relevant external influences include temperature changes [21–29], pressure changes [25, 27], presence of foreign materials [16, 18] and hidden factors in the material [17].

It can be argued that calibrating a sensor inside a sterile laboratory is all that is necessary to overcome the problem of spectral distortion. Unfortunately, this approach is not ideal, as firstly it is not always obvious when a sensor requires re-calibration and secondly, re-calibration procedures can involve the use of expensive reference materials, as well as the loss of revenue because of the need to stop the plant to calibrate the sensor [30]. This means that calibration is, at best, only a temporary solution.

Furthermore, it has been observed that specific types of sensors respond non-linearly to different materials as well as to interactions between different compounds inside the sample [31], meaning that a single calibration may not resolve the problem. There is also a growing need to use different laboratories for sampling similar material, and for this a complex and costly calibration method, called "Calibration Transfer", is required [26, 32]. The problem is further complicated by the fact that using the same equipment, in the same laboratory, with the same sampled material, has been shown to produce inconsistent results [33]. In addition, the advent of spectral analysis in real-time quality monitoring [34] highlights the need for aligned spectra in sensors located inside a plant in which conditions are time-varying. All of the described factors illustrate the difficulty of satisfying the fundamental assumption that the spectra being analyzed are perfectly aligned.

To solve this problem it is firstly necessary to understand how external influences affect spectral measurements and to then compensate for these influences accordingly. The manner in which a spectrum shape is distorted depends on the type of sensor being used, the material being inspected, and the external influence involved. For example, the influence of temperature on spectral measurements has been studied extensively, and its effects on spectral data have ranged from Spectral Shift [31, 35], to Spectral Warp [29, 36], to a non-linear distortion that could only be modelled by a second-order system [24].

With knowledge of the relationship between any external influence and the manner by which the specified measurement is distorted, it is then possible to remove the distortion. For example, the components extracted by ICA from spectral data suffering from Spectral Shift were able to be corrected by a post-processing method [37]. However, a high investment of time and energy would be required to fix current component extraction methods for every type of distortion that may be encountered.

In this paper, a new component extraction method is proposed that is able to analyze spectra suffering from various types of distortion. Its main objective is to estimate the spectral signatures of the components inside a set of distorted spectra, whilst being able to be easily modified, according to the distortion suspected of occurring. In the following section, the proposed methodology is described. In Section 3, a simulation study is defined and a comparison with ALS is provided. The proposed method is also applied to a public mass spectral data set and a spectral ice analog set; both suffer from distortions. Finally, discussions and conclusions from both studies are presented.

2 Proposed Technique

The proposed algorithm aims to extract the underlying components in a spectral data set by subtracting scaled versions of the samples from the remainder of the set, such that the samples are left with only one unique component. If the number of samples being processed is equal to the number of components to be retrieved, then the result of the algorithm would be that the features appearing in one sample would be those of only one component.

To illustrate the subtraction process, an example is given where no distortion is present. Two simple components, shown in Figure 1, were used to create the two simulated sampled spectra shown in Figure 2.

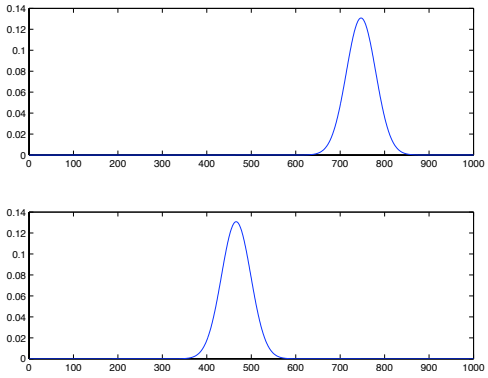


Figure 1: Reference spectra used in example.

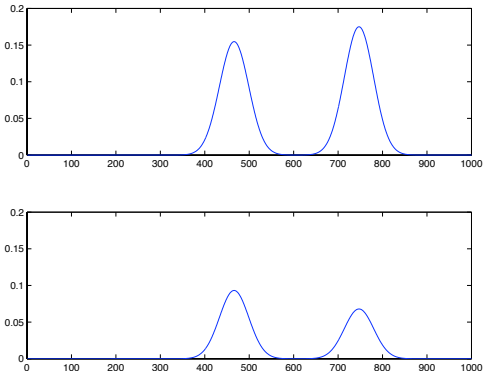


Figure 2: Two simulated spectral samples created using reference from Figure 1.

Sample 1, shown in the upper graph of Figure 2, is scaled such that the maximum amplitude is equal to the amplitude at the same location in Sample 2. In the upper graph of Figure 3, the dashed line is the scaled version of Sample 1, trended over Sample 2. Subtracting the scaled Sample 1 from Sample 2 results in the spectrum shown in the lower graph of Figure 3. The samples in this example did not suffer from any distortion, therefore the second peak has been eliminated completely from Sample 2, whilst the remaining features of the spectra persist.

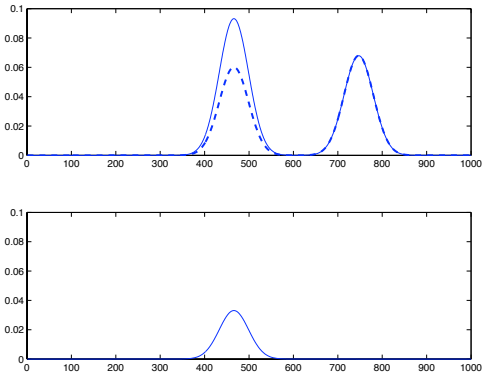


Figure 3: Result of subtracting scaled Sample 1 from Sample 2 without shift.

If this process is then carried out in reverse, scaling Sample 2 with Sample 1 and subtracting, both samples become the unique components, as shown in Figure 4.

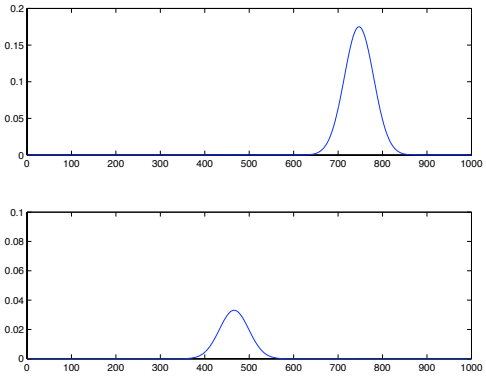


Figure 4: Samples after subtracting scaled Sample 2 from Sample 1 without shift.

Comparison of the shapes of the resulting subtracted components to the reference spectra in Figure 1 suggests that they are very similar to the sources. It is important to note that, because of the subtraction process, the amplitudes in the estimated components are not equal to the ones originally in the samples. However, scaling them back is trivial, as the original amplitudes of each peak can be carried forward. Algorithm 1 summarizes the method when no distortion is present in the spectra.

Algorithm 1 Simple Blind Source Separation Algorithm.

repeat
  for all $\hat{S} \Rightarrow \hat{S}_a$ **do**
    $\hat{S}_{a_m} = \mathbf{max}(\hat{S}_a, return\ index)$
    for all $\hat{S} \neq \hat{S}_a \Rightarrow \hat{S}_b$ **do**
      $\hat{S}_b \leftarrow \hat{S}_b - (\hat{S}_a * \hat{S}_b[\hat{S}_{a_m}]/\hat{S}_a[\hat{S}_{a_m}])$
    end for
  end for
until convergence

where $\hat{S}$ is a subgroup of the data set S which has k samples, equal to the amount of components to be retrieved. Convergence is reached when the Mean Square Error is below a pre-specified tolerance, which should be specified to avoid overshooting. For a set of four spectra, each of length 1500 points, a tolerance of 0.001 gave the best results in this study.

To illustrate how the proposed approach is able to be used when spectral distortion is present, Spectral Shift is introduced to each component, as shown in Figure 1 is introduced. When spectral samples are distorted, two issues arise. First, the peak in which the maximum value of Sample 1 lies needs to be re-scaled and aligned to the one in Sample 2 such that it is completely eliminated from Sample 2, as shown in Figure 5. The lower graph shows the result of appropriately aligning and scaling Sample 1 such that the second peak in Sample 2 is eliminated.

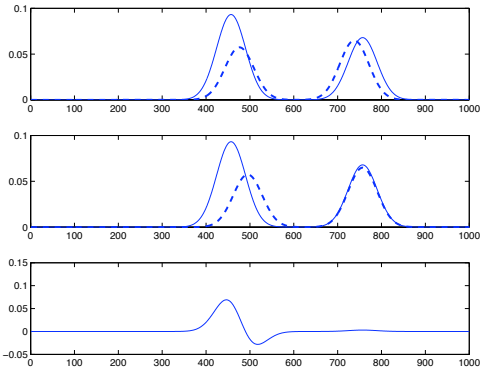


Figure 5: Result of subtracting scaled Sample 1 from Sample 2 with shift.

The second issue can be observed in the lower graph of Figure 5, where the shape of the remaining peak is considerably distorted. This can be overcome by 'adding back' to Sample 2 only 'certain parts' of the re-scaled, aligned Sample 1. In the lower graph of Figure 6, the remaining peak in Sample 2 is restored by adding back the dashed line shown in the upper graph, which is an appropriately chosen section from Sample 1.

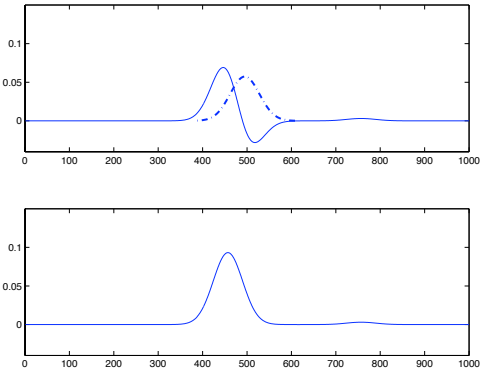


Figure 6: Restoring remaining features of Sample 2 by partially adding back Sample 1.

The alignment phase and the restoring phase can be incorporated into the definition of Algorithm 1 to obtain the final version of the proposed approach, summarized in Algorithm 2. It is assumed that a subset of randomly chosen samples $\hat{S}$ has been defined.

Algorithm 2 BSS by Substraction.

repeat
  for all $\hat{S} \Rightarrow \hat{S}_a$ **do**
    for all $\hat{S} \neq \hat{S}_a \Rightarrow \hat{S}_b$ **do**
      $\hat{S}_{a_{distorted}} = \textbf{find\_best\_alignment}(\hat{S}_a, \hat{S}_b)$
      $s_{a_m} = \textbf{max}(\hat{S}_{a_{distorted}}, return\ index)$
      $N = \hat{S}_b[s_{a_m}] / \hat{S}_{a_{distorted}}[s_{a_m}]$
      $\hat{S}_b \leftarrow \hat{S}_b - (\hat{S}_{a_{distorted}} * N)$
      $I = \textbf{find\_indexes\_to\_repair}(\hat{S}_b)$
      $\hat{S}_b[I] \leftarrow \hat{S}_b[I] + (\hat{S}_{a_{distorted}}[I] * N)$
    end for
  end for
until convergence

The **find_best_alignment** function finds the best way to temporarily distort $\hat{S}_a$ such that its maximum value best aligns with the features that are near it in $\hat{S}_b$. N is the value which normalizes the distorted $\hat{S}_a$ such that its maximum value is equal to the corresponding location in $\hat{S}_b$. The **find_indexes_to_repair** function locates the indexes I of $\hat{S}_b$ that need repairing, and, hence, defines the sections of $\hat{S}_{a_{distorted}}$ that need to be added back to $\hat{S}_b$. An explanation of both of these functions, as well as several observations made from the implementation, are now discussed.

Finding Best Alignment. The process of finding the best distorted alignment involves estimating the amount of global distortion needed to be artificially applied to $\hat{S}_a$ such that the area near its maximum value when being subtracted from $\hat{S}_b$ is close to zero. In other words, $\hat{S}_a$ is artificially distorted and subtracted from $\hat{S}_b$ to create a temporary spectrum D, from which the values near the maximum value of $\hat{S}_a$ are gathered. This is summarized in (1).

$$D = \hat{S}_b - \textbf{distort\_signal}(\hat{S}_a, distort\_measures) \tag{1}$$

An optimization algorithm, as well as a brute-force approach, can be applied to find the optimal set of *distort_measures*. However, it was found that when dealing with two or more spectral distortions at the same time, a brute-force approach was highly inefficient. The use of an optimization algorithm, such as Particle Swarm Optimization [38, 39] provided faster results.

Depending on the type of distortions, the objective function used by the optimization algorithms may differ. For example, when observing Spectral Shift, the objective function in (2) sufficed.

$$M = -\textbf{abs}(D[s_{a_m}]) \tag{2}$$

where

$$s_{a_m} = \textbf{max}(\hat{S}_{a_{shifted}}, return\ index) \tag{3}$$

and

$$\hat{S}_{a_{shifted}} = \textbf{distort\_signal}(\hat{S}_a, shift). \tag{4}$$

The **max** function, with the *return index* flag, returns the index of the maximum value of a spectrum, and M is the measure of optimality.

When dealing with a Spectral Shift coupled with a Spectral Warp, the objective function in (5) provided accurate results.

$$M = -\sum_{i=s_{a_m}-range}^{s_{a_m}+range} D[i]^2 * i \tag{5}$$

where *range* is a predefined number of points to the left and right of s_{a_m}, the location of the maximum value of $\hat{S}_{a_{shifted\_and\_warped}}$ defined by (6).

$$\hat{S}_{a_{shifted\_and\_warped}} = \textbf{distort\_signal}(\hat{S}_a, shift, warp) \tag{6}$$

Each value in D is weighted such that the values to the right side of s_{a_m} are given more weight. The reason for this is that, because of the nature of the warp distortion, the differences between the features are more predominant and more easily identifiable to the right side of the maximum value.

It is important to note that the objective function needs to be defined by using appropriate knowledge of the distortion taking place. However, being able to consider several distortions at the same time is an important quality of the proposed algorithm, as knowledge from a plant expert can be incorporated in a relatively easy manner.

Finding Locations to Repair. The process of finding which frequency locations require modification after the subtraction has taken place is relatively simple. It was found that if the value of a frequency location was outside a predefined range, then it was necessary for it to be repaired. This range does need to be tuned for the specific spectral signals being used, but, in this study, a value of 1% of the full frequency spread was found to be suitable for both limits (one negative and the other positive). If the peaks have sharp edges, having a negative limit close to zero and increasing the positive limit were found to smoothen the peaks. However, overshooting may result in slow convergence as well as thin features being created.

Post-Filtering. If there is a small amount of overlap or noise in the spectra, all the locations to be repaired may be difficult to identify. Therefore, the samples will have an increasing amount of small features, as more locations are not identified. However, because of the small nature of these features, they can be filtered out relatively easy. A moving average window filter with a window of length of 1% of the range of the spectrum was used in this work to remove the small features. This filter was found to remove a large quantity of these features, without significantly affecting the structure of the identified component features.

Number of Samples to Use. The number of samples in the subset $\hat{S}$ must be the same as the number of components to be extracted, k. Although k can be estimated using several techniques, these have been shown to be fragile against spectral distortion [37].

However, it was found that the number of components can be estimated if the proposed algorithm is applied repeatedly, with an increasing amount of samples every time, until the extracted components begin repeating themselves.

Because the algorithm only requires as much samples as there are components, it may be possible that a sample with a heavy distortion or noise may be used. Although this is not an issue in cases where there are more than one component, as the distortion/noise is 'absorbed' by the subtracting process in the other components, it becomes an important factor when dealing with only one component. To this regard, a way to circumvent this issue that has provided very good results (Section 4 provides a good example of this) is to run the algorithm with all the possible pairings between samples in the set, and average the most similar ones, which is a good estimation of the component.

3 Results with Simulated Data

To demonstrate the capabilities of the proposed method, artificial data sets that simulate the mixtures of four components are used. The reference spectra for these four components were randomly generated and are shown in Figure 7.

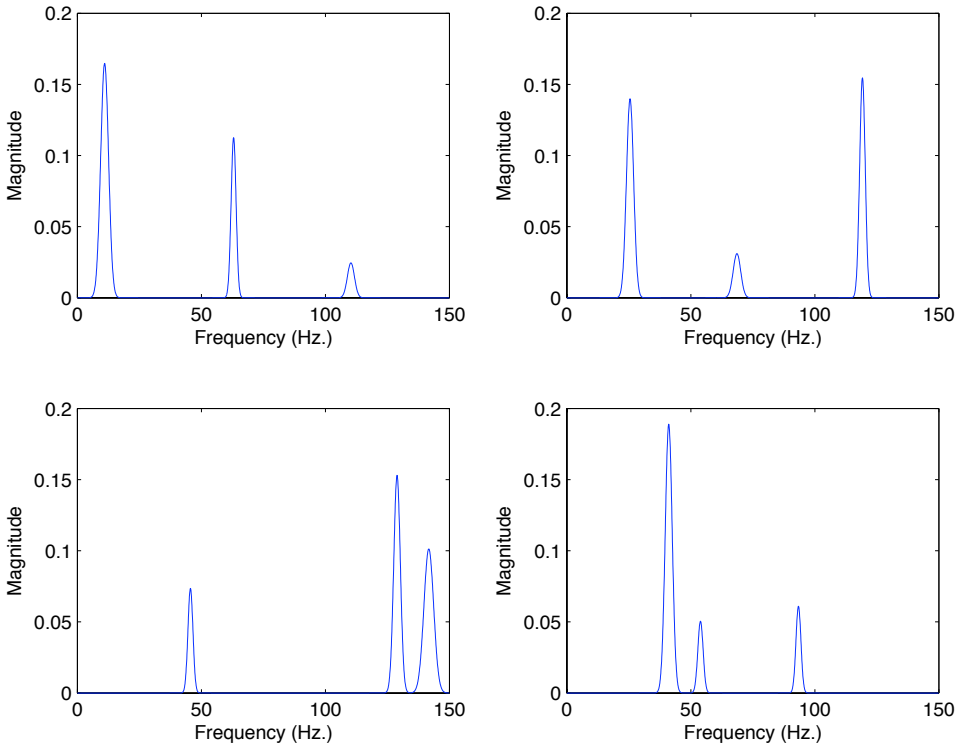


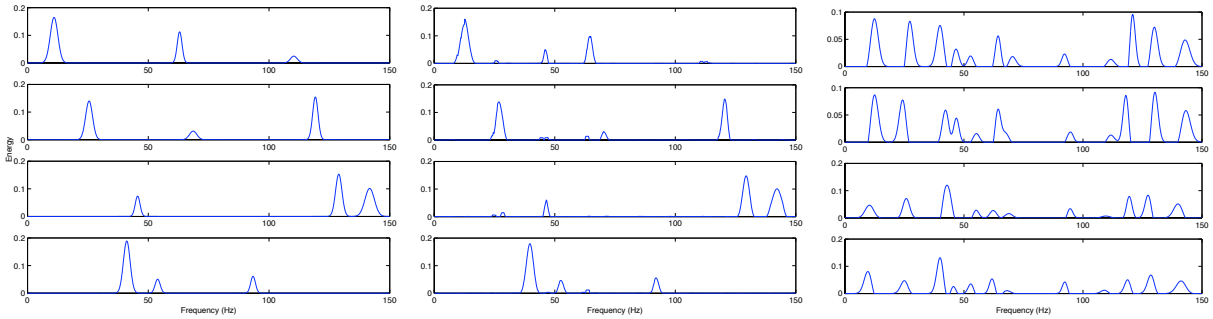
Figure 7: Reference spectra randomly generated for use in experiments.

The domain of the spectra is in Hertz and their resolution is 0.1 Hz per frequency point. The structure of these spectra were defined such that they were consistent with data observed in the Pharmaceutical and Biomedical Industry [40], as well as Mass Spectrometry[35], and other fields [41, 42].

Two 100-samples data sets were created. Each sample contained a spectrum of a simulated mixture of the four reference spectra. Before being 'mixed', each spectrum was scaled by a factor randomly chosen between 0.5 and 1.5, simulating its concentration.

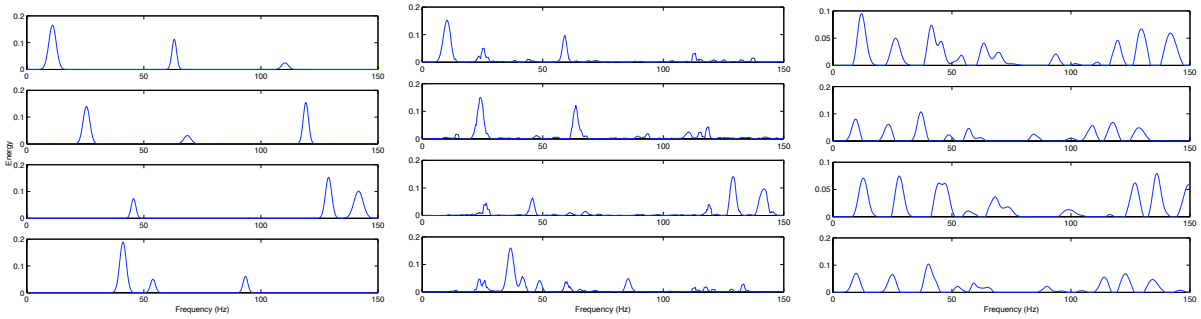
To simulate the effects of external influences, two distortions were chosen: Spectral Shift, that displaced each component by a different amount, and Spectral Warp, that stretched or compressed the spectrum by a defined percentage. These two distortions were decided upon because of their frequent occurrence in literature. One data set suffered from Spectral Shift within a range of ± 2 Hz, and the other suffered from Spectral Shift as well as a Spectral Warp that ranged between 95% to 1.05%.

Alternating Least Squares and the proposed technique, BSS by Subtraction (SubBSS), were applied to both data sets, both assuming that the correct number of components was known *a-priori*. Four randomly chosen samples were used by BSS by Subtraction, whilst the whole data set was used for ALS. The results for the first data set are shown in Figure 8 and for the second data set in Figure 9.



(a) Benchmark used. (b) Components extracted by Sub-BSS (c) Components extracted by ALS

Figure 8: Components extracted from data set suffering from local shift.



(a) Benchmark used. (b) Components extracted by Sub-BSS (c) Components extracted by ALS

Figure 9: Components extracted from data set suffering from local shift and global warp.

Figures 8 and 9 show that Alternating Least Squares was not able to extract the correct components. In fact, in both cases, ALS reached 2000 iterations without convergence. Both Figures show that the BSS by Subtraction technique was able to estimate

four components similar to the benchmark. As can be observed, when two distortions occur at the same time, as is the case of Figure 9, the accuracy of the technique is reduced. However, the extracted components are still similar to the actual source components and significantly better than those extracted by ALS.

4 Case Study: SELDI-TOF Samples

The data set that was used in this study was a publicly available data set consisting of 11 mass spectrometry samples that were measured from a study examining individuals with multiple sclerosis [35]. The samples used were those of healthy individuals, taken as a reference for quality control. It was found that, even though the spectra should be consistent, as can be seen in Figure 10, there were significant deviations in the m/z axis, the most relevant of which is Spectral Shift. Jeffries states the reason for this as being the considerable amount of calendar time between the processing of samples (even several weeks), which makes the procedure highly frail against machine drift, even with 'identical calibration procedures, personnel, equipment and sample handling techniques' [35].

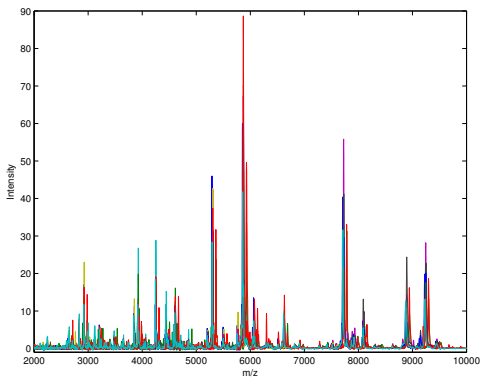
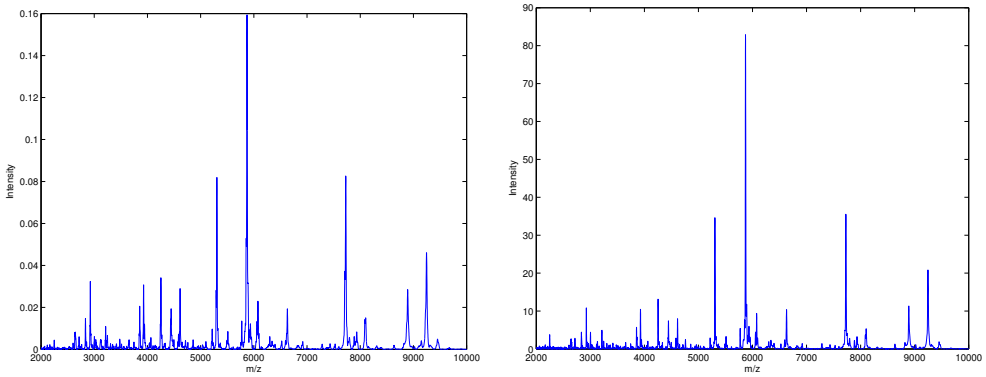


Figure 10: SELDI-TOF proteomic spectra from serum.

Using rank analysis techniques, such as Singular Value Decomposition, it was concluded that there was no redundancy in the data, i.e. it was of full rank. However, by applying the BSS-subtraction method to this data repeatedly, using an increasing number of samples, it was concluded that there was a single component in the spectra. This component is shown in Figure 11a.

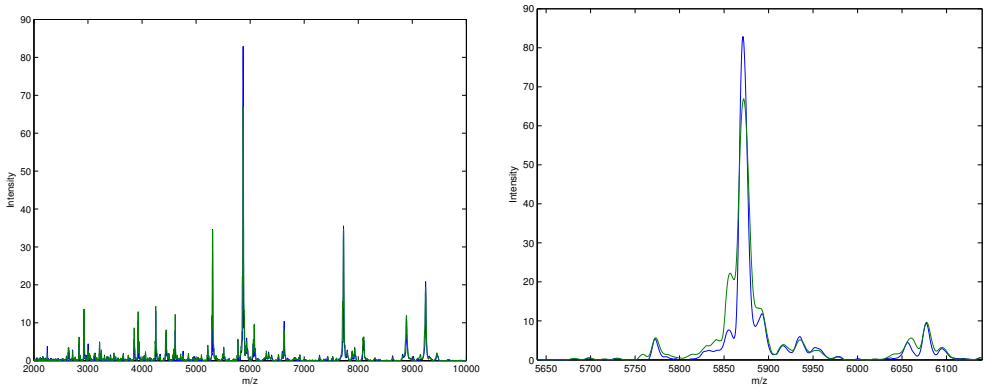
Figure 11b shows the spectral sample ultimately used in the study as reference; it required re-sampling and further alignment to create [35]. Comparison of Figures 11a and 11b illustrate that the proposed algorithm has successfully identified the reference spectrum of the active component.



(a) Component obtained by SubBSS. (b) Reference spectrum.

Figure 11: Recovered component and its reference spectrum.

In Figure 12a, it can be seen that the component[1] and the reference are aligned as the result of the proposed algorithm. In Figure 12b, a zoomed in area of the composite plot (between 5640 and 6140 m/z) indicates that the alignment is nearly exact.



(a) Component and reference overlayed. (b) Zoomed in version of Figure 12a.

Figure 12: Recovered component and its reference spectrum.

5 Case Study: Ice Analogs

The data set that was used in this study was a publicly available data set consisting of 9 NIR samples that were measured from samples of carbonized ice (H_2O+CO). These measurements were initially measured in an experiment to observe the effect that heat had on the NIR spectrum [43]. The two most relevant variations that were recorded are highlighted in Figure 13. It shows the 9 NIR spectra and highlights, with arrows, the effect that the increase in heat has on the spectra. All the spectra shown in this case study are in cm^{-1}.

[1]The component was scaled for visualization purposes.

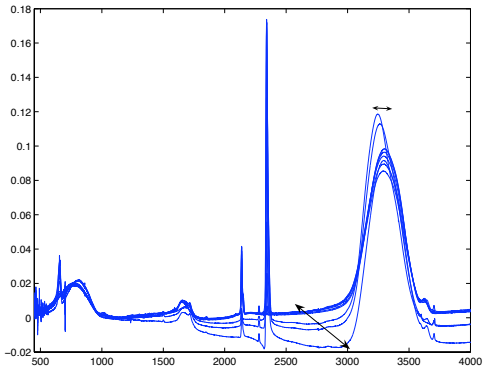


Figure 13: Spectral data of ice UV-radiated in different amounts.

As in the SELDI-TOF case study, using rank analysis to estimate the number of components did not provide sensible results, making the data set full rank. Again, the BSS-subtraction method was applied repeatedly, using an increasing number of samples, until the extracted components began repeating themselves. It was concluded that there were two components in the spectra, which are shown in Figure 14.

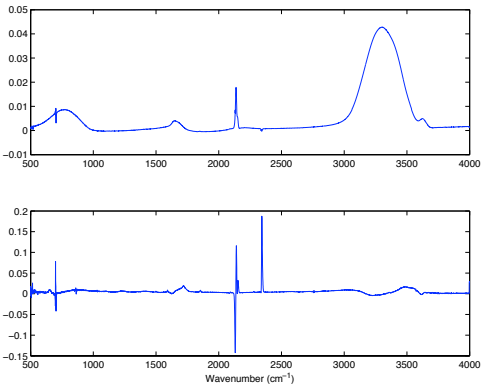
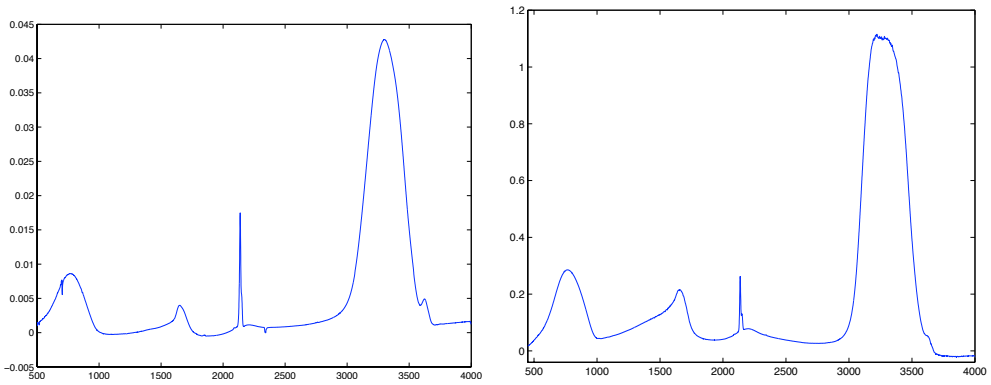


Figure 14: Components extracted using SubBSS.

Figure 15b provides the reference spectrum for carbonised ice obtained from The Cosmic Ice Laboratory at NASA [44]. Comparison of Figures 15a and 15b illustrate that the proposed algorithm has successfully identified the reference spectrum of one component. As for the second component, an important peak at 2340 cm^{-1} is found, which is close to what spectroscopy literature [45] states as the probable *presence of a type of carbonyl* (CO or CO_2), which is expected.



(a) First component obtained by SubBSS. (b) Reference spectrum from NASA.

Figure 15: Recovered component and its reference spectrum.

6 Discussion and Conclusions

This paper has proposed a novel technique for extracting components from spectral measurements, that have been distorted by external influences. The proposed method is simple in form and intuitive, which should help with acceptance in industry. The method can also cope with many different types of distortion and can be modified with relative ease to cope with an array of distortions that may be suspected of being present.

The algorithm showed great potential for resolving the problem of component extraction from distorted spectra with both a simulated data set and data distorted from practical situations. It was shown that the algorithm was able to cope with two different types of distortions taking place at the same time, and the fact that more distortions can be considered bodes well for its generalized application. This algorithm, as shown in the Ice Analogs case study, is well suited for use on composite spectra (the estimation process is straightforward and robust against noise). However, it also showed a good ability to obtain an aligned reference from a data set of distorted spectra due to machine drift, as shown in the SELDI-TOF case study.

It is important to note that the proposed method assumes that the components to be extracted are uncorrelated. If there is correlation, then shared features will appear in only one of the components, and incorrect components will be identified.

References

[1] W. W. Blaser, R. A. Bredeweg, R. S. Harner, M. A. LaPack, A. Leugers, D. P. Martin, R. J. Pell, J. Workman, and L. G. Wright, "Process analytical chemistry," *Analytical Chemistry*, vol. 67, pp. 47–70, June 1995.

[2] T. Togkalidou, M. Fujiwara, S. Patel, and R. D. Braatz, "Solute concentration prediction using chemometrics and atr-ftir spectroscopy," *Journal of Crystal Growth*, vol. 231, pp. 534–543, November 2001.

[3] R. Szostak and S. Mazurek, "Quantitative determination of acetylsalicylic acid and

acetaminophen in tablets by ft-raman spectroscopy," *The Analyst*, vol. 127, pp. 144–148, 2002.

[4] M. de Veij, A. Deneckere, P. Vandenabeele, D. de Kaste, and L. Moens, "Detection of counterfeit viagra with raman spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 46, pp. 303–309, 2008.

[5] J. W. Childers, W. J. Phillips, E. L. Thompson, D. B. Harris, D. A. Kirchgessner, D. F. Natschke, and M. Clayton, "Comparison of an innovative nonlinear algorithm to classical least-squares for analyzing open-path fourier transform infrared spectra collected at a concentrated swine production facility," *Applied Spectroscopy*, vol. 56, pp. 325–336, March 2002.

[6] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, May-June 2000.

[7] J.-H. Jiang, Y. Liang, and Y. Ozaki, "Principles and methodologies in self-modeling curve resolution," *Chemometrics and Intelligent Laboratory Systems*, vol. 71, pp. 1–12, April 2004.

[8] M. Ringnér, "What is principal component analysis?," *Nature Biotechnology*, vol. 26, no. 3, pp. 303–304, 2008.

[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization.," *Nature*, vol. 401, pp. 788–791, October 1999.

[10] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization.," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[11] E. J. Karjalainen, "The spectrum reconstruction problem : Use of alternating regression for unexpected spectral components in two-dimensional spectroscopies," *Chemometrics and Intelligent Laboratory Systems*, vol. 7, no. 1-2, pp. 31 – 38, 1989.

[12] C. Jutten and J. Herault, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, July 1991.

[13] M. Bressan, D. Guillamet, and J. Vitria, "Using an ica representation of high dimensional data for object recognition and classification," *IEEE Conference on Computer Vision & Pattern Recognition*, vol. 1, pp. 1063–6919, 2001.

[14] 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, October 2003.

[15] H. Malika, A. Khokhar, and R. Ansari, "Improved watermark detection for spread-spectrum based watermarking using independent component analysis," *Proceedings of the 5th ACM workshop on Digital rights management*, pp. 102–111, 2005.

[16] V. Kachel, O. Kempski, J. Peters, and F. Schödel, "A method for calibration of flow cytometric wavelength shift fluorescence measurements," *Cytometry*, vol. 11, pp. 913–915, 1990.

[17] Y. Wang and M. Gu, "Application of comprehensive calibration to mass spectral peak analysis and molecular screening." United States Patent 7451052, November 2008.

[18] T. I. Ivanov, M. O. Vieitez, C. A. de Lange, and W. Ubachs, "Frequency calibration of b 1σ+u–x 1σ+g (6,0) lyman transitions in h2 for comparison with quasar data," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 41, no. 3, p. 035702 (4pp), 2008.

[19] Z. Ye, G. Auner, and P. Manda, "Raman spectra calibration, extraction and neural network based training for sample identification," *Proceedings of the International Joint Conference on Neural Networks*, vol. 1, pp. 622–626, 2003.

[20] J. T. M. Pearce, T. J. Athersuch, T. M. D. Ebbels, J. C. Lindon, J. K. Nicholson, and H. C. Keun, "Robust algorithms for automated chemical shift calibration of 1d 1h nmr spectra of blood serum," *Analytical Chemistry*, vol. 80, pp. 7158–7162, August 2008.

[21] M. A. Czarnecki, Y. Ozaki, M. Suzuki, and M. Iwahashi, "Quantitative study of the dissociation of dimeric cis-9, cis-12-octadecadienoic acid in pure liquid by the ft-ir liquid film technique," *Applied Spectroscopy*, vol. 47, no. 12, pp. 2157–2161, 1993.

[22] K. H. Hazen, M. A. Arnold, and G. W. Small, "Temperature-insensitive near-infrared spectroscopic measurement of glucose in aqueous solutions," *Applied Spectroscopy*, vol. 48, no. 4, pp. 477–483, 1994.

[23] T. Iwata, J. Koshoubu, C. Jin, and Y. Okubo, "Temperature dependence of the mid-infrared oh spectral band in liquid water," *Applied Spectroscopy*, vol. 51, no. 9, pp. 1269–1275, 1997.

[24] Z.-P. Chen, J. Morris, and E. Martin, "Correction of temperature-induced spectral variations by loading space standardization," *Analytical Chemistry*, vol. 77, pp. 1376 – 1384, March 2005.

[25] Z.-P. Chen and J. Morris, "Improving the linearity of spectroscopic data subjected to fluctuations in external variables by the extended loading space standardization," *The Analyst*, vol. 133, pp. 914–922, 2008.

[26] J. Lin, "Near-ir calibration transfer between different temperatures," *Applied Spectroscopy*, vol. 52, no. 12, pp. 1591–1596, 1998.

[27] A. Inoue, K. Kojima, Y. Taniguchi, and K. Suzuki, "Effects of temperature and pressure on the near-infrared spectra of hod in d20," *Journal of Solution Chemistry*, vol. 16, no. 9, pp. 727–734, 1987.

[28] Y. Ghebremeskel, J. Fields, and A. Carton, "The use of near infrared (nir) spectroscopy to study specific interactions in polymer blends," *Journal of Polymer Science*, vol. 32, pp. 383–386, 1994.

[29] S. L. Eix, S. A. Schlueter, and A. Anderson, "Raman and infrared spectra of solid dibromodifluoromethane," *Journal of Raman Spectroscopy*, vol. 23, pp. 495–499, 1992.

[30] M. Zuppa, C. Distante, K. C. Persaud, and P. Siciliano, "Recovery of drifting sensor responses by means of dwt analysis," *Sensors & Actuators: B. Chemical*, vol. 120, pp. 411–416, January 2007.

[31] L. Cetto, "Alignment of mass spectromotry data." United States Patent: 7365311, April 2008.

[32] D. Ozdemir and R. Williams, "Multi-instrument calibration with genetic regression in uv-visible spectroscopy," *Applied Spectroscopy*, vol. 53, no. 2, pp. 210–217, 1999.

[33] K. A. Baggerly, J. S. Morris, and K. R. Coombes, "Reproducibility of seldi-tof protein patterns in serum: comparing datasets from different experiments.," *Bioinformatics*, vol. 20, pp. 777–785, Mar 2004.

[34] C. Rascon, B. Lennox, and O. Marjanovic, "Using lagged spectral data in feedback control using particle swarm optimisation," *Proceedings of the UKACC Control 2008 Conference*, September 2008.

[35] N. Jeffries, "Algorithms for alignment of mass spectrometry proteomic data," *Bioinformatics*, vol. 21, pp. 3066–3073, May 2005.

[36] H. Hoshino, S. Tajima, and T. Tuschiya, "The effect of the temperature on the mass spectra of aliphatic primary alcohols and 1-alekenes. i.," *Bulletin of the Chemical Society of Japan*, vol. 46, pp. 3043–3048, 1973.

[37] C. Rascon, B. Lennox, and O. Marjanovic, "Recovering independent components from shifted data using fastica and swarm intelligence," *Applied Spectroscopy*, vol. 63, October 2009.

[38] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceeding of the IEEE International Conference on Neural Networks*, vol. IV, pp. 1942–1948, 1995.

[39] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *Evolutionary Computation, IEEE Transactions on*, vol. 6, pp. 58–73, Feb 2002.

[40] W. P. Findlay and D. E. Bugay, "Utilization of fourier transform-raman spectroscopy for the study of pharmaceutical crystal forms," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 16, pp. 921–930, February 1998.

[41] J. C. Brown, "Musical fundamental frequency tracking using a pattern recognition method," *The Journal of the Acoustical Society of America*, vol. 92, pp. 1394–1402, September 1992.

[42] M. Castanys, M. J. Soneira, and R. Perez-Pueyo, "Automatic identification of artistic pigments by raman spectroscopy using fuzzy logic and principal component analysis authors: M. castanys,m. j. soneira, r. perez-pueyo.," *Laser Chemistry*, pp. 11–19, October 2006.

[43] P. A. Gerakines, W. A. Schutte, and P. Ehrenfreund, "Ultraviolet processing of interstellar ice analogs," *Astronomy and Astrophysics*, vol. 313, pp. 289–305, 1996.

[44] P. A. Gerakines and Moore, "Spectral database, the cosmic ice laboratory . nasa."

[45] M. P. Bernstein, D. P. Cruikshank, and S. A. Sandford, "Near-infrared laboratory spectra of solid h2o/co2 and ch3oh/co2 ice mixtures," *Icarus*, vol. 179, pp. 527–534, December 2005.