A PSEUDO-VOIGT COMPONENT MODEL FOR HIGH-RESOLUTION RECOVERY OF CONSTITUENT SPECTRA IN RAMAN SPECTROSCOPY

Tommy S. Alstrøm[*], *Mikkel N. Schmidt*[*], *Tomas Rindzevicius*[†], *Anja Boisen*[†], *Jan Larsen*[*]

[*]Department of Applied Mathematics and Computer Science, Technical University of Denmark
Richard Petersens Plads 324, 2800 Kgs. Lyngby, Denmark

[†]Department of Micro and Nanotechnology, Technical University of Denmark
Ørsteds Plads 345C, DK-2800, Kgs. Lyngby, Denmark

ABSTRACT

Raman spectroscopy is a well-known analytical technique for identifying and analyzing chemical species. Since Raman scattering is a weak effect, surface-enhanced Raman spectroscopy (SERS) is often employed to amplify the signal. SERS signal surface mapping is a common method for detecting trace amounts of target molecules. Since the method yields large amounts of data and, in the case of very low concentrations, low signal-to-noise (SNR) ratio, ability to extract relevant spectral features is crucial. We propose a pseudo-Voigt model as a constrained source separation model, that is able to directly and reliably identify the Raman modes, with performance is similar to the state of the art non-negative matrix factorization approach. The model is a step towards enabling the use of SERS in detection of trace amounts of molecules in real-life settings.

Index Terms— Raman Spectroscopy, Non-negative matrix factorization (NMF), Bayesian Modeling, Pseudo-Voigt, Multivariate Curve Resolution (MCR)

1. INTRODUCTION

An important physical phenomenon that can be utilized for chemical analysis via molecular "fingerprinting" is the so-called Raman scattering. A key issue with Raman scattering is that it is a very weak process, i.e. only 0.1% or less incident photons on a sample are inelastically (or Raman) scattered while the rest remain elastically (or Rayleigh) scattered. One straightforward method to amplify the Raman scattering signal is using corrugated, nanosized noble metal surfaces, and the effect is known as surface-enhanced Raman scattering. SERS was discovered in 1973 [1] and explained by Jeanmaire & Van Duyne in 1977 [2]. The "new" chapter for SERS begun in the late 1990s when researchers showed that in some

cases a strong surface enhancement can even enable single molecule detection [3].

Today surface-enhanced Raman scattering (SERS) is one of the most relevant and rapidly advancing spectroscopy methods for label-free and sensitive detection of target molecules [4]. A good illustration is a record low attomolar detection of TNT and DNT explosives that few methods if any can match [5]. Another advantage is that Raman instrumentation for SERS-based analysis is becoming better, cheaper, and portable [6].

A number of challenges in using the SERS technique for quantitative molecular detection in complex bioassays still remain [7]. One path for quantitative SERS is using SERS tags and it usually requires development of a reliable quantification methodology based on an ensemble of SERS spectra, e.g. SERS signal mapping [8]. Mapping of SERS signals is necessary because of generally poor reproducibility and uniformity of SERS substrates which produces significant Raman signal variations. Typically, underlying reasons for SERS signal variations are fabrication driven, i.e. it is very difficult to control accurately nanoparticle morphology and especially inter-particle separation at <10 nm distances which affect significantly electromagnetic enhancement (the so-called "hot spots") [9]. Hence, SERS intensities vary between hot spots with even smallest differences in morphology and junction dimensions.

Another issue is that target molecules are deposited or adsorbed at random locations on the SERS substrate surface area. Since the coverage is non uniform, significant SERS intensity variations over the substrate area are expected. At extremely low analyte concentrations (<nM) the probability to find target molecule exactly at the hot spot is extremely low, i.e. most hot spots are unoccupied and only few are expected to produce measurable SERS signals [10]. And finally, each SERS substrate type coupled with a target molecule display unique characteristics in terms of surface coverage, electromagnetic enhancement factors, distribution of hot spots, etc. Therefore, the statistical interpretation of SERS signals is a

crucial problem that needs to be addressed more rigorously.

When analyzing complex Raman spectra it is important to resolve overlapping line shapes [11]. While the true lineshape of a vibrational mode can be approximated by a Lorentzian the measured lineshape due to the instrument distortion is often the convolution of a Lorentzian and a Gaussian functions [12]. Therefore, to accurately determine line widths and peak positions, Raman spectra are fitted using Voigt functions which is a convolution of a Lorentzian with a Gaussian line shape.

Herein, we developed a method to automatically analyze large-area Raman or SERS signal maps for more reliable and reproducible identification of Raman modes (analyte molecules). The model is compared to the current state of the art approaches of extracting spectra, which are multivariate factor models with non-negativity constraints.

2. RELATION TO PRIOR WORK

Blind source separation techniques such as NMF has been used for spectral data on numerous occasions [13, 14, 15]. In the field of general spectroscopy the method of Multivariate Curve Resolution (MCR) using non-negativity constraints has also been applied [16]. The NMF and MCR models are identical but the inference algorithms differ. MCR uses the solution found using PCA as the initial solution and the model is then estimated using alternating least squares. NMF can also be learned in a number of ways, and was demonstrated for extracting facial features by Lee and Seung [17]. In the case of very low signal to noise ratios, blind source separation algorithms fails to recover meaningful and interpretable spectral components. Solutions include to apply additional constraints on the search for basis vectors [18], or to include some prelearned basis vectors in the model [14]. Another approach is to fit a power-law distribution at a certain wavenumber [10, 8]. However this method requires knowledge of the exact location of the Raman mode, and in addition assumes that the data is free from outliers. The multivariate approach of NMF/MCR does not suffer from these downsides, so a combination of the two approaches may be ideal. The proposed pseudo-Voigt model is considered as a combination of the two.

3. METHODS

Let X denote the observed data: An array of positive real numbers, where $X_{w,n}$ is the measured amplitude at wave number w in the observed spectrum n. We denote the number of observed wavelengths by W and the number of observed spectra by N.

We will compare two approaches for modeling such spectra: 1) A non-negative matrix factorization approach in which the spectra are modeled as scaled sums of K basis spectra which are learned from the observed data, and 2) a related approach where the basis spectra are constrained to have a pseudo-Voigt shape.

3.1. Non-negative matrix factorization

The non-negative matrix factorization approach to modeling spectroscopy data assumes that each observed spectrum can be decomposed as a weighted sum of constituent basis spectra [13, 14, 15]. Using a Normal observation noise model

$$X_{w,n}|S, A, \tau \sim \mathrm{N}([SA]_{w,n}, \tau^{-1}), \qquad (1)$$

where τ is the noise precision, the data array is modeled as product of two matrices, S and A, where the K columns of S are the basis spectra, and each column of A contain their amplitudes for a given observation spectrum.

Following [19], we assume a standard conjugate Gamma prior for the noise precision,

$$\tau \sim \mathrm{Gamma}(a_\tau, b_\tau), \qquad (2)$$

and we use independent exponential priors for both the spectra and their amplitudes,

$$A_{k,n} \overset{i.i.d}{\sim} \mathrm{Exp}(A_0), \quad S_{k,n} \overset{i.i.d}{\sim} \mathrm{Exp}(S_0) \qquad (3)$$

These priors reflect the constraints that the spectra and their amplitudes are non-negative and corresponds to a Bayesian sparse NMF model.

3.2. Pseudo-Voigt spectral peak model

As in the NMF approach, we will assume an independent Normal noise model,

$$X_{w,n}|I_{w,n}, \tau \sim \mathrm{N}(I_{w,n}, \tau^{-1}), \qquad (4)$$

with precision τ and mean $I_{w,n}$ modeled as a sum of K pseudo-Voigt shaped component plus a background component,

$$I_{w,n} = \sum_{k=1}^{K} \alpha_{k,n} V_w(c_k, \gamma_k, \eta_k) + \beta_n B_w. \qquad (5)$$

Here, V is a pseudo-Voigt shaped component centered at wave number w with parameters c, γ, and η; $\alpha_{k,n}$ is a component and observation specific amplitude; B_w is a common background spectrum; and β_n is a observation specific background amplitude. The pseudo-Voigt [20] profile is given by

$$V_w(c, \gamma, \eta) = \eta L_w(c, \gamma) + (1 - \eta)G_w(c, \gamma), \qquad (6)$$

where $L_w = \frac{\gamma \pi^{-1}}{(w-c)^2 + \gamma^2}$ and $G_w = \frac{1}{\sqrt{2\pi}\gamma} \exp\left(\frac{-(w-c)^2}{2\gamma^2}\right)$.

As in the NMF approach, we assume a standard conjugate Gamma prior for the noise precision. As priors for the pseudo-Voigt profile parameters we use the following:

$$\eta_k \overset{i.i.d}{\sim} \mathrm{Unif}(0, 1), \qquad \alpha_{k,n} \overset{i.i.d}{\sim} \mathrm{Exp}(\alpha_0), \qquad (7)$$

$$c_k \overset{i.i.d}{\sim} \mathrm{N}^+(\mu_{c_0}, \tau_{c_0}^{-1}), \qquad \gamma_k \overset{i.i.d}{\sim} \mathrm{N}^+(\mu_{\gamma_0}, \tau_{\gamma_0}^{-1}), \qquad (8)$$

where N^+ denotes the Normal distribution truncated to the positive reals. Here η_k is given an vague uniform prior, $\alpha_{k,n}$ is exponential to reflect the anticipation that a few observed spectra might contain "hot spots" with high amplitudes, and the c_k and γ_k parameters are given fairly flexible truncated Normal priors, reflecting that the parameters must be positive while making it possible to choose their hyperparameters to reflect locations and widths of the spectral peaks, or in lack of prior knowledge to set them as vague.

The prior for the background component is chosen as

$$\beta_n \overset{i.i.d}{\sim} N^+(\mu_{\beta_0}, \tau_{\beta_0}^{-1}), \qquad B_w \overset{i.i.d}{\sim} N^+(\mu_{B_0}, \tau_{B_0}^{-1}), \qquad (9)$$

allowing the hyperparameters to be chosen in accordance with prior knowledge.

3.3. Inference

The sparse NMF model is learned using a Gibbs sampler as specified in [19]. Following the same approach, we develop a similar algorithm for learning the pseudo-Voigt model. The conditional distributions for the parameters with conjugate priors (α, β, B and τ) are sampled using Gibbs sampling. First define $V_{w,k} = V(w, c_k, \gamma_k, \eta_k)$. The conditional distribution for the weights of the pseudo-Voigt functions is proportional to the product a normal distribution and an exponential distribution and the resulting distribution is then a truncated Normal distribution limited to positive values with center and scale

$$\alpha_{k,n}|X, \alpha_{\backslash k,n}, \beta, B, \tau \sim N^+(\mu_{\alpha_{k,n}}, \tau_{\alpha_{k,n}}^{-1}) \qquad (10a)$$

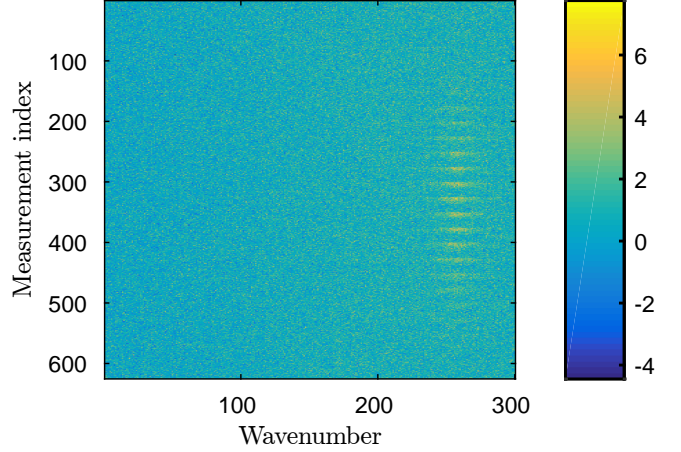$$\tau_{\alpha_{k,n}}^{-1} = \tau \sum_w^W V_{w,k}^2 \qquad (10b)$$

$$\mu_{\alpha_{k,n}} = \frac{\tau \sum_w^W V_{w,k}\left(X_{w,n} - b_n B_w - \sum_{k' \neq k}^K a_{k',n} V_{w,k'}\right) - \alpha_0}{\tau \sum_w^W V_{w,k}^2} \qquad (10c)$$

The conditional distributions of the background signals are on a similar account also truncated Normals, with the center and scale
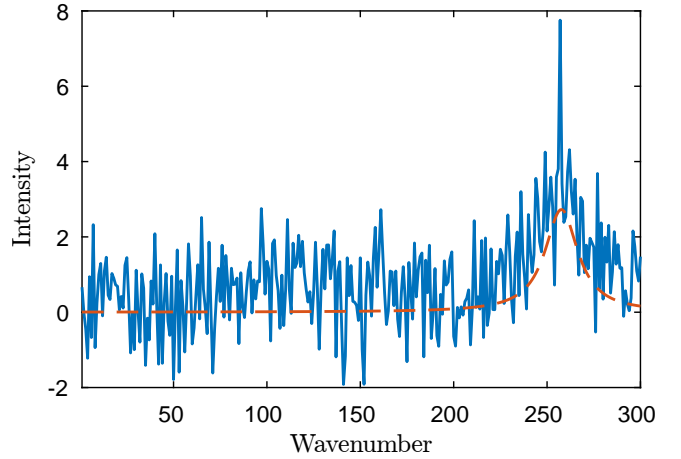
$$\beta_n|X, \alpha, \beta_{\backslash n}, B, \tau \sim N^+(\mu_{\beta_n}, \tau_{\beta_n}^{-1}) \qquad (11a)$$

$$\tau_{\beta_n}^{-1} = \tau_{\beta_0} + \tau \sum_w^W B_w^2 \qquad (11b)$$

$$\mu_{\beta_n} = \frac{\tau \sum_w^W B_w\left(X_{w,n} - \sum_k^K a_{k,n} V_{w,k}\right) + \tau_{\beta_0}\mu_{\beta_0}}{\tau_{\beta_0} + \tau \sum_w^W B_w^2} \qquad (11c)$$

(a) Data generated by simulation.

(b) Example spectrum with a SNR=2.7.

Fig. 1. Data generated by simulation. The blue line is an example spectrum with added noise and the dotted red is the true underlying spectrum.

The condition distribution of B_w is almost identical due to symmetry in the likelihood. For the precision we get the standard result

$$\tau|X, \alpha, \beta, B \sim \text{Gamma}(a_\tau, b_\tau) \qquad (12a)$$

$$a_\tau = \frac{WN}{2} + a_{\tau_0} \qquad (12b)$$

$$b_\tau = b_{\tau_0} + \frac{1}{2}\sum_w^W \sum_n^N (X_{w,n} - I_{w_n})^2 \qquad (12c)$$

The pseudo-Voigt parameters are sampled using Metropolis Hastings.

4. RESULTS AND DISCUSSION

The algorithm is tested on simulated data where the true parameters are known. The data is generated using a fixed
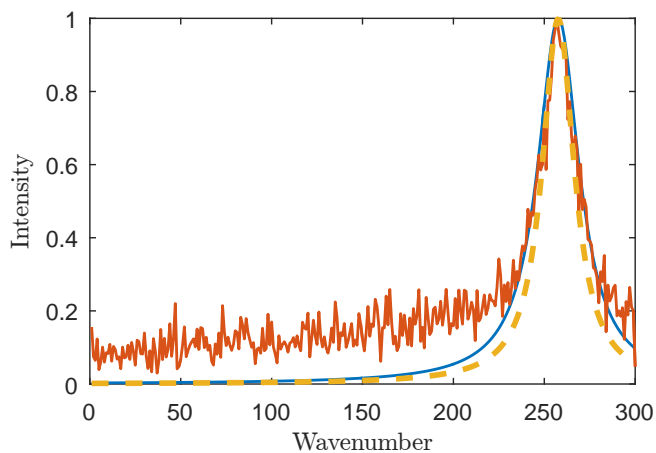
Fig. 2. Spectrum as identified by the pseudo-Voigt model (blue) and the NMF model (red) compared to the true underlying signal (dotted orange).
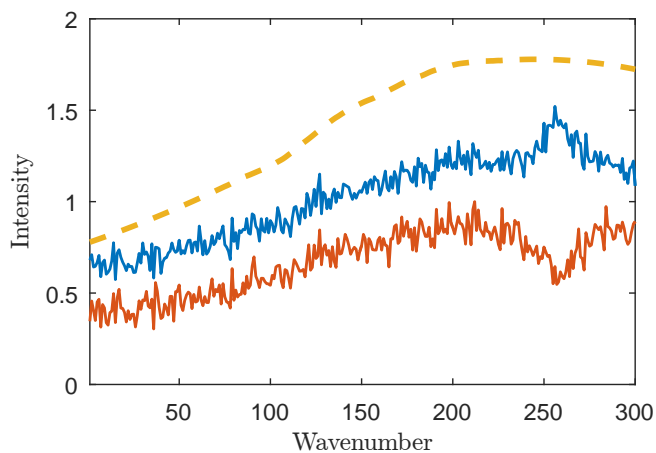

Fig. 3. Background signal recovered by the pseudo-Voigt model (blue) and the NMF model (red) compared to the true underlying background signal (dotted orange).

pseudo-Voigt profile added to a randomly generated a smooth background added to zero mean Gaussian distributed noise with precision equal to one. We generated a map of 625 measured spectra at 300 wavenumbers. We created the weights for the pseudo-Voigt profiles by simulating hot-spot behavior. A single hotspot was randomly placed on a grid and the weights of this hotspot were calculated using a 2-d Gaussian curve. At the maximum peak location the SNR is 2.7. The generated data is shown in Fig. 1.

The pseudo-Voigt model is able to recover the exact center of the Raman peak, whereas NMF recovers a noisy spectrum with the highest intensity located very close to the correct Raman peak (Fig. 2). The most important attribute of a Raman peak is the location of the center as it can be directly mapped to a specific molecular binding, and thus it is essential to extract the center of a Raman mode reliably.
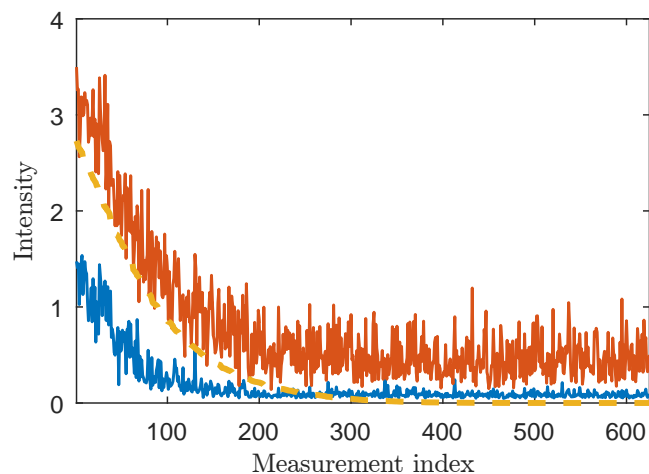

Fig. 4. Intensities as identified by the pseudo-Voigt model (blue) and the NMF model (red) compared to the true intensities (dotted orange).

Recovery of the background signal is displayed in Fig. 3 where the peak clearly interferes with the background signal estimate. The pseudo-Voigt model overestimates the background signal in the region of the Raman peak whereas the NMF model underestimates the background signal here. This is further reflected in the recovery of the Raman peak intensities which are displayed in Fig. 4. The presence of a molecule is in essence determined by setting a threshold on the intensity. The threshold should ideally be inferred from data but setting it to one yields very similar performance of the two methods.

In the presence of mixtures of peaks with overlap, the posterior densities of both models become multimodal. Determining the hyperparameters is thus essential to obtain the solution which makes physical sense. This can easily and readily be achieved in the pseudo-Voigt model, as the experimenter can intuitively select potential areas of Raman peaks. The algorithm had similar performance for mixtures ($K = 3$, not shown) as for a single Raman peak, and was able to recover the centers of Raman peaks.

5. CONCLUSION

The pseudo-Voigt approach works well for the simulated data, demonstrating comparable performance to current state of the art approaches such as NMF (and MCR). The advantage of the pseudo-Voigt model is that the center of a Raman peak is directly estimated as a model parameter. In addition, the pseudo-Voigt model offers a degree of confidence that the recovered spectrum has a Voigt shape which means that a true Raman peak has been recovered. Extracting the same information from the spectrum recovered by the NMF model requires additional post-processing.

6. REFERENCES

[1] M. Fleischmann, P. Hendra, and A. McQuillan, "Raman spectra of pyridine adsorbed at a silver electrode," *Chemical Physics Letters*, vol. 26, no. 2, pp. 163–166, may 1974.

[2] D. L. Jeanmaire and R. P. Van Duyne, "Surface raman spectroelectrochemistry," *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, vol. 84, no. 1, pp. 1–20, nov 1977.

[3] S. Nie and S. R. Emory, "Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering," *Science*, vol. 275, no. 5303, pp. 1102–1106, feb 1997.

[4] S. Schlücker, "Surface-Enhanced Raman Spectroscopy: Concepts and Chemical Applications," *Angewandte Chemie International Edition*, vol. 53, no. 19, pp. 4756–4795, may 2014.

[5] A. Hakonen, P. O. Andersson, M. Stenbæk Schmidt, T. Rindzevicius, and M. Käll, "Explosive and chemical threat detection by surface-enhanced Raman scattering: A review," *Analytica Chimica Acta*, vol. 893, pp. 1–13, sep 2015.

[6] "Serstech," www.serstech.com.

[7] M. Procházka, *Surface-Enhanced Raman Spectroscopy*, vol. 893 of *Biological and Medical Physics, Biomedical Engineering*, Springer International Publishing, Cham, sep 2016.

[8] J. Yang, M. Palla, F. G. Bosco, T. Rindzevicius, T. S. Alstrøm, M. S. Schmidt, A. Boisen, J. Ju, and Q. Lin, "Surface-Enhanced Raman Spectroscopy Based Quantitative Bioassay on Aptamer-Functionalized Nanopillars Using Large-Area Raman Mapping," *ACS Nano*, vol. 7, no. 6, pp. 5350–5359, jun 2013.

[9] E. Petryayeva and U. J. Krull, "Localized surface plasmon resonance: Nanostructures, bioassays and biosensing—A review," *Analytica Chimica Acta*, vol. 706, no. 1, pp. 8–24, nov 2011.

[10] M. Palla, F. G. Bosco, J. Yang, T. Rindzevicius, T. S. Alstrom, M. S. Schmidt, Q. Lin, J. Ju, and A. Boisen, "Mathematical model for biomolecular quantification using large-area surface-enhanced Raman spectroscopy mapping," *RSC Adv.*, vol. 5, no. 104, pp. 85845–85853, 2015.

[11] J. D. Weatherston, N. C. Worstell, and H.-J. Wu, "Quantitative surface-enhanced Raman spectroscopy for kinetic analysis of aldol condensation using Ag–Au core–shell nanocubes," *The Analyst*, 2016.

[12] T. Sundius, "Computer fitting of Voigt profiles to Raman lines," *Journal of Raman Spectroscopy*, vol. 1, no. 5, pp. 471–488, nov 1973.

[13] P. Sajda, S. Du, and L. Parra, "Recovery of constituent spectra using non-negative matrix factorization," *Proceedings of SPIE*, vol. 5207, pp. 321–331, 2003.

[14] T. S. Alstrøm, K. B. Frøhling, J. Larsen, M. N. Schmidt, M. Bache, M. S. Schmidt, M. H. Jakobsen, and A. Boisen, "Improving the robustness of Surface Enhanced Raman Spectroscopy based sensors by Bayesian Non-negative Matrix Factorization," *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2014.

[15] H. Li, T. Adali, W. Wang, and D. Emge, "Non-negative matrix factorization with orthogonality constraints for chemical agent detection in raman spectra," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2005, pp. 253–258.

[16] R. Tauler, B. Kowalski, and S. Fleming, "Multivariate curve resolution applied to spectral data from multiple runs of an industrial process," *Analytical Chemistry*, vol. 65, no. 15, pp. 2040–2047, aug 1993.

[17] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, oct 1999.

[18] H. Li, T. Adali, W. Wang, D. Emge, and A. Cichocki, "Non-negative Matrix Factorization with Orthogonality Constraints and its Application to Raman Spectroscopy," *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 48, no. 1-2, pp. 83–97, feb 2007.

[19] M. N. Schmidt, O. Winther, and L. K. Hansen, "Bayesian Non-negative Matrix Factorization," in *Independent Component Analysis and Signal Separation*, 2009, vol. 5441 of *Lecture Notes in Computer Science*, pp. 540–547.

[20] G. K. Wertheim, M. A. Butler, K. W. West, and D. N. E. Buchanan, "Determination of the Gaussian and Lorentzian content of experimental line shapes," *Review of Scientific Instruments*, vol. 45, no. 11, pp. 1369–1371, 1974.