Paper

Evaluation of Three Peak Detection Algorithms for XPS Spectra

Y. Furukawa,[1,*] M. Otsuki,[2] N. Ikeo,[2] Y. Nagatsuka,[3] M. Yoshitake,[4] and A. Tanaka[5]
1 Denka, 2-1-1 Nihonbashi-Muromachi, Chuo-ku, Tokyo 108-8338
2 JEOL Ltd., 3-1-2 Musashino, Akishima-shi, Tokyo 196-8558
3 Harmony Pj., 2-17-1 Tanaka-cho, Akishima-shi, Tokyo 196-0014
4 National Institute for Materials Science, 3-13 Sakura, Tsukuba, Ibaraki 305-0003
5 National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047
**furukawa4160@hop.ocn.ne.jp*
(Received: June 29, 2014: Accepted: September 21, 2016)

In order to evaluate three peak detection algorithms, the influences of parameter values used were examined using digitized synthetic XPS spectra with different levels of noise. The three peak detection algorithms are the Threshold Curve of the Second Derivative (2nd DER method), the Directly Calculating Peak and Background Relations at a Candidate Peak (PB method), and the Rough Estimation of Spectrum Background (BGD method). The peak detection results clearly showed that particular combination of parameter values produce the best performance for each algorithm. The validity of those parameter values was assessed by comparing the results from the peak detection algorithms with those of visual detection. On the whole, the BGD method was found to be the most practical. The importance of choosing effective parameter values and the advantages and disadvantages of these three peak detection methods were summarized.

1. Introduction

In surface chemical analysis, the first thing to be done is a qualitative analysis to identify the elements or atomic groups constituting unknown materials. This initial qualitative analysis is of decisive importance for the subsequent analysis. To avoid the burdensome analysis being dependent on analysts, utilization of automatic peak detection and identification software is recommended. In fact, the analysis is usually performed with software attached to the analysis instrument or with commercially-available analysis software [1-3]. However, such automatic software sometimes fails in reproducible peak detection. The process of peak detection should be clearly defined so that anyone can reproduce the same results. For appropriate peak detection and identification using automatic software, effective parameter values such as the number of the second derivative points should be chosen when the algorithm and parameters in the software are disclosed. This verification of the peak detection process results in high-quality analysis.

Accordingly we worked out peak detection software based on three algorithms proposed in 2008 [4] and showed that they were useful for surface analysis such as XPS and AES. These three peak detection algorithms are the Threshold Curve of the Second Derivative (2nd DER method), the Directly Calculating Peak and Background Relations at a Candidate Peak (PB method), and the Rough Estimation of Spectrum Background (BGD method). For the details of these algorithms, see the Appendix.

Studies evaluating the performance of these three algorithms for detecting peaks using digitized synthetic spectra of XPS have been reported [5]. However, previously-reported parameter values used in these algorithms were not necessarily adequate for the evaluation. In the present work, we have studied the numeric influences of values for various parameters using the digitized synthetic spectra of XPS so that these algorithms exhibit the best peak detection performance, and we evaluated these three peak detection algorithms by comparing the results with the

results of visual detection from a practical point of view. The characteristics of each peak detection method, the importance of choosing effective parameter values and the advantages and disadvantages of these three peak detection methods were summarized based on the findings of the peak detection results so that the present research report helps a surface scientist to understand how to select the best conditions to carry out peak detection effectively.

2. Algorithms Used for Detecting Peaks

The three algorithms for peak detection are described below in broad terms for convenience; details of the three algorithms proposed in 2008 [4] are shown in the Appendix.

2.1 Peak Detection Using the Threshold Curve of the Second Derivative (2nd DER method)

A spectrum was processed to obtain the second derivative curve and determine the standard deviation σi of the curve. A peak in the spectrum is judged to be detected when the local negative minimum d_{min} is smaller than $k\sigma_i$, i.e., when inequality (1) is satisfied:

$$d_{\min} < k\sigma_i \quad \text{.....(1)}$$

Here, the coefficient k is a parameter for giving a multiplying factor of σ_i,, and it is an important parameter because the criterion to judge the peak varies depending on the value of $k\sigma_i$. In this study, kσi is defined as the noise threshold curve to examine the performance for detecting peaks.

As the second derivative curve changes depending on the number of the second derivative points, the value of d_{min} also changes, leading to changes in the performance for detecting peaks. Therefore, the effective value of the number of the second derivative points was examined with its effect on the performance for detecting peaks.

2.2 Peak Detection by Directly Calculating Peak and Background Relations at a Candidate Peak (PB method)

This method directly determines the peak intensity and the background intensity corresponding to the candidate peak position to improve the accuracy of peak judgment. Specifically, the net intensity N of a peak is $N = P - B$, which is obtained by subtracting the background value B from the peak intensity P. When σ_N is the standard deviation of N, the peak is judged to be detected if inequality (2) holds:

$$N > k\sigma_N \quad \text{.....(2)}$$

The value of k in this case is a parameter for giving a multiplying factor of the standard deviation σ_N, and the performance for detecting peaks was examined by changing it. Similarly, the number of the second derivative points in the 2nd DER method is also a parameter affecting the performance for detecting peaks, and its effective value was studied.

2.3 Peak Detection Using Rough Estimation of Spectrum Background (BGD method)

The moving average value of each point of a spectrum is calculated to obtain the background intensity b_i. The standard deviation value σ_{ni} of a peak intensity n_i after subtraction of the averaged background is calculated to be $n_i = y_i - b_i$, and the peak is judged to be detected if the peak intensity y_i satisfies inequality (3):

$$y_i > b_i + k\sigma_{ni} \quad \text{.....(3)}$$

The value of k of this inequality is a parameter for giving a multiplying factor of the standard deviation σ_{ni} as described above, and the effective values of these two parameters, k and the moving average points, were examined since the number of moving average points also affects the value of σ_{ni}.

This method offers two advantages. One advantage is that minor peaks with a low peak intensity and a broad width, which are hard to detect by other algorithms, can be detected if the peak area, S_i , is 2 count·eV or more as we will discuss later in section 7.1. Another advantage is that if two or more adjacent peaks overlap, those peaks can be easily detected if the valley depth D of the local minimum of the secondary differential spectrum exceeds the noise fluctuation $k\sigma_{ni}$ of the point. Specifically, if the value of the valley of a peak is more than $k\sigma_{ni}$, it is a criterion to judge that the peak has been detected.

3. XPS Spectra Used in the Performance Evaluation

For evaluating the performance of peak detection algorithms, the spectra provided by VAMAS/TWA2/A9 (2006) [5] were used. Those spectra are synthesized from Au, Ag, and Cu spectra,

each of which was measured at 0.5 eV step using Al X-ray source. The synthesis was carried out by adding weighted Ag and Cu spectra to the Au spectrum as follows. The spectrum "b001" was synthesized as I_{Au} + I_{Ag} + I_{Cu},, where I denotes an intensity Au, Ag, and Cu spectra respectively. The spectra of "b002" and "b003" were also synthesized by the linear summation as I_{Au} + 0.1 I_{Ag} + 0.01 I_{Cu} and I_{Au} + 0.01 I_{Ag} + 0.001 I_{Cu}, respectively. (In what follows, they are called b001, b002 and b003, respectively.)

Figure 1 shows those spectra. In this figure, the abscissa axis represents the binding energy from 1200 eV to 0 eV, and the ordinate axis represents the intensities as count numbers. As can be seen from these spectra, the spectrum b001 synthesized by the equal ratios of Au, Ag and Cu shows the largest number of clear peaks; the number of peaks appears to decrease in b002 and to decrease further in b003. The component of Cu is barely visible in b002 and in b003.

Additional spectra were synthesized by superposing noise with different amplitudes to these spectra. They are b001_73 and b001_43 for b001, b002_73 and b002_33 for b002, and b003_73 and b003_23 for b003, respectively, in the order of noise amplitudes. In the

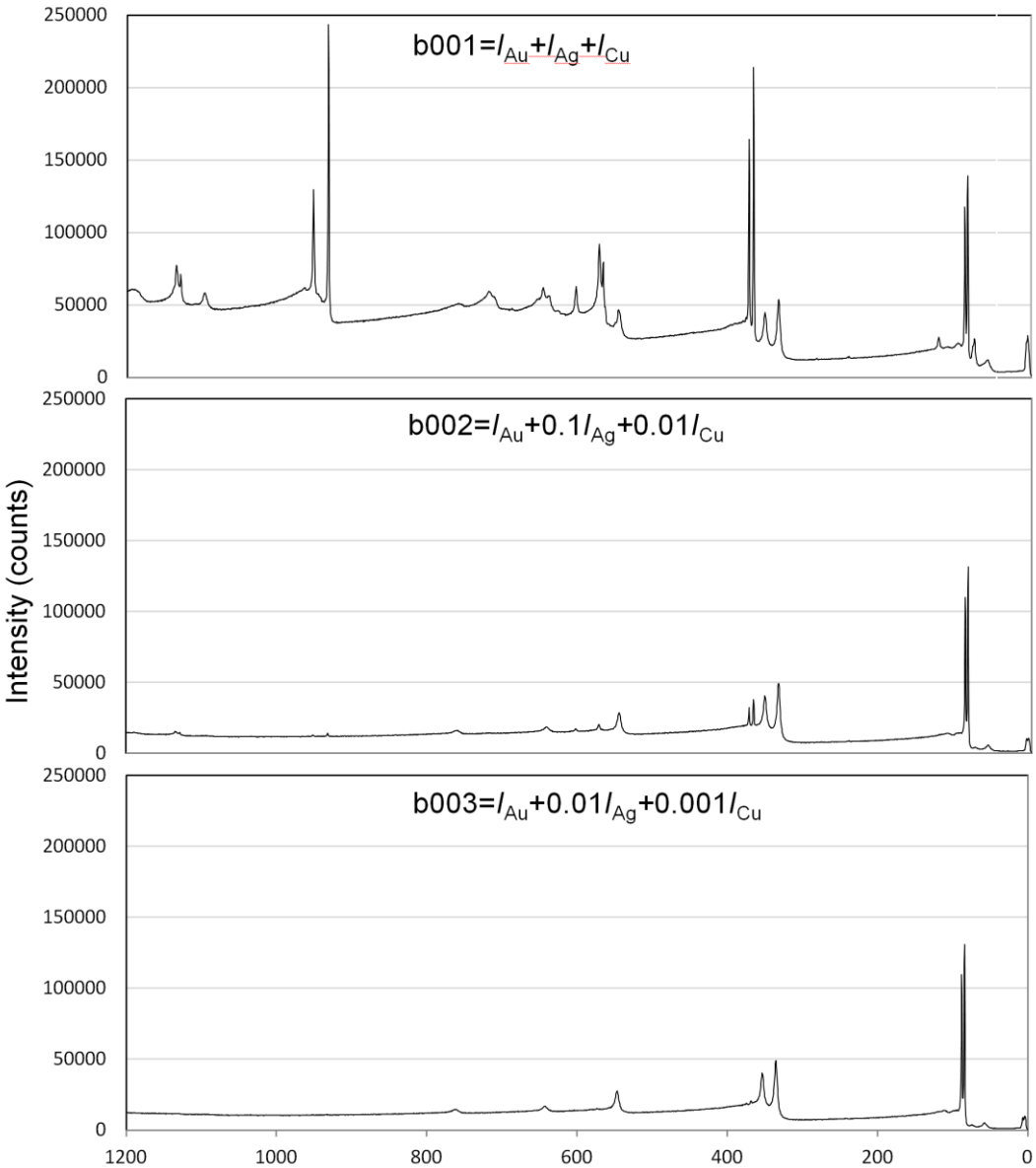

Figure 1. Synthesized XPS spectra: The spectrum "b001" was synthesized as I_{Au} + I_{Ag} + I_{Cu}, where I denotes an intensity of Au, Ag, and Cu spectra respectively. The spectra of "b002" and "b003" were also synthesized as I_{Au} + 0.1 I_{Ag} + 0.01 I_{Cu} and I_{Au} + 0.01 I_{Ag} + 0.001 I_{Cu}, respectively.

two-digit numbers appended to the spectrum names, the first digit represents the noise amplitudes, which becomes smaller when the reciprocal of the digit is smaller. The second digit represents the number of random numbers generated, as shown in Appendix B.

Figure 2 shows the examples of spectra to which noise is superposed. As is shown in the figure, as the noise increases, the spectrum intensities decrease by multiplying the scaling factor so that the spectra had signal intensity appropriate to the S/N. In the case of spectrum b003_73, the scaling factor that presents the entire spectrum intensity becomes 1/18. Similarly, as the spectrum b003_23 had a scaling factor of 1/215, the entire spectral intensity was reduced to 1/215.

The detailed information of how noise was superposed is described in Appendix B and also reported in the research paper [5]. The characteristics of the noise-superposed spectra are described on the website of the Surface Analysis Society of Japan [6].

4. Outline of Evaluation

4.1 Peak Detection by Visual Observation

At this point it seems appropriate to note that peak detection by experts with human eyes is generally highly reliable compared with current automatic peak detection software.

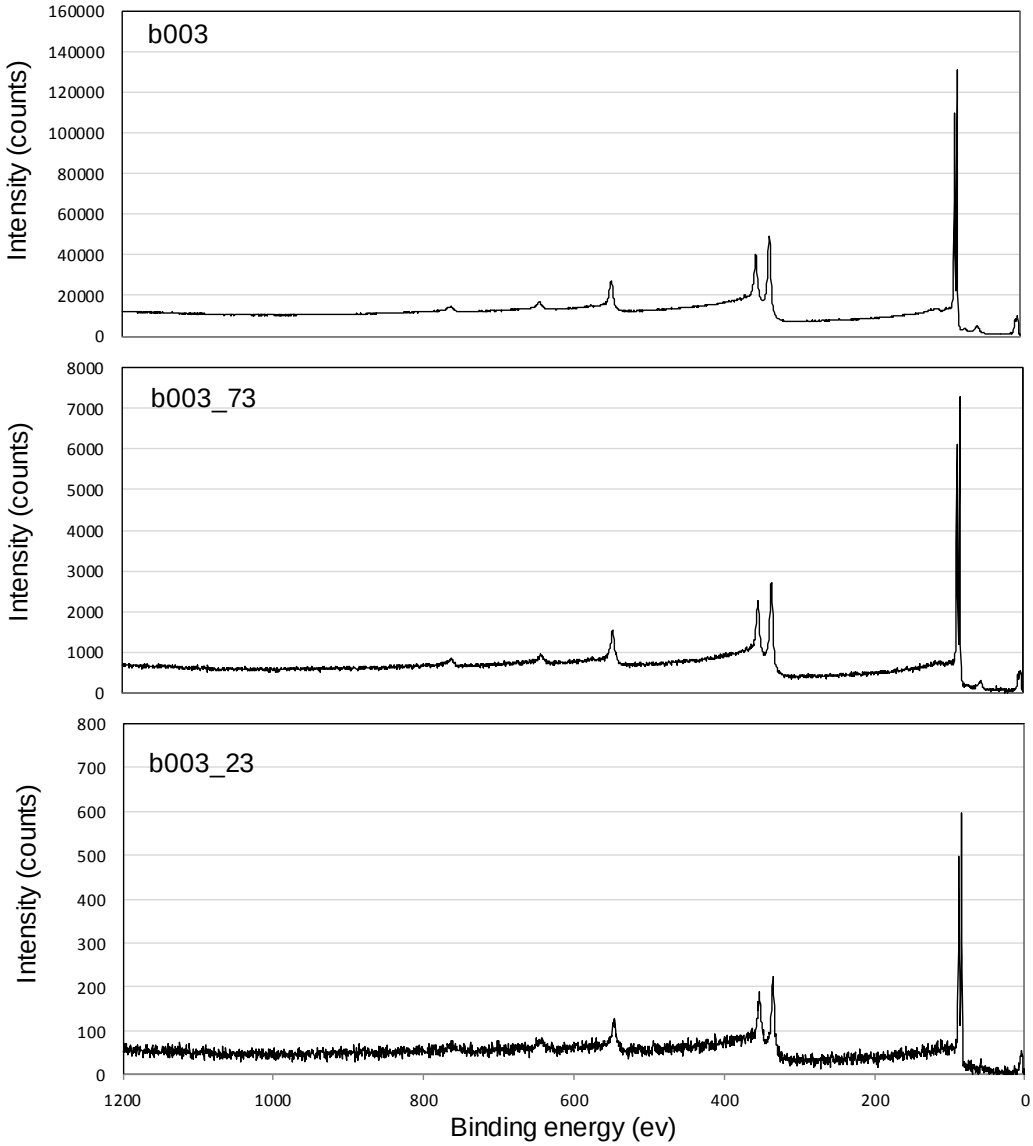

Figure 2. Examples of the original spectrum (b003) and spectra superposed with artificial noise (b003_73, b003_23): the entire spectral intensity of b003_73 is reduced to 1/18 and to 1/215 in the spectrum of b003_23 respectively according to the S/N.

The peak detection by human eyes was performed at the start of our performance analysis in order that the performance of a peak detection algorithm can be evaluated by the degree of coincidence of peak detection results by automatic algorithm and those by human eyes.

The nine spectra described above were distributed to seven judges, who are surface chemical scientists, to perform visual peak detection. Figure 3 shows an example of visual peak detection and judgment of peak intensity on b003_23 spectrum with a small number of clear peaks, and Table 1 shows an example of peak detection results reported by judges for this spectrum.

In the report of the results, the seven judges reported the characteristic features of peaks by categorizing the peak intensities into three levels: strong ("s"), medium ("m"), and weak ("w"). The numerical value 3 is given to strong peaks ("s"), "2" to medium peaks ("m"), and "1" to weak peaks ("w") to keep the score of the peaks. Peaks detected at their close binding-energy positions within ±1.5 eV were regarded as the same peak. The mean scores in Table 1 were calculated by dividing the sums of scores by the total number of judges, 7.

Figure 4 shows the relationship between the headcount of judges who detected the peak and the mean score of the same peak on the spectrum b003_23. The number of kinds of peaks to be subjected to judgment was 12 (see the last peak number in Table1),

Table 1. Example calculations of mean scores according to reports from 7 judges on spectrum b003. The peak intensities are categorized into three levels: strong ("s"), medium ("m"), and weak ("w").Peak No.: The number corresponding to each peak in Figure3 is shown for convenience. Mean scores: The score calculated by dividing the sums of scores by the total number of judges, 7.

Peak No.	Peak position(eV)	Intensity: number of judges	Mean score
1	5.5	m:7	2.0
2	59.0	w:4	0.6
3	84.5	s:7	3.0
4	88.0	s:7	3.0
5	257.0	w:1	0.1
6	283.0	w:1	0.1
7	336.0	m:2 s:5	2.7
8	354.5	m:3 s:4	2.6
9	494.0	s:2	0.3
10	546.0	w:3 m:3 s:1	1.7
11	646.0	w:5 m:1	1.0
12	763.5	w:6	0.9

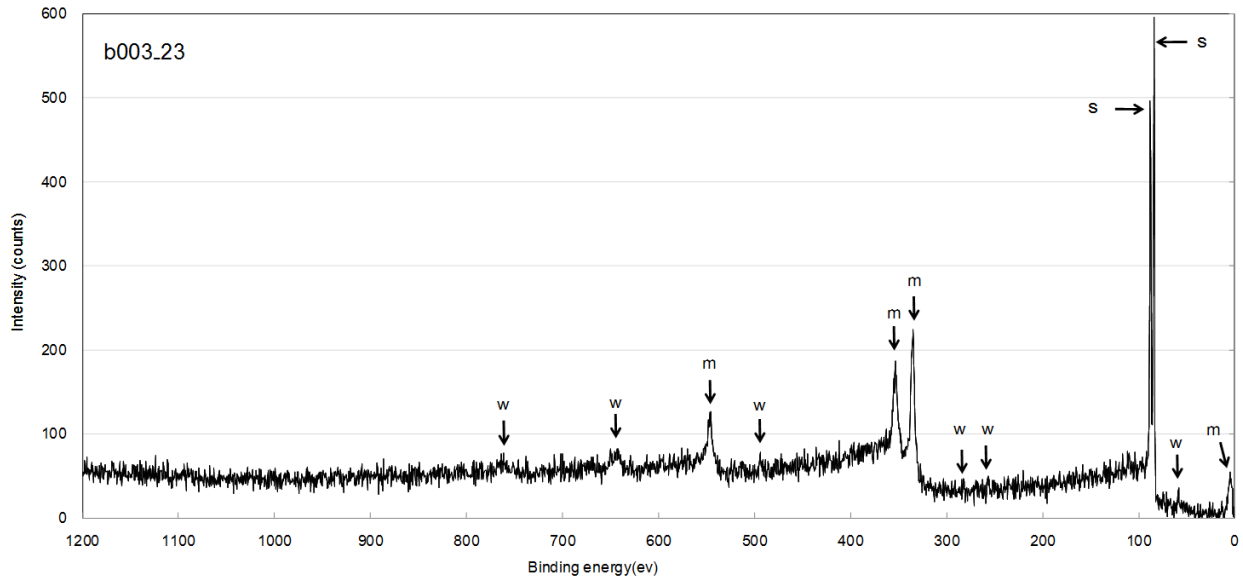

Figure 3. An example of visual peak detection and judgment of peak intensity on spectrum b003_23. The peak intensities are categorized into three levels: strong ("s"), medium ("m"), and weak ("w").

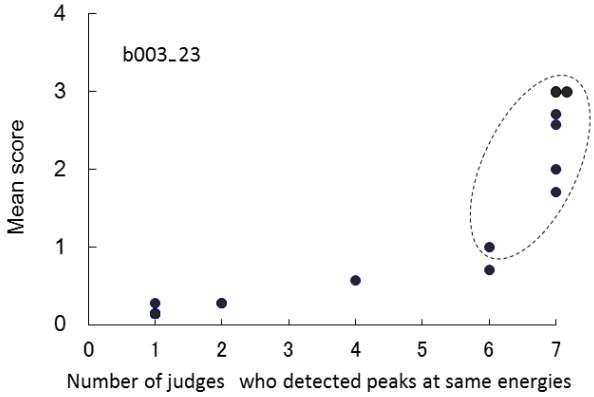

Figure 4. The relationship between the number of judges who detected peaks at the same energies and the mean scores. The region enclosed by a dotted line includes detected peaks with a mean score of more than 1.

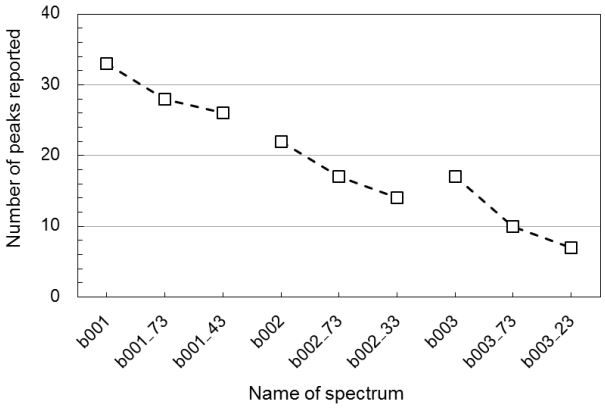

Figure 5. The number of peaks reported by visual detection for 9 synthesized XPS spectra. Each point in the same spectrum group is connected with a dotted line.

as shown by filled circles in this figure. The main 6 peaks (strong ("s") and medium ("m") peaks) which all seven judges detected have mean scores between 1.7 and 3.0. In this case, six peaks are recorded as plotted points including duplicated points at the mean score of 3.0. By contrast, in the case of smaller peaks, some judges recognized them as peaks but others didn't, leaving those peaks with mean scores of 1.0 or lower. In addition, as six judges have detected one peak with its mean score of almost 1.0, this peak is added to the record. As a result, seven peaks were judged as peaks in total for the spectrum b003_23 as shown in the region enclosed by a dotted line. Peaks with a score of less than 1 are not recorded.

Figure 5 shows the summed numbers of reported peaks after applying visual peak detection procedures described above for 9 synthesized XPS spectra. As expected in the spectrum intensity of synthesized peaks, the number of reported peaks decreases in the order of b001, b002, and b003 in accordance with the increases of noise in each spectrum. It can be inferred that this occurs because minor peaks of minor elements (Cu or Ag) are gradually buried in the noise as the noise component increases.

4.2 Evaluation of peak detection algorithms

The performance for detecting peaks using each algorithm was evaluated by counting the number of peaks detected by the automatic peak detection software based on each algorithm.

A peak in each spectrum can be classified into one of the following three types:

i. Matched peak: a peak that was judged as a peak by both visual detection and the automatic detection software,
ii. Missed peak: a peak that was judged as a peak by visual detection but not regarded as a peak by the automatic detection software,
iii. Noise peak: a peak that was not judged as a peak by visual detection but judged as a peak by the automatic detection software.

Table 2 shows a comparison between the results from the automatic detection software and those from visual detection. The "Peak area (%)" column shows the normalized peak area compared to the biggest one and is used as a reference when checking the results of judgment. The mean score M_v assigned to each peak detected by visual detection is determined by how many judges classified the peak into one of the three levels: strong ("s"), medium ("m"), and weak ("w") (see an example of Table 1). The mean matched score M_m is the mean score M_v of the selected peak which both the automatic detection and the visual detection successfully identify. The matched score M_s, that is an index of the degree of matching among all the results of automatic detection and visual detection, is defined as the ratio of TM_m (total of mean matched score M_m) and TM_v (total of mean score M_v).

Table 2. Example of peak judgment by 2nd DER method.

Name of spectrum: b002_73									
No.	Peaks		Results of visual detection			Results of 2nd DER method parameter: k=3, the number of 2nd derivative points: 11			
	Binding energy (eV)	Peak area (%)	Total number of judges who detected the peaks	Total score of each peak	M_V	Classification of peaks	M_m: mean matched score	M_{mis}: mean missed score	N_P: noise peak score
1	1134.4	1.1	7	10	1.4	Matched	1.4	-	
2	1129.2	1.2	7	9	1.3	Matched	1.3	-	
3	952.5	0.6 *	-	-	-	Noise	-	-	1.0
4	933.0	2.3	7	10	1.4	Matched	1.4	-	
5	762.9	0.6	6	9	1.5	Missed	-	1.5	
6	642.9	2.1	7	13	1.9	Matched	1.9	-	
7	604.3	1.8	7	10	1.4	Matched	1.4	-	
8	573.3	3.6	7	13	1.9	Matched	1.9	-	
9	546.6	11.1	7	16	2.3	Matched	2.3	-	
10	497.0	0.4 *	-	-	-	Noise	-	-	1.0
11	374.4	11.3	7	18	2.6	Matched	2.6	-	
12	368.4	18.8	7	18	2.6	Matched	2.6	-	
13	353.4	20.1	7	19	2.7	Matched	2.7	-	
14	335.2	34.2	7	19	2.7	Matched	2.7	-	
15	286.5	0.3 *	-	-	-	Noise	-	-	1.0
16	88.0	66.7	7	21	3.0	Matched	3.0	-	
17	84.1	100.0	7	21	3.0	Matched	3.0	-	
18	74.5	0.9 *	-	-	-	Noise	-	-	1.0
19	57.7	2.5	7	12	1.7	Matched	1.7	-	
20	44.5	0.3 *	-	-	-	Noise		-	1.0
21	38.0	0.2 *	-	-	-	Noise		-	1.0
22	36.5	0.3	-	-	-	Noise		-	1.0
23	31.5	0.3 *	-	-	-	Noise		-	1.0
24	6.6	2.5	7	12	1.7	Missed		1.7	-
25	3.9	5.3	7	16	2.3	Missed		2.3	-
Totals of each column			118	246	TM_V:35.4	-	TM_m:29.9	TM_{miss}:5.5	TN_P: 8.0
Mean value			6.9	-	-	-	M_S: 84.4	Mis: 15.6	-
TS (Total score)						60.9			

Notes
Matched: a peak identified as a peak by the program among the peaks detected by human eyes,
Missed: a peak not identified by the program among the peaks detected by human eyes,
Noise: a peak detected by the program but not detected by human eyes
*(Area of Noise peak):These peak area values are calculated after the missed peaks have been turned into matched peaks by changing the peak-detection conditions (i. e., the k value and the Ndiff value).

Therefore, the matched score (M_s) is calculated as a percentage as follows:

$$M_s = (TM_m) \times 100 / (TM_v)(\%) \quad \text{.....(4)}$$

For the spectrum b002_73 under the condition of k = 3.0 and the number of points for calculating the second derivative curve (N_{diff} = 11), M_m and M_s were given by the following calculation:

M_m for the peak No. 1 in spectrum b002_73 = 10/7 = 1.4, TM_m of spectrum b002_73: 29.9

M_s for the spectrum b002_73 = (29.9 × 100) ÷ 35.4 = 84.4.

The remaining non-matching peaks that are detected by visual detection become missed peaks (see Table 2), and a penalty should be given to the score so that the resultant score is subtracted from the matched peak score after normalization.

The missed score (*Mis*) is calculated as a percentage as follows:

$$Mis = (TMmis) \times 100 / (TM_v) \quad \text{.....(5)}$$

Here, *TMmis* represents "Totals of mean scores of missed peaks," i.e., those which are recognized not by the software but by visual detection.

Mis were given by the following calculation also:

Mis for the spectrum b002_73 = (5.5 × 100) ÷ 35.4 = 15.6.

The other peaks that are detected only by automatic detection become noise peaks. The well-defined scoring of noise peaks is difficult because the current scoring is based on the score of the peaks obtained by visual detection and scoring of noise peaks has no reference in this work. Therefore, the score of each noise peak (N_P) was arbitrarily set to 1.0 in Table 2.

The sum of these three scores for each spectrum is expressed as an index of the performance for detecting peaks.

The total score of the performance of automatic detection is defined as follows:

$$T_S(total\ score) = M_S(matched) - Mis(missed) - TN_p(noise) \quad \text{..... (6)}$$

Here TN_P represents totals of N_P.

A total score of 60.9 for the spectrum b002_73 was given by the following calculation:

T_S for the spectrum b002_73 = 84.4 - 15.6 - 8.0 = 60.9

Clearly the higher this total score is, the better the peak judgment is.

The ideal peak detection result is that all peaks are matched without missed peaks or noise peaks. The algorithm and parameters to give such a desirable result are said to be a good algorithm and good parameters.

In this study, for the three types of algorithms, the parameter values were varied to obtain the maximum matched score and minimum number of noise peaks. They are varied in the following ranges:

i 2nd DER method.
The number of points for calculating the second derivative curve (hereinafter abbreviated as "N_{diff}"): 5 to 17, value of k : 2.0, 2.5, 3.0, 3.5.

ii PB method.
N_{diff}: 5 to 17, value of k : 1.5, 2.0, 2.5, 3.0.

iii BGD method.
The number of moving average points (hereinafter abbreviated as "N_{mave}"): 11 to 51, value of k : 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, Minimum peak area: 1 to 10 count･eV.

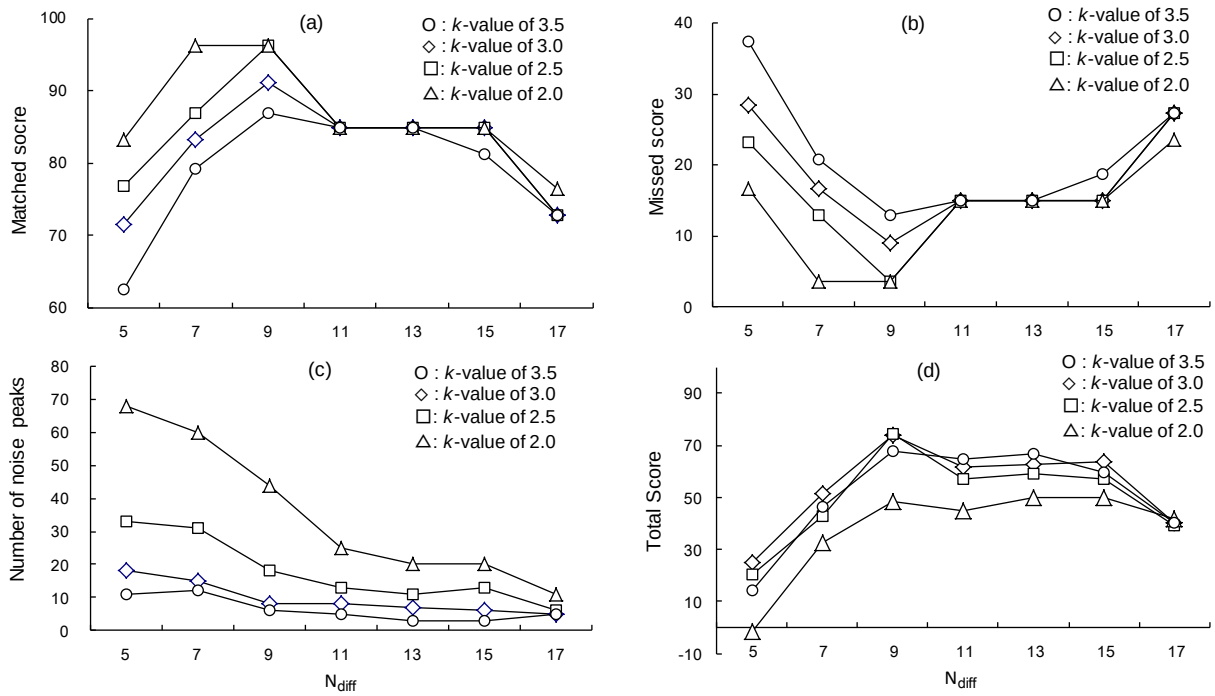

Figure 6. An evaluation example of the 2nd DER method in spectrum b002_73. The abscissa axis represents N_{diff}, the number of points used for calculating the second derivative curve. The ordinate axis represents the scores in graph (a), (b),(d) and the Number of noise peaks in graph (c). The four curves correspond to the scores at four different values of k (○: value of k = 3.5, ◇: value of k = 3.0, □: value of k = 2.5, △:value of k = 2.0).

5 Performance analysis of the 2nd DER method

The 2nd DER method, whose algorithm is relatively simple, is evaluated first. Figure 6 shows an example evaluation of b002_73. The abscissa axis represents N_{diff}, the number of points used for calculating the second derivative curve, and the ordinate axis represents the scores in Figures 6(a), 6(b) and 6(d). The four curves correspond to the scores at four different values of k. The matched score gives best results when N_{diff} is 7 or 9. The number of noise peaks decreases as the N_{diff} is increased, and it decreases as the value of *k* is increased in Figure 6(c).

The result of the 2nd DER method appears to be most affected by the N_{diff}. Viewed from the matched score, the performance appears to be slightly better when the value of *k* is 2.0.

On the other hand, the number of noise peaks increases sharply when the N_{diff} is decreased. This tendency is significant if the value of *k* is small. The total score is obtained by adding these two effects, with highest scores being obtained at an N_{diff} of 9 and scores dropping on either side of 9.

Figure 7 shows that using small values of N_{diff} results in poor performance of peak matching. Figure 7(a) shows the enlarged part of the original spectrum b003_23. Other figures show the second derivative curve and the standard deviation curves when varying the N_{diff} as 5, 11, and 17, respectively. The standard deviation curves in Figures (b), (c), and (d) were obtained by processing a spectrum to determine the standard deviation σi. The amplitude of the standard deviation curve is $k\sigma_i$, where σ_i is the standard deviation and *k* is a multiplying factor. A signal in the spectrum is judged to be detected as a peak when the local negative minimum d_{min} is smaller than $k\sigma_i$.

If the N_{diff} is small, as shown in Figure 7(b), peaks cannot be identified because they are buried in the noise. The smaller the N_{diff}, the more the noise peaks are emphasized, so that minor peaks are likely to be missed. When the N_{diff} is increased, a typical second derivative curve having negative local minimums appears in Figures 7(c) and 7(d). The bar markers in the curves show the matched peaks.

Figure 8 shows that using large values for N_{diff} results

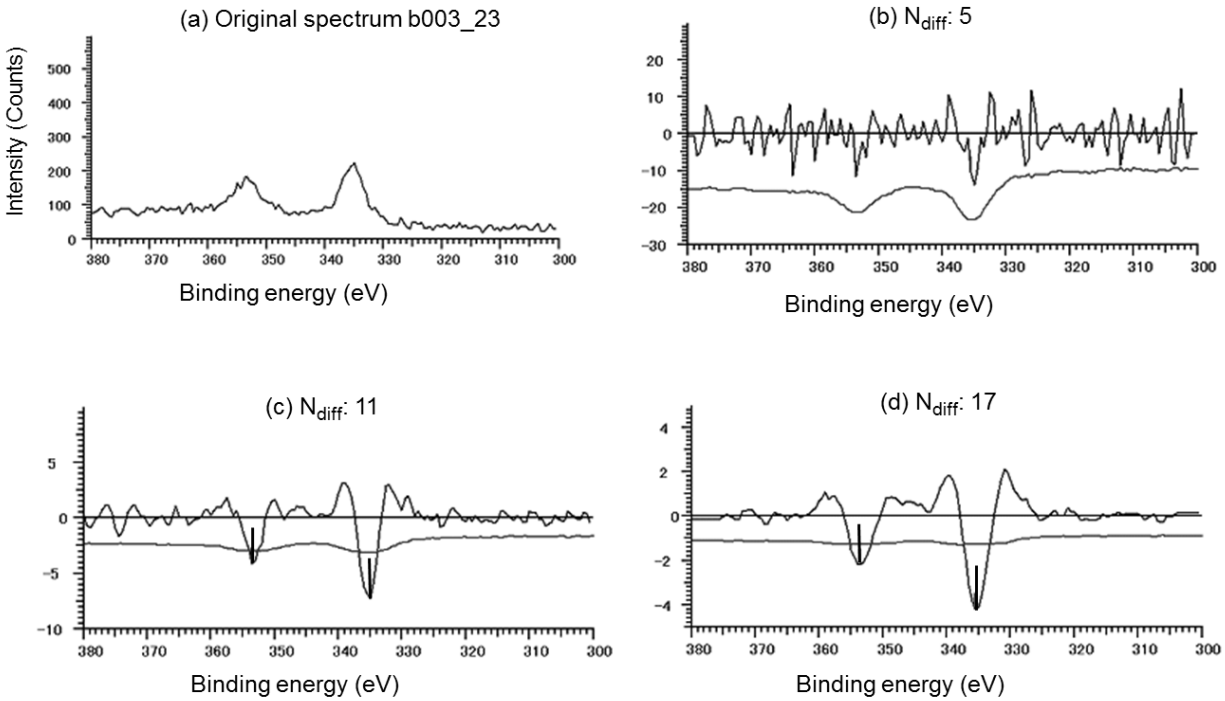

Figure 7. Influence of small N_{diff} (k = 3.0, partial expansion of spectrum b003_23). Spectrum (a) shows the enlarged part of the original spectrum b003_23. The other figures, (b), (c) and (d) represent the second derivative curve having negative local minimums when varying the N_{diff} as 5,11, and 17.

in poor performance of peak matching. The scale of binding energy in this figure is approximately five times smaller than that of binding energy in Figure 7. Figure 8(a) shows the enlarged part of the original spectrum b001, and the other figures show the second derivative curves when the N_{diff} is 5, 13, and 17, respectively. When the N_{diff} is small, as shown in Figure 8(b), two local minimums are observed, but in contrast as the N_{diff} is increased, as shown in Figures 8(c) and 8(d), two peaks interfere with each other, deforming the second derivative curve and causing it to fail to work. In these figures, the solid bar markers indicate the matched peaks and the dotted ones indicate missed peaks or noise peaks, respectively.

Figure 9(b) shows the relationship between the second derivative curve and the standard deviation curve with value of *k* for a monotonous spectrum without clear peaks shown in Figure 9(a), which is a portion of b001.

These curves were obtained under the condition that the value of *k* is changed as 2.0 and 3.0 with the N_{diff} of 9. When the value of *k* is 2.0 in Figure 9(b), noise peaks were detected because the second derivative curves crossed the standard deviation curves. On the other hand, as the value of *k* was increased, noise peaks were not detected because the two curves separated from each other.

Figure 10 shows the total scores of peak matching for all 9 spectra. The original spectrum data are b001, b002 and b003 series from the left, and they are arranged in the vertical direction in the order of noise peak intensity. Most graphs show a maximum score when the number of points N_{diff} is between 9 and 13. The effect of N_{diff} on the total score is rather weak in b001 but becomes stronger in b002 and stronger still in b003. The score variation may be alleviated because many clear peaks were detected in b001. In general, the peak matching performance is good at the value of k = 3.0; this is because the number of noise peaks decreases as the value of *k* becomes larger.

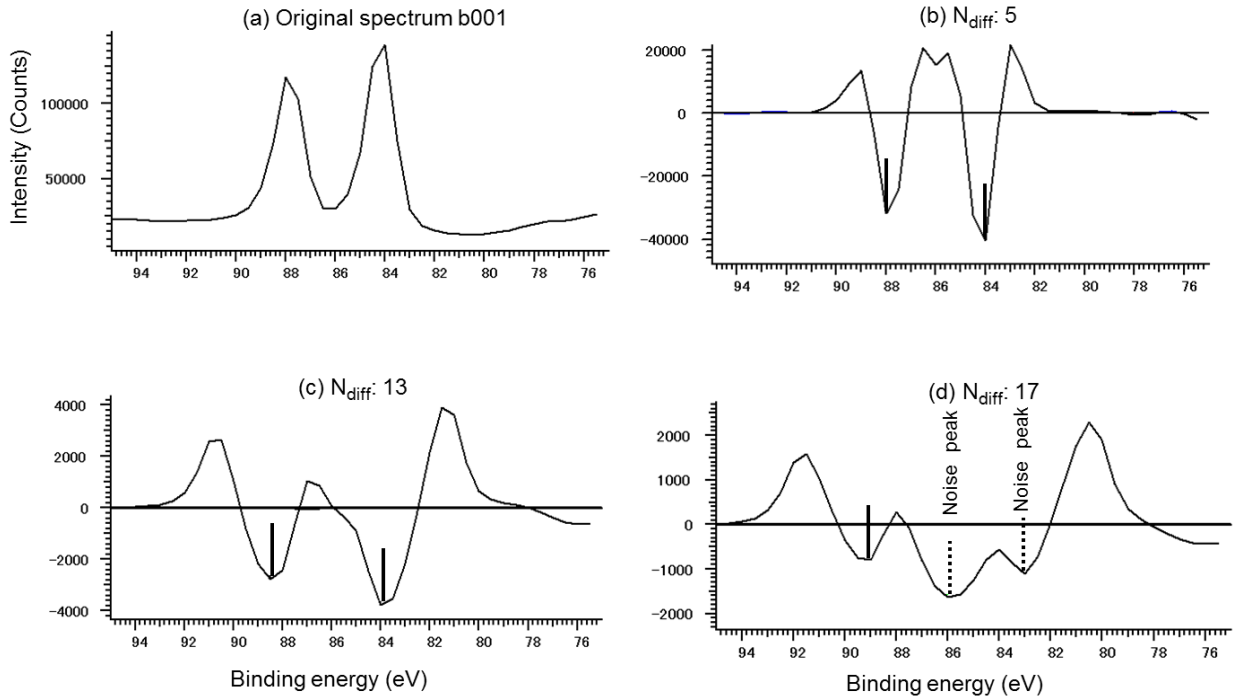

Figure 8. Influence of large N_{diff} (k = 3.0). Spectrum (a) shows the enlarged part of the original spectrum b001. The other figures, (b), (c) and (d) show the second derivative curves when the N_{diff} is 5, 13, and 17, respectively (Solid bar: matched peaks, Dotted bar: Missed peaks).

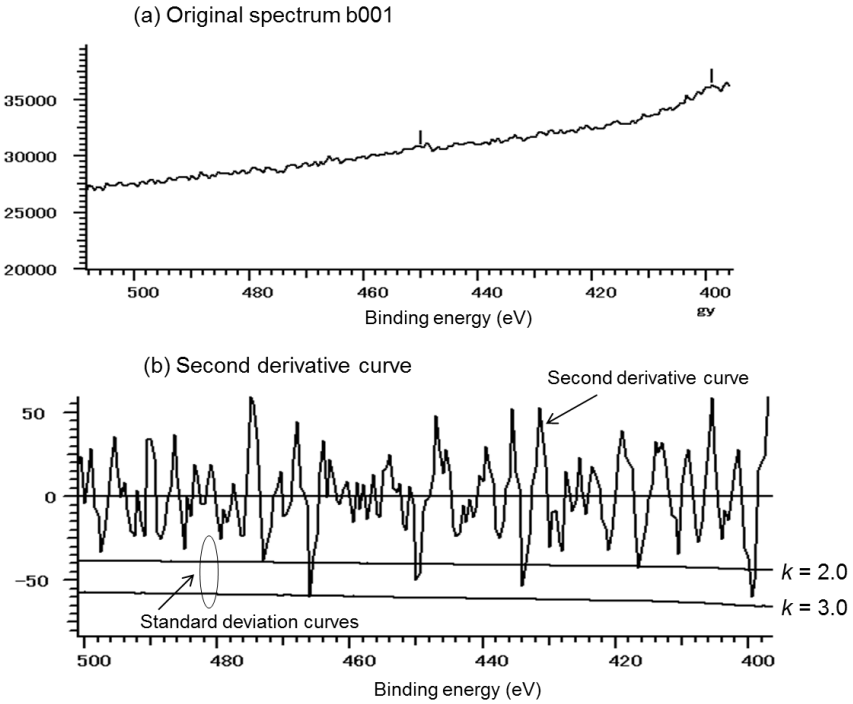

Figure 9. Relationship between the second derivative curve and the standard deviation curve. Figure (a) is a portion of b001. Figure (b) shows the relationship between the second derivative curve and the standard deviation curve with value of k = 2.0 and k = 3.0 for a monotonous spectrum.

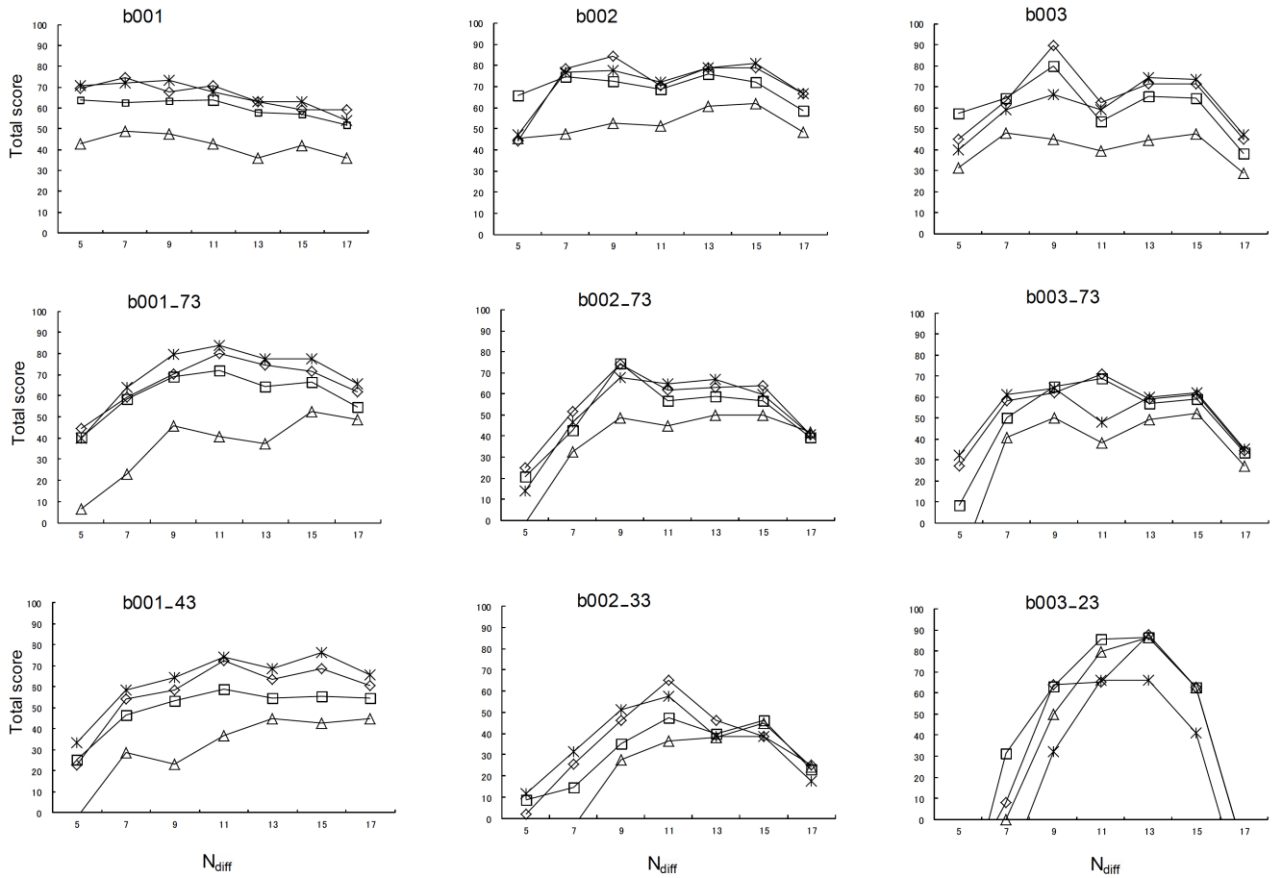

Figure 10. Total evaluations of the 2nd DER method for each spectrum (∗: value of k = 3.5, ◇: value of k = 3.0, □: value of k = 2.5, △: value of k = 2.0). The abscissa axis represents Ndiff, the number of points used for calculating the second derivative curve. The ordinate axis represents the Total score.

6 Performance analysis of the PB method

Figure 11 shows an example of the evaluation of peak matching in spectrum b002_73 by the PB method. The abscissa axis represents the N_{diff} and the ordinate axis represents the score in Figures 11(a), 11(b) and 11(d). As shown in Figure 11(c), the number of noise peaks is low as compared to those of the other two methods. It means that the process of finding candidate peaks in the PB method suppresses the generation of noise peaks. Not only was the number of noise peaks reduced by taking the second derivative, it was also reduced by selecting them on the basis of whether each candidate peak after subtracting background met the condition of peak detection by considering the statistical errors of each peak and its background. As a result, better total scores are realized at smaller *k* values compared to the 2nd DER method.

On the whole, the results produced by the PB method are less affected by the values of *k* and N_{diff} than the results produced by the 2nd DER method. The figure shows that the best results are produced when N_{diff} has a value of 9.

This algorithm uses the full width at half maximum as one of the determining factors for peak detection. Although the full width took a value from several eV to about 10 eV depending on the peak, 10 eV was adopted because stable results were obtained.

Figure 12 shows the total scores of peak matching for all 9 spectra by the PB method. The layout of the graphs is the same as that in the 2nd DER method. The total scores are best at the N_{diff} of around 9. The effect of the N_{diff} values on the total scores is the smallest when *k* has a value of 2.0.

7 Performance analysis of the BGD method

7.1 Minimum area parameter

As is described in Section 2.3, another parameter of peak area, Si, is introduced in the BGD method so as to be able to detect a low intensity small peak with a broad width. Please refer to Appendix A.3 for the method of calculating Si. As shown in inequality (A21) for the detection condition, if the peak area S_i is greater than a certain value S_{i0}, the detected candidate peak is judged to be a real peak. The effective value for this parameter S_{i0} is studied here.

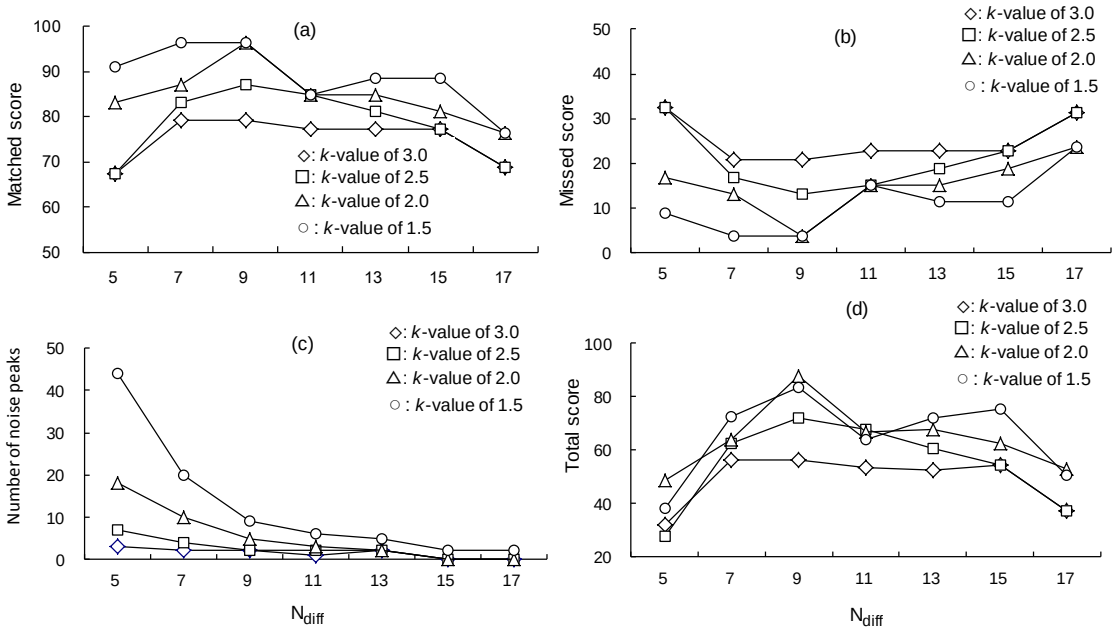

Figure 11. An evaluation example of the PB method (example of b002_73).The abscissa axis represents Ndiff, the number of points used for calculating the second Derivative curve. The ordinate axis represents the scores in graph (a), (b), (d) and the Number of noise peaks in graph (c). The four curves correspond to the scores at four different values of *k* (◇: value of *k* = 3.0, □: value of *k* = 2.5, △: value of *k* = 2.0, ○: value of *k* = 1.5).

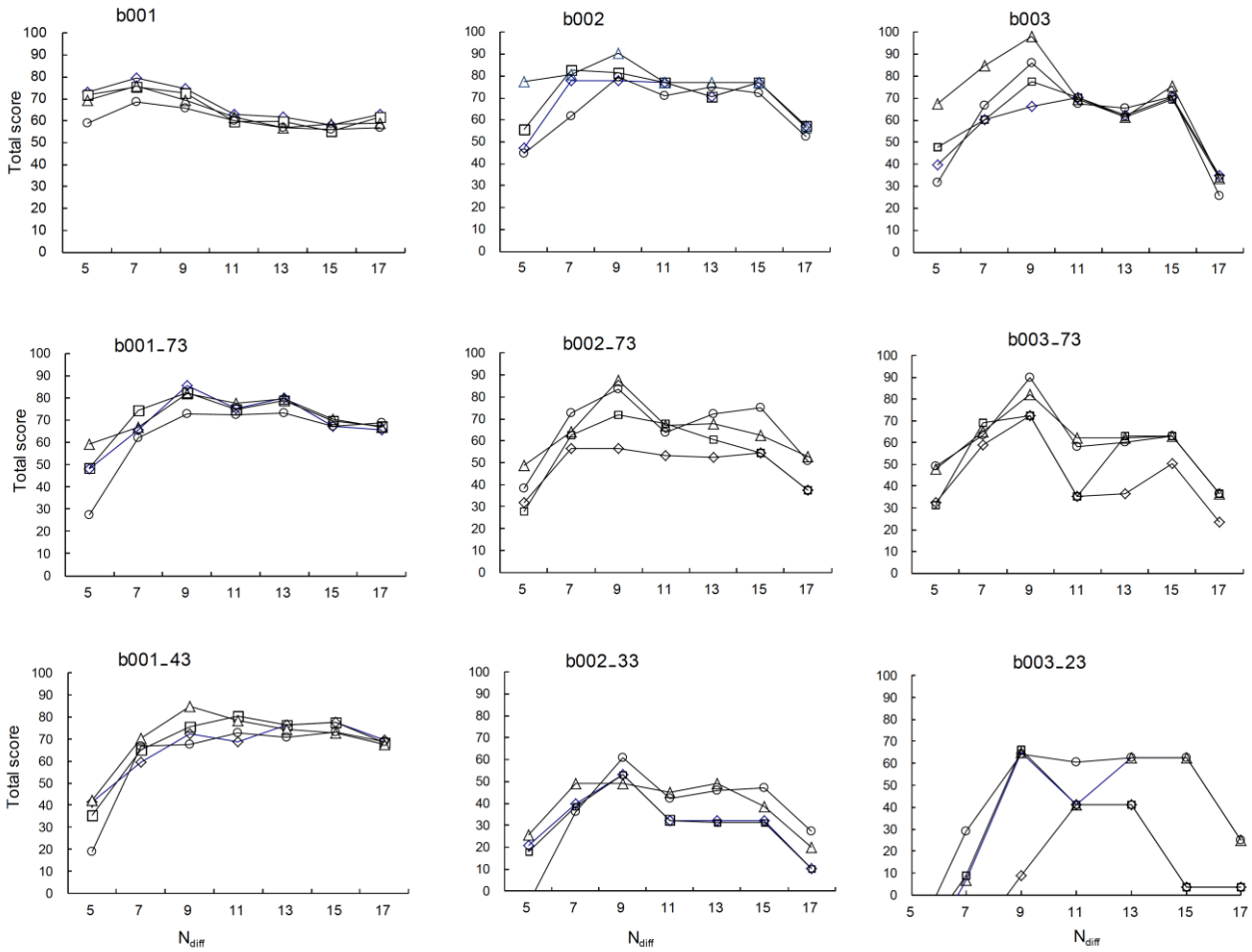

Figure 12. Total evaluation of the PB method for each spectrum (◇: value of k = 3.0, □: value of k = 2.5, △: value of k = 2.0, ○: value of k = 1.5). The abscissa axis represents Ndiff, the number of points used for calculating the second derivative curve. The ordinate axis represents the Total scores.

Figure 13 shows the relationship between the minimum area parameter and the number of noise peaks for b001, b002, and b003 spectra. When the minimum area is 1 count•eV, many noise peaks are generated, but if it exceeds 2 count•eV, the number of noise peaks decreases, and if it is 5 count•eV or more, it hardly affects the peak detection result. In this study, the minimum area was set to 2 count•eV as an effective and smaller S_{i0} for these spectra.

As is the case with the PB method, this algorithm used the full width at half maximum as one of the determining factors for peak detection. The value adopted for this was 10 eV.

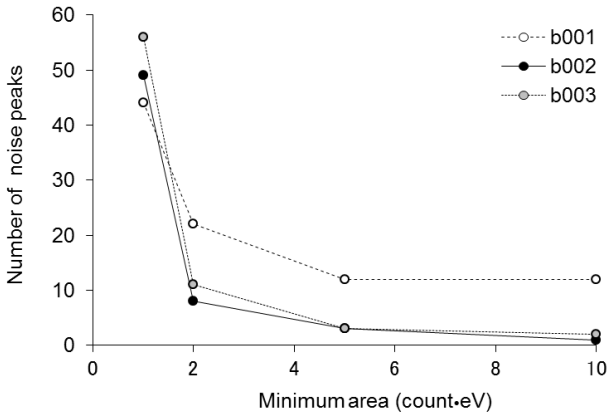

Figure 13. Variation in the number of noise peaks as a function of minimum area.

7.2 Performance analysis

Figure 14 shows an example result of the peak

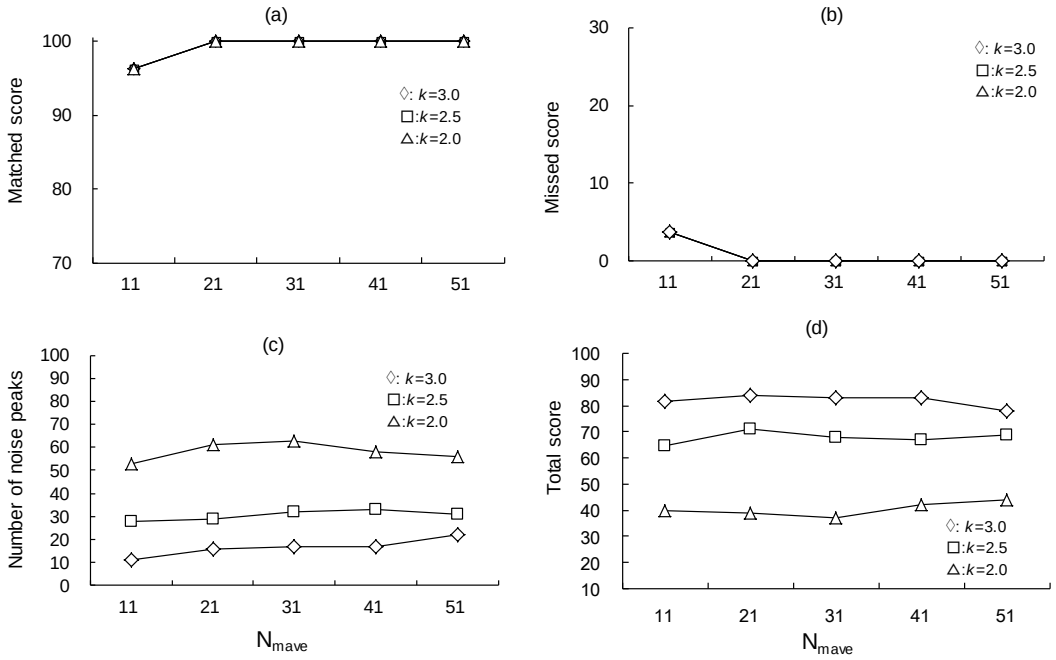

Figure 14. An evaluation example of the BGD method (example of b002_73). The abscissa axis represents the number of moving average points, Nmave, and the ordinate axis represents the scores in graph (a), (b), and (d) and the Number of noise peaks in graph (c). The four curves correspond to the scores at four different values of k.(◇: value of k = 3.0,□: value of k = 2.5, △: value of *k* = 2.0).

matching for b002_73. The abscissa axis represents the number of moving average points N_{mave}, and the ordinate axis represents the score. The three curves in Figure 14(c) and 14(d) show the number of noise peaks and total scores respectively at different *k* values, *k* = 2.0, 2.5, and 3.0.

The matched score is sufficiently high when N_{mave} is 11, and the matched scores reach 100 when N_{mave} is more than 21 for all three values of *k*. The number of noise peaks increases greatly when the value of *k* is small, and the total scores do not reach 100 even at *k* = 3.0. It seems that the total scores solely reflect the number of generated noise peaks determined by *k* under the condition of sufficiently large N_{mave}.

Figure 15 shows an example of failing to detect a small peak when the value of *k* is 3.0. Figure 15(a) shows an example of a small, noisy, isolated peak because of a small N_{mave} of 11 in b002_73, and Figure 15(b) shows an example of a small, smooth peak at the tail of a large peak under the condition of a large N_{mave} of 51 in b003. The solid arrow markers show matched peaks, and dotted ones show missed peaks. When the N_{mave} is small, some peaks cannot be detected because the noise component is larger than the minor peak. On

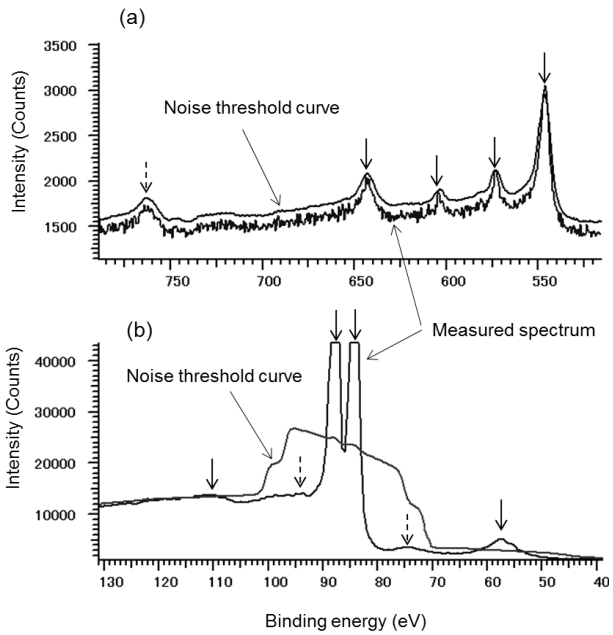

Figure 15. Examples of failing to detect a peak in the BGD method. Figure 15(a) shows an example of a small, noisy, isolated peak because of a small N_{mave} of 11 in b002_73, and Figure 15(b) shows an example of a small, smooth peak at the tail of a large peak under the condition of a large N_{mave} of 51 in b003 (Solid arrow: matched peaks, Dotted arrow: missed peaks).

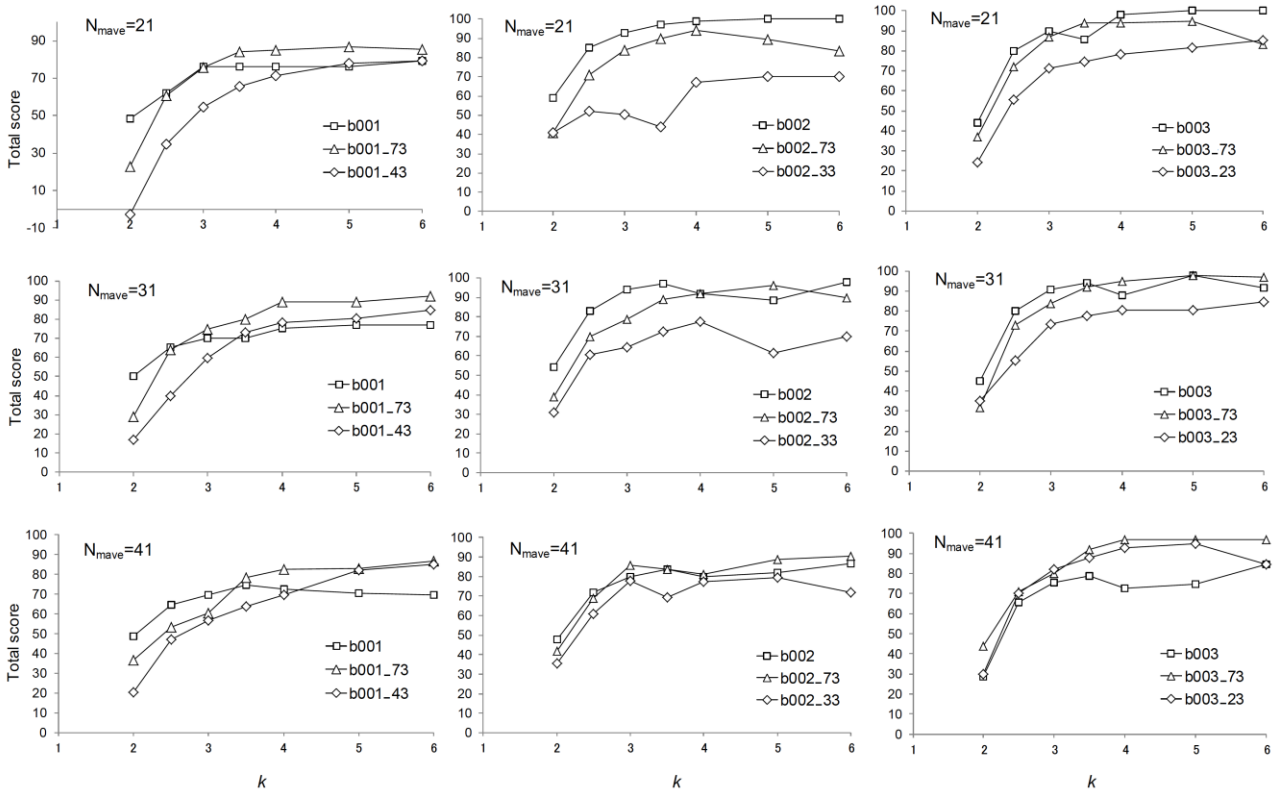

Figure 16. Effect of the *k* values in the wider range of 2 and 6 on the total scores of all the 9 spectra for the N_{mave} values of 21, 31 and 41. The abscissa axis represents *k* and the ordinate axis represents the Total score. The score of each spectrum tends to become relatively constant at the *k* values of 4 or more.

Table 3. Effect of *k* value on the sum of total score.

N_{mave}	*k* value	Total scores for each spectrum									Sum of total scores
		b001	b001_73	b001_43	b002	b002_73	b002_33	b003	b003_73	b003_23	
21	6	79	85	79	100	84	70	100	83	85	766
	5	76	87	78	100	90	70	100	95	81	777
	4	76	85	71	99	94	67	98	94	78	763
31	6	77	92	85	98	90	70	92	97	84	785
	5	77	89	81	89	96	62	98	98	80	770
	4	75	89	78	92	92	77	88	95	80	767
41	6	70	87	85	87	91	72	85	97	84	757
	5	71	83	82	82	89	79	75	97	95	752
	4	73	82	70	80	81	77	73	97	93	726

the other hand, even when the N_{mave} is large, some peaks at the tail of a large peak cannot be detected.

In Figure 16, the effect of the *k* values in the wider range of 2 and 6 on the total scores of all the 9 spectra was examined for the N_{mave} values of 21, 31 and 41, where the score became relatively stable. The abscissa axis represents *k* and the ordinate axis represents the total score. As shown in Figure16, it was found that the score of each spectrum tends to become relatively constant at the *k* values of 4 or more.

Firstly, the total scores of the most stable peak detection at the *k* values of 4 or more are summarized in Table 3. When looking at the sum of total scores, it was found that high sums of total scores were obtained at the *k* values of 5 and 6.

Secondly, the relationship between the number of detected peaks and the *k* values of 5 and 6 was examined in each spectrum as shown in Table 4. When the *k* value was 5, the number of detected peaks was slightly larger in each spectrum. If the number of detected peaks is

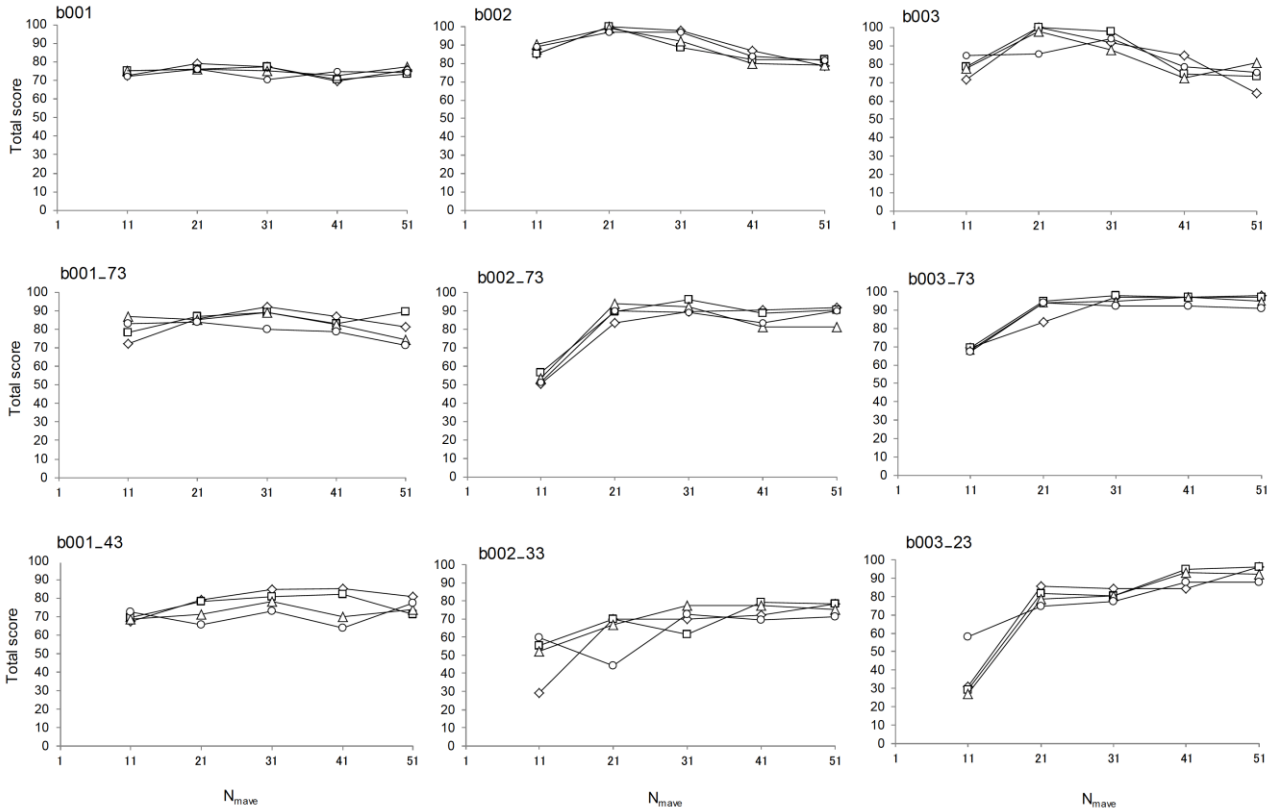

Figure 17. The total evaluation of the BGD method (◇: value of k = 6.0, □: value of k = 5.0, △: value of k = 4.0, ○: value of k=3.5). The abscissa axis represents N_{diff}, the number of point used for calculating the second derivative curve. The ordinate axis represents the scores in (a), (b) and (d).

Table 4 Effect of k value on the number of peaks detected in each spectrum.

N_{mave}	k value	Number of peaks detected in each spectrum								
		b001	b001_73	b001_43	b002	b002_73	b002_33	b003	b003_73	b003_23
21	6	31	28	25	22	16	14	17	13	8
	5	34	31	26	22	19	14	17	15	12
31	6	33	31	25	24	19	14	18	13	9
	5	33	34	29	24	21	13	19	12	13
41	6	32	27	25	22	18	14	16	13	9
	5	36	31	28	21	20	16	16	13	12

large, there is a possibility that it includes noise peaks and the score becomes lower. However, since it is also possible that small real peaks which will be identified as peaks have been detected among noise peaks, having an appropriate number of noise peaks is welcome and helps analysts.

Figure 17 shows the effect of N_{mave} values on the total scores of all the 9 spectra by the BGD method. The graph layout is the same as that of the 2nd DER method in Figure 10.The total score exhibits a slightly poorer performance in a range where the N_{mave} is small for the spectrum with a higher noise component.The reason why such a tendency occurs is that peaks are buried in the noise components when the N_{mave} is small.The BGD method attains a good score of about 70% to 100% for a relatively wide range of N_{mave} and even an example of a matched score higher than those of other methods can be observed.This good performance was achieved when N_{mave} values ranging from 21 to 41 and a k value ranging from 4 to 6 are employed.

8. Summary of Performance analysis

The characteristics of each peak detection method and the importance of choosing effective parameter values

are summarized below.

The 2nd DER method is sensitive to the N_{diff} and the value of *k*, and the matched score is highest when the N_{diff} is between 7 and 9. On the other hand, when the value of *k* is decreased, noise peaks increase rapidly in a range where the N_{diff} is small. The best total scores were obtained at the N_{diff} of 11 and the value of k = 3.0. This method was found to be the most sensitive to the parameter values among the three peak detection methods.

In the PB method, the process of finding candidate peaks suppresses the generation of noise peaks. Not only was the number of noise peaks reduced by taking the second derivative, it was also reduced by selecting them on the basis of whether each candidate peak met the condition of peak detection by considering the statistical errors of each peak and its background.

The total score is best at the N_{diff} of around 9. Because the effect of the *k* value on the performance of peak detection is small, comparatively good results were obtained at the small value of k = 2.0. However, there is a possibility that peaks sometimes were overlooked, as shown in the peak detection results in a noisy spectrum such as b003_23 (see Figure 12).

As far as the total score, the BGD method was found to be the best among the three methods. There were almost no missed peaks when N_{mave} was more than 21, as is shown in Figure 14(b), compared to the other two methods. Good performance, with a total score of around 90 %, was achieved with N_{mave} values ranging from 21 to 41. Although a high total score is obtained for a wide range of N_{mave} values, noise generation is not completely suppressed. Since the tendency of the total score to increase with the increase of value of *k* is clear as compared with the other two methods, noise peaks can be suppressed at the value of k = 4.0 or higher. There is still room for some improvements in the treatment of minor peaks when the spectrum is noisy (*i.e.*, when N_{mave} is small) or an adjacent large peak is overlapping.

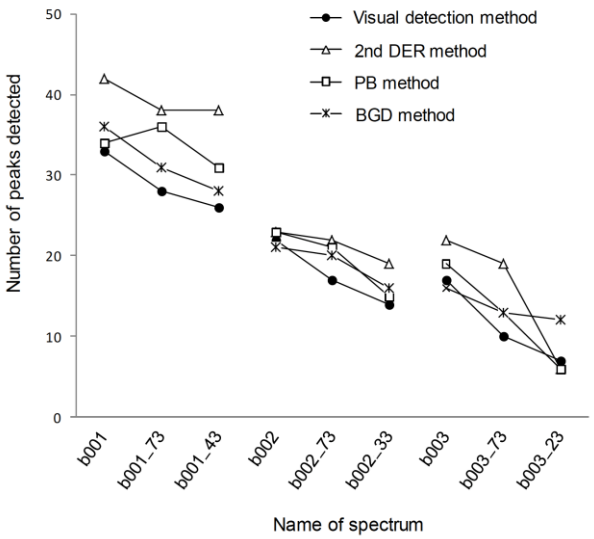

Figure 18. Comparison of proposed methods with visual detection.

9 Discussions

In this study, the performance of three algorithms for automatic peak detection was studied using nine synthetic XPS spectra with different levels of noise by changing the parameter values used for these algorithms. The peak detection results at the effective parameter values found through this process were compared with those obtained by visual detection as well.

Figure 18 shows the peak detection results obtained by visual detection and by the three methods under the respective recommended conditions. The number of peaks detected on the ordinate axis indicates the number of peaks as described below. For the visual detection method, it indicates the number of matched peaks with mean scores of 1.0 or more, and for the automatic detection methods, it indicates the number of all the detected peaks, including matched and noise peaks.

An algorithm that detects an adequate number of peaks (that is, one that is capable of detecting real peaks even if it also includes several noise peaks as candidate peaks) might be better than an algorithm that detects a lesser number of peaks (that is, one that fails to detect smaller real peaks).

From a standpoint of this, the BGD method attained mostly successful peak detection results in this study. The peak detection ability of the BGD algorithm reaches a level that allows it to be used practically as a peak detection tool.

The recommended conditions for the parameters of these three methods are as follows:

i. 2nd DER method: Number of the second derivative points: 11, value of *k*: 3.0
ii. PB method: Number of the second derivative points: 9, value of *k*: 2.0

Table 5. Features and Peak Detection Ability of Three Methods.

Method	Detection ability			
	Single large peak	Single small peak	Multiple overlapping peaks	
			Adjacent peaks	Valley (shallow)
Peak Detection Using the Threshold Curve of the Second Derivative (2nd DER method)	Good	Good but sensitive to value of *k*	Detecting adjacent peaks is sensitive to the number of the second derivative points.	(same as on the left)
Peak Detection by Directly Calculating Peak and Background Relations at a Candidate Peak (PB method)	Good	Good	Detecting adjacent peaks is sensitive to the number of the second derivative points.	(same as on the left)
Peak Detection Using Rough Estimation of Spectrum Background (BGD method)	Good	Reasonably good but difficult to detect very small peaks adjacent to a large peak	Detecting small shoulder peaks on a large peak or tail peak at both sides of a large peak is difficult.	Good

iii. BGD method: Number of moving average points: from 21 to 41, value of *k*: 5.0, minimum peak area: 2 count•eV

These recommended conditions are strictly for XPS spectra with an energy step of 0.5 eV. If the energy step width is different from this value, there is a possibility that the parameter value will change, so it is desirable to select optimum values by changing the values of N_{diff} and N_{mave}.

It is also known that even better results are achieved if N_{diff} is given a value corresponding to the number "(the full width at half maximum) / (energy step width)." As for N_{mave}, it is desirable to give it a value several times the value of N_{diff}. However, in the peak detection process, since the full width at half maximum of the peak to be detected is unknown, scientists need to find the optimum value by giving different values to N_{diff} and N_{mave}.

In order to serve as a future reference, the advantages and disadvantages of each method, for various peak shapes in a spectrum, which were found through this evaluation are summarized in Table 5. As can be seen from the table, it can be suggested that combinations of these three methods enable scientists to implement a reliable peak detection method by specifically creating automatic peak detection software.

It will be appreciated if the present research report helps a surface scientist to understand how to select the effective conditions to carry out peak detection and if it is useful in obtaining better peak detection results.

10. Conclusions

In order to evaluate three algorithms for detecting peaks, synthetic XPS spectra superposed with different levels of noise were used to find out the combination of parameter values resulting in the best performance. The three peak detection algorithms are the Threshold Curve of the Second Derivative (2nd DER method), the Directly Calculating Peak and Background Relations at a Candidate Peak (PB method), and the Rough Estimation of Spectrum Background (BGD method). The peak detection results clearly showed that particular combinations of parameter values produce the best performance for each algorithm. The results show that the recommended parameter values in the 2nd DER and PB methods were 11 and 9 for the number of the second derivative points and 3.0 and 2.0 for the value of k, respectively. In addition, in the BGD method, the

recommended value of the moving average points was between 21 and 41, and the recommended value of *k* was 5.0.

By comparing the detection results from the peak detection algorithms and those by visual detection, the BGD method was found to be the most practical if the effective parameter values are chosen.

This study showed the importance of choosing effective parameter values for the best performance. For convenience the advantages and disadvantages of these three peak detection methods are summarized in the table.

11. References

[1] Y. Nagatsuka, K. Yoshida, Y. Nagasawa and J. Suzumi, *J. Surf. Anal.*, **3**, 602 (1997).

[2] P. Gans, *Data Fitting in the Chemical Sciences*, p. 17-18, Wiley (1992).

[3] Commercially available software: GRAMS/AI, GRAMS/32, Peak Fit, Flexpro, AutoSignal, IGOR Pro, and MATLAB.

[4] Y. Furukawa, Y. Nagatsuka, Y. Nagasawa, S. Fukushima, M. Yoshitake, and A. Tanaka, *J. Surf. Anal.* **14**, 225 (2008).

[5] M. Suzuki, S. Fukushima, and S. Tanuma, *J. Surf. Anal.* **14**, 104 (2007).

[6] Available Test Spectra for Performance Investigation in XPS; http://www.sasj.jp/eng-index.html

[7] A. Savitzsky and M. J. E. Golay, *Anal. Chem.*, **36**, 1627 (1964).

[8] K. Wakimoto, T. Tarumi and Y. Tanaka, PC Statistical Analysis Handbook I, Basic Statistics Edition, P.164 , Kyoritsu Shuppan (1984).

Appendices

Details of the three kinds of algorithm for detecting peaks proposed in 2008 [4] are described below for the reader's convenience.

A.1 Peak Detection Using Threshold Curve of the Second Derivative (2nd DER method)

This method has the same effect as subtracting the background from the spectrum, by making use of the second derivative. Because it has no arbitrariness in the background subtracting procedure, it may be a relatively convenient algorithm when used with the aid of computers.

By making use of the moving polynomial approximation procedure (Savitzky-Golay method) [7], it is possible to calculate the second derivative spectrum d_i (i = 1, T) from the original spectrum y_i (i = 1, T) as follows:

$$d_i = \sum_{j=-n}^{n} g_j y_{i+j} \quad \text{..... (A1)}$$

where g_j (j = -n, n) is the Savitzky-Golay coefficients for the second derivative. If $2n+1$ points of data cover roughly the half width of the typical peak, the obtained second derivative will faithfully represent the true second derivative curve. If it is expressed as described above, the variance σ_i^2 of d_i can be obtained as follows:

$$\sigma_i^2 = \sum_{j=-n}^{n} g_j^2 (Var)_{i+j} + 2\sum_{j=-n}^{n}\sum_{l\neq j}^{n} g_j g_l (Cov)_{i+j,i+l} \quad \text{..... (A2)}$$

where $(Var)_{i+j}$ is the variance of y_{i+j} and $(Cov)_{i+j,j+l}$ the covariance of y_{i+j} and y_{j+l}. Again, if randomness of the spectrum data y_i is assumed, then:

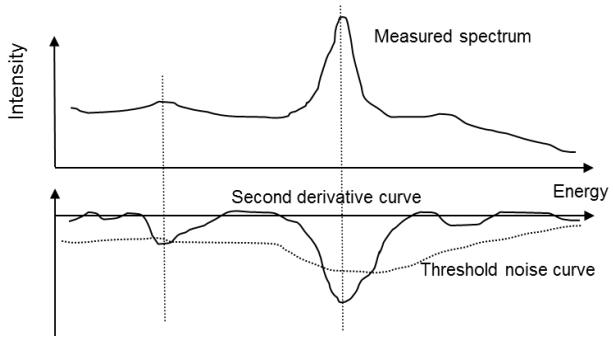

Figure A1. Schematic diagram of peak positions, the second derivative curve, and the threshold noise curve.

$$(Var)_{i+j} = y_{i+j} \quad \text{.....(A3)}$$

Further, if each y_i is independent and has no correlation with any other, then:

$$(Cov)_{i+j,\,j+l} = 0 \quad \text{.....(A4)}$$

Thus, the variance σ_i^2 of di can be approximately calculated as follows:

$$\sigma_i^2 = \sum_{j=-n}^{n} g_j^2 y_{i+j} \quad \text{.....(A5)}$$

As the peak in the spectrum corresponds to the local minimum of the spectrum, the peak is judged to be real if the local negative minimum d_{min} is less than (in absolute values, greater than) its noise fluctuation. Therefore, for the peak judge inequality, if the following inequality is satisfied, the peak is considered to have been detected at the position giving d_{min} in the second derivative spectrum (Figure A1):

$$d_{\min} < -k\sigma_i = -k\sqrt{\sum_{j=-n}^{n} g_i^2 y_{i+j}}\,. \quad \text{.....(A6)}$$

A.2 Peak Detection by Directly Calculating Peak and Background Relations at the Candidate Peaks (PB method)

In this case, the candidate peaks are detected following the steps described below:

i Calculate the second derivative curve (*e.g.*, by using the Savitzky-Golay method [6]).

ii. Pick the positions that satisfy the condition that the local minimum of the second derivative curve is negative. Picked positions are classified as candidate peak positions.

iii. Let p be a candidate peak position, and w be the typical full width at half maximum of the peak in the original spectrum, where w is usually given in the peak detection condition. (See Figure A2.) If there exists a positive local maximum at $x = p_1$ in the second derivative spectral range $p - 3w \le x < p$ in the nearest candidate peak (if it does not exist, the position $p_1 = p - 3w$ is regarded as the position), the position p_1 is regarded as a mid-position to the left-side background. Furthermore, if there exists a local minimum in the smoothed spectrum or a zero cross position in the second derivative spectrum at $x = q_1$ in the spectral range $p_1 - 2w \le x < p_1$ in the nearest candidate peak (if it does not exist, the position $q_1 = p_1 - 2w$ is regarded as the position), the distance $l_1 = p - q_1$ corresponds to the left-side background position $p - l_1$ with intensity B_1. Likewise, if there exists a positive local maximum at x in the second derivative spectral range $p < x \le p + 3w$ in the nearest candidate peak (if it does not exist, the position $p_2 = p + 3w$ is regarded as the position), the position x = p_2 is regarded as a mid-position to the right-side background. Furthermore, if there exists a local minimum peak position in the smoothed spectrum or a zero cross position in the second derivative spectrum in the spectral range $p_2 \le x < p_2 + 2w$ in the nearest candidate peak (if it does not exist, the position $q_2 = p_2 + 2w$ is regarded as the position), the distance $l_2 = q_2 - p$ corresponds to the right-side background position $p + l_2$ with intensity B_2.

iv. If the background curve near the peak can be approximated by a straight line, the background intensity B at the peak position is calculated as

$$B = (B_1 l_2 + B_2 l_1)/(l_1 + l_2) \quad \text{..... (A7)}$$

If P is denoted as the peak intensity with the background, and N as the net peak intensity, then $N = P - B$, and the variance σ_N^2 of N is given as follows:

$$\sigma_N^2 = \sigma_P^2 + \sigma_B^2 \quad \text{..... (A8)}$$

where

$$\sigma_P^2 = P, \sigma_{B1}^2 = B_1, \sigma_{B2}^2 = B_2 \quad \text{..... (A9)}$$

and

$$\sigma_B^2 = [l_2/(l_1 + l_2)]^2 \sigma_{B1}^2 + [l_1/(l_1 + l_2)]^2 \sigma_{B2}^2 \quad \text{.....(A10)}$$

then σ_N^2 is calculated as follows:

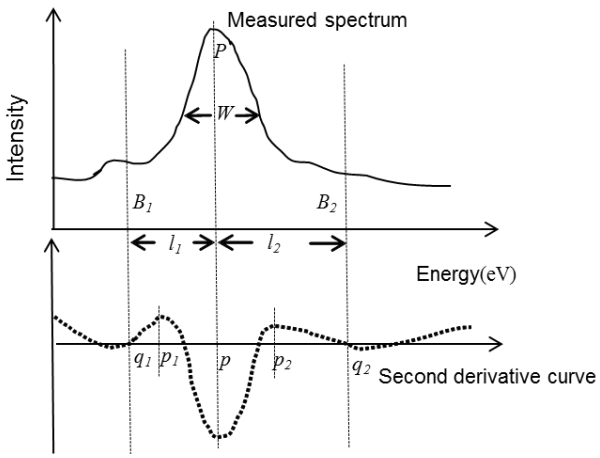

Figure A2. Schematic diagram of a peak position and its background at both sides of the peak.

$$\sigma_N^2 = P + \left(B_1 l_2^{\,2} + B_2 l_1^{\,2}\right) / \left(l_1 + l_2\right)^2 \quad \text{.....(A11)}$$

Therefore, the peak judgment condition is given as follows:

$$N > k\sigma_N \quad \text{.....(A12)}$$

A.3 Peak Detection Using Rough Estimation of Spectrum Background (BGD method)

This method firstly assumes the background curve of a spectrum is generally gentle and the total spectrum region containing peaks is much narrower than the region without peaks, and then it makes a rough estimation of the background intensity for each point of the spectrum.

As the background intensity changes rather gently compared with the intensity near the peak, it can be approximately expanded by using $2m + 1$ moving average points of data which cover the region with several times of the typical full width at half maximum of the peak w. Then the background b_i (i = 1,T) (where T is the total sampling points) can be approximately written by using the given spectrum data y_i (i = 1,T) as follows:

$$b_i = \sum_{j=-m}^{m} h_j y_{i+j} \quad \text{.....(A13)}$$

where h_j is the coefficient of the simple moving average and is expressed as

$$h_j = 1/(2m+1) \quad \text{.....(A14)}$$

The variance of b_i is expressed by using such expansion as follows [1]:

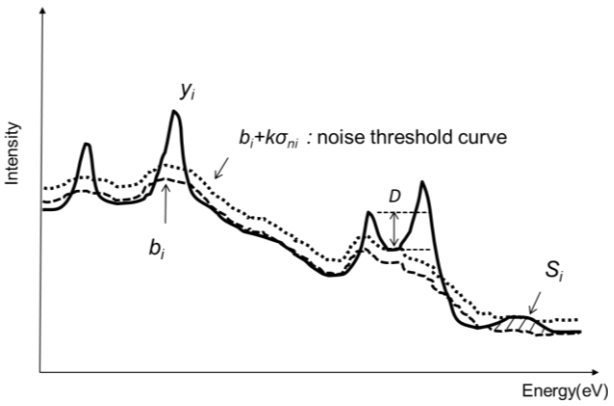

Figure A3. Separation of background intensity from the true peak intensity for the Ar 2p spectrum peak.

$$\sigma_{bi}^{\,2} = \sum_{j=-m}^{m} h_j^{\,2} (Var)_{i+j} + 2 \sum_{j=-m}^{m} \sum_{l \neq j}^{m} h_j h_l (Cov)_{i+j,i+l} \quad \text{...(A15)}$$

where $(Var)_{i+j}$ is the variance of y_{i+j} and $(Cov)_{i+j,\,j+l}$ the covariance of y_{i+j} and y_{j+l}. If randomness of the spectrum data y_i is assumed,

$$(Var)_{i+j} = y_{i+j} \quad \text{.....(A16)}$$

Further, assuming that each y_i is independent and has no correlation with other y_i values,

$$(Cov)_{i+j,\,j+l} = 0$$

Then, the variance of b_i can be approximately expressed as

$$\sigma_{bi}^{\,2} = \sum_{j=-m}^{m} h_j^{\,2} y_{i+j} = \frac{1}{(2m+1)^2} \sum_{j=-m}^{m} y_{i+j} \quad \text{.....(A17)}$$

If the variance of $ni = y_i - b_i$ is defined as $\sigma_{ni}^{\,2}$, it can be estimated as follows:

$$\sigma_{ni}^{\,2} = \sigma_{yi}^{\,2} + \sigma_{bi}^{\,2} = y_i + \frac{1}{(2m+1)^2} \sum_{j=-m}^{m} y_{i+j} \quad \text{.....(A18)}$$

Therefore, the final inequality to judge a peak is given by using the critical value $k \approx 2.0 \sim 3.0$ as

$$n_i > k\sigma_{ni} = k\sqrt{y_i + \frac{1}{(2m+1)^2} \sum_{j=-m}^{m} y_{i+j}} \quad \text{.....(A19)}$$

Or, in a more familiar expression,

$$y_i > b_i + k\sigma_{ni} \quad \text{.....(A20)}$$

The local maximum y_i satisfying the inequality (A20) is regarded as a peak. (See Figure A3.)

Furthermore, in order to apply this procedure more effectively to practical situations, some exceptional cases should be taken into consideration. In one such situation, there is a case that some especially broad peaks in the spectrum do not satisfy inequality (A20). Even in such a case, if a peak with a peak area S_i satisfies the inequality (A21), the peak will be regarded as a real peak.

$$S_i > S_{i0} \quad \text{.....(A21)}$$

where the approximate value of S_{i0} is roughly estimated by assuming it as an ideal triangle peak with the typical full width w at half maximum of the peak and height $k\sigma_{ni}$, then its peak area S_{i0} will be $1/2 \times 2 \times w \times k\sigma_{ni} = kw\sigma_{ni}$:

$$S_{i0} = kw\sigma_{ni} \quad \text{.....(A22)}$$

In fact, if the peak area S_i is greater than the above value, the peak is found to be real in most cases [5]. Sometimes, it happens that the overlapped peaks have plural peaks above the noise threshold curve $k\sigma_{ni}$ with the valleys also above (without crossing) it. In such a case, if

the valley depth D of the local minimum of the secondary differential spectrum exceeds the noise fluctuation $k\sigma_{ni}$ of the point, then it is regarded as the real peak:

$$|D| > k\sigma_{ni} \quad \text{.....(A23)}$$

By adding such exceptional cases, this peak-detection method becomes effective in the actual spectrum processing.

B. Method of superposing noise on a spectrum and the adjustment of spectral intensity

Different levels of noise were added to the spectra used in this simulation experiment based on synthetic spectra b001, b002 and b003. As a scale for adjusting the amount of noise to add, peak intensity of Ar 2p which appears in the spectra was used.

Figure B1 shows the spectrum of an enlarged peak of Ar 2p appearing in spectrum b001. After performing the smoothing processing by cubic spline for the spectrum to determine the background intensity by connecting the both-side skirts of the peak, obtaining the values corresponding to the background intensity I_B and the peak intensity A at the position for giving the maximum intensity of the peak are obtained. Whether the Ar 2p peak is hidden in the noise was made to be one of the guidelines of giving S/N. For peak intensity A, let the variance of the noise corresponding to the background intensity I_B determined for the condition of S/N given at the position of the peak intensity A be $\left(\sqrt{I_B}\right)^2 = I_B$ where the amplitude of the noise becomes $k \cdot I_B = A$ when the scaling factor is set to k. Therefore, since I_B is determined from a given k by using A, S/N based on A can be given by k. Furthermore, the square root (standard deviation) σ of the noise amplitude (variation) is given by $\sqrt{A/k}$, and the value of k can be given by how many times the value of σ^2 is set for the signal intensity A. If k is increased, since $A = k \cdot I_B$, it is found that the value of I_B decreases, degrading S/N.

Noise components to be added were generated by normalized random numbers that follow the standardized normal distribution. To avoid bias in the result due to the inclination of random numbers, five patterns of random numbers were generated independently for each S/N condition, namely for a given value of k to create new random numbers whenever the value of k is changed. The central limit theorem was used to generate the standardized normal random numbers for this noise. Actually, one normalized random number is obtained by the procedure that generates 12 uniform random numbers independently in the generable range of [0,1]; sum them and subtract 6 from the sum [8]. The figure "n" of b003_7n (the name of the synthetic spectrum) represents a number for the respective identifications of 5 patterns.

For the spectra employed in this study, conditions assuming the amounts of noise as $A = 2\sigma^2$, $A = 3\sigma^2$, $A = 4\sigma^2$ (degree of noise such that Ar 2p is almost buried in the noise), and $A = 7\sigma^2$ (slightly poorer S/N than the original data) were adopted.

On the other hand, when changing the amount of the noise to superimpose, it is also necessary to reduce the intensity of the overall spectrum corresponding to the amount of the noise. This is because, as shown in Figure B1, although the noise amplitude at a position where A is determined should be originally determined by I, I_B is determined by only k regardless of I, I_B/I must be multiplied over the entire spectrum excluding the noise. If the amount of noise to superimpose on the spectrum b001 corresponds to $A = 2\sigma^2$, the amount becomes 901 counts; however, the adjustment of only the noise component is carried out by generating the normalized normal random numbers corresponding to the difference between the noise determined by this procedure and the originally existing noise, and then carried out by superimposing them on the overall noise component

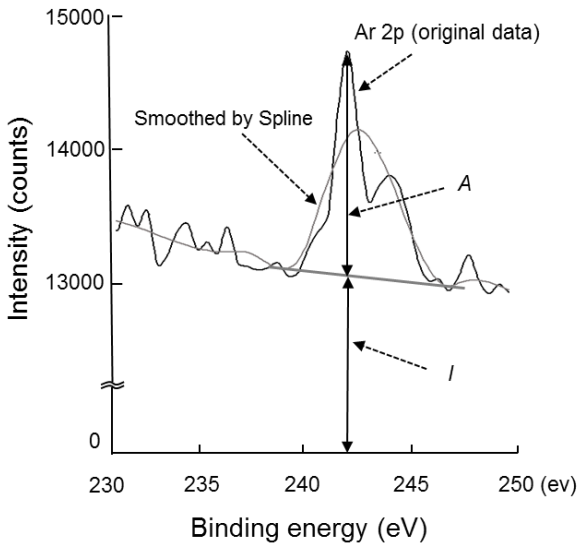

Figure B1. Separation of background intensity from the true peak intensity for the Ar 2p spectrum peak.

which was separated from the spectrum.

In the case of spectrum b001, the average background intensity becomes 38,313 counts when calculating it excluding the intensities around the peak positions, so if the amount of noise to add is defined at this time as S/N: $1/R$ = I/A, it becomes $1/R$ = 38,313/901 = 43, and therefore, the scaling factor representing the entire spectral intensity is R = 1/43. Multiplying this R value by the signal component of the spectrum excluding the noise makes the spectrum have a signal intensity appropriate to the S/N. Finally, adding the adjusted signal component to the noise component creates the synthesized spectrum.

Although spectra reduced to a scaling factor R = 1/43 (A = $2\sigma^2$) correspond to b001_21 to b001_25, the normalized normal random numbers corresponding to the respective noise differences are generated independently from each other.

Furthermore, similarly, for the spectra of b001_73 corresponding to the case of A = $7\sigma^2$, for example, since R = 1/3.4, the entire spectral intensity is adjusted to 1/3.4.

Similarly, in the case of spectrum b003, the average background intensity excluding the intensities around peak positions is found to be 12,253 counts. If the amount of noise to add is set to A = $2\sigma^2$, its amount becomes 57 counts. Therefore since $1/R$ = 12,253/57 = 215, the scaling factor that represents the entire spectrum intensity becomes R = 1/215. The spectrum corresponding to this is, for example, b003_23. Similarly, as the spectrum b003_73 has a scaling factor of R = 1/18, the entire spectral intensity is reduced to 1/18.

Acknowledgments

Spectra used in this study are those created by Dr. Sei Fukushima, chief researcher of NIMS; the authors are grateful to him for providing them for this study. The description of the superposing method of noise on the original spectrum which was presented in the Appendix is provided by Dr. Fukushima himself; all of the collaborators in the preparation of this paper would like to thank him for his kind cooperation.

査読コメント，質疑応答

査読者 1．吉原一紘（シエンタ オミクロン）

[査読者 1-1]

[7.1 Minimum area parameter] の節の中には minimum area parameter の定義が明示されていません．S_{i0}を示すのではないかと思われますが，定義が記述されてないので分かりにくくなっています．理解を助けるために定義を明示していただけないでしょうか.

[著者]

この研究報告は引用文献，[4] Y. Furukawa, Y. Nagatsuka, Y. Nagasawa, S. Fukushima, M. Yoshitake, A. Tanaka, *J. Surf. Anal.* ***14****, 225 (2008)*で報告しましたピーク検出アルゴリズムの性能評価を行っております．その中で提案しております 3 件のアルゴリズムを再度，本研究報告の本文に記述することは本報告の目的とは異なりますので，読者の利便性を考慮し，Appendix A-1、A-2、A-3 に掲載しております.

minimum area parameter の定義は Appendix,A.3 Peak Detection Using Rough Estimation of Spectrum Background（BGD method）で定義しております．本文にも注意喚起の目的で 2. Algorithms Used for Detecting Peaks の冒頭のイントロ文の最後に詳細は Appendix を参照する様に促しております.

ただ，アルゴリズムの概要を記述している 2. Algorithms Used for Detecting Peaks 2.3 節の中に the peak area とありますが，パラメーターであることを明示するために，S_i，を加筆致しました.

[査読者 1-2]

Figure 18 の縦軸の[Number of peaks detected]の定義が明示されていません．文章からは，visual detection method の場合は，mean score が 1.0 以上のピーク数（matched peak + missed peak）であり，automatic detection method を用いて求めた場合は，ノイズと判定されたピークを含めて，検出された全てのピーク数（matched peak + noise peak）としているように，推定できます.

[著者]

visual detection method の場合は，mean score が 1.0 以上のピーク数(matched peak)であり，automatic detection method を用いて求めた場合は，ノイズと

判定されたピークを含めて，検出された全てのピーク数（matched peak + noise peak）としています.

このことを Figure 18 の説明文に加筆しました.

[査読者 1-3]

この推定に基づきますと，visual detection method と automatic detection method の縦軸の値の差は missed peak の数と noise peak の数の差となります.

[著者]

missed peak は関与してきませんので automatic detection method によるピーク検出本数は noise peak の数だけ多くなります.

[査読者 1-4]

この差の大小は，必ずしも[peak detection ability]を示さないと思います．Page 19 の第３パラグラフには「As far as the total score, the BGD method was found to be the best among three methods.」と書かれていますように，最善の評価法は visual detection method で検出されたピーク数と一致した検出数を与える方法（すなわち，missed peak score と noise peak score の双方が[0]となる方法）であると思われます．したがって，Figure 18 からは「All three of these methods can attain mostly successful peak detection results.」との結論を導くことには無理があるように思えます．誤解を避けるために，[Number of peaks detected]の定義を明示し，automatic detection method の優劣の評価基準が分かるように記述していただけないでしょうか.

[著者]

我々が考えるピーク検出性能の善し悪しの基準は以下の通りです．ピークの数え落としは最悪です．一方，人間が見落とすかもしれないピーク，それは automatic detection method では noise として処理しておりますが，実は noise ピークも真のピークである可能性があります．最終的には手動でデータ集と比較しながら noise ピークがピークの候補であるか否かを検討します．その結果，実ピークであった場合は見落としピークの数が目視検出において減ることになります．これはピーク検出作業において良い結果に繋がります．従って，目視による検出本数より少し多めが好ましいと考えます．このことは Table 5 の結果説明のところにも述べております．従いまして，我々の自動ピーク検出方法による結果の判断基準を Table 6 の下に以下の様に記述しております.

“An algorithm capable of detecting real peaks without error along with a moderate number of noise peaks as candidate peaks may be an excellent peak detection method for the safety of avoiding a failure in detecting real peaks.”

この主張を強調するために，本ブロックを少し前の位置，p19 Figure18 shows….の文の次に移動しました.

ご指摘戴きました以下の点，「最善の評価法は visual detection method で検出されたピーク数と一致した検出数を与える方法（すなわち，missed peak score と noise peak score の双方が[0]となる方法）であると思われます．」に関しまして，筆者の経験では noise peak が全くない場合，上記の理由により，自動検出結果に不安を覚えます.

また，ご指摘頂いておりますが，[Number of peaks detected]の定義は明示すべきと考えますので加筆修正致します．本文の Figure18 の説明を行う箇所に加筆いたしました.

査読者 2．田沼繁夫（物質・材料研究機構）

[査読者 2-1]

"digitized synthetic XPS spectra with different levels of noise"を元にして，議論を進めています．したがって，ここで得られた結論（最適なパラメータの組み合わせ，最適アルゴリズムの選択）は使用したデータに依存しないのでしょうか？

それとも一般化できるのでしょうか? 若しくは，どのように一般化できますか？

[著者]

本論文で使用したピーク検出方法は，広く普及していると思われる代表的な２つの方法であり，それらは二次微分法と移動平均法です．さらに二次微分法では，ピークと認識する条件を候補ピーク前後のバックグランドの大きさを決めてからピーク判定を行う方法に改良し，“Peak and Background Relations at a Candidate Peak (PB method)”として新たに加えています．この方法ではノイズピーク数が減ります．また移動平均方法には，ピーク幅が広くピーク高さが低い微小ピークを拾えるように“ピーク面積”パラメータ“*s*”を，そして複数ピークの重なり合い

により谷が生成するピーク群のピークを拾える“谷深さ”パラメータ“*D*”をそれぞれ導入し，検出機能を向上させています．以上，計 3 種のピーク検出方法の性能評価を行いました．

ピーク検出方法におけるパラメータの最適値は，総てではありませんが，適用するスペクトルの質により変わります．今回用いましたスペクトルの energy step 幅は 0.5 eV ですが，これより小さい step 幅で測定されたスペクトルでは測定点の数が多くなりますから，二次微分法の微分点数や移動平均法における移動平均点数は増加させる方向になります．

一方，得られた分析データが厳密にポアソン分布に従っており，それ以外の物理的・化学的・機器的な要因を無視できれば，判断するための閾値，つまりバックグランドの標準偏差の何倍とするかを決める *k* 値は影響を受けませんし，また移動平均方法における“ピーク面積，*s*”，“谷深さ，*D*”も影響を受けないでしょう．元素の定性分析はルーチン的に行われるものですから，何度か行ううちに様々なスペクトルに対応するパラメータの最適な値は経験値として決まってくると思います．

この論文を読まれた分析者は，最良のピーク検出結果を得るためには，最適なパラメータの値を選択する必要があることに気付いて欲しいと思います．ブラックボックス型のソフトウエアを使用せざるを得ないのであれば，メーカーに選択可能なパラメータの種類の開示とパラメータ値の自由な選択を可能ならしめる様に要望することも良いと思います．

[査読者 2-2]

同じような質問で恐縮です．

ここでは Au-Ag-Cu の合成スペクトルを使い，それらの量の比を変え，さらにノイズを付加しています．したがって，微小ピークの検出のチェックには適していると思います．しかし，ピークの重なりやバックグランドの変化の大きい上にピークがある場合などについては考慮されていません．これらについても同じ結論になるのでしょうか．

[著者]

本論文の 7 章，7 Performance analysis of the BGD method の 7.2 Performance analysis において，Figure15(b)で例示しながら BGD 法では移動平均点数が大きすぎると大きなピークの裾にある微小ピークの検出は難しくなることを説明しています．ピークの重なりについては，重なりがある程度離れていれば，二次微分等の処理で分離できる場合もありますが，重なりが近接している場合には分離できないこともあります．これはデジタルデータ処理の関係上，人間の目には確認できても計算結果に数値として明らかに反映されないためで，やむを得ない結果です．同様にバックグラウンドの変化が激しい場合も，二次微分した結果によって，バックグラウンドの変化がピークの変化と比べて無視できれば，正しい結果を与えますが，両者が同程度の場合には，ピークと見誤る可能性もあります．様々なパラメータ値を選択できるピーク検出ソフトウエアを使用しながら，これらのことに留意し，ピーク検出結果と測定されたスペクトルを見比べ，見落としが無いかチェックする姿勢が大切であると思います．

ピーク検出あるいは元素の同定作業をソフトウエアにより自動的に行うのは分析者の手間を軽減するためでありますが，結果を 100%信用するのは危険です．注意深い分析者は最終的にはピーク検出結果と測定されたスペクトルを見比べ，見落としが無いかチェックすると思います．またソフトウエアに使用されるパラメータ値を適切に選択する必要があるもう一つの理由ですが，検出された結果にノイズピークの可能性があっても，実際存在するよりも少し多めにピーク検出結果を出力できる条件に設定することが無難です．ピーク検出結果に見落としがある場合，あるはずの元素を無いと判断してしまう可能性が出てきます．従って，分析者は見落としが無いかチェックする際に，疑わしきはピークの可能性があると考え，ノイズピークであってもピークの可能性を科学的な知識をもとに精査する姿勢が大切です．