Marine Pollution Bulletin 190 (2023) 114887

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

A double sliding-window method for baseline correction and noise estimation for Raman spectra of microplastics

Zijiang Yang *, Hisayuki Arakawa

Tokyo University of Marine Science and Technology, Konan 4-5-7, Minato-Ku, Tokyo 108-8477, Japan

ARTICLE INFO

Keywords:
Raman spectroscopy
Raman spectrum
Baseline correction
Microplastics
Automated identification

ABSTRACT

When measuring microplastics of environmental samples, additives and attachment of biological materials may result in strong fluorescence in Raman spectra, which increases difficulty for imaging, identification, and quantification. Although there are several baseline correction methods available, user intervention is usually needed, which is not feasible for automated processes. In current study, a double sliding-window (DSW) method was proposed to estimate the baseline and standard deviation of noise. Simulated spectra and experimental spectra were used to evaluate the performance in comparison with two popular and widely used methods. Validation with simulated spectra and spectra of environmental samples showed that DSW method can accurately estimate the standard deviation of spectral noise. DSW method also showed better performance than compared methods when handling spectra of low signal-to-noise ratio (SNR) and elevated baselines. Therefore, DSW method is a useful approach for preprocessing Raman spectra of environmental samples and automated processes.

1. Introduction

When measuring microplastics, weathering, additives and attachment of biological materials may result in strong fluorescence in Raman spectrum compared with Fourier transform infrared (FTIR) spectrum (Lenz et al., 2015; Käppler et al., 2016; Anger et al., 2018; Ghosal et al., 2018; Araujo et al., 2018; Dong et al., 2020; Song et al., 2021). As a result, small signal-to-noise ratio (SNR) and elevated baselines could lead to lower spectra similarities (Araujo et al., 2018; Renner et al., 2019) and problems in chemical identification (Song et al., 2014; Lenz et al., 2015; Ghosal et al., 2018; Dong et al., 2022). As long as unfavorite baseline is corrected, further analysis, such as chemical identification, quantification of weathering and degradation (Almond et al., 2020; Phan et al., 2022; Walfridson and Kuttainen Thyni, 2022), and cross comparison with other studies (Cowger et al., 2021) could be reliable. Therefore, it is necessary to improve quality of spectra.

Quality of Raman spectrum could be improved by instrumental method and mathematical method (Hu et al., 2018). For instrumental method, increasing exposure time and adjusting laser parameters could lower the interference from fluorescent and level of noise (Lenz et al., 2015; Anger et al., 2018; Cabernard et al., 2018; Schymanski et al., 2018). However, such method is usually time consuming and labor intense (Dong et al., 2022). In addition, measurement under the same condition is more feasible for automated processes, development of standard methods, and regulations. For mathematical method, the spectra are mathematically preprocessed to eliminate the elevated baseline and lower the noise. Compared with instrumental methods, mathematical methods are usually inexpensive in time and labor (Hu et al., 2018). Mathematical method also has some challenges of overfitting and requirement of time and computational resources, and it may also require a certain level of mathematical expertise. However, recent advancements in model validation techniques, as well as increased computational power and availability of user-friendly coding programming languages, have helped to address these limitations. As a result, the use of mathematical methods for spectral analysis is becoming more accessible. Therefore, mathematical method would be a practical solution.

Currently, baseline correction is usually included in spectrum processing software, and user can remove baseline manually (Zhang et al., 2015; Wesch et al., 2016; Yu et al., 2016; Karami et al., 2017; Almond et al., 2020). However, algorithms in the software are often unclear, and manual operations are usually time consuming, labor intense, and subjective (Jirasek et al., 2004; Renner et al., 2019). Besides, there are several other available baseline correction methods that have potential

* Corresponding author.
E-mail addresses: zyan001@kaiyodai.ac.jp (Z. Yang), arakawa@kaiyodai.ac.jp (H. Arakawa).

https://doi.org/10.1016/j.marpolbul.2023.114887
Received 10 January 2023; Received in revised form 19 March 2023; Accepted 24 March 2023
Available online 4 April 2023

to address the problems. Polynomial fitting is a widely used baseline correction method (Lieber and Mahadevan-Jansen, 2003; Zhao et al., 2007; Baek et al., 2009; Hu et al., 2018; Liu et al., 2015; Renner et al., 2019; Cowger et al., 2021), where a polynomial function is used to fit the spectrum. However, this method usually has difficulties in choosing the proper loss function (Liu et al., 2015), and the fitting results depends on the order of polynomial function (Hu et al., 2018). Wavelet decomposition is a method based on decomposition of the spectrum by fitting with wavelet functions (Xu et al., 2005; Asfour et al., 2011; Du et al., 2010; Li et al., 2014, 2015; Xi et al., 2018). This method assumes that the spectrum is a combination of multiple wavelets of different frequencies that reflect peaks and baseline (Li et al., 2015). However, it is usually difficult to select the proper scale and some information may be lost during decomposition (Liu et al., 2015; Xi et al., 2018). Least squares method is another option (Baek et al., 2015; Zhang et al., 2010; He et al., 2014), where a specified least squared model is used, and the noisy dataset is fitted by spline interpolation. However, selection of a proper least square model and determination of best polynomial orders are usually difficult (Hu et al., 2018). Recently, machine learning methods have been applied to baseline correction problems, and machine learning methods showed promising performance for baseline estimation (Liu et al., 2017; Chen et al., 2022). However, machine learning methods usually depend on a large training dataset and might be case-specific if training data is not representative, which raised difficulties of use by other researchers.

The sliding-window coupled with local minima is another useful method for baseline correction. While the term *sliding-window method* is generally used to describe various methods that employ sliding-window (Schulze et al., 2011, 2012; Gierlinger et al., 2012; Stanford et al., 2016; Sikora et al., 2019; Golotvin and Williams, 2000; Yuan et al., 2021; Schmid et al., 2022; Yang et al., 2022), in this paper sliding-window method specifically refer to the local minima-based sliding-window method for baseline correction.

Sliding window method was originally developed for baseline correction in mass spectrometry (Crecelius et al., 2011; Zimmerman et al., 2009; Fan et al., 2016; Povey et al., 2014), and it is based on the local minima of the signal intensity within each window, which are then interpolated using shape-preserving cubic interpolation techniques (Utsunomiya et al., 2014; Tanaka et al., 2014). However, the sliding-window method has also been increasingly used in Raman spectroscopic analysis, for applications in food (Amjad et al., 2018; Radzol et al., 2014), blood serum (Khan et al., 2018; Mankova et al., 2020), plasma (Bilal et al., 2015), disease detection (Yu et al., 2019; Ullah et al., 2020), biomaterials (Kaya et al., 2017; Uckermann et al., 2018; Nowacki et al., 2020; Karmenyan et al., 2020), and polymer analysis (Noack et al., 2013; Lenz et al., 2015; Zimmerer et al., 2019a,b) etc., as well as in FTIR spectroscopic analysis, for applications in sexing (Steiner et al., 2016), disease detection (Urboniene et al., 2014), chemical quantification (Muller et al., 2010; Wang et al., 2019), and polymer analysis (Genest et al., 2013; Zimmerer et al., 2019a,b) etc. Moreover, the sliding-window method has also found extensive use in signal processing applications in other fields, such as electrocardiography (Casas et al., 2015; Gonzalez et al., 2015), chromatography (Jiménez-Carvelo et al., 2017; Kumar and Cava, 2019), and neurology (Sathyanesan et al., 2012; Kellner et al., 2021) etc.

Sliding-window method can better reflect local fluctuation of baseline than polynomial and least squared methods, and it is also more intuitive than wavelet decomposition and machine learning method. However, there are two potential problems that limit its application. Firstly, the estimated baseline is biased and systematically below the spectrum signal, which was due to the fact that the sliding-window method uses interpolated local minimal values to estimate baseline. Secondly, the estimated baseline is sensitive to window size. If window size is small, the baseline estimated by using small window size can capture the fluctuations in baseline while wide and large peaks could also be identified as baseline; if window size is large, baseline estimated by using large window size can correctly estimate the baseline of wide and large peaks, but the fluctuations in baseline could not be captured.

Therefore, the purpose of the current study is to take advantages of sliding-window method, address its potential problems, and improve its feasibility in spectrum analysis. In this paper, a double sliding-window (DSW) method is developed to estimate the baseline of a spectrum and in the meanwhile estimate standard deviation of noise. In DSW method algorithm, standard deviation of noise is estimated by the most frequently appeared distance between upper and lower envelope functions of the spectrum. Next, the estimated standard deviation of noise is used to determine peaks and estimate the bias of the original sliding-window method. The determined peaks together with estimated standard deviation of noise are used for determination of the weights between baseline of small window size and baseline of large window size. Then, the bias of estimated baselines from original sliding-window method is corrected by using information of noise. Finally, the bias-free baselines are combined according to the weights. In order to evaluate the performance of DSW method, the proposed baseline correction method was compared with commonly used polynomial fitting method and least squared method with simulated spectra and experimental spectra of environmental samples.

2. Materials and method

2.1. Problem formulation

A spectrum is defined as a function between wavenumber $\boldsymbol{X}$ and signal intensity $\boldsymbol{Y}$:

$$\boldsymbol{Y} = f(\boldsymbol{X}) \tag{1}$$

where $\boldsymbol{Y}$ refers to signal intensity and $\boldsymbol{X}$ refers to wavenumber (cm^{-1}). The signal of a spectrum consists of three components (Chen and Hsu, 2019):

$$\boldsymbol{Y} = \boldsymbol{Y}_{\mathrm{pk}} + \boldsymbol{Y}_{\mathrm{bc}} + \boldsymbol{Y}_{\mathrm{ns}} \tag{2}$$

where $\boldsymbol{Y}_{\mathrm{pk}}$ is the signal of peak, and it may come from substance of interest or additives. In current study, $\boldsymbol{Y}_{\mathrm{pk}}$ is defined as the signals that are statistically different from noise. $\boldsymbol{Y}_{\mathrm{bc}}$ is the signal of baseline, and it may come from florescence and disturbance from environment. $\boldsymbol{Y}_{\mathrm{ns}}$ is the signal of noise, and the noise is assumed to follow a normal distribution (Smulko, 2019):

$$\boldsymbol{Y}_{\mathrm{ns}} \sim N\left(0, \sigma_{\mathrm{ns}}^2\right) \tag{3}$$

where σ_{ns} is the standard deviation of noise signal. Eq. (3) indicates that the center of noise in the wavenumber regions without signals of peak is the baseline, and this is a clue for evaluating if the baseline is accurately estimated.

For derivation purpose, we defined an imaginary spectrum of pure noise, and the imaginary spectrum has the same mean and standard deviation with the noise of the original spectrum:

$$\boldsymbol{Y}_0 = \boldsymbol{0} + \boldsymbol{0} + \boldsymbol{Y}_{\mathrm{ns}} \tag{4}$$

Additionally, we defined $\boldsymbol{Y}_{\mathrm{up}}$ and $\boldsymbol{Y}_{\mathrm{lw}}$ as the upper and lower envelope functions of the spectrum, respectively. Then, the center of envelope functions, $\boldsymbol{Y}_{\mathrm{ct}}$, is calculated as follows:

$$\boldsymbol{Y}_{\mathrm{ct}} = \frac{1}{2}\left(\boldsymbol{Y}_{\mathrm{lw}} + \boldsymbol{Y}_{\mathrm{up}}\right) \tag{5}$$

A spectrum can be transformed by center of its envelope functions, and such transformation is defined as envelope-center transformation (EC-transformation):

$$\boldsymbol{Y}^{'} = \boldsymbol{Y} - \boldsymbol{Y}_{\mathrm{ct}} \tag{6}$$

where $\boldsymbol{Y}'$ is the EC-transformed spectrum of $\boldsymbol{Y}$. Similarly, the imaginary

spectrum can be EC-transformed into $\boldsymbol{Y}_0'$.

In summary, based on Eqs. (2) and (3), σ_{ns}, $\boldsymbol{Y}_{bc}$, and $\boldsymbol{Y}_{pk}$ are the parameters to be estimated. In following sections, estimation of these parameters is described.

2.2. Estimation of standard deviation of noise, σ_{ns}

In Eq. (3), standard deviation of the noise signal (σ_{ns}) is estimated by quantiles of the EC-transformed imaginary spectrum:

$$\widehat{\sigma}_{ns} = \frac{1}{2 \times 1.96}\left(Q'_{0.975} - Q'_{0.025}\right)\cdot f_{sk} \tag{7}$$

where $Q'_{0.975}$ and $Q'_{0.025}$ are the 97.5 % and 2.5 % and quantiles of the EC-transformed imaginary spectrum ($\boldsymbol{Y}_0'$). 2×1.96 corresponds to the relation between standard deviation and quantiles of 97.5 % and 2.5 % for a normal distribution (Casella and Berger, 2002). f_{sk} is the skewness correction factor.

$Q'_{0.975}$ and $Q'_{0.025}$ and can be calculated as follows:

$$Q'_{0.975} = \frac{1}{2}\cdot d_{ns}\cdot f_{up} \tag{8}$$

$$Q'_{0.025} = -\frac{1}{2}\cdot d_{ns}\cdot f_{lw} \tag{9}$$

where d_{ns} is defined as *noise range*, which is calculated by the distance between the upper envelope function, $\boldsymbol{Y}_{up}$, and the lower envelope function, $\boldsymbol{Y}_{lw}$, of the spectrum:

$$\boldsymbol{d}_{ns} = \boldsymbol{Y}_{up} - \boldsymbol{Y}_{lw} \tag{10}$$

$1/2\cdot d_{ns}$ in Eqs. (8) and (9) is approximately the amplitude of noise, which is also the proximate distance between local maximum or local minimum of the noise signal and the center of noise. Thus, the center of noise can be estimated by bottom envelop plus $1/2\cdot d_{ns}$. However, due to randomness of noise signal, the envelope functions of imaginary spectrum are not straight lines, but curves with fluctuations. As a result, randomized d_{ns} values are obtained from envelope functions along $\boldsymbol{X}$. To elucidate the properties of d_{ns}, spectra of pure noise signals were simulated, and corresponding envelope functions were calculated. The results revealed that there was a functional relationship between the quantiles and $1/2\cdot d_{ns}$, and such functional relation was also related to window size, l_{ws}, which is used to calculate the envelope functions.

For imaginary spectrum, $\boldsymbol{Y}_0$, the most frequently appeared distance between envelope functions along $\boldsymbol{X}$ is used as an estimate of d_{ns}:

$$\widehat{d}_{ns} = \mathrm{argmax}(f_d(d_{ns}|\boldsymbol{Y}_0)) \tag{11}$$

where $f_d(d_{ns}|\boldsymbol{Y}_0)$ is the probability density function (PDF) of d_{ns} given the imaginary spectrum $\boldsymbol{Y}_0$. Because noise of spectrum is more likely to result in similar envelope distance than peaks, the most frequently appeared d_{ns} value of imaginary spectrum should equal to the most frequently appeared d_{ns} value of the original spectrum, $\boldsymbol{Y}$. Thus, the following relation holds:

$$\mathrm{argmax}(f_d(d_{ns}|\boldsymbol{Y})) = \mathrm{argmax}(f_d(d_{ns}|\boldsymbol{Y}_0)) \tag{12}$$

where $f_d(d_{ns}|\boldsymbol{Y}_0)$ is the PDF of d_{ns} given the real spectrum $\boldsymbol{Y}$. Combining Eqs. (11) and (12), d_{ns} can be estimated by empirical PDF of d_{ns} based on the spectrum data.

During calculation of envelope functions, the local minimal and maximal signals within a window size, l_{ws}, are interpolated by shape-preserving piecewise cubic interpolation (Lenz et al., 2015). According to the algorithm, local minimal and maximal signals depend on window size. Thus, window size could influence estimated envelope functions and further influence the relation between d_{ns} and quartiles. In order to quantify such relationship, Monte Carlo simulation (l_{ws} = 10 to 100 cm^{-1} with interval of 1 cm^{-1}, $n = 1000$) of an imaginary spectrum with pure noise was conducted. The empirical relationship between f_{up}, f_{lw} and l_{ws} were found as follows (Fig. S1a, b, $R^2 > 0.99$, $p < 0.05$ for both):

$$\widehat{f}_{up} = 3.777 \times l_{ws}^{-0.5296} + 0.8164 \tag{13}$$

$$\widehat{f}_{lw} = 0.9594 \times l_{ws}^{-0.6008} + 0.9407 \tag{14}$$

Eqs. (13) and (14) indicate a slightly asymmetric relation between d_{ns} and quantiles, i.e., normally distributed noise becomes slightly skewed after EC-transformation (a demonstration is provided in Fig. S2 with a brief explanation). Therefore, a correction factor was used to account for such skewness in Eq. (7). Based on Monte Carlo simulation ($n = 1000$, l_{ws} = 10–100 cm^{-1} with interval of 1 cm^{-1}), the empirical relationship between skewness correction factor, f_{sk}, and l_{ws} are as follows (Fig. S1c, $R^2 > 0.99$, $p < 0.05$):

$$\widehat{f}_{sk} = 0.2246 \times l_{ws}^{-0.1126} + 0.2369 \tag{15}$$

In summary, the signal noise, σ_{ns}, could be estimated by spectrum data, $\boldsymbol{Y}(\boldsymbol{X})$, and empirical values of f_{up}, f_{lw}, and f_{sk}.

2.3. Estimation of baseline, $\boldsymbol{Y}_{bc}$, and peak, $\boldsymbol{Y}_{pk}$

In Eq. (3), signal of baseline, $\boldsymbol{Y}_{bc}$, is calculated based on weighted combination of two potential baselines:

$$\widehat{\boldsymbol{Y}}_{bc} = \boldsymbol{f}_s\cdot\left(\boldsymbol{Y}_{bc|s} + f_{lw}\cdot\frac{1}{2}d_{ns}\right) + \boldsymbol{f}_l\cdot\left(\boldsymbol{Y}_{bc|l} + f_{lw}\cdot\frac{1}{2}d_{ns}\right) \tag{16}$$

where $\widehat{\boldsymbol{Y}}_{bc}$ is estimated baseline. $\boldsymbol{Y}_{bc|s} + f_{lw}\cdot 1/2d_{ns}$ and $\boldsymbol{Y}_{bc|l} + f_{lw}\cdot 1/2d_{ns}$ are the two potential baselines based on l_{ws} (small window) and 5 times of l_{ws} (large window). $\boldsymbol{Y}_{bc|s}$ and $\boldsymbol{Y}_{bc|l}$ are calculated based on shape-preserving cubic interpolation of local minimal of each window size l_{ws} and $5l_{ws}$, respectively (Lenz et al., 2015). Using local minimal values as baseline introduced bias, and the bias is approximated by the distance between center of noise and $Q'_{0.025}$. Thus, the term of $f_{lw} \cdot 1/2d_{ns}$ is added to the baseline calculated from local minimum (Eq. (7)). As a result, the bias is corrected. f_s and f_l are the corresponding two weighting factors for $\boldsymbol{Y}_{bc|s}$ and $\boldsymbol{Y}_{bc|l}$, and they were defined as follows:

$$f_l(X_i \pm x) = -\frac{x}{l_m} + 1 \quad \forall X_i \in \{X_i\}_{pk} \quad \forall x \in [0, l_m] \tag{17}$$

$$f_l(X_i) = 0 \quad \forall X_i \notin \left[\{X_i\}_{pk} \pm l_m\right] \tag{18}$$

$$f_s(X_i) = 1 - f_l(X_i) \quad \forall X_i \tag{19}$$

where l_m (cm^{-1}) is the smoothing length when combining two potential baselines, and l_m = 11 cm^{-1} is used (Nakano et al., 2021). X_i is the wavenumber of spectrum. $\{X_i\}_{pk}$ refers to a set of X_i that is corresponding to signal of peaks, which satisfies the following conditions:

$$\{X_i\}_{pk} = arg\left(\boldsymbol{Y}(X_i) - \left(\boldsymbol{Y}_{bc|s}(X_i) + \frac{1}{2}f_{lw}\cdot\frac{1}{2}d_{ns}\right) > Q'_{0.995}\right) \tag{20}$$

Given that $Q'_{0.005}$ and $Q'_{0.995}$ perform as an empirical 99 % confident interval (CI) (Martinez and Martinez, 2001). $Q'_{0.995}$ was approximated by $Q'_{0.975}$ times an exaggeration factor 1.31, which is the ratio between 95 % CI and 99 % CI (Casella and Berger, 2002). Eq. (20) indicates that the signals of peaks are equivalent to the signals that are significantly greater than signal of noise at $\alpha = 0.01$. In addition, as it is possible to have isolated noise signals above $Q'_{0.995}$, for X_i that only >4 points within $[X_i \pm l_m]$ is counted as in $\{X_i\}_{pk}$.

Eqs. (17)–(20) indicates that (1) if there is a peak, then the baseline of large window size is used, and when there is no peak, then the baseline of small window size is used, (2) in order to obtain a continuous baseline, for regions with isolated signals of peak and noise, the weight of the isolated signals is averaged within the range of l_m based on the relative abundance. As a result, the influence of isolated peak signals among noise or isolated noise signals around peaks could be minimized.

With estimated baseline $\widehat{\boldsymbol{Y}}_{\mathrm{bc}}$, signal of peak can be estimated by following:

$$\widehat{\boldsymbol{Y}}_{\mathrm{pk}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}_{\mathrm{bc}} \tag{21}$$

The algorithms were programmed in Matlab 2022b (The MathWorks, Inc. Natick, MA, USA), and the script and an example dataset are available online (https://github.com/River20104047/DSWmethod).

2.4. Simulated spectra

In order to evaluate and validate the performance of the proposed DSW method, simulation study was conducted. In practice, sample spectra varied in terms of SNR and shape of baseline. Thus, spectra of different combinations of SNR and shape of baseline were simulated for performance evaluation. In current study, SNR was classified into two categories, large SNR (N1) and small SNR (N2), which correspond to small noise and large noise in the spectrum. Baseline was classified into four categories: flat baseline (B1), elevated baseline (B2), fluctuating

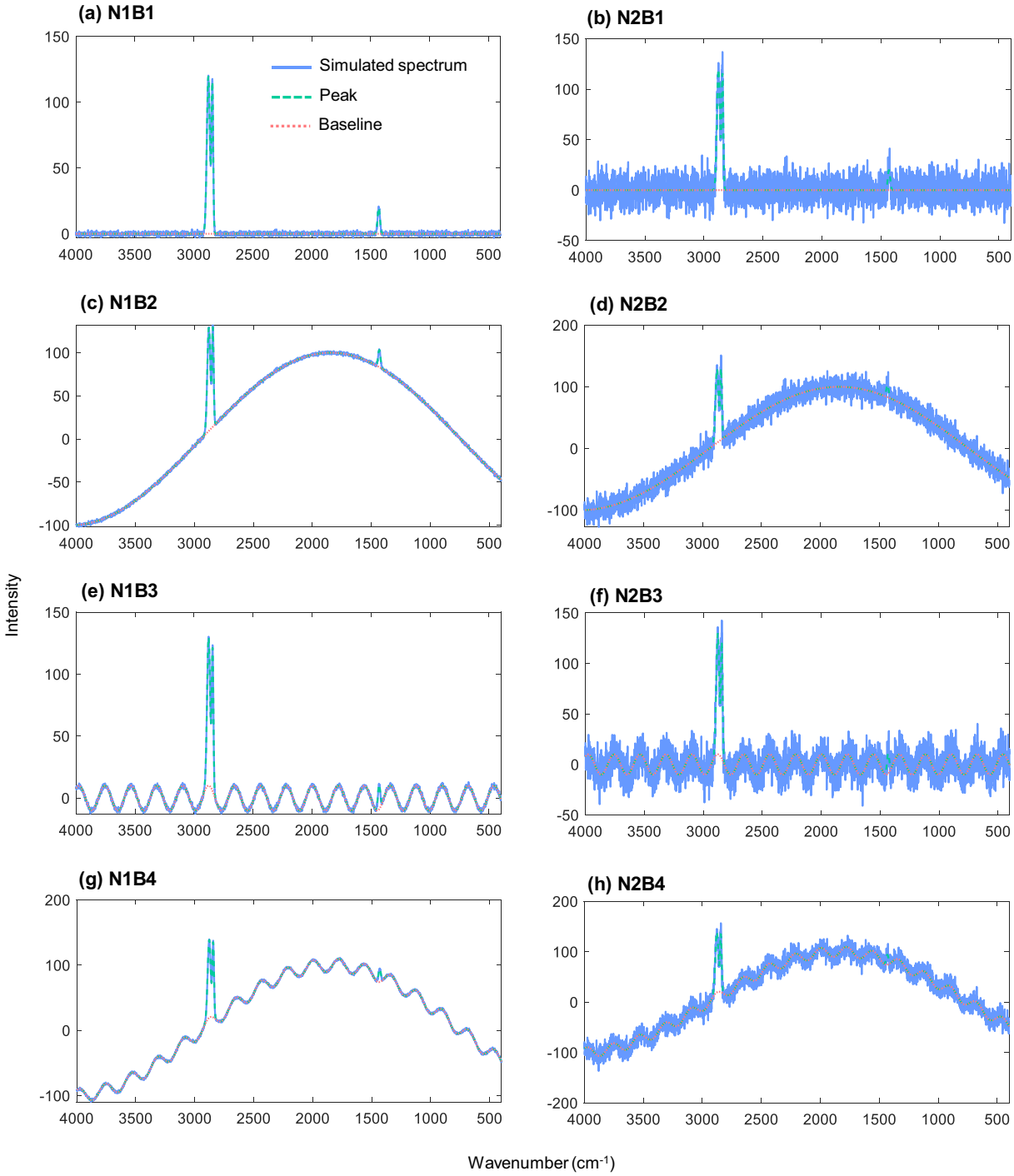

Fig. 1. Demonstration of simulated spectra with different SNR and baseline combinations. N1 refers to large SNR and N2 refers to small SNR. B1 refers to flat baseline, B2 refers to elevated baseline ($\boldsymbol{Y}_{\mathrm{e}} = 100\sin(0.00143\boldsymbol{X} + 200)$), B3 refers to fluctuating baseline ($\boldsymbol{Y}_{\mathrm{f}} = 10\sin(0.0286\boldsymbol{X} - 200)$), and B4 refers to elevated and fluctuating baseline ($\boldsymbol{Y}_{\mathrm{ef}} = \boldsymbol{Y}_{\mathrm{e}} + \boldsymbol{Y}_{\mathrm{f}}$). The parameters of baseline were adjusted and determined based on typical low-quality spectra in sample (e.g., Fig. 6a).

baseline (B3), and elevated and fluctuating baseline (B4). In current study, the combinations were expressed as the combination of the labels. For example, N1B1 refers to the simulated spectrum with large SNR and flat baseline.

Simulation of spectrum was based on the model of Eq. (2), where a single peak was assumed to be Gaussian, and the overlapped peaks were the linear combination of each peak (Oller-Moreno et al., 2014; Xu et al., 2021a). Since polyethylene (PE) is the commonly detected type of MPs (Andrady, 2017; Wang et al., 2021; Xu et al., 2022), PE spectrum was used as a reference. Thus, three peaks were used in current study, and they were α-peak at 2877 cm^{-1}, β-peak at 2843 cm^{-1}, and γ-peak at 1432 cm^{-1}. α-Peak and β-peak were tall and overlapped while the top peaks were separated, which represent the tall and closely positioned CH_3 peaks in PE spectrum (Larkin, 2017). γ-Peak was short and could be overlaid by large noise, which represents the peaks at 1000 to 1500 cm^{-1} (Lenz et al., 2015). For baseline, two sine wave functions were used for elevated baseline and fluctuation of baseline, respectively, and the relative amplitude of each sine wave function was determined based on representative sample spectra with elevated baselines. Noise consisted of the random values drawn from a normal distribution with zero mean and a specified standard deviation. Without loss of generality, standard deviation was set to 1 for large SNR, and 10 for small SNR while other parameters remain the same. Range of simulated spectra was 400 to 4000 cm^{-1}, which matched range of experimental spectra. Demonstration of simulated spectra is shown in Fig. 1.

2.5. Experimental spectra

Simulated spectra may not be able to reflect the complexity of real spectra, and thus, the performance of baseline correction methods was also evaluated by experimental spectra. Experimental spectra consist of two components, spectra of standard plastics and spectra of environmental samples from sea surface water. In current study, standard plastic samples of polyamide resin (PA), polycaprolactone (PLC), polycarbonate (PC), polyethylene (PE), polyethylene tetrphthalate (PET), polypropylene (PP), polystyrene (PS), and polyvinyl chloride (PVC) (Scientific Polymer Products Inc., USA) were used. Environmental plastic samples from a previous study (Çelik et al., 2023) were subsampled and re-analyzed, where the plastic samples were obtained from surface seawater at two sampling sites of offshore Shimane Prefectures in the Sea of Japan and Toyama Bay area. The subsamples (in total n = 413) were identified by attenuated total reflectance (ATR) FTIR spectroscopy. The analysis identified the subsamples as 39 % PE, 24 % PP, and 29 % PS. More detailed information about the sampling and identification processes can be found in the Supplementary information (Araujo et al., 2018; Cowger et al., 2021; Nakano et al., 2021; Ouyang et al., 2022; Çelik et al., 2023).

Then, the environmental plastic samples together with standard plastic samples were measured by Raman spectrometer (NRS-4500, JASCO Inc., Japan). Common parameter combinations from literature (Nava et al., 2021; Dong et al., 2022) were tested and tuned with first couple of samples before applying fixed parameters for all samples. After preliminary evaluation, the parameters were determined as wavenumber range of 400–4000 cm^{-1} with 1 s exposure, φ = 17 μm slit, 16 accumulations, 532 nm excitation laser, and 20× objective as a balance among instrument, analysis speed, and spectra quality. In order to minimize impact of local heterogeneity of the sample, Raman spectra were acquired at two different spots on sample surface. If the two spectra were substantially different, then two more spectra were acquired at another spots, and the most frequently appeared spectra pattern were kept.

Preliminary evaluation of experimental spectra showed that some spectra had small SNR with elevated baseline. These spectra may provide some information, but due to low-quality of spectra, preprocessing is needed. Thus, the ability of baseline correction methods to recover the low-quality spectra was evaluated. Some experimental spectra and spectra of standard plastic samples have large SNR with sharp peaks and flat baseline. For these cases, baseline correction may not be needed. However, within automated processes, all spectra are subject to being preprocessed under the same conditions, including standard samples that are used as a reference for comparison or identification (Xu et al., 2019; Weisser et al., 2022). Thus, it is also necessary to assess the performance of baseline correction algorithms to spectra of standard samples as a representative of sample spectra with good SNR.

In current study, experimental spectra are classified in 4 + 1 categories. The four categories include: Type A (≈N1B1) with large SNR and flat baseline, Type B (≈N1B4) with large SNR but elevated and fluctuating baseline, Type C (≈N2B1) with small SNR and flat baseline, and Type D (≈N2B3) with small SNR and elevated baseline. Another category was included as Type Z, which consists of pure noise. Type Z could be removed during manual measurement and analysis, but during automated process, this type of spectra could also be processed. Therefore, the evaluation of performance on spectrum of pure noise was necessary.

2.6. Evaluation of baseline estimation performance

The proposed DSW method was compared with two popular and readily available baseline correction methods, adaptive iteratively reweighted penalized least square method (airPLS) (Zhang et al., 2010) and improved modified multi-polynomial fit method (iModPoly) (Zhao et al., 2007). In application of DSW method, the window sizes in current study were determined based previous application of original sliding-window method (Lenz et al., 2015), and l_{ws} = 25 cm^{-1} was used. In application of airPLS method, the default parameters were used, but if needed, the square order was adjusted to obtain an optimal fit. For iModPoly method, the default polynomial order 5 was used, but the order was adjusted if needed to obtain optimal fit. According to preliminary trials, it was found that as polynomial order got large, the fitting might be better for some cases, but the fitting time also increased drastically. Thus, only polynomial order <10 was used for current test.

While evaluating the performance of baseline correction, the performance was classified into four types. *Perfect* means that estimated baseline captured the center of noise and had negligible influence on peak. *Good* means that the estimated baseline almost captured the center of noise, and it had small influence on the shape of peak. Perfect and Good were considered as *acceptable*. *Neutral* means that the estimated baseline may have problems of capturing the center of noise, or had substantial influence on peaks, but the general shape of spectrum was still recognizable if manually evaluated. In current study, we assumed that the baseline correction method would be applied in automated process. Thus, Neutral was considered as not acceptable. *Failed* refers to the situation where the center of noise was not captured, or the shape of peak had a substantial difference from original peak. Failed was also considered as not acceptable.

Besides descriptive evaluation, root mean square error (*RMSE*) was also used to evaluate the performance of different methods (Xu et al., 2021b). *RMSE* is calculated by:

$$RMSE = \sqrt{\frac{\sum (\boldsymbol{Y}_{-\mathrm{bc}} - \widehat{\boldsymbol{Y}}_{-\mathrm{bc}})^2}{N}} \tag{22}$$

where $\boldsymbol{Y}_{-\mathrm{bc}}$ is the baseline-free simulated spectrum, $\widehat{\boldsymbol{Y}}_{-\mathrm{bc}}$ is the baseline corrected spectrum, and N is the number of signals in a spectrum. Based on the definition, smaller *RMSE* value indicates better similarity and better estimate of baseline. In addition, correlation coefficient, noted as R, between corrected spectrum and reference spectrum is a common algorithm for hit quality index (HQI) calculations (Renner et al., 2019). Thus, correlation coefficient between simulated baseline-free spectrum and baseline corrected spectrum was calculated.

The above evaluation methods were also applied for spectra of

environmental samples. Since the true baselines of the environmental samples are unknown, *RMSE* is calculated by comparing the baseline corrected spectrum and standard spectrum of the same type of plastic according to ATR-FTIR identification results as follows:

$$RMSE = \sqrt{\frac{\sum (\widehat{\boldsymbol{Y}}_{\mathrm{ns}|-\mathrm{bc}} - \widehat{\boldsymbol{Y}}_{\mathrm{ne}|-\mathrm{bc}})^2}{N}} \tag{23}$$

where $\widehat{\boldsymbol{Y}}_{\mathrm{ns}|-\mathrm{bc}}$ and $\widehat{\boldsymbol{Y}}_{\mathrm{ne}|-\mathrm{bc}}$ represent normalized baseline-free spectra of standard sample and baseline-free spectra of environmental sample, respectively. Similarly, correlation coefficient between baseline corrected spectra and corresponding standard sample spectra was also calculated.

3. Results and discussion

3.1. Predictive accuracy of standard deviation of noise

For each combination of SNR and baseline, 1000 spectra were simulated to obtain good statistics. Then, DSW method was applied to the simulated spectra to estimate standard deviation of noise. The estimated standard deviation of noise ($\hat{\sigma}_{\mathrm{ns}}$) was summarized in Fig. 2. For N1 group (large SNR and relatively small noise), the estimated standard deviations are all around 1, which suggests that DSW method is able to accurately estimate the standard deviation of noise. N1B3 (large SNR with fluctuating baseline, Fig. 1e) result shows a slight overestimation. This is due to the fact that fluctuation of baseline in a high frequency could contribute to local variability of noise signal, resulting in upwards of upper envelope function or downwards of lower envelope function. Such influence increases the probability of getting large d_{ns} values, and finally increases the estimated standard deviation. For N2 group (small SNR and relatively large noise), the estimated standard deviations were around 10, and all of the differences were within one standard deviation. Better estimation performance at small SNR was probably due to smaller impact to noise variability from the fluctuation of baseline. In general, DSW method was able to estimate standard deviation of noise within acceptable accuracy.

3.2. Comparison with simulated spectra with large SNR

The results of baseline correction on spectra with large SNR are presented in Figs. S2, S3, S4 and Fig. 3 as a typical example. When SNR is large, DSW shows good performance, and the estimated baselines generally captured the center of noise and bottom of the peaks. However, there is an increasement at the bottom of the α-peak and β-peak (Figs. 3b, S3b, S4b, S5b). This is due to the fact that elevation of peak signals at the edge are also possible to be elevated baseline, and such increasement is a compromise between likelihood of peak and likelihood of elevated baseline. Since the influence is small relative to peak height, and the shapes of peaks are preserved after baseline correction, the performance of DSW method is considered as acceptable for all types of baselines under large SNR.

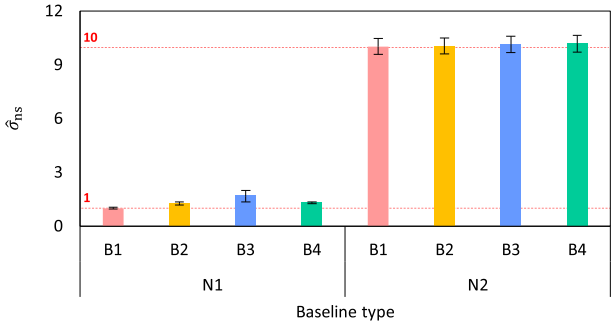

Fig. 2. Summary of estimated standard deviation of noise ($\hat{\sigma}_{\mathrm{ns}}$) of different SNR and baseline combinations. Error bar refers to standard deviation based on 1000 simulated spectra. N1 refers to large SNR and N2 refers to small SNR. B1 refers to flat baseline, B2 refers to elevated baseline, B3 refers to fluctuating baseline, and B4 refers to elevated and fluctuating baseline.

For airPLS, the estimated baseline is able to capture the center of noise for flat baseline and elevated baseline (Figs. S3b, S4b), and the shape of peaks is preserved. However, airPLS fails to fit the fluctuating baseline without elevation (Fig. S5b). On the other hand, when fluctuations come together with elevation, airPLS shows improved performance and only the edges of α-peak and β-peak are slightly tortured (Fig. 3e). In general, airPLS works well for flat and elevated baseline, but it shows some problems with handling the fluctuations of baseline without elevation.

For iModPoly, similar to airPLS, the estimated baseline fits well with flat and elevated baseline (Figs. S3b, S4b). However, when baseline had fluctuations, it fails to capture the center of noise by adjusting order of polynomial (Fig. S5b, Fig. 3b). The corrected spectra could somehow reflect the peak while the fluctuation was untouched, which would be useful for manual processes, but the corrected spectra are not feasible for automated processes. In general, iModPoly works well for spectra without fluctuations.

The summary of *RMSE* and *R* values are presented in Table 1. Based on the results, when SNR is large, *RMSE* values of DSW method are relatively small and stable. airPLS have relatively small *RMSE* with flat baseline and elevated baseline, but *RMSE* is large for fluctuating baseline without elevation. This is consistent with the observation that airPLS fails to fit baseline of the same condition. iModPoly has very small *RMSE* values for baseline without fluctuation. However, *RMSE* values become very large for fluctuated baselines, indicating unsatisfactory fits with fluctuating baseline. For *R* values, DSW method has good and stable performance for all cases with large *R* values. airPLS fails for fluctuated baseline without elevation, resulting in lower *R* value. iModPoly has good fit for baseline without fluctuation, giving *R* values close to one, while *R* values drop for fluctuated baseline due to unsatisfactory fit with fluctuations.

3.3. Comparison with simulated spectra with small SNR

For small SNR, the performance of baseline correction methods was also evaluated. In this case, the standard deviation of noise had a 10-fold increasement, and γ-peak was overlaid by noise. DSW method estimated baseline could successfully capture the center of noise for all baseline types (Figs. 4b and S6b, S7b, S8b). The estimated baselines are visually meandering curves, which is different from most baseline algorithms that are looking for visually smooth curves (Xu et al., 2021b). On the other hand, the DSW algorithm also results in increasement of baseline at the bottom of α-peak and β-peak. In such cases, the elevation is comparable to the height of peaks. However, the increasement does not alter the shape of peak. Therefore, such influence was considered acceptable.

In terms of airPLS, when baseline is flat, the estimated baseline is consistently below the spectrum (Fig. S6b). This results in an elevated spectrum after correction. However, the shape of peaks is preserved. Thus, the results are acceptable. On the other hand, when baseline elevates or fluctuates, the estimated baseline fails to capture the signal by adjusting the orders (Figs. 4b, S6b, S7b, S8b). Thus, airPLS seems to have limited performance with increased noise with fluctuations.

The performance of iModPoly is similar to large SNR situation. The estimated baseline well captures the center of noise, and the shape of peaks is preserved for flat and elevated baselines (Figs. S6b, S7b). However, if baseline has fluctuations, iModPoly fails to capture it although the peaks were unchanged (Figs. 4, S8b). Again, such a situation might be okay for manual operation but not optimal for automated process.

Due to small SNR, *RMSE* values become larger, and *R* values become

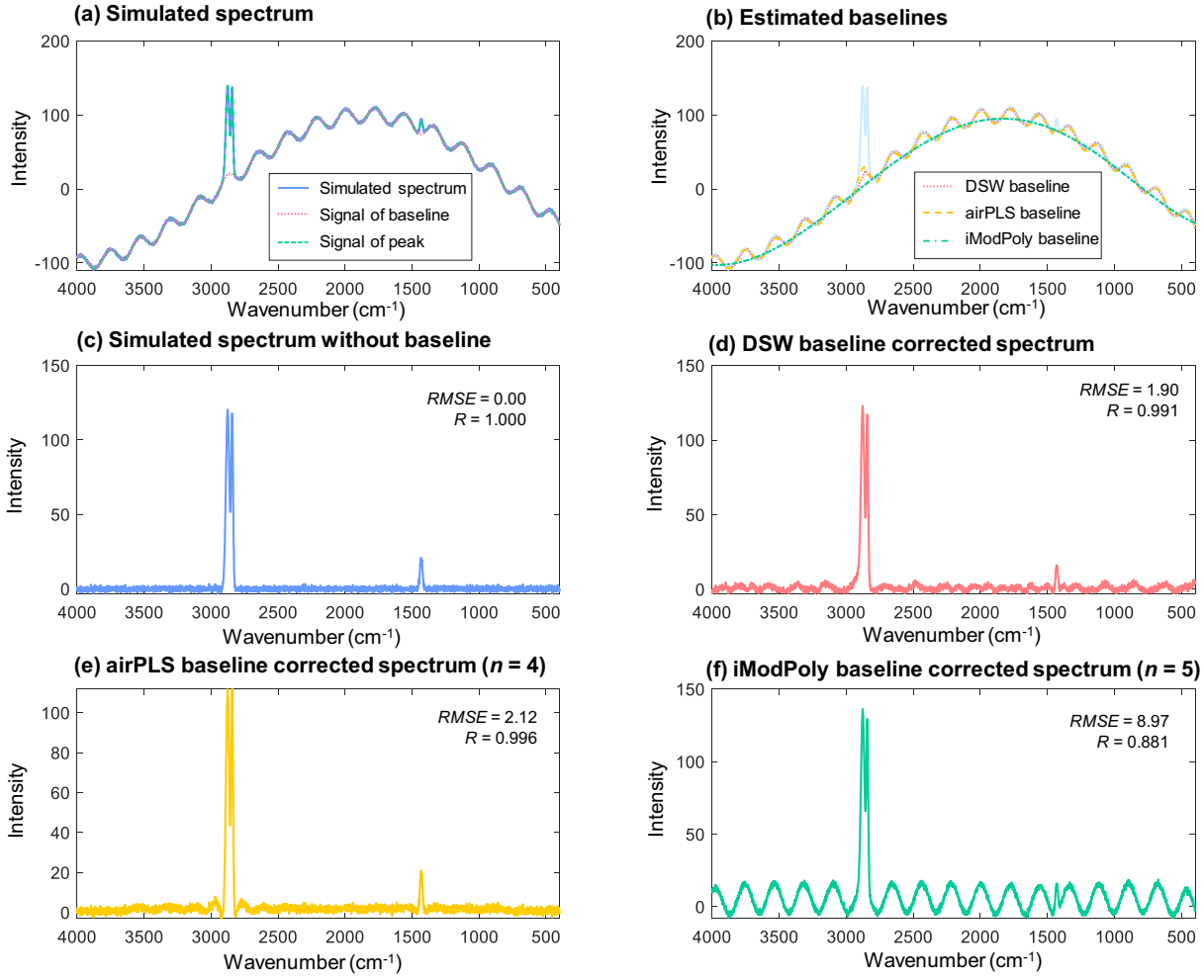

Fig. 3. Demonstration of baseline estimation of simulated spectrum with large SNR and elevated and fluctuating baseline. In (b), the light blue line refers to simulated spectrum for reference use. *n* refers to order of least square in (e) and refers to order of polynomial function in (f). *RMSE* refers to root mean square error, and *R* refers to correlation coefficient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of performance of baseline correction methods. *RMSE* is root mean square error between true baseline-free spectrum and baseline corrected spectrum for simulated spectra and between standard baseline-corrected standard spectrum and baseline corrected spectrum of environmental sample. *R* is correlation coefficient between baseline-free spectrum and baseline corrected spectrum for simulated spectra and between standard baseline-corrected standard spectrum and baseline corrected spectrum of environmental sample. DSW refers to double sliding-window method. airPLS refers to adaptive iteratively reweighted penalized least square method (Zhang et al., 2010), and iModPoly refers to improved modified multi-polynomial fit method (Zhao et al., 2007). Type refers to the types in Section 2.5. N/A refers to data not available.

		Type	*RMSE*			*R*			Descriptive		
			DSW	airPLS	iModPoly	DSW	airPLS	iModPoly	DSW	airPLS	iModPoly
Simulation	Large SNR	N1B1	1.35	2.66	0.58	0.998	0.999	1.000	Good	Perfect	Perfect
		N1B2	1.82	1.50	0.84	0.996	0.999	0.999	Good	Perfect	Perfect
		N1B3	1.56	14.53	9.02	0.994	0.700	0.882	Good	Failed	Neutral
		N1B4	1.90	2.12	8.97	0.991	0.996	0.881	Good	Good	Neutral
	Small SNR	N2B1	5.45	27.04	5.67	0.943	0.970	0.999	Good	Good	Perfect
		N2B2	5.57	84.63	5.69	0.940	0.375	0.998	Good	Failed	Perfect
		N2B3	4.89	35.20	10.06	0.952	0.875	0.919	Good	Neutral	Neutral
		N2B4	5.09	93.36	10.02	0.948	0.370	0.919	Good	Failed	Neutral
Environmental sample	A - PE		0.024	0.023	0.025	0.959	0.956	0.960	Good	Perfect	Perfect
	A - PP		0.009	0.018	0.012	0.996	0.987	0.994	Good	Neutral	Good
	B - PE		0.024	0.040	0.051	0.951	0.898	0.859	Good	Neutral	Neutral
	B - PP		0.027	0.062	0.069	0.968	0.843	0.836	Good	Neutral	Neutral
	B′ - PP		0.013	0.070	0.056	0.993	0.769	0.904	Good	Neutral	Failed
	C - PE		0.064	0.074	0.069	0.722	0.798	0.828	Good	Good	Perfect
	C - PS		0.035	0.033	0.039	0.897	0.925	0.903	Good	Neutral	Neutral
	D - PE		0.079	0.118	0.071	0.719	0.559	0.859	Good	Neutral	Perfect
	Z - unknown		N/A	N/A	N/A	N/A	N/A	N/A	Perfect	Perfect	Failed

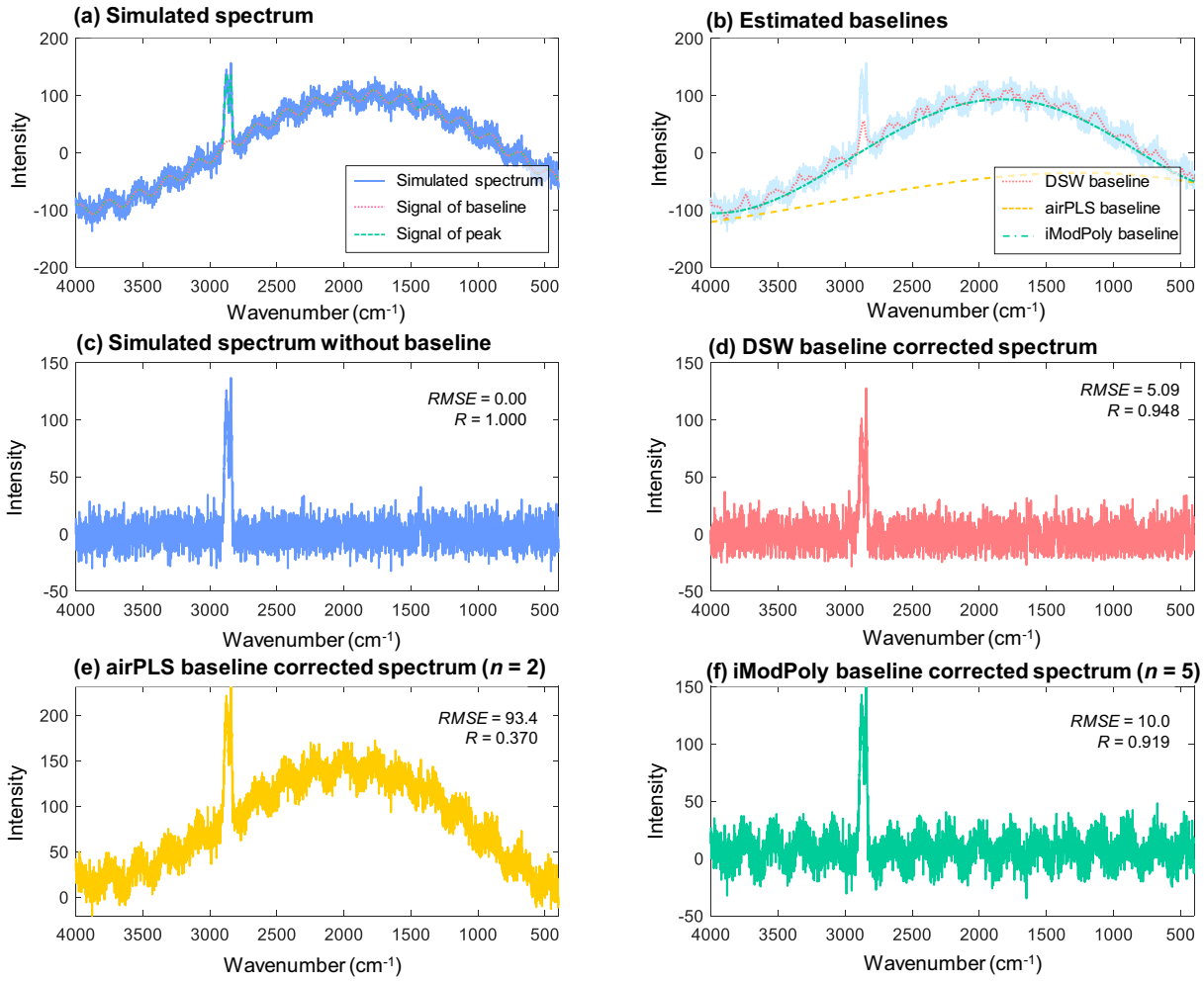

Fig. 4. Demonstration of baseline estimation of simulated spectrum with small SNR and elevated and fluctuating baseline. In (b), the light blue line refers to simulated spectrum for reference use. *n* refers to order of least square in (e) and refers to order of polynomial function in (f). *RMSE* refers to root mean square error, and *R* refers to correlation coefficient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

smaller compared with N1 groups (Table 1). For *RMSE*, DSW method results in small and stable *RMSE* values among all baseline combinations. *RMSE* values of airPLS are relatively large among all type of baselines. iModPoly has relatively small *RMSE* values for baselines without fluctuation while *RMSE* values got large for fluctuating baseline. For *R*, DSW method results in relatively large and stable *R* values, indicating DSW method could handle all baseline situations with good and stable performance even SNR is small. Spectra corrected by airPLS had relatively small correlation with true spectra for elevated baselines, but the result is okay for baseline without elevation. On the other hand, iModPoly could provide a good fit for baseline without fluctuation while performance drops for baseline with fluctuations.

3.4. Comparison with spectra of standard plastics

Spectra of standard plastics usually have sharp peaks with flat baseline and are useful during automated identification processes, such as those used as in-house library. Therefore, it is crucial to evaluate the performance of baseline correction methods when applied to standard plastic spectra. The results are presented in Figs. S9 to S15, and Fig. 5. Except for PA and PC, the baseline center was well captured, indicating the good performance of the evaluated methods. A representative example of PE is shown in Fig. 5, where the DSW method could produce a flat baseline after baseline correction.

For standard PA sample, the spectrum has elevated and fluctuating baseline (Fig. S9). This was probably due to light yellow color of the standard sample. In this case, the DSW method could capture the center of baseline, and the peaks were well preserved. However, airPLS and iModPoly had problems with capturing the center of baseline, but the main peaks could be identified. On the other hand, for standard PC spectra (Fig. S11), DSW method gave flatter baseline, but the peaks around 3100–3200 cm^{-1} and 2500–2600 cm^{-1} were eroded while other methods well preserved the peaks. Since these small peaks with low height-to-width ratio have a similar pattern to fluctuations in the baseline, DSW method treated the peaks as baseline, resulting in eroded peak signals.

3.5. Comparison with spectra of environmental samples

For Type A spectrum, which represents the samples that are relatively pure, clean, and well-focused during measurement, all methods show good performance (Figs. S16, S17). Similar to standard spectrum, DSW estimated baseline has increasement at the bottom of wide peaks due to its algorithm, but such influence is negligible. For a typical example of PE, all methods resulted in small and similar *RMSE* values and large and similar *R* values (Fig. S16). The PDF of d_{ns} could also be calculated (Fig. S16f), and due to large peaks, the PDF of d_{ns} is heavily left skewed. For this example, the estimated standard deviation of noise is 0.913. For another example of Type A, as identified as PP (Fig. S17), DSW method and iModPoly method show good performance while

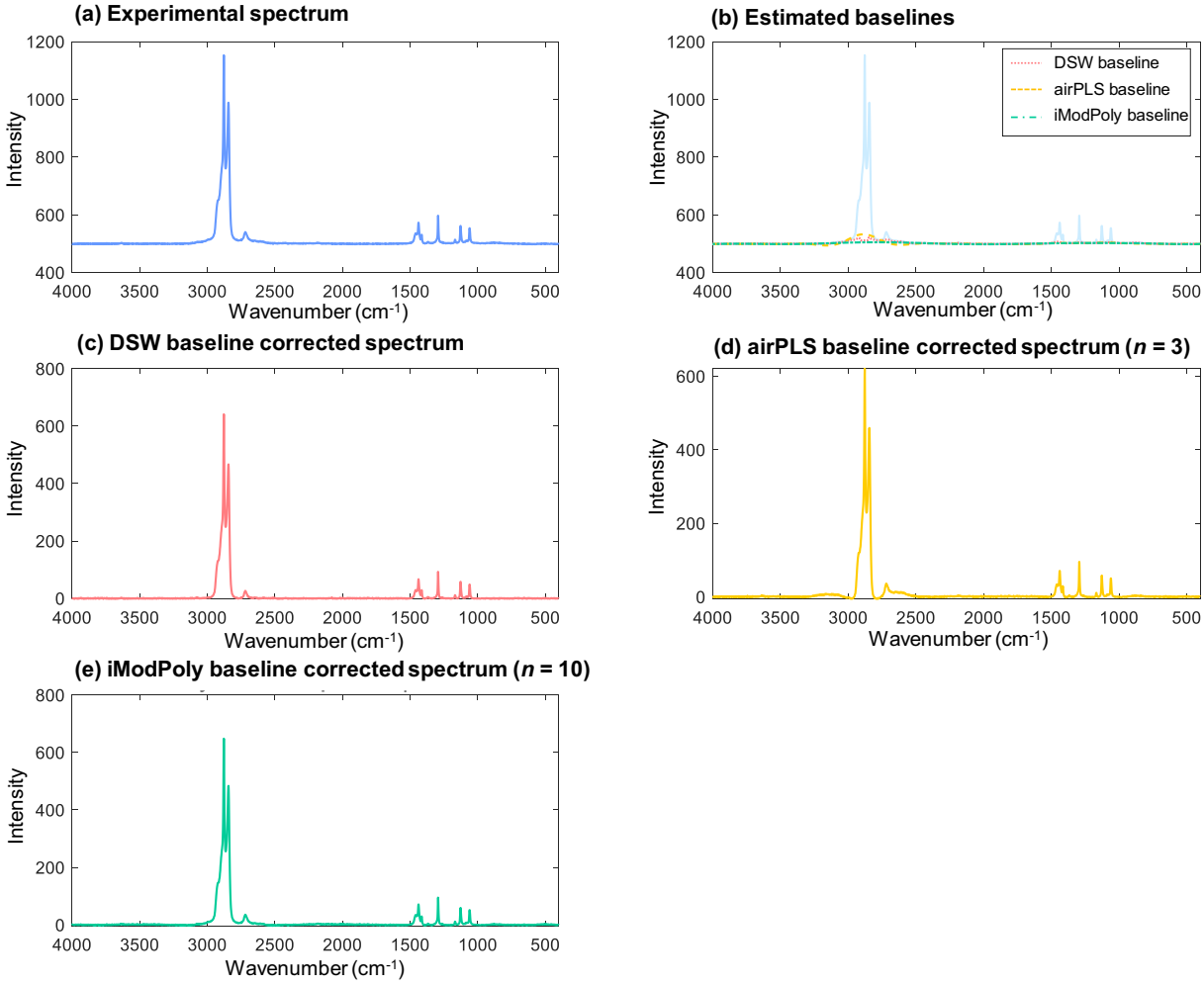

Fig. 5. Baseline estimation of experimental spectrum of standard PE. In (b), the light blue line refers to simulated spectrum for reference use. *n* refers to order of least square in (d) and refers to order of polynomial function in (e). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

airPLS has some problems with baseline around peak at 2800 cm^{-1}, resulting in larger *RMSE* value and smaller *R* value. The estimated standard deviation is 1.47 in this case.

Type B spectrum has visible peaks and elevated baseline with fluctuation, and an example of PE is presented in Fig. 6. In this situation, it would be optimal to remove the baseline and leave the peaks for further analysis. DSW estimated baseline captured center of noise and resulted in a flat and zero-centered corrected baseline when there were no peaks (Fig. 6b, c). Increasement under wide peak occurs, but the influence is minor. On the other hand, airPLS and iModPoly could preserve the peak but failed to capture the baseline fluctuations (Fig. 6b, d, e). As a result, the corrected baseline overlaid with peaks, which would have negative impact for identification process. Such phenomena were also reflected by smaller *RMSE* value and *R* value of DSW method. PDF of d_{ns} is shown in Fig. 6f, and it is a combination of a bell-shape distribution with right skewness. The estimated standard deviation is 1.66 for this example.

Another two examples of Type B are from PP, where two typical patterns of baselines are observed (Figs. S18, S19). For these two cases, DSW method could capture the baseline and preserve the peaks while airPLS and iModPoly methods had difficulties in capturing the center of baseline, resulting in lower *R* value and larger *RMSE* value.

Type C and Type D spectra could be obtained when failing to focus the surface of sample, or sample moved during measurement (Figs. S20, S21, S22). DSW method results in shortened peak height for wide peaks, but the shape is preserved. While most of the small and narrow peaks are reflected, some small peaks with low height-to-width ratio were eroded or eliminated, e.g., peaks at around 1450 cm^{-1} for PE and around 2800 cm^{-1} for PS. airPLS also captures noise center but in the meanwhile tortures the baseline at the edges of the wide peak. iModPoly also captures the centers of noise, and the peaks are well reflected. The standard deviation of noise for examples of Type C-PE, Type C-PS and Type D-PE are 0.635, 0.896 and 0.614, respectively.

Type Z spectrum represents the situation when there was no chemical of interest, and the substance was covered by other substances such as inorganic salts or biological materials that have strong fluorescence. Such spectra could be disposed during manual process, but it could be included in automated process. DSW and airPLS capture the center of noise and result in pure noise signal while iModPoly fails (Fig. S23). The estimated standard deviation of noise is 1.22.

The discussions on performance of the methods are summarized in Table 1. In general, DSW method has small influence on peaks and can capture the noise center for all situations with some small peaks with low height-to-width ratio being subject to erosion or elimination. However, the influence of erosion or elimination shows small influence on total similarity between corrected spectrum and reference spectrum. Thus, DSW method was able to handle various spectra without adjusting parameters during baseline estimation for identification purposes.

On the other hand, peak height and area under band are important for some quantitative analysis, such as calculation of carbonyl index (CI) (Andrady, 2017; Almond et al., 2020; Li et al., 2022). Since erosion or

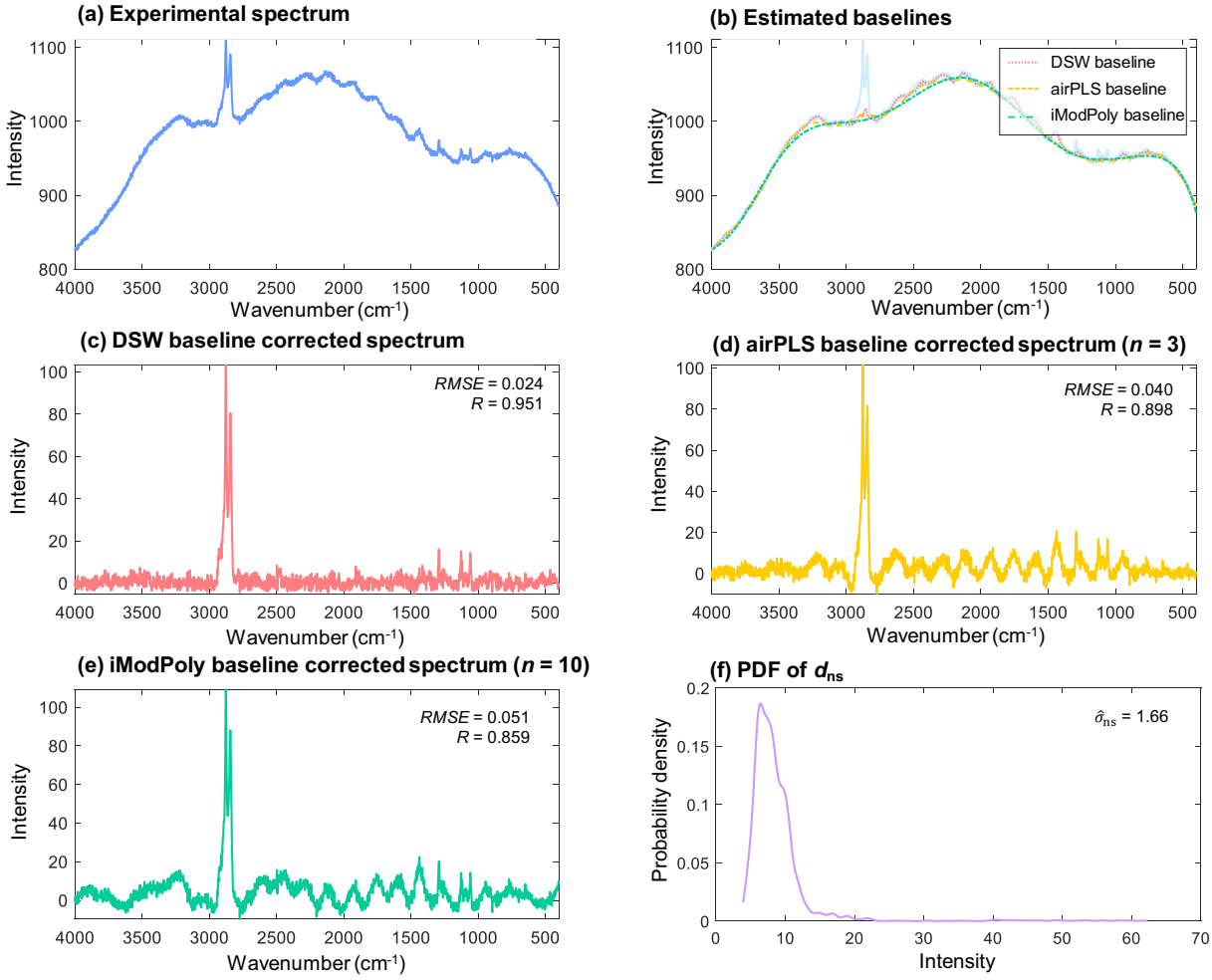

Fig. 6. Demonstration of baseline estimation of experimental spectrum (Type B) with large SNR and elevated and fluctuating baseline. In (b), the light blue line refers to simulated spectrum for reference use. *n* refers to order of least square in (d) and refers to order of polynomial function in (e). *RMSE* refers to root mean square error, *R* refers to correlation coefficient, and $\hat{\sigma}_{ns}$ refers to estimated standard deviation of noise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

elimination of peaks may have nonnegligible influence, further evaluation is needed to quantify such influence, and parameters may need adjustment to obtain the optimal fit for these types of analysis.

4. Conclusions

In this study, a double sliding-window method (DSW) was proposed to estimate the baseline and standard deviation of noise. Simulated spectra and experimental spectra were used to evaluate the performance in comparison with two popular and widely used methods. Validation with simulated spectra showed that DSW method could accurately estimate the standard deviation of noise. It could also reliably handle spectra with different situations of SNR and baseline combinations without user intervention. Evaluation on experimental spectra also revealed the stable and good performance of DSW method for handling different types of situations. Therefore, DSW method is a useful approach for preprocessing Raman spectra of environmental samples and automated processes.

It is important to note that the tested spectra in the current study may not represent all situations of Raman spectra. This method also shows some limitations in dealing with small peaks of low height-to-width ratio, as they may be eroded or eliminated. To address this issue, a two-pass strategy can be implemented, where weights are re-assigned based on the identification results of the first pass. For example, Bayesian approaches could be used to adjust the weights on condition of the identification results of the first pass, and expertise and prior knowledges could also be reflected under Bayesian framework (Gelman and Hill, 2006; Gelman et al., 2013). Future studies will be focused on address these limitations.

CRediT authorship contribution statement

Zijiang Yang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Hisayuki Arakawa:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We would like to express our gratitude to Nagashima Kyouya for

assistance with the Raman spectrometer, as well as Suzuki Nonoka, Winnie Awuor, Nakamura Fumie, and Murat Çelik for assistance with the FTIR spectrometer. We also thank Honda Motoko for support with lab equipment, and Huan Gao for valuable advice on optics. We extend our appreciation to the staff of the research vessels (Seiyo Maru and Shinyo Maru) from the Tokyo University of Marine Science and Technology for their unwavering support and dedication throughout our research activities.

Funding sources

This study was supported by the Environmental Research and Technology Development Fund (JPMEERF20211003) of the Environmental Restoration and Conservation Agency of Japan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2023.114887.

References

Almond, J., Sugumaar, P., Wenzel, M.N., Hill, G., Wallis, C., 2020. Determination of the carbonyl index of polyethylene and polypropylene using specified area under band methodology with ATR-FTIR spectroscopy. e-Polymers 20 (1), 369–381.
Amjad, A., Ullah, R., Khan, S., Bilal, M., Khan, A., 2018. Raman spectroscopy based analysis of milk using random forest classification. Vib. Spectrosc. 99, 124–129.
Andrady, A.L., 2017. The plastic in microplastics: a review. Mar. Pollut. Bull. 119 (1), 12–22.
Anger, P.M., von der Esch, E., Baumann, T., Elsner, M., Niessner, R., Ivleva, N.P., 2018. Raman microspectroscopy as a tool for microplastic particle analysis. TrAC Trends Anal. Chem. 109, 214–226.
Araujo, C.F., Nolasco, M.M., Ribeiro, A.M., Ribeiro-Claro, P.J., 2018. Identification of microplastics using Raman spectroscopy: latest developments and future prospects. Water Res. 142, 426–440.
Asfour, H., Swift, L.M., Sarvazyan, N., Doroslovački, M., Kay, M.W., 2011. Signal decomposition of transmembrane voltage-sensitive dye fluorescence using a multiresolution wavelet analysis. IEEE Trans. Biomed. Eng. 58 (7), 2083–2093. https://doi.org/10.1109/TBME.2011.2143713.
Baek, S.J., Park, A., Kim, J., Shen, A., Hu, J., 2009. A simple background elimination method for Raman spectra. Chemom. Intell. Lab. Syst. 98 (1), 24–30.
Baek, S.J., Park, A., Ahn, Y.J., Choo, J., 2015. Baseline correction using asymmetrically reweighted penalized least squares smoothing. Analyst 140 (1), 250–257. https://doi.org/10.1039/C4AN01061B.
Bilal, M., Saleem, M., Amanat, S.T., Shakoor, H.A., Rashid, R., Mahmood, A., Ahmed, M., 2015. Optical diagnosis of malaria infection in human plasma using Raman spectroscopy. J. Biomed. Opt. 20 (1), 017002.
Cabernard, L., Roscher, L., Lorenz, C., Gerdts, G., Primpke, S., 2018. Comparison of Raman and Fourier transform infrared spectroscopy for the quantification of microplastics in the aquatic environment. Environ.Sci.Technol. 52 (22), 13279–13288.
Casas, M.M., Avitia, R.L., Gomez-Caraveo, A., Reyna, M.A., Cardenas-Haro, J.A., 2015. A complete study of variability in time and amplitude of a standard ECG database. Int.J.Comput.Theory Eng. 7 (5), 366.
Casella, G., Berger, R.L., 2002. Statistical Inference. Cengage Learning.
Çelik, M., Nakano, H., Uchida, K., Isobe, Atsuhiko, Arakawa, H., 2023. Comparative evaluation of the carbonyl index of microplastics around the Japan coast. Mar. Pollut. Bull. 190, 114818 https://doi.org/10.1016/j.marpolbul.2023.114818.
Chen, Y.S., Hsu, Y.C., 2019. Effective and efficient baseline correction algorithm for Raman spectra. Lect. Notes Eng. Comput. Sci. 2239, 295–298.
Chen, T., Son, Y., Park, A., Baek, S.J., 2022. Baseline correction using a deep-learning model combining ResNet and UNet. Analyst 147 (19), 4285–4292.
Cowger, W., Steinmetz, Z., Gray, A., Munno, K., Lynch, J., Hapich, H., Primpke, S., De Frond, H., Rochman, C., Herodotou, O., 2021. Microplastic spectral classification needs an open source community: open specy to the rescue! Anal. Chem. 93 (21), 7543–7548.
Crecelius, A.C., Alexandrov, T., Schubert, U.S., 2011. Application of matrix-assisted laser desorption/ionization mass spectrometric imaging to monitor surface changes of UV-irradiated poly (styrene) films. Rapid Commun. Mass Spectrom. 25 (19), 2809–2814.
Dong, M., Zhang, Q., Xing, X., Chen, W., She, Z., Luo, Z., 2020. Raman spectra and surface changes of microplastics weathered under natural environments. Sci. Total Environ. 739, 139990.
Dong, M., She, Z., Xiong, X., Ouyang, G., Luo, Z., 2022. Automated analysis of microplastics based on vibrational spectroscopy: are we measuring the same metrics? Anal. Bioanal. Chem. 414 (11), 3359–3372.
Du, J., Wu, X., Zhang, H., Wang, S., Tan, W., Guo, X., 2010. Mass spectrometry-based proteomic analysis of Kashin-Beck disease. Mol. Med. Rep. 3 (5), 821–824.
Fan, M., Yu, Q., Wang, X., Zheng, Z., Xu, S., Chen, Z., Li, L., 2016. Identification of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry as predictors of prognosis in triple negative breast cancer. J. Nanosci. Nanotechnol. 16 (12), 12483–12488.
Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/hierarchical Models. Cambridge University Press.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. CRC Press.
Genest, S., Salzer, R., Steiner, G., 2013. Molecular imaging of paper cross sections by FT-IR spectroscopy and principal component analysis. Anal. Bioanal. Chem. 405, 5421–5430.
Ghosal, S., Chen, M., Wagner, J., Wang, Z.M., Wall, S., 2018. Molecular identification of polymers and anthropogenic particles extracted from oceanic water and fish stomach–a Raman micro-spectroscopy study. Environ. Pollut. 233, 1113–1124.
Gierlinger, N., Keplinger, T., Harrington, M., 2012. Imaging of plant cell walls by confocal Raman microscopy. Nat. Protoc. 7 (9), 1694–1708.
Golotvin, S., Williams, A., 2000. Improved baseline recognition and modeling of FT NMR spectra. J. Magn. Reson. 146 (1), 122–125.
Gonzalez, J.A., Avitia, R.L., Flores, N., Bravo, M.E., Reyna, M.A., Cetto, L.A., 2015. Reconstruction of premature atrial contraction and premature ventricular contraction on ECG traces by applying PLA as segmentation process. J.Image Graph. 3 (2).
He, S., Zhang, W., Liu, L., Huang, Y., He, J., Xie, W., Wu, P., Du, C., 2014. Baseline correction for Raman spectra using an improved asymmetric least squares method. Anal. Methods 6 (12), 4402–4407.
Hu, H., Bai, J., Xia, G., Zhang, W., Ma, Y., 2018. Improved baseline correction method based on polynomial fitting for Raman spectroscopy. Photon.Sens. 8 (4), 332–340. https://doi.org/10.1007/s13320-018-0512-y.
Jiménez-Carvelo, A.M., González-Casado, A., Pérez-Castaño, E., Cuadros-Rodríguez, L., 2017. Fast-HPLC fingerprinting to discriminate olive oil from other edible vegetable oils by multivariate classification methods. J.AOAC Int. 100 (2), 345–350.
Jirasek, A., Schulze, G., Yu, M.M.L., Blades, M.W., Turner, R.F.B., 2004. Accuracy and precision of manual baseline determination. Appl. Spectrosc. 58 (12), 1488–1499.
Käppler, A., Fischer, D., Oberbeckmann, S., Schernewski, G., Labrenz, M., Eichhorn, K.J., Voit, B., 2016. Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, Raman or both? Anal. Bioanal. Chem. 408 (29), 8377–8391.
Karami, A., Golieskardi, A., Keong Choo, C., Larat, V., Galloway, T.S., Salamatinia, B., 2017. The presence of microplastics in commercial salts from different countries. Sci. Rep. 7 (1), 1–11.
Karmenyan, A.V., Kistenev, Y.V., Perevedentseva, E.V., Krivokharchenko, A.S., Sarmiento, M.N., Barus, E.L., Cheng, C.L., Vrazhnov, D.A., 2020. Machine learning methods for the in-vitro analysis of preimplantation embryo Raman micro-spectroscopy. November. In: Fourth International Conference on Terahertz and Microwave Radiation: Generation, Detection, and Applications, Vol. 11582. SPIE, pp. 179–183.
Kaya, M., Mujtaba, M., Ehrlich, H., Salaberria, A.M., Baran, T., Amemiya, C.T., Galli, R., Akyuz, L., Sargin, I., Labidi, J., 2017. On chemistry of γ-chitin. Carbohydr. Polym. 176, 177–186.
Kellner, V., Kersbergen, C.J., Li, S., Babola, T.A., Saher, G., Bergles, D.E., 2021. Dual metabotropic glutamate receptor signaling enables coordination of astrocyte and neuron activity in developing sensory domains. Neuron 109 (16), 2545–2555.
Khan, Saranjam, Ullah, Rahat, Khan, Asifullah, Ashraf, Ruby, Ali, Hina, Bilal, Muhammad, Saleem, Muhammad, 2018. Analysis of hepatitis B virus infection in blood sera using Raman spectroscopy and machine learning. Photodiagn. Photodyn. Ther. 23, 89–93.
Kumar, K., Cava, F., 2019. Chromatographic analysis of peptidoglycan samples with the aid of a chemometric technique: introducing a novel analytical procedure to classify bacterial cell wall collection. Anal. Methods 11 (12), 1671–1679.
Larkin, P., 2017. Infrared and Raman Spectroscopy: Principles and Spectral Interpretation. Elsevier.
Lenz, R., Enders, K., Stedmon, C.A., Mackenzie, D.M., Nielsen, T.G., 2015. A critical assessment of visual identification of marine microplastic using Raman spectroscopy for analysis improvement. Mar. Pollut. Bull. 100 (1), 82–91.
Li, J., Yu, B., Zhao, W., Chen, W., 2014. A review of signal enhancement and noise reduction techniques for tunable diode laser absorption spectroscopy. Appl. Spectrosc. Rev. 49 (8), 666–691.
Li, J., Yu, B., Fischer, H., 2015. Wavelet transform based on the optimal wavelet pairs for tunable diode laser absorption spectroscopy signal processing. Appl. Spectrosc. 69 (4), 496–506.
Li, D., Sheerin, E.D., Shi, Y., Xiao, L., Yang, L., Boland, J.J., Wang, J.J., 2022. Alcohol pretreatment to eliminate the interference of micro additive particles in the identification of microplastics using Raman spectroscopy. Environ.Sci.Technol. 56 (17), 12158–12168.
Lieber, C.A., Mahadevan-Jansen, A., 2003. Automated method for subtraction of fluorescence from biological Raman spectra. Appl. Spectrosc. 57 (11), 1363–1367.
Liu, J., Sun, J., Huang, X., Li, G., Liu, B., 2015. Goldindec: a novel algorithm for Raman spectrum baseline correction. Appl. Spectrosc. 69 (7), 834–842.
Liu, J., Osadchy, M., Ashton, L., Foster, M., Solomon, C.J., Gibson, S.J., 2017. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. Analyst 142 (21), 4067–4074.
Mankova, A.A., Cherkasova, O.P., Lazareva, E.N., Bucharskaya, A.B., Dyachenko, P.A., Kistenev, Y.V., Vrazhnov, D.A., Skiba, V.E., Tuchin, V.V., Shkurinov, A.P., 2020. Study of blood serum in rats with transplanted cholangiocarcinoma using Raman spectroscopy. Opt. Spectrosc. 128, 964–971.
Martinez, W.L., Martinez, A.R., 2001. Computational Statistics Handbook With MATLAB. Chapman and Hall/CRC.

Muller, J.J., Neumann, M., Scholl, P., Hilterhaus, L., Eckstein, M., Thum, O., Liese, A., 2010. Online monitoring of biotransformations in high viscous multiphase systems by means of FT-IR and chemometrics. Anal. Chem. 82 (14), 6008–6014.
Nakano, H., Arakawa, H., Tokai, T., 2021. Microplastics on the sea surface of the semi-closed Tokyo Bay. Mar. Pollut. Bull. 162, 111887.
Nava, V., Frezzotti, M.L., Leoni, B., 2021. Raman spectroscopy for the analysis of microplastics in aquatic systems. Appl. Spectrosc. 75 (11), 1341–1357.
Noack, K., Eskofier, B., Kiefer, J., Dilk, C., Bilow, G., Schirmer, M., Buchholz, R., Leipertz, A., 2013. Combined shifted-excitation Raman difference spectroscopy and support vector regression for monitoring the algal production of complex polysaccharides. Analyst 138 (19), 5639–5646.
Nowacki, K., Stępniak, I., Machałowski, T., Wysokowski, M., Petrenko, I., Schimpf, C., Rafaja, D., Langer, E., Richter, A., Ziętek, J., Pantović, S., 2020. Electrochemical method for isolation of chitinous 3D scaffolds from cultivated Aplysina aerophoba marine demosponge and its biomimetic application. Appl.Phys.A 126, 1–16.
Oller-Moreno, Sergio, Pardo, Antonio, Jiménez-Soto, Juan Manuel, Samitier, Josep, Marco, Santiago, 2014. Adaptive asymmetric least squares baseline estimation for analytical instruments. In: 2014 IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14). IEEE, pp. 1–5.
Ouyang, X., Duarte, C.M., Cheung, S.G., Tam, N.F.Y., Cannicci, S., Martin, C., Lo, H.S., Lee, S.Y., 2022. Fate and effects of macro-and microplastics in coastal wetlands. Environ.Sci.Technol. 56 (4), 2386–2397.
Phan, S., Padilla-Gamiño, J.L., Luscombe, C.K., 2022. The effect of weathering environments on microplastic chemical identification with Raman and IR spectroscopy: part I. Polyethylene and polypropylene. Polym. Test. 116, 107752.
Povey, J.F., O'Malley, C.J., Root, T., Martin, E.B., Montague, G.A., Feary, M., Trim, C., Lang, D.A., Alldread, R., Racher, A.J., Smales, C.M., 2014. Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-ToF mass spectrometry fingerprinting and PLS-DA modelling. J. Biotechnol. 184, 84–93.
Radzol, A.R.M., Lee, K.Y., Wahab, N.A., 2014. Low concentration melamine detection with surface enhanced Raman spectroscopy. April. In: 2014 IEEE REGION 10 SYMPOSIUM. IEEE, pp. 555–559.
Renner, G., Nellessen, A., Schwiers, A., Wenzel, M., Schmidt, T.C., Schram, J., 2019. Data preprocessing & evaluation used in the microplastics identification process: a critical review & practical guide. TrAC Trends Anal. Chem. 111, 229–238.
Sathyanesan, A., Ogura, T., Lin, W., 2012. Automated measurement of nerve fiber density using line intensity scan analysis. J. Neurosci. Methods 206 (2), 165–175.
Schmid, M., Rath, D., Diebold, U., 2022. Why and how Savitzky-Golay filters should be replaced. ACS Meas.Sci.Au 2 (2), 185–196.
Schulze, H.G., Foist, R.B., Okuda, K., Ivanov, A., Turner, R.F., 2011. A model-free, fully automated baseline-removal method for Raman spectra. Appl. Spectrosc. 65 (1), 75–84.
Schulze, H.G., Foist, R.B., Okuda, K., Ivanov, A., Turner, R.F., 2012. A small-window moving average-based fully automated baseline estimation method for Raman spectra. Appl. Spectrosc. 66 (7), 757–764.
Schymanski, D., Goldbeck, C., Humpf, H.U., Fürst, P., 2018. Analysis of microplastics in water by micro-Raman spectroscopy: release of plastic particles from different packaging into mineral water. Water Res. 129, 154–162.
Sikora, K.N., Hardie, J.M., Castellanos-García, L.J., Liu, Y., Reinhardt, B.M., Farkas, M.E., Rotello, V.M., Vachet, R.W., 2019. Dual mass spectrometric tissue imaging of nanocarrier distributions and their biochemical effects. Anal. Chem. 92 (2), 2011–2018.
Smulko, J., 2019. Methods of trend removal in electrochemical noise data–overview. Measurement 131, 569–581.
Song, Y.K., Hong, S.H., Jang, M., Kang, J.H., Kwon, O.Y., Han, G.M., Shim, W.J., 2014. Large accumulation of micro-sized synthetic polymer particles in the sea surface microlayer. Environ.Sci.Technol. 48 (16), 9014–9021.
Song, Y.K., Hong, S.H., Eo, S., Shim, W.J., 2021. A comparison of spectroscopic analysis methods for microplastics: manual, semi-automated, and automated Fourier transform infrared and Raman techniques. Mar. Pollut. Bull. 173, 113101.
Stanford, T.E., Bagley, C.J., Solomon, P.J., 2016. Informed baseline subtraction of proteomic mass spectrometry data aided by a novel sliding window algorithm. Proteome Sci. 14, 1–15.
Steiner, G., Preusse, G., Zimmerer, C., Krautwald-Junghanns, M.E., Sablinskas, V., Fuhrmann, H., Koch, E., Bartels, T., 2016. Label free molecular sexing of monomorphic birds using infrared spectroscopic imaging. Talanta 150, 155–161.
Tanaka, S., Fujita, Y., Parry, H.E., Yoshizawa, A.C., Morimoto, K., Murase, M., Yamada, Y., Yao, J., Utsunomiya, S.I., Kajihara, S., Fukuda, M., 2014. Mass++: a visualization and analysis tool for mass spectrometry. J. Proteome Res. 13 (8), 3846–3853.
Uckermann, O., Yao, W., Juratli, T.A., Galli, R., Leipnitz, E., Meinhardt, M., Koch, E., Schackert, G., Steiner, G., Kirsch, M., 2018. IDH1 mutation in human glioma induces chemical alterations that are amenable to optical Raman spectroscopy. J. Neuro-Oncol. 139, 261–268.
Ullah, R., Khan, S., Chaudhary, I.I., Shahzad, S., Ali, H., Bilal, M., 2020. Cost effective and efficient screening of tuberculosis disease with Raman spectroscopy and machine learning algorithms. Photodiagn. Photodyn. Ther. 32, 101963.
Urboniene, V., Pucetaite, M., Jankevicius, F., Zelvys, A., Sablinskas, V., Steiner, G., 2014. Identification of kidney tumor tissue by infrared spectroscopy of extracellular matrix. J. Biomed. Opt. 19 (8), 087005.
Utsunomiya, S.I., Fujita, Y., Tanaka, S., Kajihara, S., Aoshima, K., Oda, Y., Tanaka, K., 2014. Signal processing algorithm development for Mass++ (Ver. 2): platform software for mass spectrometry. IPSJ Trans.Bioinf. 7, 24–29.
Walfridson, M., Kuttainen Thyni, E., 2022. Automation of carbonyl index calculations for fast evaluation of microplastics degradation. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1679036&dswid=3138. Available from.
Wang, K., Yuan, Y., Han, S., Yang, Y., 2019. Application of FTIR spectroscopy with solvent-cast film and PLS regression for the quantification of SBS content in modified asphalt. Int.J.Pavement Eng. 20 (11), 1336–1341.
Wang, Y., Nakano, H., Xu, H., Arakawa, H., 2021. Contamination of seabed sediments in Tokyo Bay by small microplastic particles. Estuar. Coast. Shelf Sci. 261, 107552.
Weisser, J., Pohl, T., Heinzinger, M., Ivleva, N.P., Hofmann, T., Glas, K., 2022. The identification of microplastics based on vibrational spectroscopy data–a critical review of data analysis routines. TrAC Trends Anal. Chem. 148, 116535.
Wesch, C., Barthel, A.K., Braun, U., Klein, R., Paulus, M., 2016. No microplastics in benthic eelpout (Zoarces viviparus): an urgent need for spectroscopic analyses in microplastic detection. Environ. Res. 148, 36–38.
Xi, Y., Li, Y., Duan, Z., Lu, Y., 2018. A novel pre-processing algorithm based on the wavelet transform for Raman spectrum. Appl. Spectrosc. 72 (12), 1752–1763.
Xu, Y., Lin, Q., Wang, L., Wang, Q., 2005. The prediction of nitrogen concentration in soil by VNIR reflectance spectrum. July. In: Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05, Vol. 6. IEEE, pp. 4451–4454.
Xu, X., Huo, X., Qian, X., Lu, X., Yu, Q., Ni, K., Wang, X., 2021a. Data-driven and coarse-to-fine baseline correction for signals of analytical instruments. Anal. Chim. Acta 1157, 338386.
Xu, Y., Du, P., Senger, R., Robertson, J., Pirkle, J.L., 2021b. ISREA: an efficient peak-preserving baseline correction algorithm for Raman spectra. Appl. Spectrosc. 75 (1), 34–45.
Xu, H., Nakano, H., Tokai, T., Miyazaki, T., Hamada, H., Arakawa, H., 2022. Contamination of sea surface water offshore the Tokai region and Tokyo Bay in Japan by small microplastics. Mar. Pollut. Bull. 185, 114245.
Xu, J.L., Thomas, K.V., Luo, Z., Gowen, A.A., 2019. FTIR and Raman imaging for microplastics analysis: state of the art, challenges and prospects. TrAC Trends Anal. Chem. 119, 115629.
Yang, Y., O'Riordan, A., Lovera, P., 2022. Highly sensitive pesticide detection using electrochemically prepared silver-gum arabic nanocluster SERS substrates. Sensors Actuators B Chem. 364, 131851.
Yu, X., Peng, J., Wang, J., Wang, K., Bao, S., 2016. Occurrence of microplastics in the beach sand of the Chinese inner sea: the Bohai Sea. Environ. Pollut. 214, 722–730.
Yu, M., Yan, H., Xia, J., Zhu, L., Zhang, T., Zhu, Z., Lou, X., Sun, G., Dong, M., 2019. Deep convolutional neural networks for tongue squamous cell carcinoma classification using Raman spectroscopy. Photodiagn. Photodyn. Ther. 26, 430–435.
Yuan, Y., Liu, X., Pu, G., Wang, T., Guo, Q., 2021. Corrosion features and time-dependent corrosion model of Galfan coating of high strength steel wires. Constr. Build. Mater. 313, 125534.
Zhang, Z.M., Chen, S., Liang, Y.Z., 2010. Baseline correction using adaptive iteratively reweighted penalized least squares. Analyst 135 (5), 1138–1146.
Zhang, K., Gong, W., Lv, J., Xiong, X., Wu, C., 2015. Accumulation of floating microplastics behind the Three Gorges Dam. Environ. Pollut. 204, 117–123.
Zhao, J., Lui, H., McLean, D.I., Zeng, H., 2007. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. Appl. Spectrosc. 61 (11), 1225–1232.
Zimmerer, C., Häußler, L., Arnhold, K., Ziegler, L., Heinrich, G., 2019a. Molecular structure of reactive polycarbonate-amine interfaces characterized by IR-spectroscopy and differential scanning calorimetry. January. In: AIP Conference Proceedings, 2055. AIP Publishing LLC, p. 130001. No. 1.
Zimmerer, C., Matulaitiene, I., Niaura, G., Reuter, U., Janke, A., Boldt, R., Sablinskas, V., Steiner, G., 2019b. Nondestructive characterization of the polycarbonate-octadecylamine interface by surface enhanced Raman spectroscopy. Polym. Test. 73, 152–158.
Zimmerman, T.A., Rubakhin, S.S., Romanova, E.V., Tucker, K.R., Sweedler, J.V., 2009. MALDI mass spectrometric imaging using the stretched sample method to reveal neuropeptide distributions in aplysia nervous tissue. Anal. Chem. 81 (22), 9402–9409.