Article

Applied Spectroscopy

Applied Spectroscopy
2019, Vol. 73(12) 1370–1379

DOI: 10.1177/0003702819871330
journals.sagepub.com/home/asp
SAGE

An Open Source, Iterative Dual-Tree Wavelet Background Subtraction Method Extended from Automated Diffraction Pattern Analysis to Optical Spectroscopy

Robert B. Chevalier and Jason R. Dwyer

Abstract

Background subtraction is a general problem in spectroscopy often addressed with application-specific techniques, or methods that introduce a variety of implementation barriers such as having to specify peak-free regions of the spectrum. An iterative dual-tree complex wavelet transform-based background subtraction method (DTCWT-IA) was recently developed for the analysis of ultrafast electron diffraction patterns. The method was designed to require minimal user intervention, to support streamlined analysis of many diffraction patterns with complex overlapping peaks and time-varying backgrounds, and is implemented in an open-source computer program. We examined the performance of DTCWT-IA for the analysis of spectra acquired by a range of optical spectroscopies including ultraviolet–visible spectroscopy (UV–Vis), X-ray photoelectron spectroscopy (XPS), and surface-enhanced Raman spectroscopy (SERS). A key benefit of the method is that the user need not specify regions of the spectrum where no peaks are expected to occur. SER spectra were used to investigate the robustness of DTCWT-IA to signal-to-noise levels in the spectrum and to user operation, specifically to two of the algorithm parameter settings: decomposition level and iteration number. The single, general DTCWT-IA implementation performs well in comparison to the different conventional approaches to background subtraction for UV–Vis, XPS, and SERS, while requiring minimal input. The method thus holds the same potential for optical spectroscopy as for ultrafast electron diffraction, namely streamlined analysis of spectra with complex distributions of peaks and varying signal levels, thus supporting real-time spectral analysis or the analysis of data acquired from different sources.

Keywords

Automated baseline determination, background subtraction, wavelet transforms, spectroscopic background, iterative background subtraction

Date received: 26 April 2019; accepted: 23 July 2019

Introduction

Background subtraction remains a difficult challenge across a wide range of spectroscopies and other techniques. The problem reduces to how to extract the signal of interest, $S_{interest}(x)$, from a measured signal, $S_{measured}(x) = S_{interest}(x) + S_{background}(x)$,[1,2] when the background signal, $S_{background}(x)$, can emerge by a variety of often difficult to predict mechanisms including variability in time and from sample to sample.[3,4] Background subtraction may be itself relegated to the background of data analysis procedures, if details are provided at all. While perhaps a prosaic topic, background subtraction can strongly influence downstream analyses,[5] and the issue thus forms an essential part of a larger focus on how similar foundational considerations can profoundly influence measurement reliability.[6] Simple approaches, such as polynomial fitting, have been adopted and adapted for background subtraction in spectroscopy.[2,7–9] The possibility in some data to cleanly separate signal peaks from background allows the development of more complex, but still relatively straightforward and efficient, approaches to background subtraction.[7,10,11] Such

Department of Chemistry, University of Rhode Island, Kingston, RI, USA

Corresponding author:
Jason R. Dwyer, University of Rhode Island, 140 Flagg Rd, Kingston, RI 02881, USA.
Email: jason_dwyer@uri.edu

methods can of course be augmented with additional steps, including iteration, to enhance performance while preserving accessibility, as illustrated in the auto-adaptive method of Xie et al.[11] Baek et al. constructed a few-parameter approach with iterative weighting of measurement noise, that similarly preserved implementation accessibility.[12] Many approaches to background subtraction remain of the ad hoc type, though, when the details of this important analysis step are specified at all. In these approaches one must frequently make assumptions about where the spectrum is expected to generate a signal corresponding entirely to background ($S_{interest}(x) = 0$).[2,13] In high-throughput data analysis where background, matrix composition, and even analyte identities may change, such as in real-time environmental monitoring,[7,8] the use of such custom and even sample-specific algorithms (and parameter values) can be tedious, time-consuming, and introduce bias into analyses. Certainly the likelihood that frequent user-intervention would be required would make such methods incompatible with high-volume data analysis.[2] The need for dealing with large data sets with variable backgrounds was a strong practical motivation given for an iterative wavelet-transform-based background subtraction method developed for high-data-volume applications such as Raman mapping.[1] This technique joined a sophisticated suite of approaches that have been successfully used for signal processing,[14,15] de-noising,[16–19] and background subtraction.[1,2,18] Fourier transform methods figure prominently amongst these approaches,[2,14,20,21] but wavelet transforms offer the benefit that the wavelet basis functions are localized in space and frequency, providing a more suitable analysis of sharp peaks and non-periodic noise.[1,22,23] Discrete wavelet transform (DWT) has shown good results for background subtraction of different spectroscopic techniques such as energy dispersive X-ray fluorescence (EDXRF),[24] Raman,[1,19] near-infrared spectroscopy,[25] high-resolution electron energy-loss spectroscopy (HREELS),[21,26,27] and X-ray photoelectron spectroscopy (XPS),[21,26,27] but it has been noted that DWT has performance limitations arising from its lack of shift invariance and poor directional selectivity.[15,23]

We were seeking a method requiring minimal user input capable of performing background subtraction for surface-enhanced Raman spectroscopy (SERS). We were particularly interested in eliminating the frequent need to specify regions—if they exist—of a spectrum where $S_{interest}(x) = 0$. This would minimize the amount of user intervention in the background subtraction and, more importantly, allow for the analysis of congested spectra where overlapping peaks might make this step difficult if not impossible. We were thus excited by a new background subtraction method, the dual-tree complex wavelet transform with an iterative algorithm (DTCWT-IA), that was motivated by the need to cope with large time-varying data sets with appreciable segments of data with no significant range of background-only signal.[22,28] Such a signal characteristic challenges curve-fitting and Fourier methods, and looms forbidding to the prospect of batch processing. Beyond this technical benefit, we were pleased with the practical motivation that the development was directed by a desire to reliably perform background subtractions across large series of measurements with variability in signal and signal-to-noise (S/N) levels, and with minimal user intervention. This was tested using simulated time-dependent data with Gaussian noise and with backgrounds with a known form to introduce qualitatively consistent background structure across spectra that varied in magnitude in time.[22] The most captivating aspect of the background subtraction method is that the basic algorithm, as outlined below, requires no a priori input of peak location and extent. The method combines the earlier promise of iterative wavelet methods applied to Raman spectral analysis[1] with the dual-tree approach, but was developed for analysing ultrafast electron diffraction patterns.[22,28] The signal in these studies was generated by a mechanism orthogonal to those of spectroscopy, but given the favorable constellation of features and the formal equivalency of (powder) diffraction pattern and optical spectrum as consisting of peaks arrayed along an abscissa,[29–31] we undertook to explore the viability of the DTCWT-IA background subtraction applied to a range of spectroscopies: ultraviolet–visible (UV–Vis), SERS, and XPS. We performed background subtractions for these spectral types and compared the results to commonly used background subtractions often available as part of commercial analysis packages.

There is a hierarchy of wavelet transforms upon which the DTCWT-IA is based. Wavelets, themselves, are localized in space and frequency, and their type and order affects their shape and complexity.[1,2] Given wavelet spatial localization, wavelet transforms can thus fit featureless regions with greater ease than Fourier methods which may require large numbers of high-frequency terms for sufficient cancellation.[1,2] Discrete wavelet transforms (DWT) have widespread use, but have some performance drawbacks addressed by the dual-tree complex wavelet transform (DTCWT). Specifically, DTCWT eliminates wavelet coefficient instability near sharp signal features, and aliasing artefacts that can arise from the use of real-valued wavelets.[22,28] It provides shift invariance and good directional selectivity.[15,23] Complex-valued wavelets are implemented by having one DWT tree, each, for the real and imaginary components, yielding the dual-tree approach.[22,28] Wavelet decomposition is a recursive process in which a spectrum is processed into "smooth" and "detailed" components, with further processing applied to the smooth components remaining after each recursion step.[2] The decomposition level of the method necessary to recover the (physical) background is determined by the frequency composition of the background.[1] De Cotret and Siwick extended an iterative method for DWT to

DTCWT.[1,22] Iteration acts to remove the influence of peaks in regions where the peak and background overlap. In brief, an initial background profile at a desired decomposition level is used as a low-pass intensity filter (spectral magnitudes—predominantly peak magnitudes—greater than the background magnitude proposed at this stage are set equal to this initial background profile). The background fitting is then repeated on this filtered spectrum until it converges to what Galloway et al. described as "the most physical representation (without including the peaks)".[1] A step-by-step pictorial illustration of this process is provided in Galloway et al.,[1] who emphasize that the number of iterations needed will depend on the spectral characteristics and can be decided on a case-by-case basis.

In practical terms, the basis of DTCWT-IA applied to background subtraction can be understood by considering a spectrum as being comprised of three frequency regimes: high (noise), medium ($S_{interest}$), and low ($S_{background}$).[1] At a very simple level, wavelet analysis leverages these frequency regimes to decompose $S_{measured}$ into $S_{interest}$ and $S_{background}$; this process is controlled, for a given choice of wavelets, by the decomposition level (and for iterative methods, also by the iteration number). Overlap between the frequency regimes means that the wavelet transform cannot adequately distinguish between $S_{interest}$ and $S_{background}$ without the further iterations that underpin the wavelet and DTCWT iterative methods.[1,22] The quality of a background spectrum can be generally evaluated by the degree to which it does not introduce nonphysical structure (spurious peaks) into the spectrum, and the degree to which it isolates the physically meaningful structure (peaks) without perturbing them.[1] We introduce two convenient metrics for illustrating how these two criteria change during optimization of the fitting parameters of decomposition level and iteration number. These simple metrics do not replace inspection of the spectrally resolved profile of background and spectrum. We calculated the root mean square average of the numerically differentiated background ($S_{background}$) across the entire spectrum (RMS_D) and used it as a metric for the amount of structure present in each calculated background spectrum. We also numerically integrated the area of the peak centered at $\sim$1300 cm^{-1} between limits of 1250 and 1405 cm^{-1}. These peak areas were normalized relative to the peak areas of $S_{measured}$, so that when $S_{background} = 0$ across the specified peak range, a value of one is recovered. Such simple metrics are an economical means for quickly screening algorithm parameters for the experimental case where the background form and magnitude are unknown, and are useful for illustrating the algorithm behavior for the sake of this report, but should not replace the inspection of the spectrally resolved results that we also present.

De Cotret and Siwick[22] have provided a streamlined package for DTCWT-IA currently configured to have only five user parameters: First stage wavelet, dual-tree wavelet, extension mode, decomposition level, and iteration. The availability of this software addresses a serious concern raised by Xie et al.[11] that the adoption of sophisticated background subtraction algorithms can be limited by the complexity of implementation. In addition to exploring the robustness of the algorithm to spectrum type, we explore the effects of decomposition and number of iterations on the background-subtraction process. We focus on only these two parameter settings in order to illustrate the performance of the algorithm and to highlight potential pitfalls with incorrect parameter settings. In general practice, the other parameters should be adjusted, as well, with attention to the same behaviors that we explore in this manuscript.

Materials and Methods

We used a distributable implementation of the DTCWT-IA for all background subtractions unless noted otherwise.[32] For all figures, the quality of the background profile was evaluated by a spectrally resolved comparison of background profile and spectrum (e.g., Fig. 1a), and the optimization of parameters was carried out conservatively. We selected the lowest iteration number that minimized background overlap with peak-like features, in particular overlap that disrupted the smooth nature of the background with small, but noticeable bumps. For these applications, this approach thus included the deliberate choice to avoid the linear background limit of a large iteration number. A standard approach was taken for the background subtractions in Fig. 4a to f. We outline it here while noting that a more thorough illustration of key aspects of the process is presented in the Results and Discussion section. The parameters used for all backgrounds (SERS, UV–Vis, and XPS) were first stage wavelet: Sym6; dual tree wavelet: qshift5; extension mode: constant. With iteration set to 1, the decomposition level was changed incrementally until the calculated background spectrum assumed the desired profile. At this stage, this meant that the decomposition level was increased until the calculated background spectrum profile smoothly followed the profile in regions apart from the peak maxima of the spectrum being analyzed. The iteration number was then increased to improve the convergence between the background spectrum and these peak-free regions, without introducing peaks into the background profile. If the background spectrum failed to converge sufficiently, then dual-tree wavelet (or other program settings) could be changed and the process repeated. At no point in the analysis or background subtraction was any input required for peak position or extent.

Comparison background subtractions were performed using a variety of methods implemented in commercial analysis programs. For SERS, a localized windowed minimum subtraction method (LWMSM) denoted TBB Baseline in

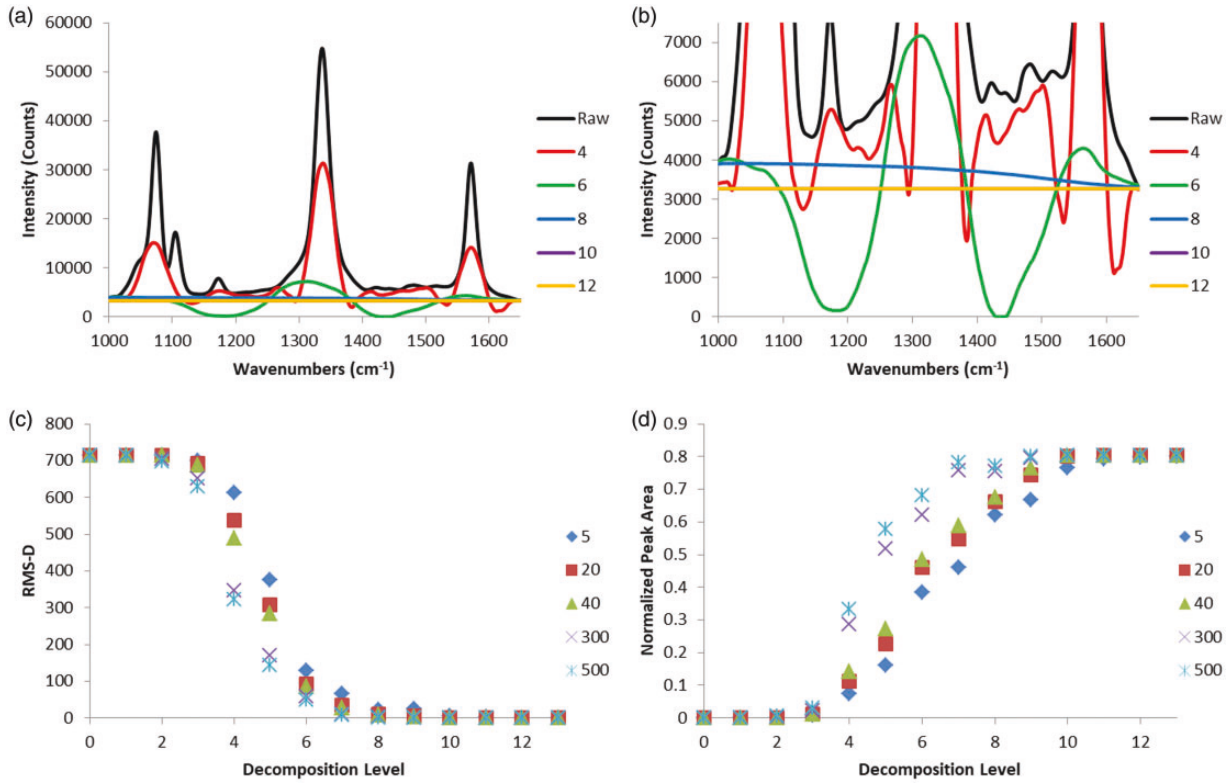

Figure 1. (a) SER spectrum (black) of 5×10^{-4} M 4-nitrothiophenol (NBT) and (b) with calculated backgrounds of varying decomposition levels (red—4, green—6, blue—8, (overlapped) purple—10, orange—12 decomposition level) using DTCWT-IA. RMS_D and Integral analyses are shown in (c) and (d) at both varying decomposition levels (*x*-axis) and iteration number.

Peak (Snowy Range Instruments, USA) was used for background subtraction. Two approaches implemented in Origin 8 Peak Analyzer (OriginLab, USA) were used for UV–Vis background subtraction: A linear interpolation and a spline, both fit to spectral positions automatically generated by the fitting algorithms in Origin. For XPS spectra, Shirley, Tougaard, and Smart background spectra were calculated using Avantage 5.9904 (Thermo Fisher Scientific Inc.). SERS and UV–Vis data were provided by Karawdeniya et al.[33,34] Representative XPS data of a NCM111 Pristine battery fresh electrode were provided by the Lucht Group.[35]

Results and Discussion

The Effect of Decomposition Level on the Calculated Background of the DTCWT-IA

The decomposition level in the DTCWT-IA controls the frequency composition of the background spectrum, and was the first parameter that we chose to address. Figure 1a to d depict the effect of the decomposition level on the calculated background, with the number of iterations set without optimization (see below) to 500 so that for this illustration the background spectrum quality would be unaffected by an insufficient number of iterations.

Figure 1a and b show five different DTCWT-IA backgrounds calculated at varying decomposition levels: 4 (red), 6 (green), 8 (blue), 10 (purple), and 12 (orange). At a low decomposition level, the medium frequencies (in essence, the peaks) were still prevalent in the calculated background, while at a high decomposition level, the background became a featureless straight line. The consequences of an incorrect background spectrum determination from insufficient decomposition will be the imposition of spurious structure on the analyte spectrum (e.g., for a decomposition level of four in Fig. 1a to d), incorrect peak intensities (and thus incorrect concentration determinations), and even incorrect peak positions. While the DTCWT-IA method has the benefit of not requiring often subjective inputs such as peak positions and extents, or positions where zero signal intensity would be expected, the determination of the optimum decomposition level remains largely subjective. We thus wanted to more concretely explore the sensitivity of the background profile to the particular decomposition level.

A suitable background spectrum should not produce an overly structured profile with peaks in the background where signal peaks were present (i.e., it should have a lower RMS_D given fewer extrema), and should not detrimentally lower the peak intensity (as in Fig. 1a with a decomposition level of four). At low decomposition

levels, a decrease in the RMS_D began at a decomposition level of approximately four. The fairly large RMS_D values at this and lower decomposition levels correspond to spectral overlap between the background and the raw signal, due to insufficient decomposition of the signal. The peak intensity is correspondingly too low at these decomposition levels, as observed in Fig. 1d. An increase in the normalized peak area occurred at a decomposition level of four where the RMS_D began to decrease. It is clear by examining the data that intermediate values of the decomposition level are necessary unless one desires to trace the measured spectrum or do little more than perform a linear subtraction. In fact, while the full implemented range of decomposition level settings was 1–99, the range was restricted here to 1–13 due to the onset of asymptotic behavior near 12. At the arbitrary iteration number of 500, it is clear that the optimal decomposition level to produce a background without evident peaks was nearer to the asymptote of the RMS_D and integral plots at higher decomposition level than to the inflection point of the curves. Final optimization of the decomposition level for a particular spectrum or set of spectra should include careful evaluation of the (spectrally resolved) background and signal spectrum in conjunction with consideration of metrics such as the RMS_D and peak intensity. The trend of the latter two metrics with decomposition level at this iteration level shows regions of insensitivity bracketing a central region with high sensitivity. Overall, a conservative approach of choosing the minimal decomposition level that achieves the desired level of smoothing is recommended unless asymptotic behavior (such as a purely linear fit) is desired.

Effect of Iterations on the Calculated Background of DTCWT-IAs

The addition of an iterative algorithm to a wavelet transform for background subtraction was done to provide greater control over minimizing the peak contribution to the calculated background.[1] Determining a suitable number of iterations then becomes a crucial step in calculating the background. To explore the effect of iteration number, we calculated spectral backgrounds at a decomposition level of eight, where the background (Fig. 1a) was not overly structured and did not appreciably overlap with the spectral peaks at the fixed iteration level (500) used for Fig. 1a to d. Even at this decomposition level, at low iteration number the background profile infiltrated the peaks (Fig. 2a and b). Our initial inclination to simply increase the number of iterations to the maximum level implemented in the software was not fruitful however, as the background profile then encompassed only very little of the seemingly broad background underneath the peaks, and began to resemble a straight line.

We calculated RMS_D and background-subtracted peak integrals as a function of iteration. As before, RMS_D and the peak area provide simple metrics for illustrating the degree to which the background profile inappropriately followed the signal profile and cut into the peaks, allowing a survey of its sensitivity to iteration number. A large value of RMS_D and a lower value of the peak intensity are indicative of such unsatisfactory background profile behaviors. The sigmoidal trend observed with changes of decomposition level in the RMS_D analysis was replaced by a monotonic decline with fairly rapid onsets of asymptotic behavior. A reasonable balance of background profile smoothness and peak area was obtained at 40 iterations for the decomposition level of eight. Once above a minimum threshold, however, the asymptotic trends of both parameters indicate that there is less sensitivity to the iteration number than to decomposition level. In determining the final iteration number, we used the same conservative approach of choosing the first iteration number that minimized the appearance of bumps in an otherwise smooth background. This was carried out by comparison of the (spectrally resolved) background profile to the spectrum.

Given the relatively low value of 40 iterations compared to the fixed 500 iterations used to explore decomposition level (Fig. 1), we repeated the analysis of iteration number at decomposition levels of 4 and 6. The RMS_D and integral data are superimposed on the Fig. 2c and d plots, and the full spectral representations are shown in Fig. 2e and f. The hallmark of these incorrect background assignments due to insufficient iteration was a background profile that was overly structured and that traced all the curves of the spectrum, including (small magnitude) sharp ones. Considered together, the data in Fig. 1a to d and 2a to f show the importance of performing the DTCWT-IA background subtraction iteratively, and that while an iteration number threshold should be exceeded, the quality of the background profile depends far more strongly on the decomposition level.

Sensitivity to Signal-to-Noise Ratio (S/N)

The results in Figs. 1a to d and 2a to f showed that it was possible to use the DTCWT-IA to define reasonable backgrounds for SER spectra, but also indicated that incorrect settings would lead to analyte peak intensity being incorrectly assigned to background. In Fig. 2b, this can be seen most prominently in a background profile closely following even sharp, small-magnitude curves ($\sim$1420–1460 cm^{-1}). Given the frequent use of SERS for trace analysis, we thus wanted to explore the effect of the signal-to-noise ratio (S/N) on the background subtraction. Figure 3a to f show experimental and simulated SER spectra. Experimental spectra were acquired from samples containing analyte at 5×10^{-4} M and 1×10^{-6} M concentrations, respectively. Simulated spectra were constructed by taking the background-subtracted (by DTCWT-IA) 5×10^{-4} M NBT spectrum, adding in Gaussian noise at 0.1, 1.0, and

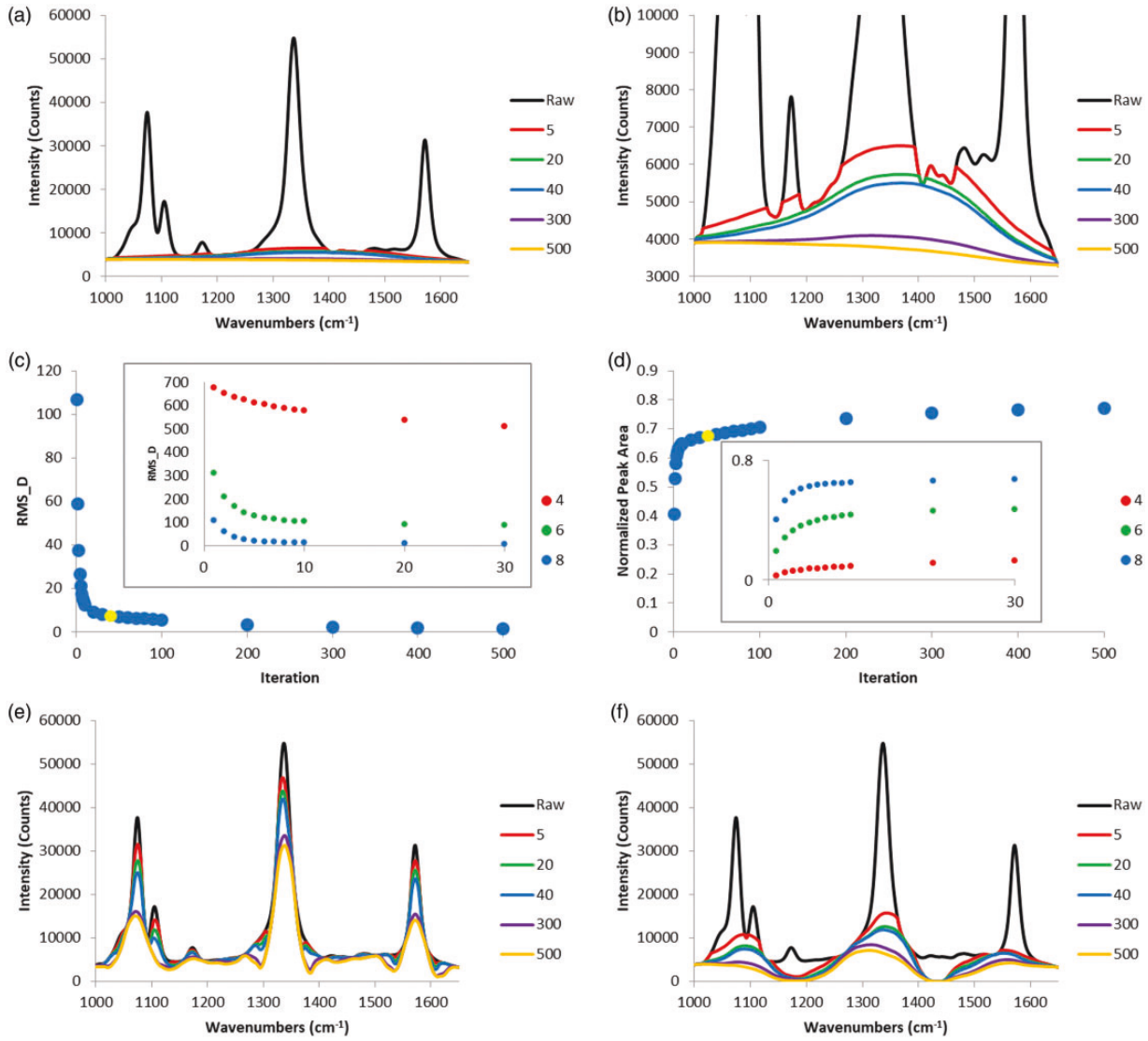

Figure 2. (a) SER spectrum (black) of 5×10^{-4} M NBT and (b) with calculated DTCWT-IA backgrounds of varying iteration number and fixed decomposition level of 8. RMS_D and integral analyses of the backgrounds in (a) are presented in (c) and (d), with the yellow data point denoting a favorable iteration number providing a background with a reasonable balance of smoothness and peak area. The insets compare the results of these metrics at decomposition levels of 4, 6, and 8. To observe the effect on the background profile of iteration at varying decomposition levels, calculated backgrounds of varying iterations at decomposition (e) level 4 and (f) 6 are shown.

2.5% of the maximum peak intensity, and adding the calculated background back in.

Since both experimental spectra were obtained under identical conditions aside from analyte concentration, the background subtractions were performed with identical settings determined from analysis of the spectrum with better S/N, namely: first stage wavelet: Sym6; dual-tree wavelet: qshift5; extension mode: constant; iteration: 44; decomposition level: 8. The resulting background profiles in Fig. 3c were different in intensity but similar in profile in both cases with the particular benefit that the background subtraction did not overlap with the small peaks as seen in the incorrectly optimized background subtractions of Fig. 2b. More importantly, though, the background subtraction required no user input to define areas where peaks might be expected: This is a significant advantage over many ad hoc methods in common use. In addition, the successful use of common analysis parameter settings across a $>10\times$ difference in peak intensity, with concomitant insensitivity of the background profile quality to S/N, indicates the convenience and robustness of DTCWT-IA.

As an added test of the method, we simulated spectra with controlled levels of noise. We background-subtracted the SER spectrum of 5×10^{-4} M NBT using the DTCWT-IA algorithm (first stage wavelet: Sym6; dual-tree wavelet: qshift5; extension mode: constant; iteration: 44;

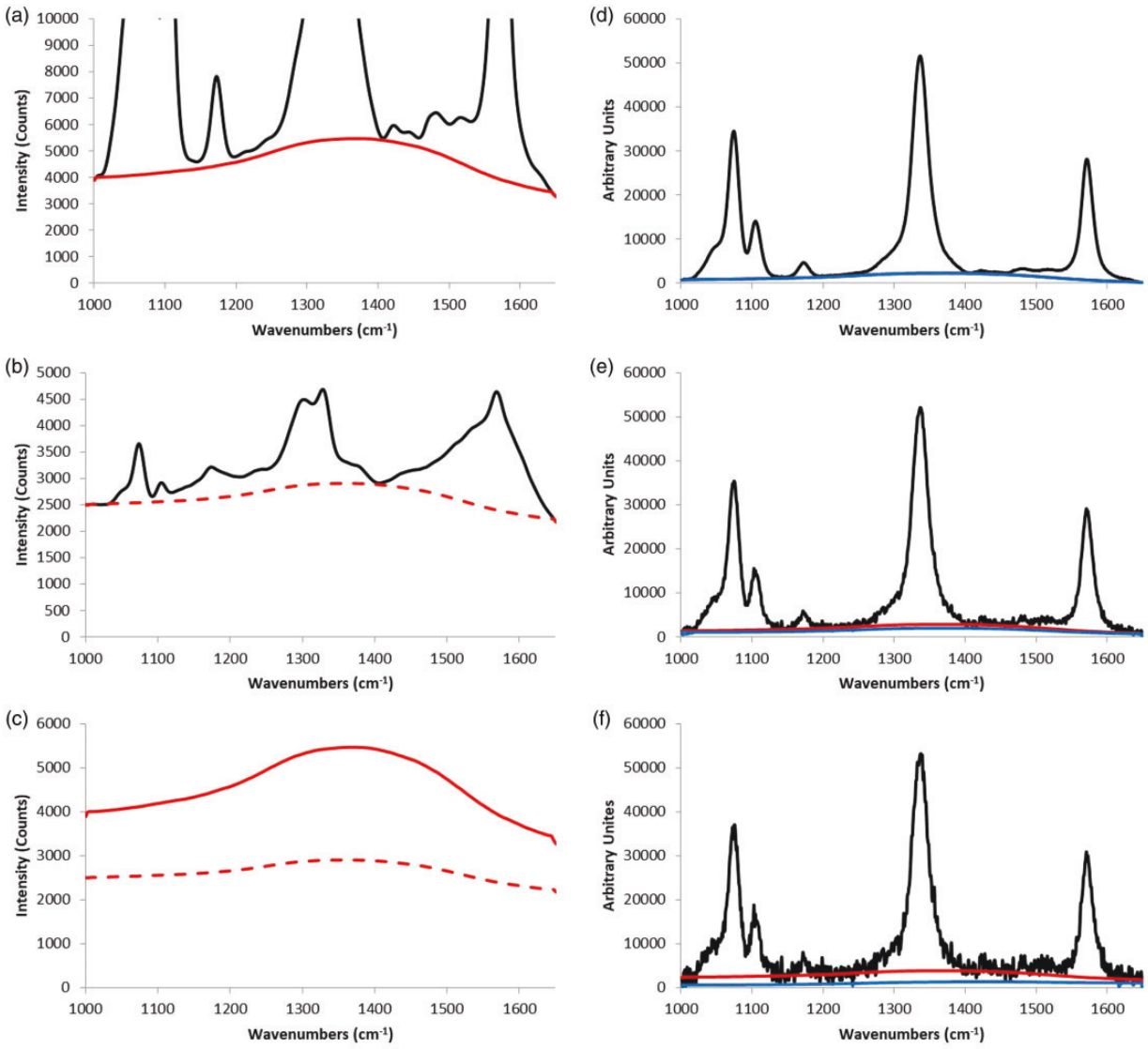

Figure 3. Experimental SER spectra of (a) 5×10^{-4} M and (b) 10^{-6} M NBT (black) with calculated backgrounds (red and dashed red) via DTCWT-IA. Both backgrounds 5×10^{-4} M (red) and 10^{-6} M (dashed red) are plotted in (c) to show the similarity in profile. Simulated SER spectra with Gaussian noise at (d) 0.1, (e) 1.0 and (f) 2.5% of the maximum peak intensity (black) with both the known background (red) and the calculated DTCWT-IA background (blue).

decomposition level: 8 (Fig. 3d to f in red)), added in Gaussian noise at 0.1, 1.0, and 2.5% of the maximum peak intensity, and then added the subtracted background back (Fig. 3d to f, in black). Each of the three resulting spectra were then individually analyzed by the DTCWT-IA (Fig. 3d to f in blue). For a fit using: first stage wavelet: Sym6; dual tree wavelet: qshift5; extension mode: constant; iteration: 44, 96, 177; decomposition level: 8 for 0.1, 1.0, and 2.5% noise levels, respectively. The DTCWT-IA performed well at recovering the known background, and required change only of iteration number. At the lowest level of noise, no change of iteration number was required for recovery. As the noise level increased, greater iteration number was required. The results slightly underestimate the known background, but this was a trade-off against the introduction of higher frequency noise components into the spectrum as a consequence of the asymmetry of the intensity filtering in the iteration implementation. Nevertheless, the disparity does not exceed the noise levels.

Sensitivity to Spectroscopy Type

With suitable precautions, the DTCWT-IA approach developed for ultrafast electron diffraction performed well in SERS spectral analysis. We extended the exploration to spectra acquired using UV–Vis and XPS, and compared DTCWT-IA performance to background subtractions by more conventional methods for all three optical

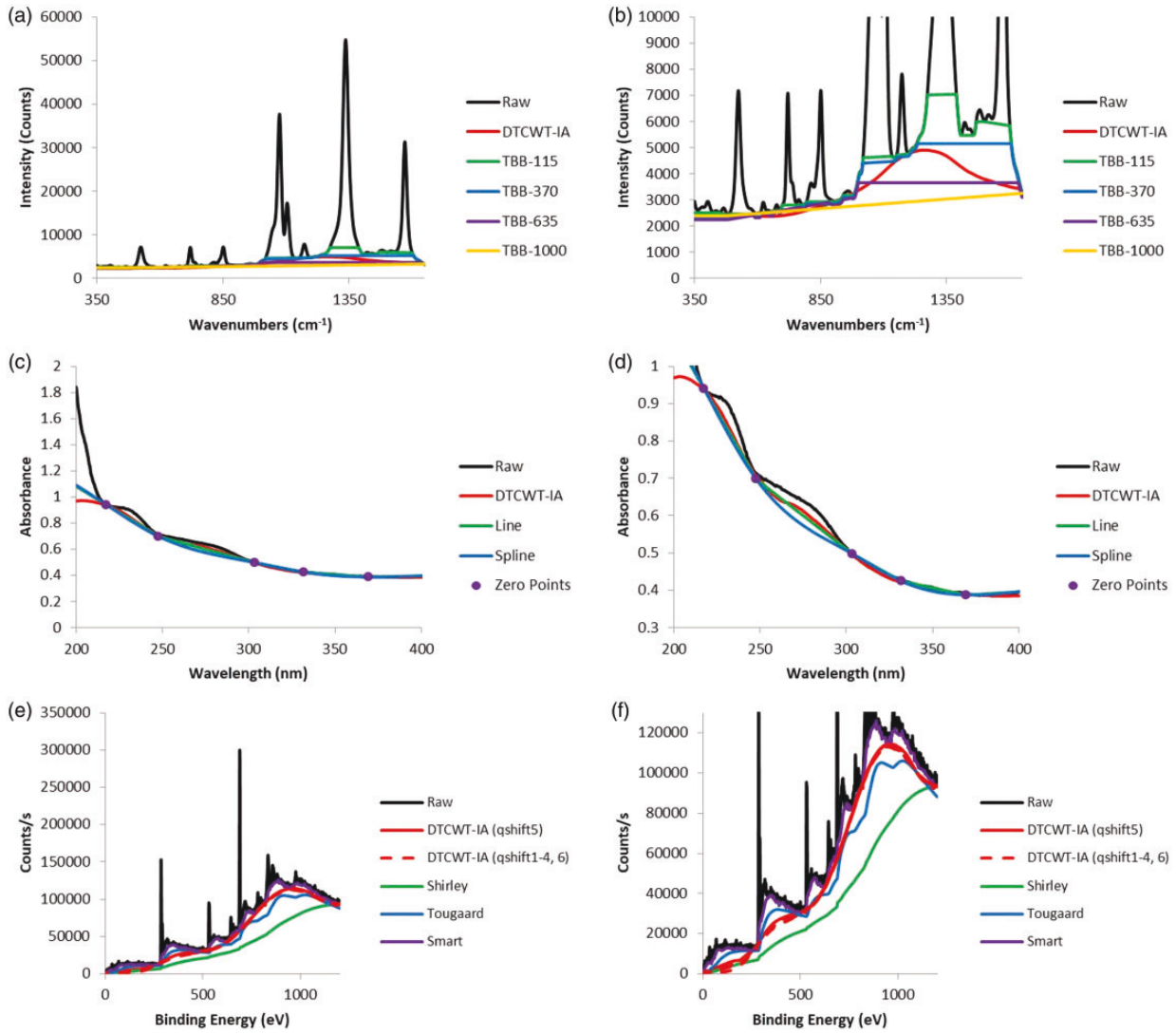

Figure 4. Calculated backgrounds for (a) and (b) SERS, (c) and (d) UV–Vis, and (e) and (f) XPS using DTCWT-IA, along with conventional background subtraction methods for each technique: TBB Baseline (SERS), line and spline (UV–Vis), and Shirley, Tougaard, and Smart (XPS). In (c) and (d), the dots show the points where the comparison background subtraction algorithms required specification where $S_{background}(x) = 0$, which were set automatically by the Origin Baseline analysis software. The dotted line in (e) and (f) shows DTCWT-IA backgrounds using different dual-tree wavelets, with all other parameters (first stage wavelet, extensions mode, decomposition level, and iteration) held constant.

spectroscopies. Optimized settings were: first stage wavelet: Sym6; dual-tree wavelet: qshift5; extension mode: constant; iteration: 64, 29, 123; decomposition level: 8, 6, 7, respectively.

Figure 4a and b compare the calculated backgrounds via DTCWT-IA and TBB Baseline at varying sensitivity (TBB-sensitivity). While the TBB method requires no user input other than sensitivity (1–1000), and the background-subtracted spectra (not shown) do not appear unusual, Fig. 4b shows that the TBB-Baseline calculated backgrounds not only overlap with the raw spectrum, but also cut directly into the raw signal (TBB-115, TBB-370, TBB-635).

Figure 4c and d compare conventional line and spline background subtraction methods to the DTCWT-IA background for the UV–Vis spectrum. Line and spline both need a set of points where $\mathrm{S_{background}(x)} = 0$, and this equates at the least to requiring peak-free regions of the spectrum. This circumstance may not exist in all spectral regions and the user may have no firm basis for the assignment, but such regions are nevertheless typically assigned by inspection. The particular background function is used to interpolate between each of these points. As observed in Fig. 4c and d, the line and spline methods both produce reasonable backgrounds for the UV–Vis spectrum, although with the need for the user to perform an initial spectral analysis to try to

find peak-free regions. The DTCWT-IA background closely resembles those produced by the two other methods, but without the necessity of having to identify spectral ranges free of peaks (and any other background contributions).

In the case of XPS, the problem of background subtraction has a robust theoretical foundation and numerous different algorithms can be found implemented in commercial XPS analysis software. In Fig. 4e and f, we compare the general DTCWT-IA background subtraction to the following XPS-specific background subtraction methods: Shirley, Tougaard, and a proprietary method known as Smart (using Avantage 5.9904 Thermo Fisher Scientific Inc.). The Shirley background explicitly accounts for material-dependent properties and inelastic scattering,[36,37] the Tougaard background[37–39] uses a universal approximation to calculate inelastic scattering signals, and the Smart background is a proprietary extension of the Shirley method with additional parameters to ensure the background does not have a greater intensity than the raw spectrum.[39] By varying decomposition level and iteration number, we achieved good fits (not shown) to Shirley, Tougaard, and Smart backgrounds. While in this work we kept the first stage wavelet, extensions mode, decomposition level, and iteration fixed, Fig. 4f shows that use of wavelets from qshift1 to qshift6 produced only minor differences in the background.

The DTCWT-IA background clusters well with the Tougaard and Smart backgrounds, with the Shirley underestimating the background relative to these other methods. The DTCWT-IA background contains less structure than its nearest neighbors, however, and is more similar in that regard to the Shirley background. In contrast to the other three background subtraction methods, DTCWT-IA does not require any input of material-dependent properties and the quality of the background profile is competitive with three widely used, commercially integrated background profiles that themselves have a fairly wide spread in intensity and form. Evaluations of background-subtraction methods frequently rely upon synthesized data where $S_{measured}(x)$ is constructed using known $S_{interest}(x)$ and $S_{background}(x)$, and the quality of the fit is established by how faithfully $S_{background}(x)$ is recovered. Here, with multiple "accepted" fits for an unknown $S_{background}(x)$, DTCWT-IA is competitive while placing low burdens on the user.

Conclusion

The DTCWT-IA technique has demonstrated utility for background subtraction in ultrafast electron diffraction, and we wanted to investigate its viability as a background subtraction tool for optical spectroscopy. The iterative component of the method was essential to ensuring the quality of the background subtraction, but the ability to differentiate (structured) peaks from (unstructured) background was more strongly affected by the decomposition level. The most obvious indications of insufficient iteration number or decomposition level are observed in an overly structured background profile that (i) overlaps small peaks and noise-like features of the measured spectrum (conveniently indicated by a high value of the RMS average of the profile derivative (RMS_D)), and (ii) causes the background to contain peaks aligned with those corresponding to the sample (conveniently indicated by the RMS_D and a peak integral that is too low). Agreement with conventional background subtraction methods was excellent even changing only two algorithm parameters (iteration number and decomposition level). The appearance of the background profile produced by at least one commercially implemented algorithm underscored the need to qualitatively assess the appearance of the background profiles, not just the background-subtracted spectra. In addition, we showed that while incorrect settings of the DTCWT-IA iteration number or decomposition level could contaminate the background profile with small, sharp peaks, suitable values could allow DTCWT-IA backgrounds to fit backgrounds to spectra with over $10\times$ changes in peak intensities. DTCWT-IA successfully generated background profiles for SERS, UV–Vis, and XPS, so that this algorithm holds the promise of providing background subtraction for a variety of optical spectroscopies with a low burden on the user for parameter selection or input, thus providing compatibility with straightforward processing of high-volumes of data.

Acknowledgments

We would like to acknowledge Professor Brett Lucht and his group at the University of Rhode Island, specifically Satu Heiskanen for providing representative XPS data to be used with the DTCWT-IA program. We thank Laurent P. René de Cotret (Siwick Group, Dept. of Physics, McGill University) for assistance with software implementation and algorithm performance.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by NSF OIA-1655221.

ORCID iD

Jason R. Dwyer https://orcid.org/0000-0002-2938-2888

References

1. C.M. Galloway, E.C. Le Ru, P.G. Etchegoin. "An Iterative Algorithm for Background Removal in Spectroscopy by Wavelet Transforms". Appl. Spectrosc. 2009. 63(12): 1370–1376.

2. G. Schulze, A. Jirasek, M.M.L. Yu, et al. "Investigation of Selected Baseline Removal Techniques as Candidates for Automated Implementation". Appl. Spectrosc. 2005. 59(5): 545–574.
3. H. Aitchison, J. Aizpurua, H. Arnolds, et al. "Analytical SERS: General Discussion". Faraday Discuss. 2017. 205: 561–600.
4. M.L. Weber, J.P. Litz, D.J. Masiello, et al. "Super-Resolution Imaging Reveals a Difference Between SERS and Luminescence Centroids". ACS Nano. 2012. 6(2): 1839–1848.
5. M.T. Moores, J. Carson, K. Gracie, et al. "Bayesian Modelling and Quantification of Raman Spectroscopy". Annals Appl. Statistics. 2018. arXiv:1604.07299v2.
6. A.C. Crawford, A. Skuratovsky, M.D. Porter. "Sampling Error: Impact on the Quantitative Analysis of Nanoparticle-Based Surface-Enhanced Raman Scattering Immunoassays". Anal. Chem. 2016. 88(12): 6515–6522.
7. J. Zhao, H. Lui, D.I. McLean, et al. "Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy". Appl. Spectrosc. 2007. 61(11): 1225–1232.
8. C.A. Lieber, A. Mahadevan-Jansen. "Automated Method for Subtraction of Fluorescence from Biological Raman Spectra". Appl. Spectrosc. 2003. 57(11): 1363–1367.
9. T.J. Vickers, R.E. Wambles Jr., C.K. Mann. "Curve Fitting and Linearity: Data Processing in Raman Spectroscopy". Appl. Spectrosc. 2001. 55(4): 389–393.
10. S.-J. Baek, A. Park, J. Kim, et al. "Chemometrics and Intelligent Laboratory Systems: A Simple Background Elimination Method for Raman Spectra". Chemom. Intell. Lab. Syst. 2009. 98(1): 24–30.
11. Y. Xie, L. Yang, X. Sun, et al. "An Auto-Adaptive Background Subtraction Method for Raman Spectra". Spectrochim. Acta. Part A. 2016. 161: 58–63.
12. S. Baek, A. Park, Y. Ahn, et al. "Baseline Correction Using Asymmetrically Reweighted Penalized Least Squares Smoothing". Analyst. 2015. 140(1): 250–257.
13. N. Wattanavichean, E. Casey, R.J. Nichols, et al. "Discrimination Between Hydrogen Bonding and Protonation in the Spectra of a Surface-Enhanced Raman Sensor". Phys. Chem. Chem. Phys. 2018. 20(2): 866–871.
14. D. King, W.B. Lyons, C. Flanagan, et al. "An Optical Fibre Ethanol Concentration Sensor Utilizing Fourier Transform Signal Processing Analysis and Artificial Neural Network Pattern Recognition". J. Opt. A Pure Appl. Opt. 2003. 5(4): S69–S75.
15. N.G. Kingsbury. "The Dual-Tree Complex Wavelet Transform: A New Technique for Shift Invariance and Directional Filters". In: Proceedings of the 8th IEEE DSP Workshop. Bryce Canyon, UT: August 9–12, 1998. Pp. 2543–2560.
16. F. Ehrentreich, L. Summchen. "Spike Removal and Denoising of Raman Spectra by Wavelet Transform Methods". Anal. Chem. 2001. 73(17): 4364–4373.
17. W. Cai, L. Wang, Z. Pan, et al. "Application of the Wavelet Transform Method in Quantitative Analysis of Raman Spectra". J. Raman Spectrosc. 2001. 32(3): 207–209.
18. P.M. Ramos, I. Ruisánchez. "Noise and Background Removal in Raman Spectra of Ancient Pigments Using Wavelet Transform". J. Raman Spectrosc. 2005. 36(9): 848–856.
19. Y. Hu, T. Jiang, A. Shen, et al. "A Background Elimination Method Based on Wavelet Transform for Raman Spectra". Chemom. Intell. Lab. Syst. 2007. 85(1): 94–101.
20. P.A. Mosier-Boss, S.H. Lieberman, R. Newbery. "Fluorescence Rejection in Raman Spectroscopy by Shifted-Spectra, Edge Detection, and FFT Filtering Techniques". Appl. Spectrosc. 1995. 49(5): 630–638.
21. C. Charles, G. Leclerc, P. Louette, et al. "Noise Filtering and Deconvolution of XPS Data by Wavelets and Fourier Transform". Surf. Interface Anal. 2004. 36(1): 71–80.
22. L.P. René de Cotret, B.J. Siwick. "A General Method for Baseline-Removal in Ultrafast Electron Powder Diffraction Data Using the Dual-Tree Complex Wavelet Transform". Struct. Dyn. 2017. 4(4): 1–6.
23. I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury. "The Dual-Tree Complex Wavelet Transform: A Coherent Framework for Multiscale Signal and Image Processing". IEEE Signal Proc. Mag. 2005. 22(6): 123–151.
24. F. Zhao, A. Wang. "A Background Subtraction Approach Based on Complex Wavelet Transforms in EDXRF". X-ray Spectrom. 2015. 44(2): 41–47.
25. H.-W. Tan, S.D. Brown. "Wavelet Analysis Applied to Removing Non-Constant, Varying Spectroscopic Background in Multivariate Calibration". J. Chemom. 2002. 16(5): 228–240.
26. C. Charles, G. Leclerc, J.J. Pireaux, et al. "Introduction to Wavelet Applications in Surface Spectroscopies". Surf. Interface Anal. 2004. 36(1): 49–60.
27. C. Charles, G. Leclerc, J.P. Rasson, et al. "HREELS Signal Processing via Wavelets". Surf. Interface Anal. 2004. 36(1): 61–70.
28. L.P.R. De Cotret, M.R. Otto, M.J. Stern, et al. "An Open-Source Software Ecosystem for the Interactive Exploration of Ultrafast Electron Scattering Data". Adv. Struct. Chem. Imaging. Springer International Publishing, 2018. 4(1): 11.
29. J.R. Dwyer, R.E. Jordan, C.T. Hebeisen, et al. "Femtosecond Electron Diffraction: An Atomic Perspective of Condensed Phase Dynamics". J. Mod. Opt. 2007. 54(7): 905–922.
30. J.R. Dwyer, C.T. Hebeisen, R. Ernstorfer, et al. "Femtosecond Electron Diffraction: 'Making the Molecular Movie'". Phil. Trans. R. Soc. A. 2006. 364(1840): 741–778.
31. A.A. Ischenko, P.M. Weber, R.J.D. Miller. "Capturing Chemistry in Action with Electrons: Realization of Atomically Resolved Reaction Dynamics". Chem. Rev. 2017. 117(16): 11066–11124.
32. L.P. René de Cotret, B.J. Siwick. "Software Projects: The Siwick Research Group". http://www.physics.mcgill.ca/siwicklab/software.html [accessed Mar 18 2018].
33. B.I. Karawdeniya, Y.M.N.D.Y. Bandara, J.C. Whelan, et al. "General Strategy to Make an On-Demand Library of Structurally and Functionally Diverse SERS Substrates". ACS Appl. Nano Mater. 2018. 1(2): 960–968.
34. B.I. Karawdeniya, Y.M.N.D.Y. Bandara, J.W. Nichols, et al. "Surveying Silicon Nitride Nanopores for Glycomics and Heparin Quality Assurance". Nat. Commun. 2018. 9(3278): 1–8.
35. B.L. Lucht, M. Tippawan, L. Yang. "Thermal Stability of Lithium Ion Battery Electrolytes". In: R. Crabtree, editor. Energy Production and Storage: Inorganic Chemical Strategies for a Warming World. Chichester, UK: Wiley, 2010. Pp. 333–340.
36. D.A. Shirley. "High-Resolution X-ray Photoemission Spectrum of the Valence Bands of Gold". Phys. Rev. B. 1972. 5(12): 4709–4714.
37. S. Tougaard. "Quantification of Nano-Structures by Electronspectroscopy". In: D. Briggs, J.T. Grant, editors. Surface Analysis by Auger and X-ray Photoelectron Spectroscopy. Chichester, UK: IM Publications, 2003. Pp. 295–344.
38. S. Tougaard. "Quantitative Analysis of the Inelastic Background in Surface Electron Spectroscopy". Surf. Interface Anal. 1988. 11(9): 453–472.
39. T. Carney. "Basic and Advanced Processing". http://www.thermo.com.cn/Resources/201411/2412740468.pdf [accessed Aug 13 2018].