A wavelet-based Gaussian method for energy dispersive X-ray fluorescence spectrum

Pan Liu, Xiaoyan Deng, Xin Tang, Shijian Shen

Abstract

This paper presents a wavelet-based Gaussian method (WGM) for the peak intensity estimation of energy dispersive X-ray fluorescence (EDXRF). The relationship between the parameters of Gaussian curve and the wavelet coefficients of Gaussian peak point is firstly established based on the Mexican hat wavelet. It is found that the Gaussian parameters can be accurately calculated by any two wavelet coefficients at the peak point which has to be known. This fact leads to a local Gaussian estimation method for spectral peaks, which estimates the Gaussian parameters based on the detail wavelet coefficients of Gaussian peak point. The proposed method is tested via simulated and measured spectra from an energy X-ray spectrometer, and compared with some existing methods. The results prove that the proposed method can directly estimate the peak intensity of EDXRF free from the background information, and also effectively distinguish overlap peaks in EDXRF spectrum.

Keywords: Statistical physics, Applied mathematics, Analytical chemistry

1. Introduction

X-ray fluorescence analysis is widely used for elemental and chemical analysis which provides both qualitative and quantitative compositional information of various samples, particularly in the study of metal, materials, ceramics,
archaeology, geology and many other areas (Van Grieken and Markowicz, 2002). However, the EDXRF spectrum is inevitably interfered by background because of the complicated condition of X-ray spectrum generation. Main background sources, including the Compton scattering, the Rayleigh scattering, the interaction between the characteristic X-rays and instrument noise etc, considerably complicate spectra processing. The focus of EDXRF analysis is to develop simple, efficient and economic methods, which can reproduce the target X-ray spectrum free from the background.

The current idea for EDXRF analysis is to estimate the target X-ray spectrum by subtracting background from the original EDXRF spectrum. Since the background is usually viewed as a continuous and smooth curve, it can be estimated by curve fitting methods after using peak striping approach to remove the rapidly varying structure in a spectrum. A number of approaches for background subtraction have been reported, such as peak clipping algorithms (Ryan et al., 1988; Morhác and Matousek, 2008), interpolation or least squares fitting algorithms (Yi et al., 2015; Vekemans et al., 1995; Brunetti and Steger, 2000; Garratt-reed and Bell, 2013), physical model approximate methods (Zhang et al., 2012; Tan and Brown, 2002), digital filters and wavelet methods (Tan and Brown, 2002; Shao et al., 2003; Galloway et al., 2009; Zhao and Wang, 2014; Daubechies, 1992; Stoffer and Holschneider, 1997). Especially wavelet method, it shows great prospect in spectral analysis, a comprehensive overview of applications of wavelet transform and wavelet-packet transform in spectral analysis were provided by Hoang (2014). The grate advantages of wavelet analysis are noise removal and resolution enhancement. Zhao et al. proposed a background subtraction approach based on complex wavelet transform, where the background is obtained at one specific low-frequency scale (Zhao and Wang, 2014). Ryan et al. proposed a statistics-sensitive non-linear iterative peak-clipping algorithm (SNIP) by a multi-pass peak clipping loop, which is considered as one of the most efficient approach to eliminate the background (Ryan et al., 1988). Despite the promising performance, most of those methods are relied on the iterative calculation or complicated processing. In particular, these wavelet-based methods require selecting suitable decomposition scales manually. Therefore, the previous methods are not suitable for automatic batched analyzing spectrum.

In this study, a local algorithm is provided for estimating the target spectral peaks from the original EDXRF spectrum by using the Mexican hat wavelet and Gaussian function. Particularly, the proposed algorithm is constructed by using two wavelet coefficients of Gaussian peak point and looks very simple, and also reveals a robust effectiveness on controlling the background and distinguishing overlap peaks.
2. Methods

The wavelet transform (WT) (Daubechies, 1992) provides a way of analyzing the local behavior of a function. One of the main aspects of the WT which is also the great advantage for our purpose is the ability to reveal the scaling behavior. The continuous wavelet transform (CWT) of \( f(t) \) is defined as follows

\[
W_f(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}^*(t) dt
\]  

(1)

where \( \psi_{a,b}(t) \) is the conjugate of \( \psi_{a,b}(t) \), and \( \psi_{a,b}(t) \) is the family of functions defined as

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t - b}{a}\right)
\]  

(2)

In the Eq. (2), \( \psi(t) \) is called the wavelet function, the scaling \( a(>0) \) adapts the width of the wavelet kernel to the microscopic resolution required, and the parameter \( b \) determines the location of analyze wavelet. The WT is sometimes referred as the "mathematical microscope" due to ability to capture the time-frequency local behavior of the function \( f(t) \).

Our objective is to estimate spectral peak by using the Gaussian curve, which referred as a function of the following form:

\[
f(t) = A \cdot \exp\left(-\frac{(t - t_0)^2}{2\sigma^2}\right)
\]  

(3)

This study proposes a wavelet-based algorithm to estimate the height parameter \( A \) and the width parameter \( \sigma \). Our algorithm is obtained via the Mexican hat wavelet which is defined as

\[
\psi(t) = \frac{2}{\sqrt{3}} \pi^{-\frac{1}{4}}(1 - r^2)e^{-\frac{r^2}{2}}
\]  

(4)

The \( \psi(t) \) is the second derivative of Gaussian function with support on interval \([-5, 5]\), and satisfied as

\[
\int_{-\infty}^{\infty} \psi(t) dt = 0
\]  

(5)

Submit Eqs. (3) and (4) to Eq. (1), and let \( b = t_0 \), by computing the integral Eq. (1), it is easily to obtain that

\[
W_f(a, t_0) = C A \frac{a^{5/2} \sigma}{(a^2 + \sigma^2)^{3/2}}
\]  

(6)

where \( C = \left(\frac{8}{3}\right)^{1/2} \pi^{1/4} \) and \( W_f(a,t_0) \) is the wavelet transform of Gaussian peak point \( t_0 \). Suppose that \( s \) is another scale factor satisfying \( s \neq a \), we have that
\[
\frac{W_f(a, t_0)}{W_f(s, t_0)} = \left(\frac{a_s}{s}\right)^{3/2}\left(\frac{s^2 + \sigma^2}{a^2 + \sigma^2}\right)^{3/2}
\]  

(7)

It follows from Eqs. (6) and (7) that

\[
\sigma = \left(\frac{s^2 - a^2}{\exp(R(a, s)) - 1} - a^2\right)^{1/2}
\]

where

\[
R(a, s) = \frac{2}{3}\left(\ln\frac{W_f(a, t_0)}{W_f(s, t_0)} - \frac{5}{2}\ln\frac{a}{s}\right)
\]  

(8)

It is found that the Gaussian parameters can be calculated by any two different wavelet coefficients at the peak point \(t_0\). Our objective is to estimate EDXRF peak which is always interfered by background and adjacent peaks. Here let \(f(t)\) be the target peak and \(t_0\) the peak point, and \(C(t)\) express the background and adjacent peaks. It is well known from the properties of wavelet transform that \(W_f + C(a, t_0) = W_f(a, t_0) + WC(a, t_0)\). If approximates to a constant in the neighborhood of peak point \(t_0\), then \(WC(a, t_0) \approx 0\) and \(W_f + C(a, t_0) \approx WC(a, t_0)\), i.e. the effect of background and adjacent peaks on wavelet coefficient is better inhibited.

To guarantee the local information used as little as possible, the small scales \(a\) and \(s\) are preferred. Let \(a = 1; s = 2\), we obtain the following formulations

\[
\sigma = \left(\frac{3}{\exp(R(1, 2)) - 1} - 1\right)^{1/2}
\]  

(9)

And

\[
A = \frac{W_f(1, t_0)(1 + \sigma^2)^{3/2}}{C\sigma}
\]  

(10)

Where

\[
R(1, 2) = \frac{2}{3}\left(\ln\frac{W_f(1, t_0)}{W_f(2, t_0)} + \frac{5}{2}\ln2\right)
\]  

(11)

The above formulations provide us with a local method to estimate the parameters of Gaussian curve. Assuming that there is a spectral series \(\{x_k, k = 1, 2, \ldots, n\}\), the wavelet-based Gaussian method (WGM) can be summarized as follows:

1. Smooth the original spectrum and determine the positions of spectral peaks. Here a bi-directional smoothness method is proposed:

\[
\dot{x}_i = \frac{1}{m} \sum_{k=i}^{i+m-1} x_k, \ddot{x}_i = \frac{1}{m} \sum_{k=i-m+1}^{i} \dot{x}_k
\]  

(12)

In this study we let \(m = 5\);
(2) Implement the continuous wavelet transform on the scales 1 and 2, and obtain the wavelet coefficients of the peak points;

(3) Estimate the width parameter $\sigma$ and height parameter $A$ at every peak point by the Eqs. (9)–(11);

(4) Reproduce the estimated peaks with the estimated Gaussian parameters, and calculate the peak intensities i.e. the net peak areas by

$$S = \sqrt{2\pi}A\sigma.$$  

Note that the presented algorithm avoids iterative computation, and also does not need to select wavelet decomposition scales manually. For EDXRF spectrum, $x_k$ corresponds to the number of counts in channel $k$, and $\sigma$ is given in channels. It is easily found that the key of the proposed algorithm lies on the estimate of the width $\sigma$, which determines the estimate accuracy of the height $A$. By the way, the estimated Gaussian parameters are expressed by $\hat{\sigma}$ and $\hat{A}$ throughout this paper.

3. Results and discussions

3.1. Gaussian parameter estimates to single peak

In order to verify the validity of the method presented in this paper, we firstly test the interference of the width $\sigma$ of the simulated spectrum on the estimated $\hat{\sigma}$ and $\hat{A}$. The relative error used in this study is defined as

$$E(D) = \frac{|\hat{D} - D|}{D}$$  \hspace{1cm} (13)

where $\hat{D}$ is the estimate of $D$.

The simulated spectrums are obtained from Gaussian curves with $A = 60$ and $\sigma$ ranging from 2 to 1000, the estimated parameters $\hat{\sigma}$ and $\hat{A}$ are calculated by the proposed algorithm in this paper. The results are shown in Fig. 1, and also part of numerical results is provided in Table 1. The relative errors of the width and height parameters are less than 5%, and less than 1% when the $\sigma$ ranges from 5 to 450. It can be noted that the height $A$ of the simulated spectrum has no effect on the estimated results. Generally, the relative error increases as the width increasing for the simulated spectrum width $\sigma > 74$.

3.2. Gaussian parameter estimates to overlap peaks

The next thing we do is to test the capacity of distinguishing overlap peaks. Seven Gaussian peaks are chosen as the target peaks, as shown in Fig. 2(a), and the simulated spectrum is constructed by superposing the seven Gaussian peaks which is shown in Fig. 2(b). The estimated peaks are obtained by the WGM method presented in this paper, as shown in Fig. 2(c). For simplicity, we would explain the situation by the estimate of the width. In Fig. 2, the isolated peak P1 is reproduced
most accurately with the relative error is less than 0.2%. The peak P6 is completely superimposed by the peak P7, however a relative error of 0.4% is obtained. The overlap peaks P2, P3, P4, P5 and P7 are effectively reproduced with the relative errors are less than 2%. The mutual influence of overlapping peaks is too complex to give a quantitative description, some numerical results are provided in Table 2. Generally, the estimate $\hat{\sigma}$ with a large width $\sigma$, gives a relatively large error when two overlap peaks have similar magnitude, P6 and P7 in Fig. 2 are in the case. It is also found that the estimate $\hat{\sigma}$ with small height $A$ generates a larger error when its widths are similar, as P4 and P5 shown in Fig. 2.

Since the Eq. (8) involves 21 adjacent channels for a specified peak, the distance between two adjacent peaks may create a significant effect on the estimates of

Table 1. The relative errors of the estimated parameters with $A = 60$.

| $\delta$ | 2   | 3   | 4   | 5   | 10  | 20  | 40  | 80  | 160 | 320 |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $\hat{\sigma}$ | 2.089 | 3.06 | 4.05 | 5.04 | 10.03 | 20.03 | 40.05 | 80.09 | 160.25 | 321.1 |
| $E(\sigma)(\%)$ | 4.4  | 1.97 | 1.15 | 0.76 | 0.27 | 0.14 | 0.113 | 0.116 | 0.156 | 0.343 |
| $\hat{A}$ | 57.64 | 58.93 | 59.39 | 59.61 | 59.9 | 59.98 | 59.99 | 60.02 | 60.07 | 60.3 |
| $E(A)(\%)$ | 4.4  | 1.783 | 1.017 | 0.65 | 0.167 | 0.003 | 0.003 | 0.003 | 0.117 | 0.5 |
overlap peaks. Consider the peak P2 as an example and let D represent the distance between P2 and P3, the relative errors of the estimated parameter $\sigma$ are 0.28%, 0.54%, 6.9%; 11.6% and 18.3% corresponding to the distance $D = 50, 40, 30, 20$ and 10, respectively. Generally, the proposed method can provide a better result when the distance between two adjacent peak points is greater than 40 channels.

![Fig. 2.](image)

**Fig. 2.** (a) The target peaks and its parameters shown in Table 2. (b) The simulated spectrum by superposing seven Gaussian peaks shown in Fig. 2(a). (c) The estimated peaks by the algorithm presented in this paper.

| P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|----|----|----|----|----|----|----|
| $\sigma$ | 50 | 9 | 10 | 25 | 25 | 20 | 90 |
| $A$ | 140 | 100 | 100 | 90 | 140 | 60 | 60 |
| $\hat{\sigma}$ | 50.0577 | 8.9517 | 9.9624 | 24.5999 | 24.8296 | 19.9227 | 88.6625 |
| $E(\sigma)(\%)$ | 0.1154 | 0.5367 | 0.376 | 1.6604 | 0.6815 | 0.39 | 1.4861 |

Table 2. The parameters of target peaks, the estimated $\hat{\sigma}$ and the relative errors.
3.3. Peak intensity estimates to simulated spectrum with background

To further test the proposed method, another two simulated spectrums are created by superposing background and the above seven Gaussian peaks. Here Gaussian background and polynomial background are chosen for testing, which are defined as follows

\[ b_1(i) = 100 \exp\left(-\frac{(i - 1350)^2}{2 \times 400^2}\right) \] (14)

and

\[ b_2(i) = \begin{cases} 0.004(i + 50)(i - 1000)(i - 1150)/60000, & 0 < i \leq 1000 \\ -0.033(i + 1000) + 105.5, & 1000 < i \leq 2000 \end{cases} \] (15)

The obtained spectrums are shown in Figs. 3(a) and 4(a), respectively.

Two simulated spectrums are respectively processed by the wavelet-based Gaussian method (WGM), Gaussian mixture modeling (GMM) (Andrzej et al., 2015) and SNIPG (Ryan et al., 1988). The estimated peaks are shown in Figs. 3

![Fig. 3](image-url)

Fig. 3. (a) The simulated spectrum by superposing the seven Gaussian peaks shown in Fig. 2(a) and the Gaussian background defined by Eq. (14). (b) The original peaks (black) and the estimated peaks (blue) by WGM. (c) The original peaks (black) and the estimated peaks (blue) by GMM. (d) The original peaks (black) and the estimated peaks (blue) by SNIP-G.
and 4, and the estimates of peak intensities are given in Tables 3 and 4, where Fig. 3 and Table 3 are associated with Gaussian background, Fig. 4 and Table 4 are for polynomial background.

Table 3. Estimated intensities of peaks for WGM, GMM and SNIP-G with Gaussian background, and their relative errors expressed by E(W), E(G) and E(S) respectively.

|     | P1  | P2  | P3  | P4  | P5  | P6  | P7  |
|-----|-----|-----|-----|-----|-----|-----|-----|
| True| 17546 | 2256 | 2507 | 5640 | 8773 | 3008 | 13536 |
| WGM | 17567 | 2194 | 2458 | 5319 | 8542 | 2929 | 12913 |
| E(W)(%) | 0.1 | 2.7 | 2 | 5.7 | 2.6 | 2.6 | 4.6 |
| GMM | 17647 | 2383 | 2625 | 6053 | 8452 | 3394 | 11576 |
| E(G) | 0.5 | 5.6 | 4.7 | 7.3 | 3.7 | 12.8 | 14.5 |
| SNIP-G | 17089 | 2197 | 2458 | 5467 | 8344 | 2351 | 11341 |
| E(S)(%) | 2.6 | 2.7 | 2 | 3.1 | 4.9 | 21 | 16 |
Note that the GMM method is implemented after the baseline is corrected using the Matlab function ‘msbackadj’ with routine and default parameters, and the SNIP-G method consists of the background subtraction by SNIP and the peak intensity estimation by Gaussian fitting. Next we provide some analysis on the experimental results. Firstly, It is found from Figs. 3(b) and 4(b) that the WGM method can effectively eliminate background interference. Compared with the non-background case, a simple computation shows that only the relative errors $E(\hat{\sigma})$ of three peaks are larger, the relative error varies from 0.1154% to 0.3994% at P1 and from 0.5367% to 0.54% at P2 with polynomial background, and from 1.4861% to 1.5027% at P7 with Gaussian background. Compared with the GMM and SNIP-G methods, WGM yields better accuracy than GMM and SNIP-G, apart from at the peak point P5 with SNIP-G (see Tables 3 and 4).

Secondly, it can be seen from Tables 1 and 2 that WGM achieves the same estimates of peak intensities for both Gaussian background and polynomial background, and GMM and SNIP-G are sensitive to the backgrounds which reveal a big deviation between Gaussian background and polynomial background.

Thirdly, it is also found that WGM is obviously superior to GMM and SNIP-G for overlapping peak identification, especially for the peak with a big width parameter $\sigma$. For example, WGM, GMM and SBIP-G respectively yield the relative errors $E(\hat{\sigma})$ of three peaks are larger, the relative error varies from 0.1154% to 0.3994% at P1 and from 0.5367% to 0.54% at P2 with polynomial background, and from 1.4861% to 1.5027% at P7 with Gaussian background. Compared with the GMM and SNIP-G methods, WGM yields better accuracy than GMM and SNIP-G, apart from at the peak point P5 with SNIP-G (see Tables 3 and 4).

Note that the GMM method is implemented after the baseline is corrected using the Matlab function ‘msbackadj’ with routine and default parameters, and the SNIP-G method consists of the background subtraction by SNIP and the peak intensity estimation by Gaussian fitting. Next we provide some analysis on the experimental results. Firstly, It is found from Figs. 3(b) and 4(b) that the WGM method can effectively eliminate background interference. Compared with the non-background case, a simple computation shows that only the relative errors $E(\hat{\sigma})$ of three peaks are larger, the relative error varies from 0.1154% to 0.3994% at P1 and from 0.5367% to 0.54% at P2 with polynomial background, and from 1.4861% to 1.5027% at P7 with Gaussian background. Compared with the GMM and SNIP-G methods, WGM yields better accuracy than GMM and SNIP-G, apart from at the peak point P5 with SNIP-G (see Tables 3 and 4).

Secondly, it can be seen from Tables 1 and 2 that WGM achieves the same estimates of peak intensities for both Gaussian background and polynomial background, and GMM and SNIP-G are sensitive to the backgrounds which reveal a big deviation between Gaussian background and polynomial background.

Thirdly, it is also found that WGM is obviously superior to GMM and SNIP-G for overlapping peak identification, especially for the peak with a big width parameter $\sigma$. For example, WGM, GMM and SBIP-G respectively yield the relative errors $E(\hat{\sigma})$ of three peaks are larger, the relative error varies from 0.1154% to 0.3994% at P1 and from 0.5367% to 0.54% at P2 with polynomial background, and from 1.4861% to 1.5027% at P7 with Gaussian background. Compared with the GMM and SNIP-G methods, WGM yields better accuracy than GMM and SNIP-G, apart from at the peak point P5 with SNIP-G (see Tables 3 and 4).

Note that the GMM method is implemented after the baseline is corrected using the Matlab function ‘msbackadj’ with routine and default parameters, and the SNIP-G method consists of the background subtraction by SNIP and the peak intensity estimation by Gaussian fitting. Next we provide some analysis on the experimental results. Firstly, It is found from Figs. 3(b) and 4(b) that the WGM method can effectively eliminate background interference. Compared with the non-background case, a simple computation shows that only the relative errors $E(\hat{\sigma})$ of three peaks are larger, the relative error varies from 0.1154% to 0.3994% at P1 and from 0.5367% to 0.54% at P2 with polynomial background, and from 1.4861% to 1.5027% at P7 with Gaussian background. Compared with the GMM and SNIP-G methods, WGM yields better accuracy than GMM and SNIP-G, apart from at the peak point P5 with SNIP-G (see Tables 3 and 4).

Secondly, it can be seen from Tables 1 and 2 that WGM achieves the same estimates of peak intensities for both Gaussian background and polynomial background, and GMM and SNIP-G are sensitive to the backgrounds which reveal a big deviation between Gaussian background and polynomial background.

Thirdly, it is also found that WGM is obviously superior to GMM and SNIP-G for overlapping peak identification, especially for the peak with a big width parameter $\sigma$. For example, WGM, GMM and SBIP-G respectively yield the relative errors

### Table 4. Estimated intensities of peaks for WGM, GMM and SNIP-G with polynomial background, and their relative errors expressed by $E(W)$, $E(G)$ and $E(S)$ respectively.

|     | P1    | P2    | P3    | P4    | P5    | P6    | P7    |
|-----|-------|-------|-------|-------|-------|-------|-------|
| True| 17546 | 2256  | 2507  | 5640  | 8773  | 3008  | 13536 |
| WGM | 17567 | 2194  | 2458  | 5319  | 8542  | 2929  | 12913 |
| $E(W)($%)| 0.1 | 2.7 | 2 | 5.7 | 2.6 | 2.6 | 4.6 |
| GMM | 16906 | 2583  | 2725  | 6153  | 8152  | 3694  | 12076 |
| $E(G)$| 3.6 | 14.5 | 8.7 | 9.1 | 7.1 | 22.8 | 10.8 |
| SNIP-G | 16748 | 2194  | 2458  | 5256  | 8613  | 2861  | 11902 |
| $E(S)($%)| 4.5 | 2.7 | 2 | 6.8 | 1.8 | 4.9 | 12.1 |

### Table 5. Lead content of samples measured by ICP-AES.

| Sample Number | No.1   | No.2   | No.3   | No.4   | No.5   | No.6   | No.7   |
|---------------|--------|--------|--------|--------|--------|--------|--------|
| Content(mg/kg)| 9.9923 | 30.7566| 48.8089| 61.4237| 73.6556| 78.6347| 89.2233|
**Fig. 5.** Relationship between channel \((x)\) and energy \((y)\): \[ y = 0.021x + 0.001 \] with determination coefficient \(R^2 = 0.9998\).

**Fig. 6.** (a) The original EDXRF spectrum for lead content detection. (b) The bi-directional smoothness spectrum with \(m = 5\).
Fig. 7. The estimated peaks by WGM method presented in this paper.

Fig. 8. The estimated peaks by GMM method.
4.6%; 14.5% and 16% at the peak point P7 with Gaussian background (see Table 3). However, the accuracies of GMM and SNIP-G is not acceptable.

3.4. Peak intensity estimates to experimental spectrum

To further verify the effectiveness of the proposed method, the WGM method is tested on actual spectrums from an EDXRF detector with lead content detection, and also compared with the GMM and SNIP-G methods. The tested sample consists of seven gypsum slices with different lead content, which are specially prepared for this experiment by mixing the gypsum in molten state with lead acetate. The weight of each gypsum slice ranges from 0.397 to 0.404 g and thickness from 0.128 to 0.13 cm, the lead content is pre-measured by inductively coupled plasma atomic emission spectrometry (ICP-AES), shown in Table 5.

Table 6. Estimated net peak areas with WGM, GMM and SNIP-G methods, respectively.

| Sample  | No. 1  | No. 2  | No. 3  | No. 4     | No. 5   | No. 6   | No. 7   |
|---------|--------|--------|--------|-----------|---------|---------|---------|
| WGM     | 138.92 | 332.08 | 511.45 | 641.8181  | 783.03  | 838.56  | 945.41  |
| GMM     | 171.38 | 344.72 | 526.2  | 618.7621  | 725.65  | 775.64  | 892.32  |
| SNIP-G  | 155.36 | 294.15 | 498.87 | 628.421   | 739.65  | 848.29  | 966.42  |
The EDXRF system consists of a mini-X ray tube (10-50 kv, 5-200 A), a Si-PIN detector (145–230 eV FWHM at 5.9 KeV) and a computer for data acquisition and evaluation. The mini-X ray tube includes the X-ray fluorescence probe and the preamplifier, the probe is equipped with collimator and filter. The collimator will avoid noise, and improve spectral quality. The filter can effectively reduce the interference of the background and other characteristic spectrum. The measured X-ray spectrums are obtained with 30 kV voltage, 100 μ A current and 120 s test time. The critical excitation energy of is 10.55 KeV. The relationship between channel and energy is shown in Fig. 5.

The linear regression models between lead content and peak intensity corresponding to the WGM, GMM and SNIP-G methods, and the determination coefficients $R^2 = 0.9972$, 0.9824 and 0.9824, respectively.

![Fig. 10.](image)

**Fig. 10.** The linear regression models between lead content and peak intensity corresponding to the WGM, GMM and SNIP-G methods, and the determination coefficients $R^2 = 0.9972$, 0.9824 and 0.9824, respectively.

Table 7. Predictive lead contents for GWM, GMM and SNIP-G methods, and their relative errors expressed by E(W), E(G) and E(S) respectively.

| Sample | No. 2     | No. 4     | No. 6     |
|--------|-----------|-----------|-----------|
| Actual content | 30.7566   | 61.4237   | 78.6347   |
| WGM(E(W)) | 31.3749(2.01%) | 60.5523(1.42%) | 78.7086(0.57%) |
| GMM(E(G)) | 34.0505(10.71%) | 61.0435(0.62%) | 76.4961(2.72%) |
| SNIP-G(E(S)) | 28.0381(8.84%) | 59.7937(2.65%) | 80.6841(2.61%) |
The original spectrums are firstly to implement a bi-directional smoothness with \( m = 5 \) as shown in Fig. 6, and then processed by WGM, GMM and SNIP-G, respectively. The estimated peaks are shown in Figs. 7, 8 and 9, where Fig. 7 is for WGM method, Fig. 8 for GMM method and Fig. 9 for SNIP-G method.

The estimates of net peak areas are given in Table 6. Note that the net peak area of each sample is computed by the average of three measurements.

The linear regression models between lead content and net peak area corresponding to WGM, GMM and SNIP-G methods are respectively established by the use of sample No. 1, No. 3, No. 5 and No. 7, and the related results can be found in Fig. 10, in which the net peak area is the abscissas and lead content is the ordinate. These models are used to predict the lead contents of the sample No. 2, No. 4 and No. 6, and the results are shown in Table 7.

Despite all of the WGM, GMM and SNIP-G methods achieve high accurate prediction, it can be found from Table 7 that the WGM is more stable than GMM and SNIP-G methods. Both of GMM and SNIP-G methods are not suitable for automatic batched analyzing spectrum since they require iterative calculation, and Fig. 8 corresponding to GMM method includes some interference peaks, which require a manually-operated method to estimate the target peak.

4. Conclusions

In this paper we propose a wavelet-based Gaussian method (WGM) for the peak intensity estimate of EDXRF spectrum. The proposed method has determined formula and is suitable for automatic batched analyzing spectrum. Experiments on simulated and measured spectrums demonstrate that WGM can eliminate background and distinguish overlap peaks effectively. Since WGM involves 21 adjacent channels for a specified peak, it may lead to a large estimate error when the distance between two adjacent peak points is smaller than 20 channels. In many cases, however, this disadvantage can be overcome by the interpolation method to increase the number of channels. Meanwhile, smooth pre-treatments are usually necessary for reducing pseudo peaks and noise since WGM is relied on the high-frequency wavelet coefficients.

Declarations

Author contribution statement

Pan Liu: Performed the experiments; Analyzed and interpreted the data.

Xiaoyan Deng: Conceived and designed the experiments; Wrote the paper.

Xin Tang: Conceived and designed the experiments.
Shijian Shen: Contributed reagents, materials, analysis tools or data.

**Funding statement**

This work was supported by Fundamental Research Funds for the Central Universities of China under grant No. 2662015PY046 and National Natural Science Foundation of China (NSFC) under grant No. 11671161.

**Competing interest statement**

The authors declare no conflict of interest.

**Additional information**

No additional information is available for this paper.

**References**

Andrzej, P., Michal, M., Monika, P., Piotr, W., Joanna, P., 2015. Signal parting algorithm for highly efficient Gaussian mixture modeling in mass spectrometry. Plos One 10, 7.

Brunetti, A., Steger, T., 2000. X-Ray spectra background fitting by projection onto convex sets. Nucl. Instrum. Methods Phys. Res. Sect. 441 (3), 504–509.

Daubechies, I., 1992. Ten lectures on wavelets. CBMSNSF regional conference series in applied mathematics. SIAM.

Galloway, C.M., Ru, E.L., Etchegoin, P.G., 2009. An iterative algorithm for background removal in spectroscopy by wavelet transforms. Appl. Spectrosc. 63 (12), 1370–1376.

Garratt-reed, A.J., Bell, D.C., 2013. Energy Dispersive X-ray Analysis in the Electron Microscope. Garland Science.

Hoang, V.D., 2014. Wavelet-based spectral analysis. Trends Anal. Chem. TrAC 62, 144–153.

Morhác, M., Matousek, V., 2008. Peak clipping algorithms for background estimation in spectroscopic data. Appl. Spectrosc. 62 (1), 91–107.

Ryan, C.G., Clayton, E., Griffin, W.L., Sie, S.H., Cousens, D.R., 1988. A statistics sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. Nucl. Instr. Meth. Phys. Res. B 34 (3), 396–402.

Tan, H.W., Brown, S.D., 2002. Wavelet analysis applied to removing non-constant, varying spectroscopic background in multivariate calibration. J. Chemometr. 16 (5), 228–240.
Shao, X., Leung, K.M., Chau, F.T., 2003. Wavelet: a new trend in chemistry. Acc. Chem. Res. 36 (4), 276–283.

Stoffer, D.S., Holschneider, M., 1997. Wavelets: an analysis tool. Phys. Today 92 (439), 1215.

Van Grieken, R.E., Markowicz, A.A., 2002. Handbook of X-Ray Spectrometry. 29, Marcel Dekker, Inc.

Vekemans, B., Janssens, K., Vincze, L., Adams, F., Espen, P.V., 1995. Comparison of several background compensation methods useful for evaluation of energy-dispersive X-ray fluorescence spectra. Spectrochim. Acta B 50 (2), 149–169.

Yi, L., Liu, Z., Wang, K., Chen, M., Peng, S., Zhao, W., Zhao, J., 2015. A new background subtraction method for energy dispersive x-ray fluorescence spectra using a cubic spline interpolation. Nucl. Instrum. Methods Phys. Res. 775 (12), 12–14.

Zhang, Q., Ge, L., Gu, Y., Lin, Y., Zeng, G., Yang, J., 2012. Background estimation based on fourier transform in the energy-dispersive x-ray fluorescence analysis. X-Ray Spectrom. 41 (2), 75–79.

Zhao, F., Wang, A., 2014. A background subtraction approach based on complex wavelet transforms in EDXRF. X-Ray Spectrom. 44 (2), 41–47.