Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 325 (2025) 125061

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

An automatic method for accurate signal-to-noise ratio estimation and baseline correction of Raman spectra of environmental microplastics

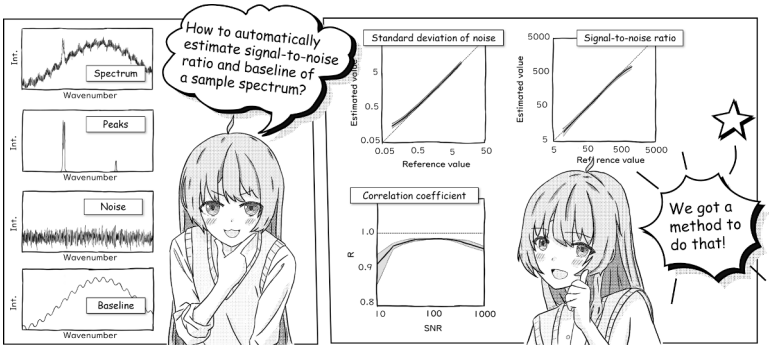
Zijiang Yang *, Hiroya Nagashima, Celik Murat, Hisayuki Arakawa

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

HIGHLIGHTS

- A *k*-iterative double sliding-window (DSW^k) method was proposed.
- DSW^k method outperforms traditional methods in estimating spectral noise and SNR.
- Estimating spectral noise: 1.01 to 1.08 times of the reference value on average.
- Estimating SNR: 0.89 to 0.93 times of the reference value on average.
- DSW^k method demonstrated good performance in baseline correction.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Raman spectroscopy
Baseline correction
Microplastics
Signal-to-noise ratio

ABSTRACT

In this study, we introduced a *k*-iterative double sliding-window (DSW^k) method for the estimation of spectral noise, signal-to-noise ratio (SNR), and baseline correction. The performance was evaluated using simulated spectra and compared against other commonly employed methods. Convergent evaluation determined that a *k* value of 20 strikes an optimal balance between convergence and computational intensity. The DSW^k method demonstrated outstanding performance across different spectral types (flat baseline, baseline with elevation, baseline with fluctuation, baseline with elevation and fluctuation) coupled with SNR values from 10 to 1000, achieving results that ranged from 1.01 to 1.08 times of the reference value in estimating spectral noise. It also showed that the estimated SNR values are 0.89 to 0.93 times of the reference value, demonstrating a 74.5 % – 131.7 % improvement over the conventional method in spectra with elevated and/or fluctuating baselines. Additionally, the DSW^k method proved effective in correcting baselines and identifying polymers in environmental samples of polyethylene (PE), polypropylene (PP), and polystyrene (PS), despite the limitation of reducing the peak height in spectra with low SNR. This method offers the potential to enhance the automatic and accurate evaluation of spectral quality and could assist in the development of guidelines for more rapid parameter adjustments in Raman measurements.

* Corresponding author.
E-mail addresses: zyan001@kaiyodai.ac.jp (Z. Yang), m233041@edu.kaiyodai.ac.jp (H. Nagashima), 8muratcelik@gmail.com (C. Murat), arakawa@kaiyodai.ac.jp (H. Arakawa).

https://doi.org/10.1016/j.saa.2024.125061
Received 1 May 2024; Received in revised form 2 August 2024; Accepted 25 August 2024
Available online 27 August 2024

1. Introduction

Microplastics, commonly defined as plastic particles with a diameter smaller than 5 mm [1–3], have emerged as a significant environmental concern [4–6]. It is reported that microplastics can negatively impact both ecosystems [7,8] and human health [9,10]. Due to their ubiquitous presence and the potential risks to the environment, detection and monitoring of microplastics in environment are of great importance.

The commonly used methods for detecting microplastics include Fourier Transform Infrared (FTIR) spectroscopy and Raman spectroscopy [11,12]. Raman spectroscopy is more sensitive to symmetric vibrations of non-polar groups while FTIR spectroscopy is more effective at detecting asymmetric vibrations of polar groups [13]. Due to its properties, Raman spectroscopy can directly analyze microplastics in water [14]. However, Raman spectra tend to be noisier and more sensitive to surface-attached chemicals on microplastics than FTIR spectra, often resulting in lower spectral quality [12,15]. As a result, acquiring a Raman spectrum that can provide sufficient amount of information usually requires additional effort.

In order to obtain a satisfactory Raman spectrum, acquisition parameters may need adaptive adjustments during measurement [12,16,17], such as exposure time. Ideally, it would be useful to develop a universally applicable parameter, so that no adjustment is needed. However, it has been reported that the physical properties of microplastics, like color, may be related to spectral quality [15,18]. Thus, it would be possibility for development of standardized guidelines for microplastic parameter adjustment as an alternative solution. With such guidelines, parameters could be adjusted more efficiently, which can save measurement time for obtaining a satisfactory sample spectrum.

However, no such standards or guidelines have been established yet. To develop universally optimal parameters or models that predict these parameters, a reliable criterion is needed. Given that the signal-to-noise ratio (SNR) is a widely used quantitative indicator of spectral quality [15,17,19–28], it presents itself as a good candidate for this criterion. In this study, we define SNR as [23,29]:

$$SNR = \frac{H_{\mathrm{pk}}}{\sigma_{\mathrm{ns}}} \tag{1}$$

where H_{pk} is the max peak height in a spectrum, and σ_{ns} is the standard deviation of spectral noise.

In practice, H_{pk} is usually calculated after manual or automatic determination of baseline. σ_{ns} is usually calculated by selecting a region of the spectrum without peaks [23,30]. However, manually selecting these regions poses challenges for automatic analysis, especially for a set of spectra from different substances. Although there is a recently developed automated method, double sliding-window (DSW) method, that can conduct baseline correction for H_{pk} estimation and estimation of SNR automatically [29], the validation was based on SNR $\lessapprox$ 100, and we found that the estimation accuracy of SNR decreases in spectra with very high SNR (i.e., SNR $\gtrapprox$ 500). This limitation highlights the need for more accurate automated SNR calculation methods to handle varying spectral conditions.

Therefore, the primary goal of this study is to improve the accuracy of estimating SNR values of spectral data for a better determination of spectral quality. Specifically, our objectives are to (1) modify the original DSW method to more accurately estimate σ_{ns} and SNR, and (2) validate the modified method using simulated spectra with known σ_{ns} and SNR values, comparing its performance with the original method and other commonly used techniques. This paper initiates with mathematical definitions of spectra, followed by a brief overview of the original DSW method. Subsequently, we introduce a *k*-iterative enhancement of the DSW method. After determining the optimal value of *k*, we assessed the estimation accuracy of the new method. Additionally, performance of baseline correction was evaluated. Finally, we demonstrated two potential applications of the proposed method in analyzing environmental microplastic samples.

2. Materials and method

2.1. Definitions

A spectrum is defined as [29]:

$$\boldsymbol{Y} = f(\boldsymbol{X}) \tag{2}$$

where $\boldsymbol{X}$ is the wavenumber (cm^{-1}), and $\boldsymbol{Y}$ is the signal intensity. $\boldsymbol{Y}$ has three components [31]: peaks, denoted as $\boldsymbol{Y}_{\mathrm{pk}}$, baseline, denoted as $\boldsymbol{Y}_{\mathrm{bc}}$, and spectral noise, denoted as $\boldsymbol{Y}_{\mathrm{ns}}$:

$$\boldsymbol{Y} = \boldsymbol{Y}_{\mathrm{pk}} + \boldsymbol{Y}_{\mathrm{bc}} + \boldsymbol{Y}_{\mathrm{ns}} \tag{3}$$

Given that baseline is denoted as $\boldsymbol{Y}_{\mathrm{bc}}$, we use $\boldsymbol{Y}_{\text{-bc}}$ to denote baseline-free spectrum:

$$\boldsymbol{Y}_{\text{-bc}} = \boldsymbol{Y} - \boldsymbol{Y}_{\mathrm{bc}} = \boldsymbol{Y}_{\mathrm{pk}} + \boldsymbol{Y}_{\mathrm{ns}} \tag{4}$$

Peak height specifically refers to the signal intensity of a peak measured from the top of the peak to its corresponding baseline [23,29]). Considering that a spectrum may have multiple peaks, each with its own peak height, in this paper, we use *peak height* and *maximum peak height* interchangeably, focusing on the highest peak for SNR calculations. As mentioned in **Eq. (1)**, peak height is denoted as H_{pk}.

Spectral noise is assumed to be approximately uniform among the spectrum [32], and it follows a normal distribution:

$$\boldsymbol{Y}_{\mathrm{ns}} \sim N\left(0, \sigma_{\mathrm{ns}}^2\right) \tag{5}$$

where σ_{ns} is the standard deviation of spectral noise. With H_{pk} and σ_{ns}, SNR can be calculated according to **Eq. (1)**.

2.2. Double-sliding window (DSW) method

In this section, we briefly describe DSW method, and in the next section, we introduce modifications of DSW method.

The DSW method [29], developed from the sliding window method used in mass spectroscopy [33–36], involves sliding a window along the spectrum. This method utilizes the smoothed line of local minimum points in the window as the baseline. In application to Raman spectra, it was found that a smaller window was more effectively capturing baseline fluctuations, while a larger window was more adept at identifying the underlying baseline of peaks [29]. Thus, DSW method took advantages of both window sizes by combining the two baselines.

DSW method is an open source (https://github.com/River20104047/DSWmethod) automated baseline correction technique that also estimates σ_{ns}, H_{pk}, and SNR of the spectrum [29]. A flowchart of DSW method is shown in Fig. 1**a**. The input for DSW method is the spectrum, $\{\boldsymbol{X}, \boldsymbol{Y}\}$. There are four outputs: peak height, $H_{\mathrm{pk}|1}$, standard deviation of spectral noise, $\sigma_{\mathrm{ns}|1}$, signal-to-noise ratio, SNR_1, and corrected spectrum, $\{\boldsymbol{X}, \boldsymbol{Y}_{\text{-bc}|1}\}$, where $H_{\mathrm{pk}|1}$ is calculated by $\boldsymbol{Y}_{\text{-bc}|1}$, i.e., $H_{\mathrm{pk1}} = \max(\boldsymbol{Y}_{\text{-bc}|1})$, and SNR is calculated by $H_{\mathrm{pk}|1}$ and $\sigma_{\mathrm{ns}|1}$. Subscript 1 indicates that these variables are from the first iteration of the DSW method, which is used to differentiate from the output after *k*-th iteration of the new method.

Previous study has demonstrated the efficacy of the DSW method in accurately capturing elevated and fluctuating baselines, particularly in spectra with low SNR values (SNR $\lessapprox$ 100). Although there might be a slight reduction in peak height, the general shape of the peak is well preserved. The performance of DSW method was validated through analyses of both simulated spectra and Raman spectra from environmental microplastic samples [29].

2.3. k-iterative double-sliding window (DSW^k) method

Although DSW method performs well in its validation with SNR $\lessapprox$

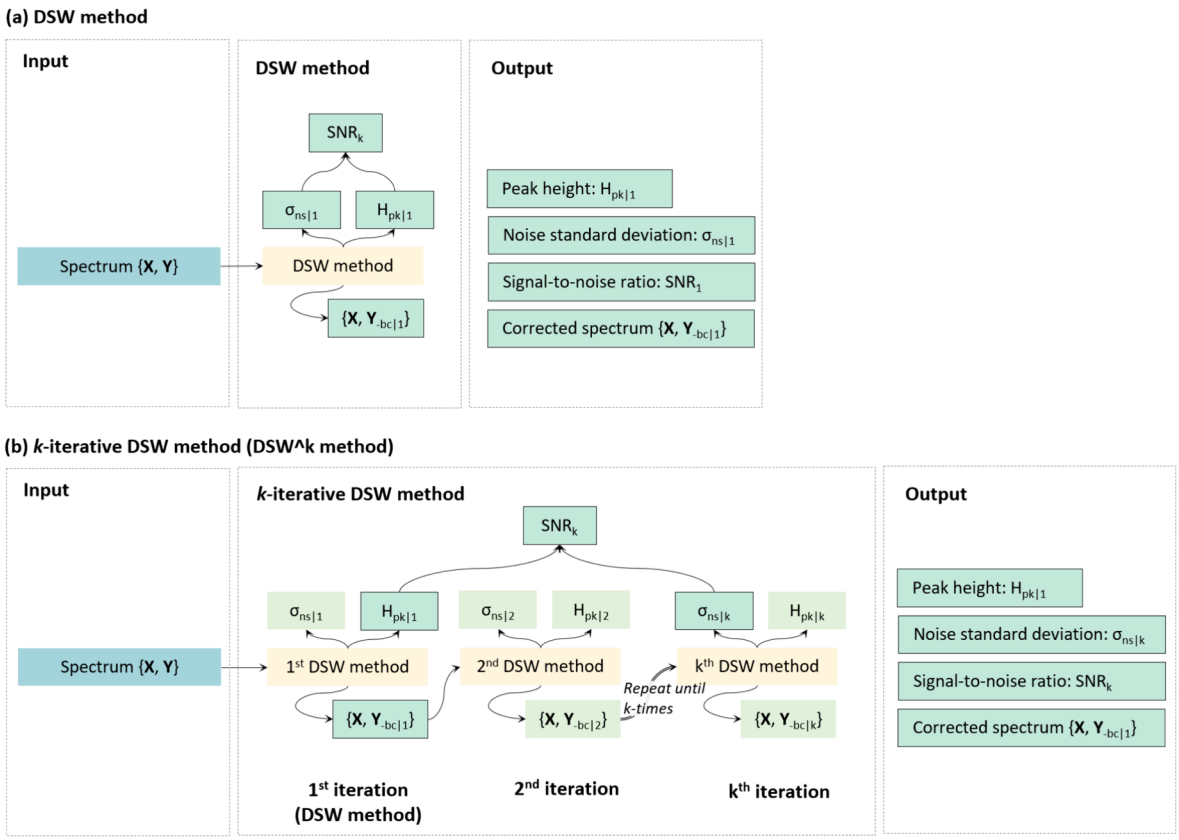


Fig. 1. Flowchart of DSW method (Yang et al., 2023d) and *k*-iterative DSW method (DSW^k method). {***X***, ***Y***} refers to original spectrum. {***X***, ***Y***$_{-bc}$}is the baseline corrected spectrum. σ_{ns} is standard deviation of spectral noise, H_{pk} is the peak height of a spectrum, and SNR refers to signal-to-noise ratio. Subscript of these variables with 1 refers to it is calculated by the 1st iteration of DSW method, and subscript with *k* refers to it is calculated by the k^{th} iteration of DSW method. Arrows indicate data flow.

100, we found that it tends to overestimate the standard deviation of spectral noise as SNR $\gtrapprox$ 500, especially in cases with fluctuating baselines. Preliminary trials indicated that elevated and fluctuating baselines may influence the estimation of σ_{ns}. This effect is due to the curvature of the baseline altering the probability density function of the distances between the upper and lower envelopes, which are determined by local maximum and minimum values in the sliding window. However, it is also noted that the initial application of DSW baseline correction could mitigate such curvature, improving the accuracy of envelopes capture. Therefore, as the method is iteratively applied, the baseline converges towards an approximately flat line, leading to σ_{ns} estimation that is less influenced by the original baseline fluctuations.

Based on the idea above, we developed a *k*-iterative DSW method (DSW^k method) to improve the accuracy of σ_{ns} and SNR estimation. The flowchart of DSW^k method is illustrated in Fig. 1**b**. In DSW^k method, a spectrum, {***X***, ***Y***}, is the input. This spectrum undergoes processing using a series of DSW methods.

The first intermediate output includes the peak height, $H_{pk|1}$, the standard deviation of spectral noise, $\sigma_{ns|1}$, and the corrected spectrum, {***X***, ***Y***$_{-bc|1}$}. This corrected spectrum then serves as the input for the second iteration of the DSW method. Subsequently, we obtain the second intermediate outputs: peak height, $H_{pk|2}$, standard deviation of spectral noise, $\sigma_{ns|2}$, and corrected spectrum, {***X***, ***Y***$_{-bc|2}$}. This iterative process continues until the *k*-th iteration, at which point we obtain the peak height, $H_{pk|k}$, standard deviation of spectral noise, $\sigma_{ns|k}$, and corrected spectrum, {***X***, ***Y***$_{-bc|k}$}.

Preliminary results indicate that the spectral noise estimation improves with additional iterations, making $\sigma_{ns|k}$ from the final iteration the most accurate. However, it was also found that the 1st iteration of the DSW^k method provided a more accurate estimate of peak height compared to subsequent iterations, making {***X***, ***Y***$_{-bc|1}$} and H_{pk1} the better estimate for the corrected spectrum and peak height. Consequently, the SNR for the DSW^k method is calculated using the formula:

$$SNR_k = \frac{H_{pk|1}}{\sigma_{ns|k}} \tag{6}$$

As a result, the four outputs of the DSW^k method are: peak height, $H_{pk|1}$, standard deviation of spectral noise, $\sigma_{ns|k}$, signal-to-noise ratio, SNR_k, and the corrected spectrum, {***X***, ***Y***$_{-bc|1}$} (Fig. 1**b**). Therefore, the primary distinction between the original DSW method and the DSW^k method is in the enhanced estimation of spectral noise and SNR.

DSW^k method is programmed in Matlab script and available online (https://github.com/River20104047/DSW-k-method) with a demonstration spectra dataset in csv format.

2.4. Simulated spectra

In order to evaluate and validate the performance of the DSW^k method, simulated spectra were used. Since the focus of this method is on microplastics, the simulated spectra are designed to be reflective of microplastics. Here, we adopted the simulated spectra from the previous study [29], which are analogous to the Raman spectrum of polyethylene (PE). However, to improve clarity in interpretation, we made revisions to the original formula, adjusting the peak height to a standardized value of 100.

The expression for the simulated spectra follows the same format as **Eq. (3)**. For the baseline component, $\boldsymbol{Y}_{bc}$, it is expressed as:

$$\boldsymbol{Y}_{bc} = A \bullet \sin(a \bullet \boldsymbol{X} + \psi_a) + B \bullet \sin(b \bullet \boldsymbol{X} + \psi_b) \tag{7}$$

where *A* represents the amplitude of the hill-shape elevation in the

baseline, and B is the amplitude of small fluctuations within the baseline. a and b are the frequency components of the baseline, and ψ_a and ψ_b are the phase shift components.

For the peak components, $\boldsymbol{Y}_{pk}$, the expression combines three distinct peaks to mimic Raman spectrum of PE:

$$\boldsymbol{Y}_{pk} = \sum_{i=1}^{3} \boldsymbol{Y}_{pki} \tag{8}$$

Each peak, $\boldsymbol{Y}_{pki}$, is modelled using a normal distribution pattern [20,27]:

$$\boldsymbol{Y}_{pki} = w_{pki} \bullet h_{pki} \bullet \frac{1}{w_{pki}\sqrt{2\pi}} e^{\frac{\left(X - x_{pki}\right)^2}{2 \bullet w_{pki}}} \tag{9}$$

where w_{pki} and h_{pki} are width and height factors of each peak, and x_{pki} is the center position of the peak.

For the spectral noise component, $\boldsymbol{Y}_{ns}$, random samples are generated using **Eq. (5)**, which models the signals as normally distributed with a standard deviation of $\sigma_{ns|ref}$. The subscript *ref* indicates that this value serves as a reference, i.e., known value, in the validation process. Similarly, the reference peak height, denoted as $H_{pk|ref}$, is calculated using the maximum value of $\boldsymbol{Y}_{pk}$:

$$H_{pk|ref} = \max\left(\boldsymbol{Y}_{pk}\right) \tag{10}$$

The relationship among reference SNR, denoted as SNR_{ref}, $H_{pk|ref}$, and $\sigma_{ns|ref}$ is the same as **Eq. (1)**.

All parameters used to simulate the spectra are summarized in **Table S1**. These parameters were adjusted empirically to closely represent the characteristics of PE [29], with the reference peak height set at 100, i.e., $H_{pk|ref} = 100$. The values for A and B were specifically varied to reflect four typical types of spectral baselines (Fig. 2): α-type (A=0, B=0), which represents a spectrum with flat baseline, (2) β-type (A=0, B=10), which represents a spectrum with elevated baseline, (3) γ-type (A=100, B=0), which represents a spectrum with fluctuated baseline, and (4) δ-type (A=100, B=10), which represents a spectrum with baseline that exhibits both elevation and fluctuation. Additionally, spectra with a wide range of SNR values were also considered. For this study, we focused primarily on SNR values between 10 and 1000, which represents the common range of SNR values for Raman spectra [15,20,21,37].

It is important to note that while some of the simulated spectra realistically reflect the characteristics of actual Raman spectra from environmental or standard samples, others, such as the δ-type with an SNR value of 1000, are less realistic. An SNR value of 1000 generally indicates very high-quality spectra, which are unlikely to exhibit significant baseline elevation and fluctuation. However, these less realistic combinations are included in the study to provide a comprehensive overview of the performance across a wide range of potential scenarios.

2.5. Evaluation of σ_{ns} and SNR estimation

The performance of the DSW^k method was evaluated and compared with the original DSW method, as well as empirical mode decomposition (EMD) [38] and variational mode decomposition (VMD) [39]. EMD, a well-established technique in signal analysis, has been widely used in spectral processing [40–43]. It decomposes a signal into a set of intrinsic mode functions (IMFs) and a residue, sequentially extracting signal components from high to low frequencies. In contrast, VMD, a newer approach in signal processing [44,45], has recently been applied to Raman spectra processing [21]. VMD decomposes a signal by adaptively segregating it into a predefined number of band-limited IMFs and a residue, isolating different frequency components. A demonstration of EMD and VMD is shown in **Fig. S2**. Considering that the high-frequency

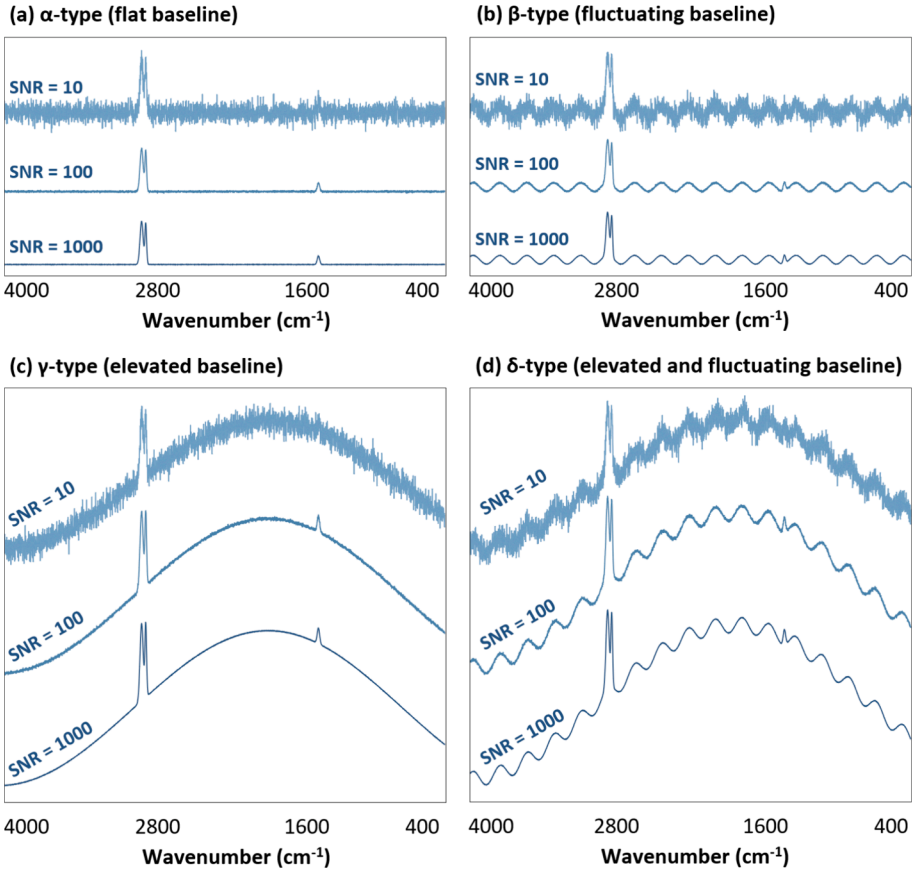


Fig. 2. Demonstration of the four types of simulated spectra with different SNR values. For a more detailed review, please refer to **Fig. S1**.

signals and residues are potential candidates for estimating σ_{ns}, the first three IMFs, which are those with the highest frequencies, and the residues from both EMD and VMD were selected as candidates for σ_{ns} estimation. Besides, for the SNR comparison, it was limited to the DSW method and the DSWˆk method, as EMD and VMD do not estimate peak height.

To evaluate performance of estimating σ_{ns} and SNR, we used SNR values from 10 to 1000, in increments of 10, i.e., $SNR_{ref} \in \{SNR_{ref|i} \mid SNR_{ref|i} = 10i, i = 1,2,3\ldots100\}$. Given the reference peak height $H_{pk|ref} = 100$, the corresponding spectral, $\sigma_{ns|ref}$, ranges from 0.1 to 10, i.e., $\sigma_{ns|ref} \in \{\sigma_{ns|ref|i} \mid \sigma_{ns|ref|i} = 10/i, i = 1,2,3\ldots100\}$, where $\sigma_{ns|ref|i}$ is the i-th $\sigma_{ns|ref}$ value.

For the i-th SNR_{ref} or $\sigma_{ns|ref}$, considering the significant change of magnitude in these values, the performance of σ_{ns} and SNR estimation was evaluated using the ratio difference under log-scale [46]:

$$d_i = 10 \bullet \log_{10}\left(\frac{z_{est|i}}{z_{ref|i}}\right) \tag{11}$$

where d_i is the decibel difference, a measure of the deviation between the estimated value, $z_{est|i}$, and the reference value, $z_{ref|i}$. **Eq. (10)** can be transformed into $10^{0.1d_i} = z_{est|i}/z_{ref|i}$. Thus, d_i can provide information about the accuracy of estimation.

To evaluate the overall performance across different SNR_{ref} or $\sigma_{ns|ref}$ values, we define the average decibel difference as:

$$d = \frac{\sum_{i=1}^{I} w_i \bullet d_i}{\sum_{i=1}^{I} w_i} \tag{12}$$

where w_i is the weight for each decibel difference d_i, calculated as the logarithmic distance between consecutive reference values: $w_i = \log_{10}(z_{ref|i+1}) - \log_{10}(z_{ref|i})$.

Additionally, we define the average absolute decibel difference as:

$$|d| = \frac{\sum_{i=1}^{I} w_i \bullet |d_i|}{\sum_{i=1}^{I} w_i} \tag{13}$$

In Eqn (12), d measures the average logarithmic bias between estimated and reference values. This metric indicates whether estimates tend to systematically deviate above or below the reference values, suggesting the presence of systematic error if d is substantially different from zero. In Eqn (13), $|d|$ measures the average magnitude of deviations, disregarding their direction, between estimated and reference values. It serves as a measure of the overall precision of the estimates. For an optimal model, both d and $|d|$ values should ideally be zero, indicating no systematic bias and minimal random error, respectively.

2.6. Evaluation of baseline correction

Although the estimated baseline of the DSWˆk method remains the same with original DSW method, as previously evaluated [29], we conducted additional evaluations under more complex and challenging conditions to gain a more comprehensive understanding of the performance of DSWˆk method, with range of spectra of SNR changing from 10 to 1000.

For a given SNR value, i.e., $SNR_{ref|i}$, the performance of DSW was evaluated by root mean square error (*RMSE*) [47]:

$$RMSE = \sqrt{\frac{\sum (\boldsymbol{Y}_{-bc} - \widehat{\boldsymbol{Y}}_{-bc})^2}{N}} \tag{14}$$

where $\boldsymbol{Y}_{-bc}$ is the baseline-free simulated spectrum, and $\widehat{\boldsymbol{Y}}_{-bc}$ is the baseline-corrected spectrum. N is the number of signals in a spectrum. In addition to *RMSE*, the correlation coefficient between $\boldsymbol{Y}_{-bc}$ and $\widehat{\boldsymbol{Y}}_{-bc}$, noted as R, was also utilized to assess performance [29]. Again, these evaluations were based a range of SNR_{ref} values from 10 to 1000, specifically $SNR_{ref} \in \{SNR_{ref|i} \mid SNR_{ref|i} = 10i, i = 1,2,3\ldots100\}$.

In addition to the evaluation using simulated spectra, the performance of DSWˆk method was further assessed using environmental samples. Here, we utilized dataset from a recent study [15], which included 413 microplastic samples collected from surface seawater off the Shimane Prefectures in the Sea of Japan. The composition of these samples was previously identified using FTIR, with 39 % consisting of polyethylene (PE), 24 % polypropylene (PP), and 29 % polystyrene (PS), before undergoing Raman spectroscopy analysis. Unlike the adaptive approach, all samples were consistently measured using a Raman spectrophotometer (NRS-4500, JASCO Inc., Japan) under fixed conditions: wavenumber range of 400–4000 cm^{-1}, 1 s exposure, φ = 17 μm slit, 16 accumulations, 532 nm excitation laser, 41.4 mW laser power, and a 20 × objective. This approach ensured that the spectral data encompassed a broad range of SNR values, making it an ideal dataset for evaluation of DSWˆk method.

Given the varying scale and intensity of the spectra, *RMSE* was deemed unsuitable for our analysis. Instead, we utilized the correlation coefficient, R, as our primary metric for evaluation. In the context of environmental microplastic samples, R is defined as:

$$R = \frac{Cov(\boldsymbol{Y}_s, \boldsymbol{Y}_l)}{\sigma_s \bullet \sigma_l} \tag{15}$$

where R is the correlation coefficient between sample spectrum and spectrum of corresponding pristine standard sample, i.e., library spectrum. $\boldsymbol{Y}_s$, and $\boldsymbol{Y}_l$ are the signal intensities of sample spectrum and library spectrum. σ_s and σ_l refers to standard deviation of signal intensities. The library spectrum includes pristine standard samples of polyethylene (PE), polypropylene (PP), and polystyrene (PS) (Scientific Polymer Products Inc., USA), acquired using the same device and measurement parameters except for with 64 accumulations. For Eqn (15), R corresponds to the ability of polymer type identification, as the correlation coefficient is a common algorithm for polymer type identification in microplastic analysis [48].

2.7. Data analysis

Simulations with the DSW method, DSWˆk method, EMD, and VMD were conducted using Matlab 2024a (The MathWorks, Inc., Natick, MA, USA). For each $SNR_{ref|i}$ value, simulations were run for $n = 1000$ iterations to obtain good statistics. The mean of these 1000 iterations was used to represent the estimate, in the meanwhile, the 2.5 % and 97.5 % quantiles were calculated to reflect the uncertainty in the estimates. Additionally, statistical significance was assessed at $\alpha = 0.05$, with analyses performed using Matlab 2024a.

3. Results and discussion

3.1. Determination of iteration number (k)

We first determined the number of iterations needed to obtain a convergent estimation of $\sigma_{ns|k}$ and SNR_k using $SNR_{ref} = 10$, 100, and 1000. Preliminary results showed that $k = 50$ could ensure the convergence for all scenarios. Thus, we evaluated $\sigma_{ns|k}$ and SNR_k, as well as $H_{pk|k}$, as a function of k.

For spectra with a low quality, i.e., $SNR_{ref} = 10$ (**Fig. S3a-k**), $\sigma_{ns|k}$ values align closely with reference values, indicating good accuracy. However, while the estimation of peak height $H_{pk|k}$ is initially accurate at the first iteration, i.e., $k = 1$, it tends to decrease and deviate from the reference value as k increases. This deviation is attributed to the tendency of DSW method to slightly reduce the height of wide and large peaks [29]. Given that $H_{pk|k}$ demonstrates the best accuracy at $k = 1$, the $H_{pk|1}$ value from the initial pass is used as the estimated peak height for the DSWˆk method, as previously mentioned in **Section 2.3**. Consequently, SNR_k also exhibits good performance, benefiting from the accurate estimation results of $H_{pk|1}$ and $\sigma_{ns|k}$. Notably, both $\sigma_{ns|k}$ and SNR_k

show negligible sensitivity to changes in k values.

For spectra with $SNR_{ref} = 100$ (Fig. S3**l-w**), which represent better-quality spectra, the SNR_k values align more closely with the reference values by $k = 3$. This indicates that the 3nd iteration of the DSW method can provide accurate estimations. This improvement is attributed to the more precise estimation of spectral noise, $\sigma_{ns|k}$ while $H_{pk|1}$ continues to serve as an accurate estimate of peak height.

For high-quality spectra with $SNR_{ref} = 1000$ (Fig. S3**x-ι**), increasing the k value to around 10 generally results in good SNR estimations, although there is a tendency to underestimate SNR in cases with fluctuated baselines.

In general, elevated and fluctuating baselines in the DSW method initially led to an overestimation of σ_{ns}. However, repeated application of the method tends to mitigate this overestimation, thus enhancing the accuracy of σ_{ns} estimations. Simultaneously, $H_{pk|1}$ provides a reliable estimate of peak height, thereby ensuring that SNR_k remains satisfactory estimation of SNR_{ref}. In addition, a k value of around 10 results in approximate convergence for $\sigma_{ns|k}$ and SNR_k, suggesting a near-stable condition but without full certainty. Thus, to balance convergence assurance and computational intensity, a k value of 20 was determined to be a practical value for the DSWˆk method.

3.2. Performance of estimating spectral noise (σ_{ns})

Since σ_{ns} is critical for SNR estimation and reflects the properties of a spectrum, its estimation performance was evaluated. In this section, we compare the σ_{ns} estimation performance of the DSWˆk method, DSW method, EMD, and VMD. The evaluation is based on $SNR_{ref} \in \{SNR_{ref|i} \mid SNR_{ref|i} = 10i,\ i = 1,2,3\ldots100\}$.

We started by comparing EMD and VMD. For EMD method (Fig. S4), IMF01 effectively captures σ_{ns} when $\sigma_{ns} \gtrapprox 2$ (i.e., SNR $\lessapprox$ 50) (Fig. S4**a-c**), indicating that the first IMF of EMD can effectively handle noisy spectral signals. However, subsequent IMFs, such as IMF02 and IMF03, show poor performance, as evidenced by much larger d and $|d|$ values (Table S2). Similarly, for VMD method (Fig. S5), IMF01 captures σ_{ns} effectively under the same conditions but performs poorly when σ_{ns} value is small. In contrast, the residues of VMD demonstrate better performance across all ranges of σ_{ns} (Fig. S5**d, h, l, p**). The satisfactory performance of VMD residues could be attributed to the fact that these residues represent the remainder of the signal after all the significant modes have been extracted, that are more representative of noise [39]. Combining the results from EMD and VMD, the residues of the VMD method emerge as the best candidate for capturing σ_{ns}, which is consistent with better performance of VMD in decomposition of Raman spectra data [21]. Therefore, the VMD residue is chosen for further comparison with the DSWˆk method. We use $\sigma_{ns|VMDr}$ to denote the σ_{ns} estimated by residue of VMD method.

The performance of estimating σ_{ns} by DSW method, DSWˆk method, and VMD method is shown in Fig. 3.

For the α-type spectra (Fig. 3**a**), $\sigma_{ns|k}$ and $\sigma_{ns|1}$ overlap, both showing good alignment with the 1:1 line, which indicates good performance. In contrast, $\sigma_{ns|VMDr}$ consistently falls below this line, suggesting a systematic underestimation. This observation is supported by the smaller d and $|d|$ values (Table S2**a**). Given $d_{DSWk} = 0.06$, it suggests that, on average, the DSWˆk method estimates the σ_{ns} at 1.01 times of the actual value, indicating very good accuracy.

For β-type (Fig. 3**b**), $\sigma_{ns|k}$ curve stays along 1:1 line while $\sigma_{ns|1}$ has good performance with large σ_{ns} values, i.e., $\sigma_{ns|ref} \gtrapprox 5$. In addition, it tends to overestimate σ_{ns} at small σ_{ns} values. This is due to fact that as σ_{ns} decreases, the influence elevated baseline becoming more profound, resulting in larger influence on σ_{ns} estimation by DSW method. However, after multiple iterative use of DSW method, the influence can be substantially mitigated. For VMD, it shows consistent underestimation. Given $d_{DSWk} = 0.31$, it suggests that, on average, the DSWˆk method estimates the σ_{ns} at 1.07 times of the actual value, indicating a slight overestimation.

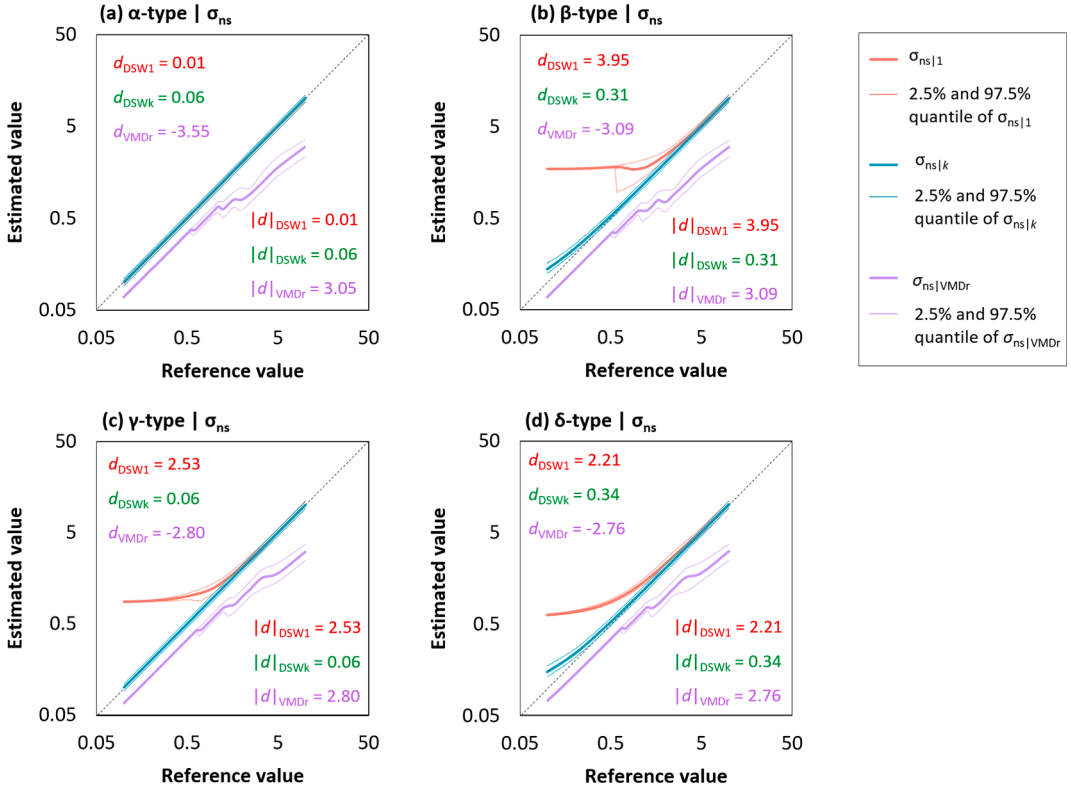


Fig. 3. Performance of prediction for spectral noise (σ_{ns}) among different methods. Dashed line refers to x: y = 1: 1. d and $|d|$ refer to average decibel between estimated value and reference value. For subscript, DSW1 refers to original DSW method, DSWk refers to DSWˆk method, and VMDr refers to residue of variational mode decomposition method. Note that in **(a)**, $\sigma_{ns|1}$ and $\sigma_{ns|k}$ overlaps. For individual plots, please refer to Figs. S5, S6.

For the γ-type and δ-type spectra, similar trends in performance were observed as with other types. The DSW method provides accurate estimations of σ_{ns} for larger values ($\sigma_{ns|ref} \gtrapprox 5$), but it tends to overestimate σ_{ns} when dealing with high-quality spectra. In contrast, the VMD method consistently underestimates σ_{ns} across all ranges. These patterns are in consistent with the relatively smaller d and $|d|$ values for the DSWˆk method, indicating that it maintains closer adherence to the reference values and shows less deviation compared to the original DSW method. Given $d_{DSWk} = 0.06$ and 0.34, it suggests that, on average, the DSWˆk method estimates the SNR at 1.01 and 1.08 times of the actual value for γ-type and δ-type spectra, respectively.

In general, while the VMD method shows potential for estimating σ_{ns} if its systematic bias could be addressed, the DSWˆk method consistently outperforms other methods across all tested $\sigma_{ns|ref}$ values, particularly when dealing with elevated and/or fluctuating baselines. The performance of the DSWˆk method underscores its reliability and effectiveness in handling complex baseline conditions more effectively than other methods.

3.3. Performance of estimating SNR

We then conducted further comparison of the performance in estimating SNR values between the DSW method and the DSWˆk method. The comparison between reference value and estimated value are presented in Fig. 4. For individual plots, refer to **Fig. S6**.

For α-type (Fig. 4**a**), the data of two methods overlap and are close to 1:1 line, suggesting that both methods show good predictive accuracy. Given $d_{DSWk} = -0.31$ and $d_{DSW1} = -0.26$, it suggests that, on average, the DSWˆk method and DSW method estimates the SNR at 0.93 and 0.94 times of the reference value, respectively, indicating a good performance for both methods.

For β-type, γ-type, and δ-type spectra (Fig. 4**b, 4c, 4d**), which exhibit either elevated, fluctuating, or both types of baselines, there is a consistent pattern of good accuracy at lower SNR values, i.e., SNR $\lessapprox$ 500. However, as the SNR increases, both methods tend to underestimate SNR. Notably, the SNR estimations from the DSWˆk method remain closer to the 1:1 line, indicating smaller deviations. In contrast, the estimations from the original DSW method begin to deviate under these conditions. This discrepancy is largely due to the DSW method's overestimation of σ_{ns} because curvature in the baseline negatively impacts the accuracy of noise probability density function identification, leading to overestimated σ_{ns} and consequently underestimated SNR. While the DSWˆk method also shows a tendency to slightly underestimate SNR, this minor deviation is considered acceptable, particularly since SNR evaluations often focus on lower SNR scenarios.

Given the d_{DSWk} values, it shows that, on average, the DSWˆk method estimates the SNR at 0.95, 0.89, and 0.95 times of the actual value, indicating a slight underestimation. In contrast, the DSW method estimates the SNR at 0.41, 0.51, and 0.62 times the actual value, indicating a notable underestimation. Thus, DSWˆk method demonstrates a 74.5 % − 131.7 % improvement over the DSW method in spectra with elevated and/or fluctuating baselines.

In general, DSWˆk method outperformed DSW method in estimating SNR values of a spectra, and DSWˆk method is a reliable tool for calculating σ_{ns} and SNR of a spectrum.

3.4. Performance of baseline correction

In this section, we assess the efficacy of the DSWˆk method for baseline correction using simulated spectra that span a broad range of SNR values, specifically from 10 to 1000, i.e., $SNR_{ref|i}$. Following this analysis, the subsequent section will focus on evaluation of the performance in baseline correction using environmental microplastic samples.

The correlation coefficient between baseline corrected spectra and

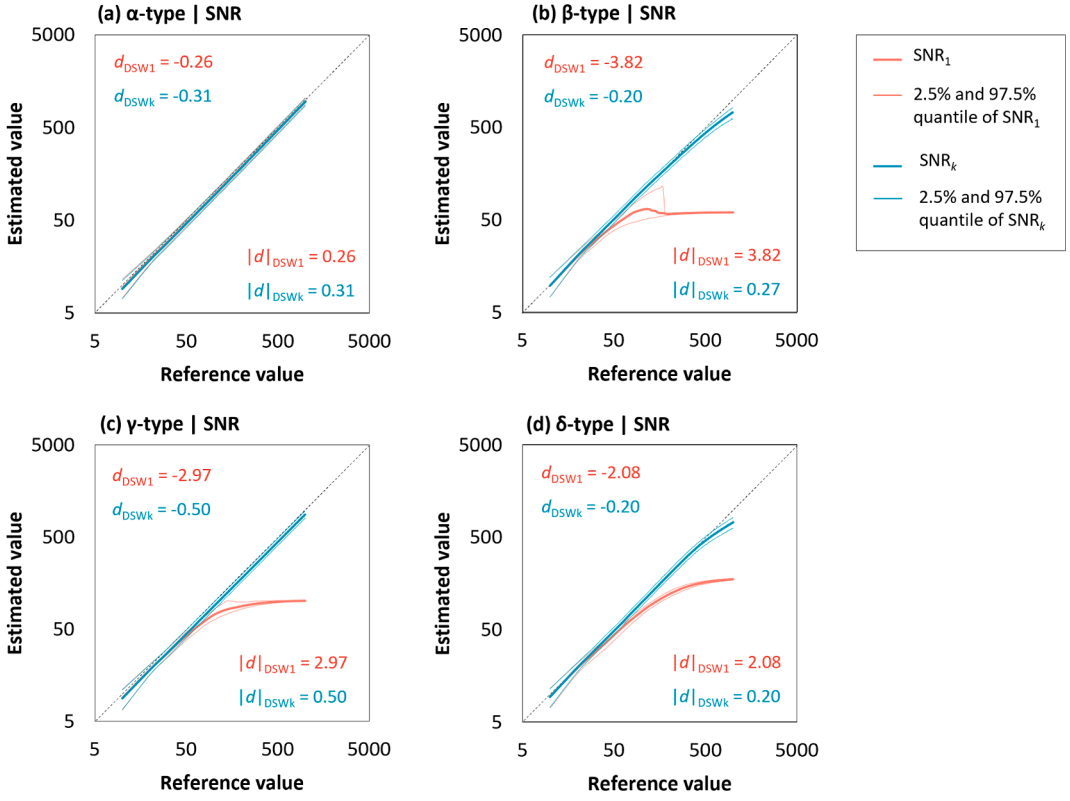


Fig. 4. Performance of prediction for signal-to-noise ratio (SNR) among different methods. Dashed line refers to x: $y = 1$: 1. d and $|d|$ refer to average decibel between estimated value and reference value. For subscript, DSW1 refers to original DSW method and DSWk refers to DSWˆk method. Note that in **(a)**, SNR_1 and SNR_k overlaps. For individual plots, please refer to **Fig. S6**.

original spectra, *R*, as a function of SNR is presented in Fig. 5**a** (for individual plot, please refer to **Fig. S7**). For α-type and γ-type spectra, the *R* value exhibits an increasing trend, converging to approximately 1 at SNR ≈ 100, with a corresponding decrease in uncertainty. At lower SNR values around 10, which typically indicate unsatisfactory spectral quality, the *R* value is around 0.9. For β-type spectra, *R* values converge to a slightly lower yet still robust value of 0.99. Conversely, for δ-type spectra, *R* values initially reach 0.99 but drop after *SNR* ≈ 200. This decline is attributed to challenges in handling large SNR values coupled with baseline fluctuations, which can lead to misidentification of baseline components as peaks (Fig. 2**d** SNR=1000). Although such cases of spectra are uncommon in practice, and the performance decrease is noticeable. Given that a similarity measure of 0.6 is commonly utilized for microplastic polymer type identification [48], the performance of the DSWˆk method in baseline correction is regarded as useful and can be considered as a reliable tool for baseline correction.

RMSE as a function of SNR is presented in Fig. 5**b** (for individual plot, please refer to **Fig. S8**). For all types except δ-type, *RMSE* values decrease as *R* values increase, indicating that as the quality of the spectra improves, so does the performance of the baseline correction. This trend aligns with the results observed for *R* values, confirming consistent improvement across most spectral types. However, similar to the behavior observed with *R* values, the performance for δ-type spectra declines after SNR ≈ 200.

Overall, DSWˆk method demonstrates satisfactory performance across a wide range of SNR values for each type of spectra, indicating its utility as a reliable tool for automatic baseline correction.

3.5. Performance of baseline correction – Environmental samples

The performance of baseline correction was further evaluated with *R* values derived from environmental samples. Given the predominance of PE ($n = 152$), PP ($n = 110$), and PS ($n = 119$) in the dataset, analysis primarily focused on these three polymer types. The correlation coefficients before baseline correction (R_{NBC}) and after baseline correction (R_{BC}) for each of these polymers are presented in Fig. 6.

In terms of PE (Fig. 6**a**), when R_{NBC} value is low, i.e., $R_{NBC} \lessapprox 0.15$, baseline correction does not show improvements after baseline correction, which is due to the original spectra lacking characteristic peaks. As R_{NBC} increases, i.e., $0.15 \lessapprox R_{NBC} \lessapprox 0.25$, a statistically noticeable improvement in the *R* values after baseline correction was observed. This improvement occurs because baseline correction could substantially improve spectral similarity by removing the interference from baseline (**Fig. S9a-d**). As R_{NBC} continues to rise, i.e., $0.25 \lessapprox R_{NBC} \lessapprox 0.90$, the low similarity is primarily due to elevated or fluctuating baselines (**Fig. S9e-l**). In these instances, the DSWˆk method effectively captures and corrects the baseline, dramatically increasing the similarity to the spectrum of pristine standard PE. However, when R_{NBC} value approaches 1.00, i.e., $0.90 \lessapprox R_{NBC} \lessapprox 1.00$, the uncorrected spectra already closely resemble the pristine PE spectrum, hence baseline correction has only a minor influence on *R* values (**Fig. S9m, n**).

Further evaluation suggests that the DSWˆk method can effectively capture the baseline of environmental PE (**Fig. S9**). However, it tends to lower the height of low and wide peaks, particularly in spectra with relatively low SNR values (**Fig. S9a-d**). This reduction is attributed to the fact that such peaks are less statistically distinguishable from noise due to their low height, and such observation aligns with the findings discussed in **Section 3.1**. While the DSWˆk method also reduces the height of higher peaks, the absolute reduction is minimal and generally negligible (**Fig. S9e, n**).

In terms of PP (Fig. 6**b**), the scatter plot shows a pattern similar to PE, where baseline correction does not enhance spectra with very small R_{NBC} values, i.e., $R_{NBC} \lessapprox 0.15$. However, as $0.25 \lessapprox R_{NBC} \lessapprox 0.90$, baseline correction can substantially increase the similarity to the spectrum of pristine standard PE. For spectra that already show good similarity, i.e., $0.90 \lessapprox R_{NBC} \lessapprox 1.00$, the influence of baseline correction is minimal (**Fig. S10**). Additionally, similar to the observations with PE, the DSWˆk method can effectively capture the baselines but tends to reduce the peak height in cases where peaks are low (**Fig. S10a-f**). This reduction is less influential as the peak height increases (**Fig. S10g-n**).

In terms of PS (Fig. 6**c**), the scatter plot pattern is consistent with those observed for PE and PP, and the baseline can be effectively captured (**Fig. S11**). However, R_{NBC} values for PS are generally lower than those for PE and PP, with the maximum R_{NBC} peaking at only 0.23. This is likely due to the intrinsic properties of PS, which tend to exhibit greater noise under the current measurement parameters. This suggests that a longer exposure time may be necessary for PS to achieve better identifiable spectra.

Overall, DSWˆk method demonstrates robust performance in capturing baselines and enhancing the identifiability of spectra. However, it should be noted that the method may slightly reduce the peak height in spectra with low SNR.

3.6. Application of DSWˆk method (I) – Optimal exposure time for environmental microplastic samples

In this section and the next section, we demonstrate two practical applications of the DSWˆk method, showcasing the usefulness of accurately estimate spectral noise and SNR in environmental microplastic samples.

One critical parameter in Raman spectroscopy is exposure time, which can influence the spectra quality, as well as the total measurement time. Thus, proper adjustment of exposure time is essential for achieving high-quality spectral data, as well as for ensuring efficient and rapid data collection [17,23].

Here, we analyzed a white-colored PE fragment from a previous study [49], a type of environmental microplastic commonly found in marine environments [50]. Measurements were carried out using a

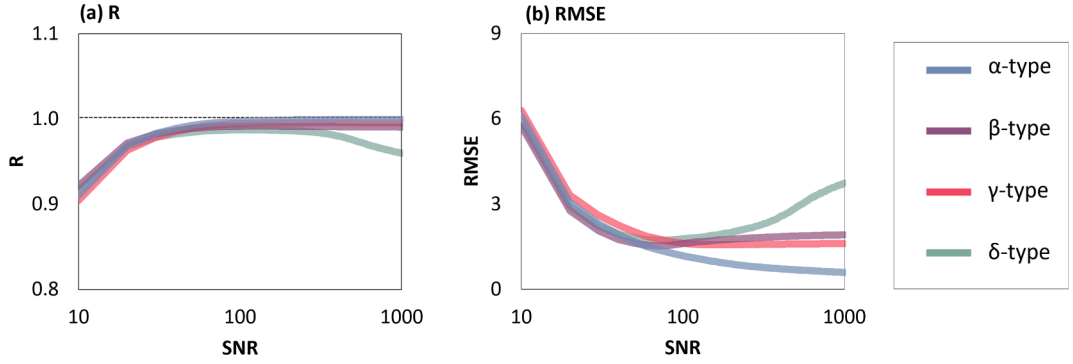


Fig. 5. Correlation coefficient (*R*) and root mean square error (*RMSE*) as a function of SNR. For individual plots with 2.5 % and 97.5 % quantiles, please refer to **Figs. S7, S8**.

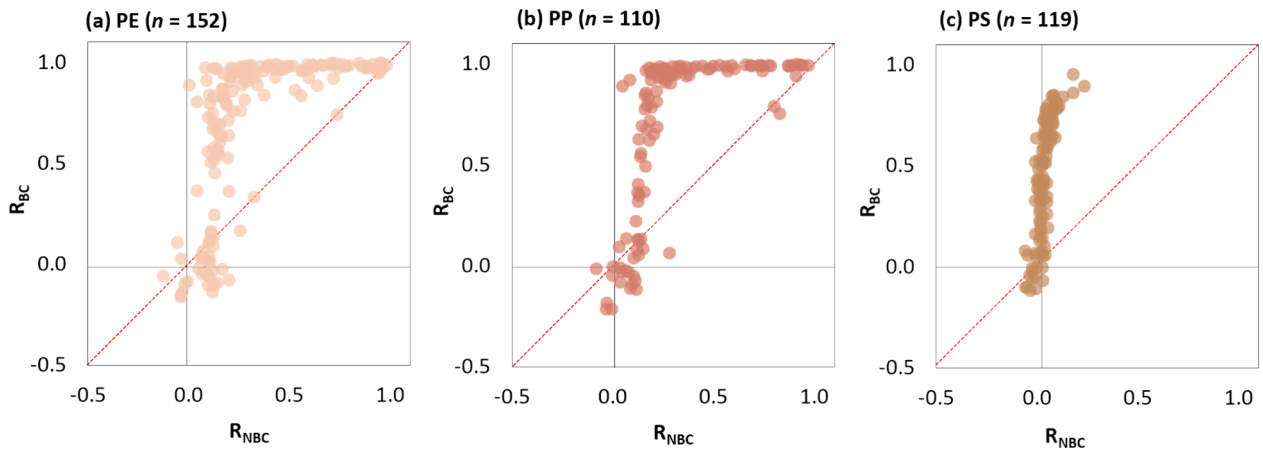


Fig. 6. Correlation coefficient (R) between environmental microplastic samples and spectrum of the corresponding pristine standard sample. For x-axis, R_{NBC} refers to R calculated by original spectra, i.e., no baseline correction, and for y-axis, R_{BC} refers to R calculated by baseline-corrected spectra. n refers to sample size. The red dashed line refers to 1:1 ratio.

Raman spectrophotometer (NRS-4500, JASCO Inc., Japan), utilizing common parameters including wavenumber range of 400–4000 cm^{-1}, φ = 17 μm slit, 2 accumulations, and 20 × objective [29], with varying exposure times of 0.1 s, 0.2 s, 1 s, 2 s, 4 s, 8 s, 16 s, 32 s, and 64 s, and with varying laser types of red (785 nm, 76.5 mW) or green (532 nm, 41.4 mW) lasers. These exposure durations and lasers are chosen for being practically used parameters [12,51]. To ensure robust statistical results, three measurements were conducted on closely situated points on the surface of the sample (**Fig. S12d**).

The spectra obtained from the Raman spectrophotometer were subsequently analyzed using the DSWˆk method to determine H_{pk}, σ_{ns}, and SNR values. The results are presented in Fig. 7**a, 7b, 7c**, which demonstrates a positive linear relationship between $\log_{10}$(SNR) and $\log_{10}$ of exposure time. This linear relationship is consistent with theoretical expectations [19,52]. This relationship allows for the estimation of the minimal required exposure time to achieve some minimal requirement of SNR values, facilitating the determination of optimal acquisition parameters for future analyses. Given that these findings are based on measurements from a single particle, it is important to expand the dataset to include more particles to ensure that the results are representative of a broader sample.

3.7. Application of DSWˆk method (II) – Comparing SNR among different categories of environmental microplastic samples

Given that spectral quality may be influenced by microplastic colors or shapes [15,18], such physical properties of microplastic samples could be an indicator for initial parameter, which could accelerate the process of finding the optimal acquisition parameter. Moreover, understanding the relationship between polymer type and SNR could guide adaptive parameter adjustments for subsequent samples, utilizing a Bayesian approach to leverage prior measurement data [53].

To illustrate the potential benefits of adaptively adjusting Raman spectroscopy parameters strategy, we utilized the dataset from the previous study [15], as mentioned in **Section 2.6**. In addition to identified polymer types by FTIR, these microplastic samples were classified by both shape (fragment, foam, line, film, microbead and pellet) and color (transparent, white, black, brown, blue, green, orange, red, pink, violet, and yellow) providing a basis for analyzing the influence of polymer type and the physical characteristics on the effectiveness of fixed measurement settings.

We first evaluated the relationship between polymer type and SNR, and the results are presented in Fig. 7**d**. It was observed that PE and PP exhibit significantly higher SNR values than PS ($p < 0.05$), consistent with observations in **Section 3.5**. In terms of color (Fig. 6**e**), it was observed that microplastics in blue typically exhibit higher spectral quality, as indicated by higher SNR values, whereas microplastics in black consistently show significantly lower SNR values. In terms of shape (Fig. 6**f**), foam microplastics are found to have notably lower SNR values compared to other shapes.

While the specific causes of these variations are beyond the scope of this study, the findings suggest that microplastics with characteristics like black color, foam shape, or a tendency to be PS, may require longer exposure times to achieve optimal Raman spectra. Recognizing these trends could reduce the time required to find optimal exposure times. Moreover, these comparisons could be expanded to include combinations such as shape and color, potentially leading to the development of parameter guidelines that quickly estimate the most likely optimal exposure times based on these physical characteristics.

4. Conclusions

In this study, we introduced a k-iterative double sliding-window (DSWˆk) method designed for the estimation of spectral noise, SNR, and baseline correction. The performance was evaluated using simulated spectra and compared against other commonly employed methods. Convergent evaluation determined that a k value of 20 strikes an optimal balance between ensuring convergence and managing computational intensity. The DSWˆk method demonstrated outstanding performance across different spectral types (flat baseline, baseline with elevation, baseline with fluctuation, baseline with elevation and fluctuation) coupled with SNR values from 10 to 1000, achieving results that ranged from 1.01 to 1.08 times of the reference value in estimating spectral noise. It also showed that the estimated SNR values are 0.89 to 0.93 times of the reference value, demonstrating a 74.5 % – 131.7 % improvement over the DSW method in spectra with elevated and/or fluctuating baselines. Additionally, the DSWˆk method proved effective in correcting baselines and identifying polymers in environmental samples of PE, PP, and PS, despite the limitation of reducing the peak height in spectra with low SNR.

This method offers the potential to enhance the automatic and accurate evaluation of spectral quality and could assist in the development of guidelines for more rapid parameter adjustments in Raman measurements. Such improvements can help accelerate the analytical process and increase data accuracy. Ultimately, these advancements could contribute to a deeper understanding of the environmental fate of microplastics.

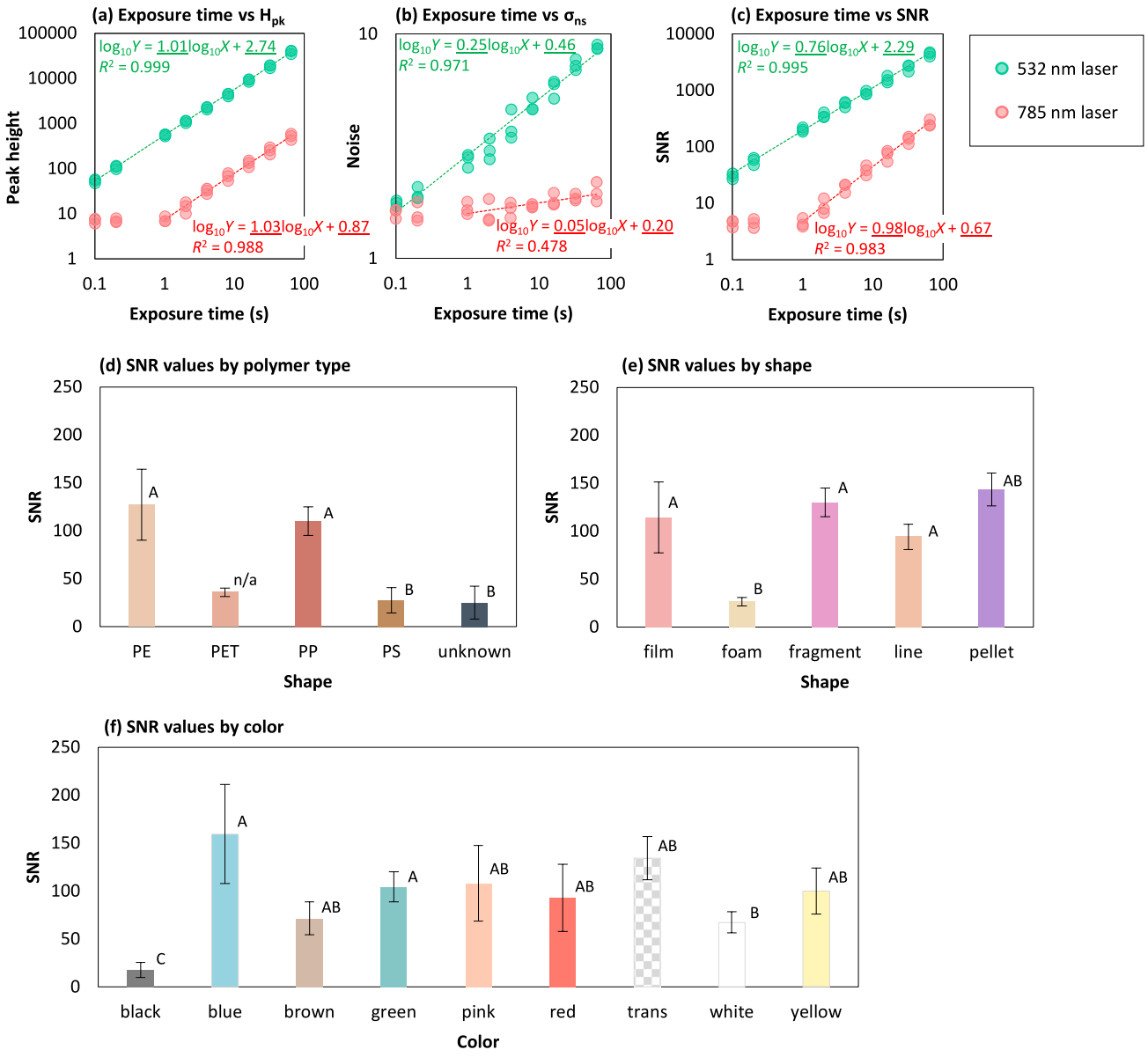


Fig. 7. Relationship between exposure time and spectra by color/shape. In **(a)**, **(b)**, **(c)**, H_{pk} is to peak height, σ_{ns} is spectral noise, and SNR is signal-to-noise ratio. When conducting linear regression, all data points were used for 532 nm laser while only 1 s and above data points were sued for 785 nm laser. In regression equations, underlined regression factors refer to significance ($p < 0.05$). In **(d)**, **(e),** and **(f)**, different letters refer to significant difference ($p < 0.05$), and the significance tests were based on log-scale. n/a refers to not applicable, where sample size is smaller than 3.

Funding sources

This work funding sources are by the Environmental Research and Technology Development Fund (JPMEERF20211003) of the Environmental Restoration and Conservation Agency provided by the Ministry of the Environment, Japan. The sponsors had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

CRediT authorship contribution statement

Zijiang Yang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Hiroya Nagashima:** Conceptualization, Investigation, Formal analysis, Investigation, Writing – review & editing. **Celik Murat:** Resources, Data curation, Writing – review & editing. **Hisayuki Arakawa:** Conceptualization, Resources, Writing – review & editing, 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 Dr. Heng Liu (Massachusetts Institute of Technology) for suggestions for the spectral analysis. We would like to thank Pin Wang (Kyoto University) with visualization

of the results and providing valuable comments. 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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2024.125061.

References

[1] T. Tokai, K. Uchida, M. Kuroda, A. Isobe, Mesh selectivity of neuston nets for microplastics, Marine Pollution Bulletin 165 (2021) 112111.
[2] H. Nakano, H. Arakawa, T. Tokai, Microplastics on the sea surface of the semi-closed Tokyo Bay, Marine Pollution Bulletin 162 (2021) 111887.
[3] H. Xu, H. Nakano, T. Tokai, T. Miyazaki, H. Hamada, H. Arakawa, Contamination of sea surface water offshore the Tokai region and Tokyo Bay in Japan by small microplastics, Marine Pollution Bulletin 185 (2022) 114245.
[4] M. Pirsaheb, H. Hossini, P. Makhdoumi, Review of microplastic occurrence and toxicological effects in marine environment: Experimental evidence of inflammation, Process Safety and Environmental Protection 142 (2020) 1–14.
[5] K. Zhang, A.H. Hamidian, A. Tubić, Y. Zhang, J.K. Fang, C. Wu, P.K. Lam, Understanding plastic degradation and microplastic formation in the environment: A review, Environmental Pollution 274 (2021) 116554.
[6] A. Hollerová, N. Hodkovicová, J. Blahová, M. Faldyna, P. Maršálek, Z. Svobodová, Microplastics as a potential risk for aquatic environment organisms–a review, Acta Veterinaria Brno 90 (1) (2021) 99–107.
[7] H. Ma, S. Pu, S. Liu, Y. Bai, S. Mandal, B. Xing, Microplastics in aquatic environments: Toxicity to trigger ecological consequences, Environmental Pollution 261 (2020) 114089.
[8] N. Khalid, M. Aqeel, A. Noman, M. Hashem, Y.S. Mostafa, H.A.S. Alhaithloul, S. M. Alghanem, Linking effects of microplastics to ecological impacts in marine environments, Chemosphere 264 (2021) 128541.
[9] L.G.A. Barboza, A.D. Vethaak, B.R. Lavorante, A.K. Lundebye, L. Guilhermino, Marine microplastic debris: An emerging issue for food security, food safety and human health, Marine Pollution Bulletin 133 (2018) 336–348.
[10] E. Winiarska, M. Jutel, M. Zemelka-Wiacek, The potential impact of nano-and microplastics on human health: Understanding human health risks, Environmental Research (2024) 118535.
[11] A. Käppler, D. Fischer, S. Oberbeckmann, G. Schernewski, M. Labrenz, K. J. Eichhorn, B. Voit, Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, Raman or both? Analytical and Bioanalytical Chemistry 408 (2016) 8377–8391.
[12] M. Dong, Z. She, X. Xiong, G. Ouyang, Z. Luo, Automated analysis of microplastics based on vibrational spectroscopy: are we measuring the same metrics? Analytical and Bioanalytical Chemistry 414 (11) (2022) 3359–3372.
[13] P. Larkin, Infrared and Raman spectroscopy: principles and spectral interpretation, Elsevier, 2017.
[14] Z. Yang, H. Arakawa, A beaker method for determination of microplastic concentration by micro-Raman spectroscopy, MethodsX 11 (2023) 102251.
[15] Z. Yang, Ç. Murat, H. Nakano, H. Arakawa, Accessing the intrinsic factors of carbonyl index of microplastics: Physical and spectral properties, baseline correction, calculation methods, and their interdependence, Marine Pollution Bulletin 197 (2023) 115700.
[16] R. Lenz, K. Enders, C.A. Stedmon, D.M. Mackenzie, T.G. Nielsen, A critical assessment of visual identification of marine microplastic using Raman spectroscopy for analysis improvement, Marine Pollution Bulletin 100 (1) (2015) 82–91.
[17] G. Lopez-Reyes, F. Rull Pérez, A method for the automated Raman spectra acquisition, Journal of Raman Spectroscopy 48 (11) (2017) 1654–1664.
[18] Y. Luo, R. Naidu, C. Fang, Toy building bricks as a potential source of microplastics and nanoplastics, Journal of Hazardous Materials (2024) 134424.
[19] K. Banas, A.M. Banas, S.P. Heussler, M.B. Breese, Influence of spectral resolution, spectral range and signal-to-noise ratio of Fourier transform infra-red spectra on identification of high explosive substances, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 188 (2018) 106–112.
[20] S.J. Barton, T.E. Ward, B.M. Hennelly, Algorithm for optimal denoising of Raman spectra, Analytical Methods 10 (30) (2018) 3759–3769.
[21] X. Bian, Z. Shi, Y. Shao, Y. Chu, X. Tan, Variational mode decomposition for Raman spectral denoising, Molecules 28 (17) (2023) 6406.
[22] S. Chen, X. Lin, C. Yuen, S. Padmanabhan, R.W. Beuerman, Q. Liu, Recovery of Raman spectra with low signal-to-noise ratio using Wiener estimation, Optics Express 22 (10) (2014) 12102–12114.
[23] S.B. Dutta, H. Krishna, K.M. Khan, S. Gupta, S.K. Majumder, Fluorescence photobleaching of urine for improved signal to noise ratio of the Raman signal–An exploratory study, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 247 (2021) 119144.
[24] I.J. Jahn, A. Grjasnow, H. John, K. Weber, J. Popp, W. Hauswald, Noise sources and requirements for confocal Raman spectrometers in biosensor applications, Sensors 21 (15) (2021) 5067.
[25] H. Wang, X. Ma, Y. Wang, D. Chen, W. Chen, Q. Li, Enhancing the signal-to-noise ratio of FTIR spectrometers by a digital J-Stop, Optics Express 25 (16) (2017) 19077–19082.
[26] Z. Yang, M. Çelik, H. Arakawa, Challenges of Raman spectra to estimate carbonyl index of microplastics: A case study with environmental samples from sea surface, Marine Pollution Bulletin 194 (2023) 115362.
[27] X.Y. Zhao, G.Y. Liu, Y.T. Sui, M. Xu, L. Tong, Denoising method for Raman spectra with low signal-to-noise ratio based on feature extraction, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 250 (2021) 119374.
[28] S. Zhu, X. Cui, W. Xu, S. Chen, W. Qian, Weighted spectral reconstruction method for discrimination of bacterial species with low signal-to-noise ratio Raman measurements, RSC Advances 9 (17) (2019) 9500–9508.
[29] Z. Yang, H. Arakawa, A double sliding-window method for baseline correction and noise estimation for Raman spectra of microplastics, Marine Pollution Bulletin 190 (2023) 114887.
[30] H. Liu, Y. Zhang, Deep learning based crack damage detection technique for thin plate structures using guided lamb wave signals, Smart Materials and Structures 29 (1) (2019) 015032.
[31] Y.S. Chen, Y.C. Hsu, Effective and efficient baseline correction algorithm for Raman spectra, Lect Notes Eng Comput Sci 2239 (2019) 295–298.
[32] J. Smulko, Methods of trend removal in electrochemical noise data–Overview, Measurement 131 (2019) 569–581.
[33] A.C. Crecelius, T. Alexandrov, U.S. Schubert, Application of matrix-assisted laser desorption/ionization mass spectrometric imaging to monitor surface changes of UV-irradiated poly (styrene) films, Rapid Communications in Mass Spectrometry 25 (19) (2011) 2809–2814.
[34] T.A. Zimmerman, S.S. Rubakhin, E.V. Romanova, K.R. Tucker, J.V. Sweedler, MALDI mass spectrometric imaging using the stretched sample method to reveal neuropeptide distributions in aplysia nervous tissue, Analytical Chemistry 81 (22) (2009) 9402–9409.
[35] J.F. Povey, C.J. O'Malley, T. Root, E.B. Martin, G.A. Montague, M. Feary, C. Trim, D.A. Lang, R. Alldread, A.J. Racher, C.M. Smales, 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, Journal of Biotechnology 184 (2014) 84–93.
[36] M. Fan, Q. Yu, X. Wang, Z. Zheng, S. Xu, Z. Chen, L. Li, Identification of Surface-Enhanced Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry as Predictors of Prognosis in Triple Negative Breast Cancer, Journal of Nanoscience and Nanotechnology 16 (12) (2016) 12483–12488.
[37] Y. Guo, W. Jin, Z. Guo, Y. He, Iterative differential autoregressive spectrum estimation for Raman spectrum denoising, Journal of Raman Spectroscopy 53 (1) (2022) 148–165.
[38] Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, *454* (1971), pp.903-995.
[39] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Transactions on Signal Processing 62 (3) (2013) 531–544.
[40] Y. Kopsinis, S. McLaughlin, Development of EMD-based denoising methods inspired by wavelet thresholding, IEEE Transactions on Signal Processing 57 (4) (2009) 1351–1362.
[41] Q. Li, G.P. Zhang, Y. Liu, A study of Raman spectra denoising based on empirical mode decomposition, Spectroscopy and Spectral Analysis 29 (1) (2009) 142–145.
[42] F. León-Bejarano, M.O. Méndez, M.G. Ramírez-Elías, A. Alba, Improved Vancouver Raman Algorithm Based on Empirical Mode Decomposition for Denoising Biological Samples, Applied Spectroscopy 73 (12) (2019) 1436–1450.
[43] X. Zhang, Y. Bai, Y. Ma, P. He, Y. Tang, X. Lv, Denoising of Raman Spectra Using a Neural Network Based on Variational Mode Decomposition, Empirical Wavelet Transform, and Encoder-Bidirectional Long Short-Term Memory, Applied Sciences 13 (21) (2023) 12046.
[44] J. Fu, F. Cai, Y. Guo, H. Liu, W. Niu, An improved VMD-based denoising method for time domain load signal combining wavelet with singular spectrum analysis, Mathematical Problems in Engineering 2020 (2020) 1–14.
[45] C. Li, Y. Wu, H. Lin, J. Li, F. Zhang, Y. Yang, ECG denoising method based on an improved VMD algorithm, IEEE Sensors Journal 22 (23) (2022) 22725–22733.
[46] D.M. Pozar, Microwave engineering, John wiley & sons, 2011.
[47] Y. Xu, P. Du, R. Senger, J. Robertson, J.L. Pirkle, ISREA: an efficient peak-preserving baseline correction algorithm for Raman spectra, Applied Spectroscopy 75 (1) (2021) 34–45.
[48] G. Renner, A. Nellessen, A. Schwiers, M. Wenzel, T.C. Schmidt, J. Schram, Data preprocessing & evaluation used in the microplastics identification process: a critical review & practical guide, TrAC Trends in Analytical Chemistry 111 (2019) 229–238.
[49] M. Celik, H. Nakano, K. Uchida, A. Isobe, H. Arakawa, Comparative evaluation of the carbonyl index of microplastics around the Japan coast, Marine Pollution Bulletin 190 (2023) 114818.
[50] Yang, Z., Zhang, J., Haruka, N., Murat, C. and Arakawa, H., 2024. Spectral analysis of environmental microplastic polyethylene (PE) using average spectra. *Science of The Total Environment*, p.171871.
[51] V. Nava, M.L. Frezzotti, B. Leoni, Raman spectroscopy for the analysis of microplastics in aquatic systems, Applied Spectroscopy 75 (11) (2021) 1341–1357.
[52] P.R. Griffiths, Fourier transform infrared spectrometry, Science 222 (4621) (1983) 297–302.
[53] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, CRC Press, 2013.