Taylor & Francis
Taylor & Francis Group

Science and Technology of Advanced Materials: Methods

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Automated analysis of photoelectron yield spectroscopy spectra interpreted via power laws

Shinjiro Yagyu, Michiko Yoshitake, Takahiro Nagata, Takeshi Yasuda, Takatsugu Wakahara, Yubin Liu & Yoshiyuki Nakajima

To cite this article: Shinjiro Yagyu, Michiko Yoshitake, Takahiro Nagata, Takeshi Yasuda, Takatsugu Wakahara, Yubin Liu & Yoshiyuki Nakajima (2025) Automated analysis of photoelectron yield spectroscopy spectra interpreted via power laws, Science and Technology of Advanced Materials: Methods, 5:1, 2465257, DOI: 10.1080/27660400.2025.2465257

To link to this article: https://doi.org/10.1080/27660400.2025.2465257

Published online: 12 Mar 2025.

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS: METHODS
2025, VOL. 5, NO. 1, 2465257
https://doi.org/10.1080/27660400.2025.2465257

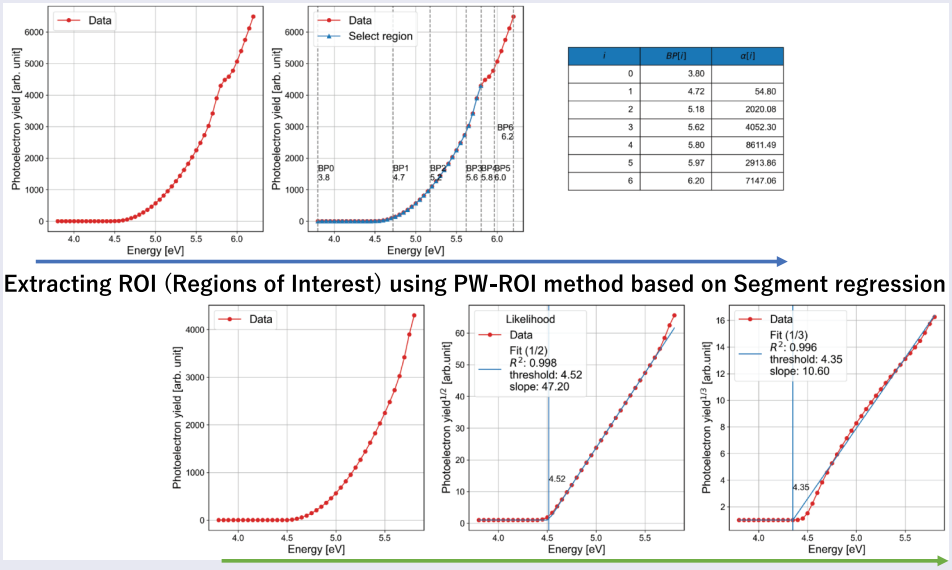
Automated analysis of photoelectron yield spectroscopy spectra interpreted via power laws

Shinjiro Yagyu[a], Michiko Yoshitake[a], Takahiro Nagata[a], Takeshi Yasuda[b], Takatsugu Wakahara[b], Yubin Liu[c] and Yoshiyuki Nakajima[c]

[a]Functional Materials Field, Research Center for Electronic and Optical Materials, National Institute for Materials Science, Tsukuba, Ibaraki, Japan; [b]Macromolecules Field, Research Center for Macromolecules and Biomaterials, National Institute for Materials Science, Tsukuba, Ibaraki, Japan; [c]Riken Keiki Co., Ltd, Itabashi-ku, Tokyo, Japan

ABSTRACT

We propose an automated method for threshold analysis in photoelectron yield spectroscopy (PYS), a technique used for measuring the ionization energy of inorganic and organic semiconductors. PYS spectra can be interpreted using the power law. First, the analysis region, the region of interest (ROI), within the measured spectrum must be identified and isolated for power law interpretation. Segment regression is employed to estimate the change point of the spectrum and slope on either side. The ROI is extracted using an algorithm that accounts for the slope variation (PW-ROI method). The inverse of the power number (n) derived from the physical model is then applied to the spectrum, and fitting is performed using the mean absolute error as the minimization function with the extended rectified linear unit function ($1/n$ plot method). In cases with multiple candidate models, the optimal model is chosen based on the coefficient of determination using the $1/n$ plot method. This automated analysis method significantly reduces the processing time and ensures that the values are objective and human independent. Results of automated analysis closely align with those of manual analysis, with their differences limited within the range of ± 0.1 eV.



i	$BP(i)$	$\alpha(i)$
0	3.80	
1	4.72	54.80
2	5.18	2020.08
3	5.62	4052.30
4	5.80	8611.49
5	5.97	2913.86
6	6.20	7147.06

Automatic analysis of spectrum interpreted via power laws

IMPACT STATEMENT

This study proposes an automated method for analyzing photoelectron yield spectra, reducing the analysis time by 90% and providing values that are objective and human independent.

ARTICLE HISTORY

Received 14 November 2024
Revised 3 February 2025
Accepted 7 February 2025

KEYWORDS

Material informatics; automated analysis; power law; photoelectron yield spectroscopy

CONTACT Shinjiro Yagyu YAGYU.Shinjiro@nims.go.jp Functional Materials Field, Research Center for Electronic and Optical Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan

1. Introduction

Recently, advances in materials informatics have driven the need for extensive data. Additionally, advances in robotics have led to the automation of experiments [1,2] and the generation of numerous measurement spectra. Consequently, methods for automatically analyzing and processing extensive amounts of spectral data are desired. Although automated analysis is applied to some types of spectral data, such as photoelectron spectroscopy data [3,4], others still rely on manual interpretation. The manual approach is time-consuming and subjected to variability among analysts, making quantitative evaluation difficult. We performed an automated photoelectron yield spectroscopy (PYS) analysis to determine the ionization energy values of semiconductor materials [5]. PYS spectra are interpreted using a power law. Similar power law spectra include voltage–current (V – I) spectra, which determine the critical current density of a superconducting material and its nonlinear performance index (n value) and also transmission and reflection UV spectra, which estimate its band gap. The characteristic parameters in such spectra interpreted using the power law are the power number and threshold at which the output value begins to increase rapidly. To estimate these parameters, analyses of PYS [6–8] and UV spectra [9] are performed with one fixed power law (given as prior information) using a theoretical model that considers various properties of the material under study, with the threshold treated as a critical physical property. However, for some newly created materials, the appropriate power is uncertain because their properties are unknown. From a model application perspective, the most linearized power number is considered as the correct solution. Additionally, analysis regions in real measurement spectra must be identified and isolated visually by analysts, posing challenges for automation. In this study, we introduce the $1/n$ plot method [10] as an automated approach for spectral analysis using a single power law. To identify the analysis region, we describe a method for capturing spectral change points (PW method), and an algorithm is proposed for automatically segmenting the analysis region (PW-ROI method) [11]. By integrating these methods, we present examples comparing automated analysis with manual analysis performed by analysts on PYS data, where the properties of the measured materials are well defined. This integrated approach significantly reduces processing time and enhances the reproducibility and objectivity of the analysis

2. Current PYS analysis method (manual method)

The PYS spectrum, used to determine the ionization potential of semiconductor materials, generally rises following a power law as the photon energy (input value) irradiating the sample rises. This relationship can be expressed by a power law, as shown in Equation 1, where the output value (y) denotes the power law spectrum for the input value (x).

$$y\begin{cases} bg & (x-th<0) \\ a\cdot(x-th)^n+bg & (x-th\geq 0) \end{cases} \quad (1)$$

where a is a normalized factor (a is a constant containing the light intensity and the absorption efficiencies for the production of excited electrons); th is the threshold; n is the power number; bg is the background; y is the photoelectron yield; x is the photon energy. The manual analysis procedure [6,8,12] to determine the threshold value of the obtained PYS spectrum is shown in Figure 1.

(1) *Power number selection (physical model selection)*;
According to the power law, the PYS spectrum is assumed to rise with energy. The physical model determines the power number, usually squared or cubed.
(2) *Graph plotting according to the selected model (1/n plot)*;
To linearize the PYS spectrum, the inverse of the selected power number is applied, generating a graph aligned with the chosen model.
(3) *Identifying background and linear slope regions*;
Those regions from the graph where the background is constant and the slope straight are selected. A line representing the slope is drawn using either the least-squares method or a visual estimate.
(4) *Determining the intersection of two lines*;
The intersection point determined in step (3) is the threshold value. The threshold may be overestimated in the presence of background as will be explained later. Manual analysis can incorporate human experience. Although squares are commonly used, cubes may be avoided because of insufficient comparative measures. When multiple slope lines are feasible, there is no clear guideline for their selection. One approach is to select the line closest to the threshold value. However, for amorphous, nanostructured materials or thin-film samples, the spectral analysis should consider the effects of the Urbach tail [13], as seen in optical absorption. Further, it is necessary to consider that the chosen power number may differ from the assumed model (interpreted with a single power number).

3. 1/*n* plot method

The spectrum is linearized to the power of $1/n$ to define the analysis regions of the background and linear slope. This process involves obtaining the intersection point of two straight lines via manual analysis.

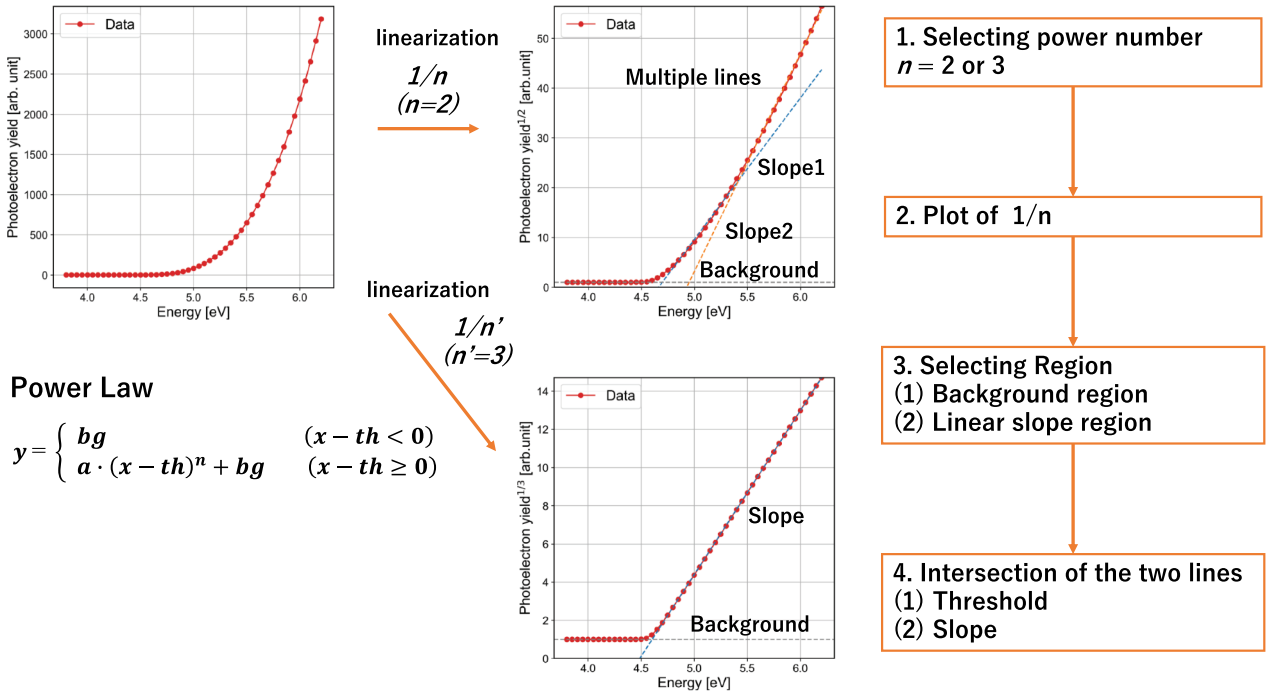


Figure 1. Manual analysis procedure. The manual analysis procedure for PYS spectra is analyzed using the power law model.

The function represented by these two lines extends the rectified linear unit (*ReLU*) function [14] commonly used in deep-learning activation functions.

$$exReLU\begin{cases} bg & (x - th < 0) \\ a'x + bg & (x - th \geq 0) \end{cases} \quad (2)$$

The extended *ReLU* (*exReLU*) function increases with a constant background value (*bg*) up to a threshold (*th*); beyond the threshold, it progresses with slope (*a'*), introducing a discontinuity in the derivative at this threshold.

In the 1/*n* plot method, (1) multiple analysis spectra are generated by applying the inverse of candidate multipliers (e.g. 1/2 and 1/3) to the spectra. (2) Each analysis spectrum is fitted using the *exReLU* function, with the absolute error used as the minimization evaluation function. (3) Because a coefficient of determination closer to 1 suggests a correspondingly better linear fit, the power number that achieves the optimal coefficient of determination and its corresponding fitting value are selected.

The error distribution is Laplacian rather than Gaussian in the mean absolute error method (*MAE*) and is robust against outliers; therefore, the *MAE* method is used to perform robust fitting, including the area around the threshold value, which deviates from the stepwise variation of the *exReLU* function due to various factors. Using the power law equation (Equation 1) over *x* values ranging 3.8–6.2 (the energy range commonly used in PYS measurements) and parameter values (*a,th, n,bg*) = (700,4.5,2,1), we calculate the spectra using the inverse power numbers of 1/2 and 1/3, with 1/1 being utilized as an original spectrum. The resulting fits are shown in Figure 2. The PYS calculated spectrum (a-1), spectrum calculated using a power of 1/2 (henceforth, 1/2 power) (a-2), and spectrum calculated using a power of 1/3 (henceforth, 1/3 power) (a-3) are shown. In Figure (a-2), the inverse of the employed power yields a linear power-function component, resulting in the best coefficient of determination (R^2). However, in the high-energy region (over 6 eV), there are discrepancies between the fitting line and data: the data points for the 1/1 power are above the fitting line, while those for 1/3 power are below the fitting line. The degree and direction of these deviations enable the qualitative determination of the appropriateness of the power value. A comparison of the coefficients of determination across fits allows us to evaluate the optimal power parameter; the linearity increases as *n* decreases, suggesting a more significant threshold. Interestingly, these results do not agree with the threshold value of 4.5 set in the initial calculations, likely due to background effects [10]. The fitted parameters help identify the energy axis intersection where the background is zero, offering a threshold estimate. To enhance the accuracy, however, the background components must be subtracted, as shown in equation (3), before applying the 1/*n* plot method.

$$y - bg = a \cdot (x - th)^n \quad (3)$$

Depending on the analysis policy, the background is included in many PYS threshold analyses. For example, in the PYS spectra measured with Riken Keiki's AC device [15] (Fc sample data shown in Figure 5(d)), when the maximum intensity is 500 and the background is approximately 2, resulting in a threshold

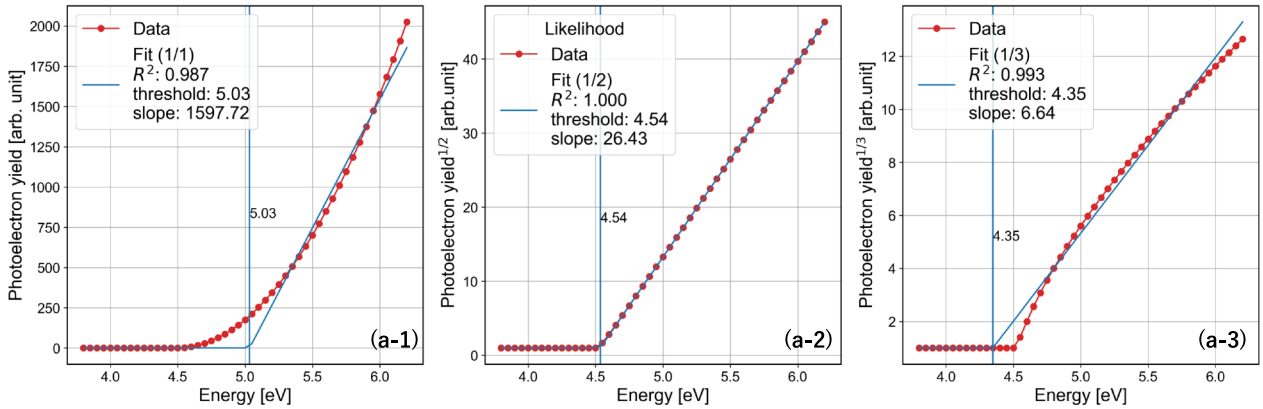


Figure 2. Results of 1/*n* plot analysis on a PYS spectrum. The spectrum is computed using parameters (*a,th,n,bg*) = (800,4.5,2,1) in Equation (1) over the energy range 3.8–6.2. The inverse of each power number, i.e. 1/2 and 1/3, is applied separately to this spectrum. The fitting and its quality (i.e. coefficient of determination (R^2)) are evaluated using the *exReLU* function. Results are shown for the (a-1) original spectrum ((1/1) power), (a-2) 1/2 power, and (a-3) 1/3 power.

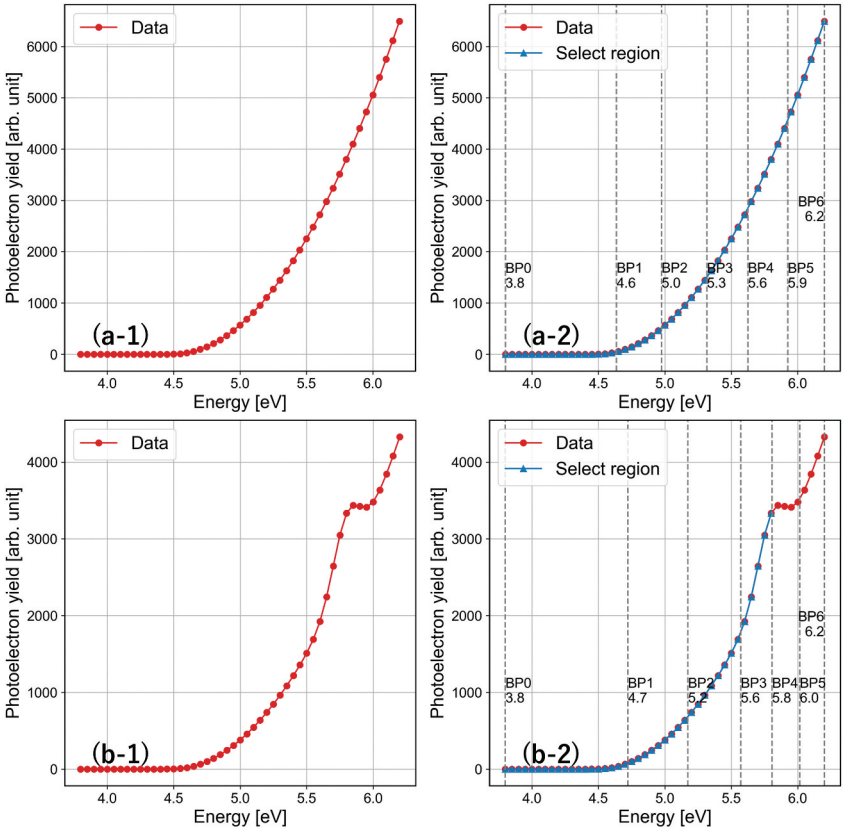
about 0.04 eV larger than when pre-processed with Equation (3).

4. PW method

When using the 1/*n* plot method, it is often essential to identify and isolate the region of interest (ROI) for analysis. To achieve this automatically requires an understanding of the overall spectrum characteristics. In general, there are several methods for characterizing spectra, including differentiation, statistics, and Fourier transformation. By obtaining the derivative (slope) of spectra, changes and characteristics of spectra values can be captured. However, due to noise in the actual spectra, grasping the characteristics of the spectrum using numerical differentiation is difficult (the difference between two points in the calculation). To overcome this problem, smoothing techniques, such as moving averages, are typically employed to reduce noise before numerical differentiation, though selecting the appropriate degree of smoothing can be challenging. Segment regression is a method used for capturing the characteristics of noisy data by modeling them with multiple regression lines. This method enables us to estimate change points and slopes, even with noisy spectra, before and after each change. The spectrum can be captured, and ROIs can be extracted by evaluating change points and slopes.

The segment regression algorithm operates as follows: Taking the number of segments (i.e. number of change points) as 1, the initial segment positions are determined randomly or from given initial values. Then, linear regression is performed on each side of this segment to calculate the residuals (i.e. the difference between the data and the regression line). The process is repeated with different segment positions, identifying the configuration with the smallest residual. Generally, increasing the number of segments improves the fit (i.e. reduces the residuals), but selecting the optimal number of segments requires expert judgment or, preferably, an information criterion-based approach. In such an approach, the number of segments that minimize an information criterion is regarded optimal. The segment regression is implemented using the open-source software R [16] and Python [17,18]. In this analysis, the optimal number of segments and segment positions (*BP*) and the slope between segments (α) evaluated by the Bayesian information criterion (BIC) were obtained using a publicly available Python module [17] (Piecewise regression, which also evaluates the number of segments and fit using the BIC, which we call the PW method after the name of the module). Automated ROI extraction requires an index and algorithm to evaluate and judge the obtained *BP* and α values.

Simulations of typical spectra seen in PYS measurements and the results of applying the PW method are shown in Figure 3. The figure shows a power spectrum with a constant background ($n = 2$ in the calculation; a normal spectrum) (a-1), the analysis region where the PW method is applied to estimate BP, the decision algorithm used (a-2), and a table summarizing the segment positions (*BP*) and slopes (α) when PW is applied (a-3). A spectrum whose intensity first rises with the power, then decreases, and increases again (the peak spectrum) (b-1), the analysis region estimated by applying the PW method (b-2), and a list of *BP* and α (b-3) are also shown. In both spectra, the intervals of *BP*s are narrower in the region where the spectrum rises with energy. The PYS analysis should extract regions with the spectral shape shown in (a-1). In the peak spectrum shown in (b-1), the ROI is the region before *BP4* (see chapter 5).



i	*BP*[*i*]	*α*[*i*]
0	3.80	
1	4.64	15.02
2	4.98	1346.74
3	5.32	2917.90
4	5.62	4376.84
5	5.92	5723.57
6	6.20	7070.29

(a-3)

i	*BP*[*i*]	*α*[*i*]
0	3.80	
1	4.72	36.54
2	5.17	1346.72
3	5.57	2689.09
4	5.81	7250.74
5	6.01	227.24
6	6.20	4639.60

(b-3)

Figure 3. Simulated PYS spectra and application of the PW method. The typical spectra observed in PYS measurements are simulated using the results obtained from the PW method. (a-1) Power spectrum with a constant background (power set to 2, referred to as the normal spectrum). (a-2) Analysis area where the PW method was applied to estimate BP and the decision algorithm, and (a-3) a table summarizing the classification position (*BP*) and slope (*α*) when applying the PW method. (b-1) the spectrum shows an initial increase with energy, followed by a decrease, and then another increase (the peak spectrum). (b-2) Analytical domain estimated using the PW method. (b-3) List of *BP* and (*α*) values.

5. Decision indicators and algorithms (PW-ROI method)

The following algorithms can automatically identify the spectral regions that follow the power law as ROIs in the spectrum shown in Figure 3.

(1) *Application of the PW method to spectrum*;
The PW method is applied to a spectrum to determine the optimal number of segments, positions (*BP*), slopes (*α*) (where the slope array α[i] represents each segment for *n* segments; *i* varies from 1 to $n+1$), and the segment position *BP*[*j*] (j varies from 0 to $n+1$; *BP*[0] and *BP*[$n+1$] are the minimum and maximum *x* values, respectively).

(2) *Peak or shoulder removal*;
Because slopes near peaks and shoulders are smaller than the previous slopes, gradients are compared from the beginning of the slope array and *BP*[i] and onward, where $\alpha[i+1] < \alpha[i]$ are identified as peak regions and removed.

In Table b-3 of Figure 3, for instance, the value of $\alpha[5]$ is smaller than that of $\alpha[4]$; hence, values up to $\alpha[4]$ are adopted. *BP*[4] is used as ROI, and the rest is removed. In Figures 3(b–2), the decision algorithm is employed, and ROI is indicated by the plot symbol ▲. Thus, ROI can be obtained using this decision algorithm. Furthermore, if the spectrum assumes a shape other than that shown in Figure 3, an automated decision algorithm that uses slope information must be added.

6. Automated analysis of PYS spectra across various materials and methods

We performed an automated analysis on actual PYS spectra that were previously analyzed manually [19–21]. We compared our results with those of manual analysis. The power and threshold values were estimated from the coefficient of determination by assuming $n = 2$ and 3 during automated analysis. Additionally, manual spectrum analysis requires about 30 s to 1 min, whereas automated analysis takes only a few seconds (CPU time: 1/*n* plot analysis: 100–200 ms, PW-ROI +1/*n* plot analysis: 2.5–3 s in Python Jupyter Notebook), thus reducing the analysis time to about 1/10 of the time required for manual analysis.

6.1. Cases excluding PW-ROI preprocessing

The following examples illustrate instances where preprocessing using the PW-ROI method is unnecessary. For instance, Figure 4 shows the analysis results for a rolled plate of metal gold (Au) and

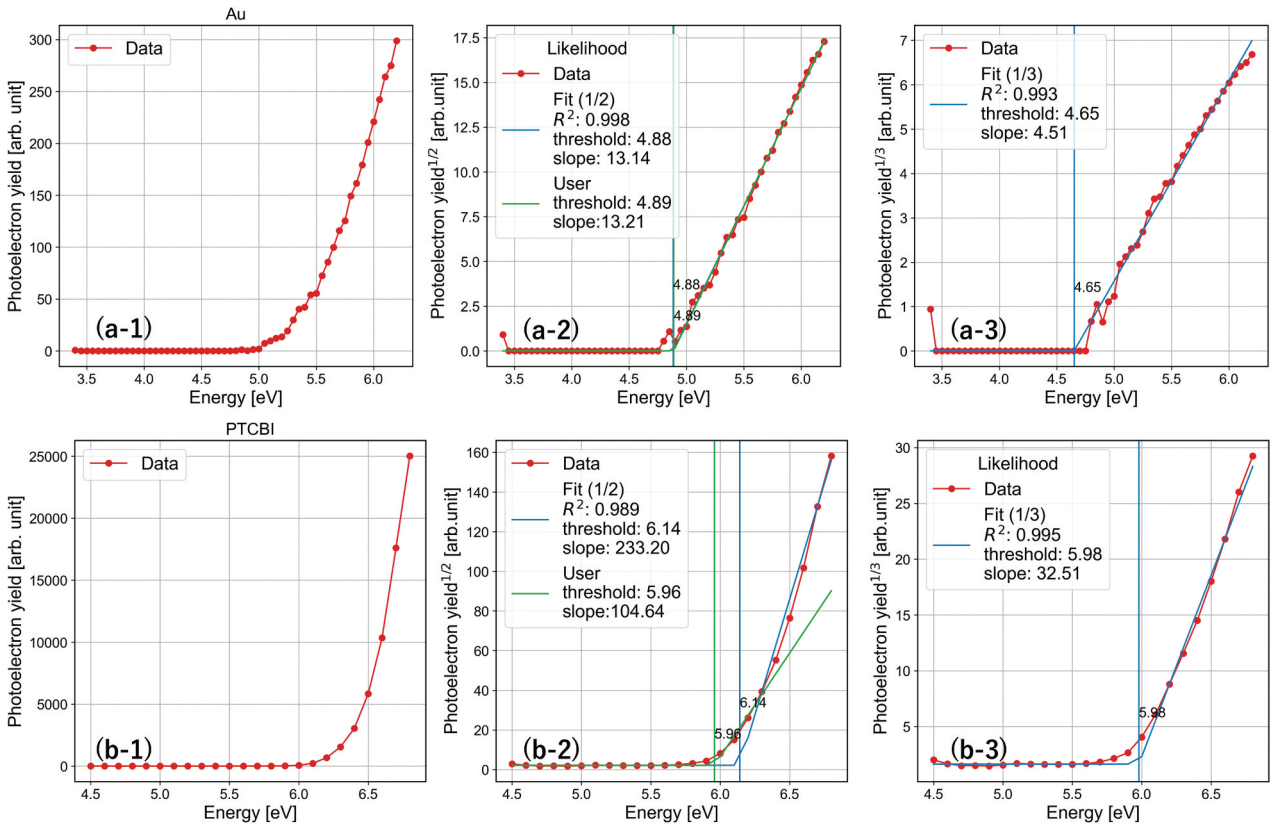


Figure 4. PYS spectra of a rolled Au plate and PTCBI. PYS spectra of (a) a rolled Au plate, a metallic material, and (b) PTCBI, an organic semiconductor, with results from automated analysis using 1/2 and 1/3 power. Manual analysis is performed on the spectrum with 1/2 power, and results are also shown on the graph for the 1/2 power.

3,4,9,10-perylenetetracarboxylic bisbenzimidazole (PTCBI), an organic semiconductor. Both spectra were obtained through the manual analysis of the squared spectra, demonstrating a clean, straight-line pattern for Au (a-2). Here, manual analysis yields almost no fluctuations, and automated analysis selects the 1/2 power spectrum (a-2), which agrees closely with the manual results. Conversely, the 1/3 power spectrum (a-3) falls below the measured data on the high-energy side of the fitting line, consistent with the trend shown in Figure 2. In the manual analysis of PTCBI, the analytical line extends over the linear region near the threshold value. However, the 1/2 power spectrum (b-2) deviates from linearity at high-energy levels. This trend is common when small power numbers are employed in simulations, as shown in Figure 2. Conversely, the 1/3 power spectrum (b-3) generated through automated analysis exhibits more linearity, with the fitting results closely matching the manual analysis results. An expert analyst conducted manual analysis and likely recognizing that the 1/2 power spectrum was nonlinear, drew the line by selecting a linear region as close to the threshold as possible. This approach can be considered the same as unconsciously assuming 1/3 power and drawing a line.

6.2. *Cases requiring PW-ROI preprocessing*

ROI processing is essential in the spectral analysis of organic semiconductor materials using ferrocene (Fe$(C_5H_5)_2$). Figure 5 presents the results of the 1/*n* plot method performed after the PW-ROI processing of a spectrum. Graph (a) contains regions where the intensity increases and then remains constant or slightly decreases. The PW-ROI method is used to identify regions where the slope begins to decrease, and these regions are excluded from the ROI. The threshold is derived from the coefficient of determination by applying 1/2 and 1/3 power functions. Graph (b) shows the change points detected by the PW-Plot method. In contrast, graphs (c) and (d) display the results obtained after trimming the region to the transition point of the slope. Graphs (e) and (f) illustrate the application of 1/2 and 1/3 power fitting on the selected region. The 1/2 power number was chosen based on the results of the coefficient of determination. The outcomes closely align with those from the manual analysis.

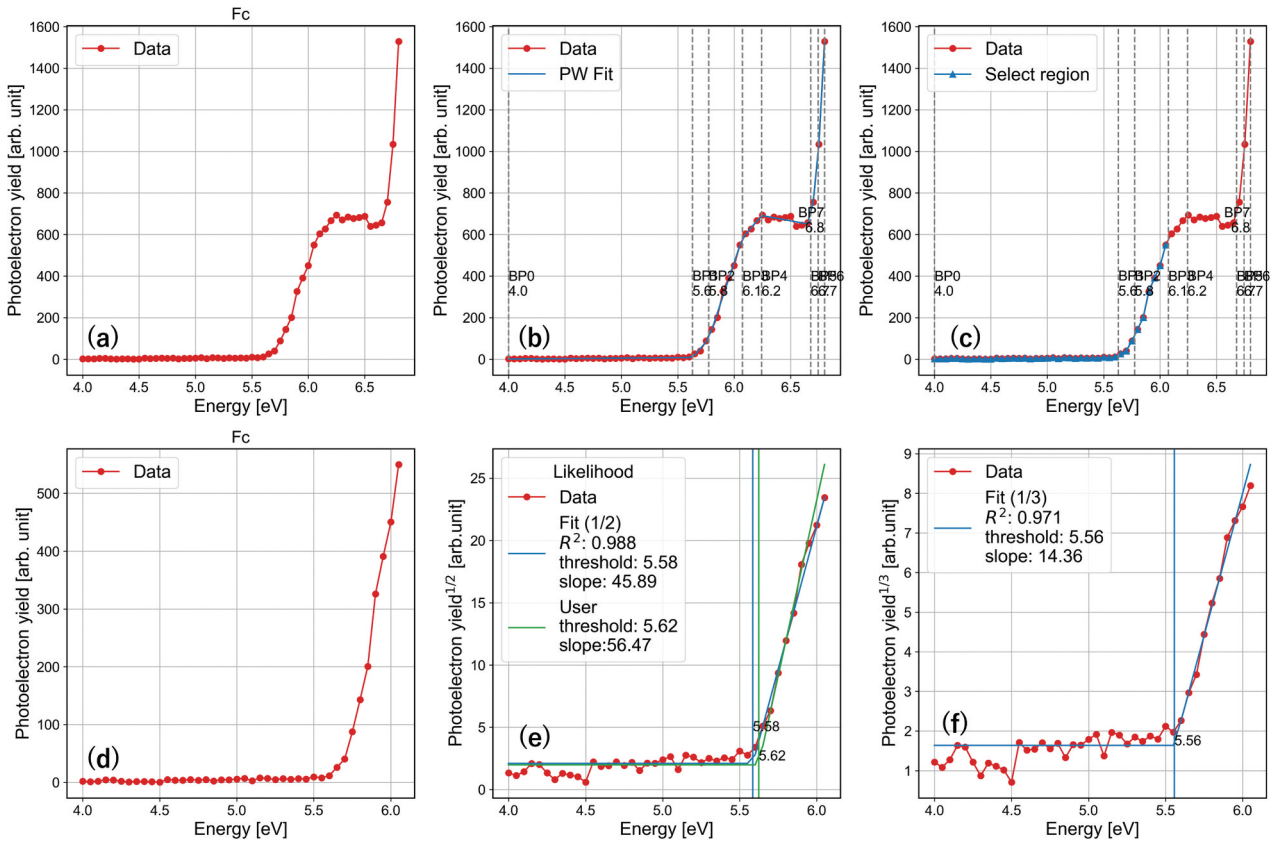


Figure 5. Application of PW-ROI and 1/*n* plot methods. Results for the PW-ROI and 1/*n* plot methods applied to the PYS spectra of FC (Ferrocene: $Fe(C_5H_5)_2$), an organic semiconductor material. (a) Original PYS spectra; (b) PW method used to detect change points; (c) and (d) indicate the analysis target area reduced to a blunted region by comparing slopes; (e) and (f) show the fitting results obtained by applying 1/2 and 1/3 power to the segmented region.

6.3. Comparison of manual and automated analyses

Figure 6 compares the results of automated analysis and manual analysis, in which the PW-ROI and 1/*n* plot methods were applied to our manually analyzed data [19–21] (complete spectral data, consisting of 48 measurements and including multiple measurements on the same sample, are provided in the supplementary materials). The diagonal lines in the graph indicate consistency between the automated and manual analyses, and the lines above and below the diagonal lines have a range of ±0.1 eV. Of the 48 spectra, four data points lie outside the ±0.1 eV range: (a) MEH-PPV (Poly[2-methoxy-5-(2'-ethylhexyloxy)-1,4-phenylene vinylene]), (b) BCFN (*N*-[[1,1'–Biphenyl]-4-yl]-9,9-dimethyl-N-[4-(9-phenyl-9 h-carbazol-3-yl)phenyl]-9 h-fluoren-2-amine), (c) $DTS(FBTh_2)_2$ (2,6-Bis{5-fluoro-7-(5"-hexyl-2,2"-bithiophen-5-yl)benzo[c][1,2,5]thiadiazol-4-yl}-(4,4'-bis(2-ethylhexyl)dithieno[3,2-b:2',3'-d]silole), and (d) PBDB-T (Poly[(2,6-(4,8-bis(5-(2-ethylhexyl)thiophen-2-yl)-benzo[1,2-b:4,5-b']dithiophene))-alt-(5,5-(1',3'-di-2-thienyl-5',7'-bis(2-ethylhexyl)benzo[1',2'-c:4',5'-c']dithiophene-4,8-dione)]), as shown in Figure 7. These outliers fall into two categories: data points (a), (b), and (c) are inflection points in the 1/2 power spectra, making it difficult to fit them with a single linear fit. Such data, as reported in the Si spectrum, require consideration of models with two transition processes that have different thresholds and powers [7]. In plot (d), the 1/2 power spectrum shows a downward shift with respect to the fitting line in the high-energy region; hence, the power number to be fitted is considered large. Therefore, in this case, 1/1 power is better-suited for analysis. Currently, extracting these outliers automatically is not possible, so judgments rely on human expertise. In the current analysis of the 1/*n* plot, comparisons are made using 1/2 and 1/3 powers. Although physical interpretation becomes challenging, it is possible to compare using other values. For example, in models incorporating two transition processes, as reported in the Si spectrum [5], it may be feasible to include powers like 1/1.5 that enable a relatively linear analysis of the two excitation components, as well as those in models with unclear physical representations, such as 1/1. It is also possible to consider cases where powers unexplained by clear physical models of 1/2 and 1/3 serve as indicators of anomalies. Generally, 1/3 power fitting is selected for most data on organic semiconductor materials. Additionally, spectra with threshold values above

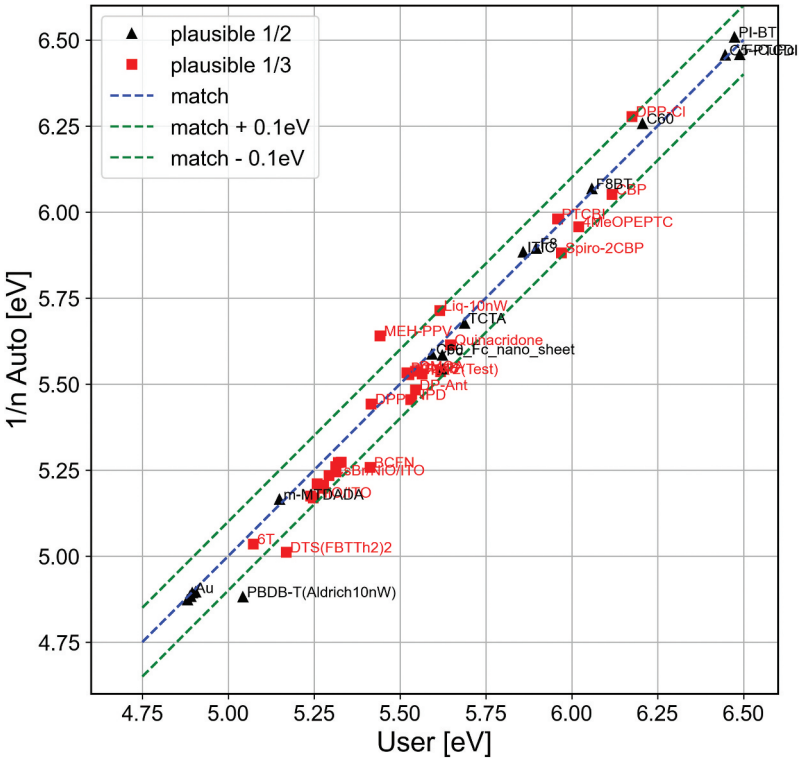


Figure 6. Comparison of the automated and manual analyses. Plot shows the results of the automated analysis of areas delineated via the PW-ROI and 1/*n* plot methods compared with the manual analysis results. Triangles (▲) indicate a fit to the power of 1/2, and squares (■) indicate a fit to the power of 1/3. These values and sample names are plotted on the graph. To prevent obscurity from overlapping, each sample name is displayed only once (e.g. in the case of Au).

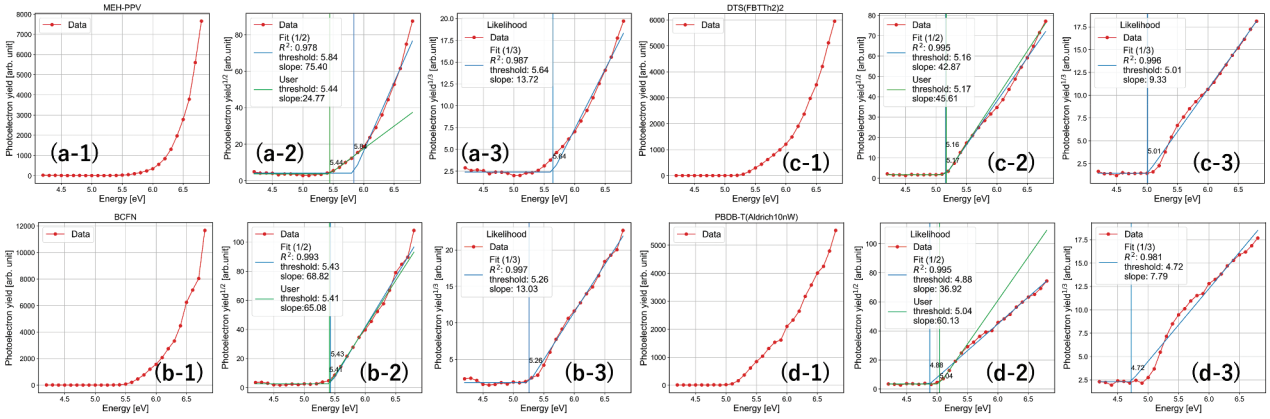


Figure 7. Data beyond the ±0.1 eV range and analysis results. Data beyond the ±0.1 eV range, as shown in Figure 6, with their analysis results. (a) MEH-PPV, (b) BCFN, (c) DTS $(FBTh_2)_2$, and (d) PBDB-T.

6.5 eV are fitted using the 1/2 power, likely because of limited data points utilized to obtain the slope in the power analysis (often performed in the range of 5–7 eV) of organic semiconductors. Although the data points corresponding to a few materials shown in Figure 7 have exceeded the ±0.1 eV range, the proposed method is reliable for analyses where minor variances are acceptable.

7. Conclusion

We propose an automated analysis method combining the PW-ROI and 1/*n* plot methods for analyzing thresholds in PYS, a technique used for measuring the ionization energy of inorganic and organic semiconductors. Additionally, the PYS spectra can be interpreted using the power law. The PW-ROI method selects the ROI of the spectrum that can be analyzed using a power law model, creating a graph based on the inverse of potential power numbers. Subsequently, fitting is performed using the *exReLU* function. The 1/*n* plot method identifies the coefficient of determination closest to 1. The integration of these two methods enables fully automated PYS analysis. The results demonstrate high consistency, all falling within a range of ±0.1 eV and thereby aligning with our

existing datasets, including manual analysis results. Our automated approach substantially reduces the analysis time to one-tenth of that required for manual analyses and ensures objectivity in analyses, along with eliminating manual dependency and facilitating reliable comparisons across different datasets.

7.1. 1/n plot method

This method applies multiple potential power numbers to the measured spectrum, generating several spectra for analysis. An *exReLU* function, with *MAE* as the minimization evaluation function, was used for fitting. The coefficients of determination were compared, selecting the power number closest to 1, along with its corresponding fitting value. Any background noise may increase the estimated threshold by influencing analysis.

7.2. PW method

This method applies segment regression on a spectrum to identify change points and slopes before and after these points, which are the characteristic of the respective spectrum. This implementation is available as open-source data. In segment regression, determining the number of change points can be challenging, but the optimal number can be derived from the degree of fit and information criterion.

7.3. PW-ROI method

The PW method provides insights into change points and slopes before and after these points. For a spectrum that rises with the power number, the slope derived from segment regression generally increases with energy. However, under the influence of other components, the slope may temporarily decrease. Therefore, the slope is compared from a low energy value to the change point just before the slope decreases significantly to determine and delineate the analysis region.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Research Center for Electronic and Optical Materials in National Institute for Materials Science (NIMS) and the JST-Mirai Program 'Materials Exploration space Expansion Platform (MEEP)' [Grant No. JPMJMI21G2].

Data availability statement

Data supporting the findings of this study will be made publicly available in the MDR of the NIMS repository.

References

[1] Shimizu R, Kobayashi S, Watanabe Y, et al. Autonomous materials synthesis by machine learning and robotics. APL Mater. 2020;8(11):111110. doi: 10.1063/5.0020370
[2] Burger B, Maffettone PM, Gusev VV, et al. A mobile robotic chemist. Nature. 2020;583(7815):237–241. doi: 10.1038/s41586-020-2442-2
[3] Shinotsuka H, Nagata K, Yoshikawa H, et al. Development of spectral decomposition based on Bayesian information criterion with estimation of confidence interval. Sci Technol Adv Mat. 2020;21(1):402–419. doi: 10.1080/14686996.2020.1773210
[4] Matsumura T, Nagamura N, Akaho S, et al. Spectrum adapted expectation-maximization algorithm for high-throughput peak shift analysis. Sci Technol Adv Mat. 2019;20(1):733–745. doi: 10.1080/14686996.2019.1620123
[5] Gobeli GW, Allen FG. Direct and indirect excitation processes in photoelectric emission from silicon. Phys Rev. 1962;127(1):141–149. doi: 10.1103/physrev.127.141
[6] Fowler H. The analysis of photoelectric sensitivity curves for clean metals at various temperatures. Phys Rev. 1931;38(1):45–56. doi: 10.1103/physrev.38.45
[7] Kane EO. Theory of photoelectric emission from semiconductors. Phys Rev. 1962;127(1):131–141. doi: 10.1103/physrev.127.131
[8] Kochi M, Harada Y, Hirooka T, et al. Photoemission from organic crystal in vacuum ultraviolet region. IV. Bull Chem Soc Jpn. 1970;43(9):2690–2702. doi: 10.1246/bcsj.43.2690
[9] Tauc J, Grigorovici R, Vancu A. Optical properties and electronic structure of amorphous germanium. Phys Status Solidi B. 1966;15(2):627–637. doi: 10.1002/pssb.19660150224
[10] Yagyu S, Yoshitake M, Nagata T. Automatic analysis method for the threshold and power exponent of spectra interpreted by the power law. J Surf Anal. 2023;30(2):98–112. doi: 10.1384/jsa.30.98
[11] Yagyu S, Yoshitake M, Nagata T. Preprocessing method for spectra interpreted by the power law using segment regression. J Surf Anal. 2024;31(1):2–10.
[12] Online Manual [Internet]. RIKEN KEIKI Co.Ltd. [cited 2025 Jan 5]. Available from: https://product.rikenkeiki.co.jp/assets/pdf/ac/ac-5/PT5E-0505_AC-5.pdf
[13] Urbach F. The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids. Phys Rev. 1953;92(5):1324. doi: 10.1103/physrev.92.1324
[14] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS); Fort Lauderdale, (FL), USA; 2011 June 14.
[15] Online Catalog [Internet]. RIKEN KEIKI Co.Ltd. [cited 2024 Nov 14]. Available from: https://product.rikenkeiki.co.jp/english/ac/ac-3/

[16] Vito MR. Muggeo [Internet]. Manual 2.1–3 package 'segmented'. [cited 2024 Nov 14]. Available from: https://cran.r-project.org/web/packages/segmented/segmented.pdf
[17] Online code repository [Internet] Github. [cited 2024 Nov 14]. Available from: https://github.com/chasmani/piecewise-regression
[18] Online code repository [Internet] Github. [cited 2024 Nov 14]. Available from: https://github.com/cjekel/piecewise_linear_fit_py
[19] Mahdaoui D, Hirata C, Nagaoka K, et al. Ambipolar to unipolar irreversible switching in nanosheet transistors: the role of ferrocene in fullerene/ferrocene nanosheets. J Mater Chem C. 2022;10(10):3770–3774. doi: 10.1039/d1tc05545c [dataset] 2024 Nov. 14 [cited 2024 Nov 14] In Material Data Repository [Internet] NIMS [cited 2024 Nov 14]. Available from: https://doi.org/10.48505/nims.3954.
[20] Pant N, Kulkarni A, Yanagida M, et al. Passivation of bulk and interface defects in sputtered-NiO x-based planar perovskite solar cells: a facile interfacial engineering strategy with alkali metal halide salts. ACS Appl Energy Mater. 2021;4(5):4530–4540. doi: 10.1021/acsaem.1c00032 [dataset]2024 Nov. 14 [cited 2024 Nov 14] In Material Data Repository [Internet] NIMS [cited 2024 Nov 14]. Available from: https://doi.org/10.48505/nims.3955.
[21] Yagyu S, Yoshitake M, Nagata T, et al. Create the database of the photoemission yield spectroscopy (PYS). J Surf Anal. 2023;29(3):146–154. doi: 10.1384/jsa.29.146