Curvature-Guided Derivative Peak Filtering for Robust Initialization of Sequential XPS Spectral Deconvolution

Sayid Bajrai Abdul Nasir[1], Narimah Kasim[2]

[1]Senior Software Engineer, Blue Ocean Data Solution Sdn. Bhd. (JB Branch), Malaysia

[2]Professor, Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1306000524

Received: 04 July 2026; Accepted: 10 July 2026; Published: 22 July 2026

ABSTRACT

Accurate initialization of peak positions is essential for reliable X-ray Photoelectron Spectroscopy (XPS) spectral deconvolution because nonlinear peak-fitting algorithms are highly sensitive to their starting parameters. Conventional derivative-based peak detection often generates excessive false-positive candidates in noisy or overlapping spectra, reducing fitting stability and reproducibility. This work introduces Curvature-Guided Derivative Peak Filtering (CG-DPF), an interpretable human-in-the-loop framework that combines first-derivative zero-crossing detection with tunable second-derivative curvature thresholding. Unlike classical derivative spectroscopy, which primarily uses the second derivative to identify extrema, CG-DPF applies curvature as a continuously adjustable post-detection acceptance criterion, allowing analysts to balance peak recall against false-positive suppression while preserving transparent parameter selection. The contribution is a reproducible filtering strategy and initialization workflow rather than a new differentiation or smoothing algorithm. CG-DPF was evaluated using a primary benchmark comprising one in-house survey spectrum and seven published XPS survey spectra, with additional external transfer assessments on published oxide thin-film and HAXPES-related datasets. Performance was assessed using reconstruction quality, FWHM compliance, peak count, computational efficiency, and statistical comparison with adaptive multi-scale detection under identical preprocessing and fitting conditions. On the primary benchmark, CG-DPF reduced false-positive initialization from 47 to 5 candidate peaks while maintaining complete FWHM compliance. Across seven independent survey spectra, it achieved a mean Phys. Ratio of 0.66 compared with 0.30 for the Adaptive Medium baseline (Wilcoxon signed-rank test, $p = 0.031$). External datasets further showed that survey-default parameters do not universally transfer across different materials and acquisition conditions, highlighting the need for analyst-guided parameter tuning.CG-DPF provides an optimizer-independent preprocessing strategy that integrates with conventional Gaussian, Lorentzian, Voigt, and pseudo-Voigt fitting workflows, improving the robustness, reproducibility, and transparency of sequential XPS spectral deconvolution.

Keywords: X-ray photoelectron spectroscopy, peak detection, derivative spectroscopy, spectral deconvolution, curvature filtering

INTRODUCTION

XPS is a cornerstone technique for surface chemical analysis. Extracting chemical-state information requires deconvolution of overlapping photoemission features through peak fitting, typically with Gaussian, Lorentzian, or Voigt profiles optimized to reconstruct the measured spectrum [2].

Fitting quality is notoriously sensitive to initialization [2,4]. Poor initial estimates of peak position, width, amplitude, or even peak count can lead nonlinear optimizers such as Levenberg–Marquardt or Trust Region Reflective to converge to local minima, produce unstable solutions, or fail to converge.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Traditional initialization is either manual (subjective, labor-intensive, poorly reproducible) or automatic via derivative spectroscopy [9,15] or general-purpose peak finders (e.g., SciPy [14]). Automatic methods frequently over-detect: the first derivative of a noisy spectrum yields numerous zero-crossings from noise as well as genuine peaks [4,15], and unconstrained peak finders applied to survey-scale data can return hundreds of candidates [5,14]. A peak produces a positive-to-negative crossing in dI/dE and negative curvature in d^2I/dE^2, but so do minor fluctuations and baseline irregularities.

Contribution

The objective of this work is to develop an interpretable derivative-based filtering framework that reduces false-positive peak initialization while preserving reproducibility and analyst control for sequential XPS spectral deconvolution. CG-DPF is proposed, a filtering framework that post-processes derivative-based candidates with an interpretable curvature threshold. The contribution is primarily methodological: a new filtering strategy and reproducible workflow that combines derivative spectroscopy, curvature-based candidate rejection, analyst-guided threshold selection, and auditable export (not a novel smoothing kernel or derivative estimator).

Classical derivative spectroscopy uses d^2I/dE^2 primarily to identify extrema and confirm peak locations [9,15]. CG-DPF instead treats curvature as a continuously adjustable acceptance criterion applied after first-derivative candidate generation, allowing analysts to trade recall against precision while preserving a reproducible, logged threshold value. This workflow is generally not provided by standard derivative plots or typical off-the-shelf peak finders that return a fixed candidate set without an interpretable, spectrum-specific curvature gate.

Novel aspects relative to classical derivative spectroscopy and generic peak picking are listed below. Unlike multi-scale consensus detectors, CG-DPF deliberately reduces candidate count before nonlinear optimization rather than increasing candidate recall:

1. Post-crossing curvature filtering (algorithmic): $\theta = \mu - k\sigma$ applied to interpolated crossings, with transparent control on the d^2I/dE^2 chart.

2. FWHM compliance screening (workflow): configurable width bounds integrated into initialization and export.

3. Sequential fitting workflow integration (workflow): prominence-ranked peak addition with supplied initial centers.

4. Reproducible audit trail (software): exported workbooks with parameters, metrics, and chart PNGs.

Items (2) to (4) are engineering integrations that make the method usable in routine XPS analysis. Item (1) is the core detection rule that distinguishes CG-DPF from unfiltered derivative crossings.

RELATED WORK

Derivative spectroscopy uses dI/dE sign changes and d^2I/dE^2 extrema to locate peaks [9]. Pre-smoothing via Savitzky-Golay [10] or Gaussian convolution is essential because differentiation amplifies noise. Multi-scale approaches detect peaks across several σ values and merge by consensus [5]. High-sensitivity settings can yield hundreds of candidates with low FWHM compliance under survey-scale screening.

General signal-processing peak finders, including scipy.signal.find_peaks (implemented in SciPy [14]), MATLAB findpeaks, and wavelet-based methods, are widely used outside XPS. These approaches generally optimize signal characteristics such as peak height, prominence, width, or multi-scale consistency on raw or lightly processed intensity data, rather than combining first-derivative zero-crossing detection with an interpretable second-derivative curvature-gating strategy tailored for sequential XPS spectral deconvolution. Consequently, their objectives differ from that of CG-DPF, which emphasizes conservative, reproducible initialization with analyst-controlled filtering. Quantitative comparisons in this study are therefore limited to

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

the in-tool Adaptive Detection presets, which provide a strong multi-scale baseline under identical preprocessing, fitting, and export conditions. Overall, CG-DPF differs from existing derivative-based and general-purpose peak detection methods by introducing a transparent post-detection curvature-filtering workflow that integrates FWHM compliance screening, sequential fitting initialization, and reproducible parameter logging for XPS analysis.

Sequential fitting, adding peaks incrementally with re-optimization, is standard in XPS practice [2,4]. CG-DPF is positioned as an interpretable filtering layer for initialization rather than a fully autonomous detector. Adaptive multi-step baseline estimation, when used in the same software, addresses a complementary problem (background modeling) and is not combined with the present peak-initialization study.

Conventional XPS peak fitting commonly employs Gaussian, Lorentzian, or Voigt line-shape models optimized using nonlinear least-squares methods, with the Levenberg–Marquardt (LM) algorithm being widely used in established XPS software packages. Trust Region Reflective (TRF) optimization is also widely used, particularly when parameter bounds are required to enforce physically meaningful constraints. Regardless of the optimization algorithm, convergence remains highly dependent on the quality of the initial peak estimates.

METHODOLOGY

Algorithm

CG-DPF operates in two stages:

1. Candidate detection: First-derivative zero crossings on a Gaussian-smoothed spectrum (configurable σ, default 8 eV).
2. Curvature filtering: Retain only candidates whose d^2I/dE^2 at the interpolated crossing lies below threshold θ.

The CG-DPF workflow consists of five sequential stages. First, the raw spectrum is smoothed using Gaussian convolution to reduce noise while preserving peak shapes. Next, candidate peaks are identified from positive-to-negative zero crossings of the first derivative. Each candidate is then evaluated using the second derivative, where a curvature threshold is applied to reject weak or noise-induced detections. The remaining filtered peaks constitute the final candidate set used to initialize the sequential XPS spectral deconvolution process.

Mathematical Formulation

Given energy axis *E* and intensity *I*, the smoothed spectrum is:

$$\tilde{I}(E,\sigma) = G(E,\sigma) * I(E)$$

where *G* is a Gaussian kernel. Derivatives use central finite differences on the smoothed data:

$$I'(E_i) = \frac{\tilde{I}(E_{i+1}) - \tilde{I}(E_{i-1})}{E_{i+1} - E_{i-1}}, \quad I''(E_i) = \frac{I'(E_{i+1}) - I'(E_{i-1})}{E_{i+1} - E_{i-1}}$$

Zero crossings occur where $I'(Ei) > 0$ and $I'(Ei+1) \leq 0$. The crossing energy is refined by linear interpolation:

$$E_0 = E_i - I'(E_i)\,\frac{E_{i+1} - E_i}{I'(E_{i+1}) - I'(E_i)}$$

The curvature threshold is $\theta = \mu I'' - k \cdot \sigma I''$, where $\mu I''$ and $\sigma I''$ are the mean and standard deviation of d^2I/dE^2 on the smoothed spectrum. The dimensionless multiplier is:

$$k = \frac{\mu_{I''} - \theta}{\sigma_{I''}}$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

A candidate at *E0* is retained if $I''(E0) < \theta$. Default $k = 1$ sets θ = μ - σ, i.e., a conventional one-standard-deviation tail threshold familiar from robust statistics. On the in-house benchmark this default maximized FWHM compliance while retaining the major core levels. For routine survey analysis, $k \approx 0.3$-0.5 often balances recall and FWHM compliance better than $k = 1$.

Implementation

In the reference implementation (v2.3) [5]: smoothing uses scipy.ndimage.gaussian_filter1d with mode reflect, derivatives use numpy.gradient and second-derivative values at interpolated crossings use linear interpolation.

Sequential Fitting Integration

Filtered peaks supply initial centers for sequential deconvolution. The procedure is:

1. Ordering: CG-DPF candidates are sorted by descending local prominence before fitting [5].
2. Initialization: For peak i, the center is fixed to the i-th supplied energy and amplitude is taken from the current residual at that center and width is initialized from a survey-span heuristic (≥ ΔE × 10, bounded relative to the energy axis).
3. Optimization: A single Gaussian is fit to the current residual (background-corrected intensity minus previously fitted components) using Trust Region Reflective (TRF) nonlinear least-squares optimization (scipy.optimize.curve_fit, method trf, ftol = xtol = 10^{-6}, maxfev ≥ 10^4). TRF was selected because it supports bounded optimization, allowing physically meaningful constraints on peak parameters during sequential fitting.
4. Update: The fitted component is subtracted from the residual (clipped to non-negative values).
5. Stopping: The loop terminates after all N detected centers have been fit. There is no additional residual-based peak insertion in the benchmark pipeline.

FWHM compliance is evaluated post-fit from FWHM = $2\sqrt{(2 \ln 2)}\ \sigma \approx 2.355\sigma$ for Gaussian components, compared against configurable bounds (the Metrics section). CG-DPF does not assign chemical states. Binding-energy labels in the Peak Assignments section are illustrative comparisons to tabulated references [6].

Benchmark Spectra

Table 1. CG-DPF is evaluated on a primary corpus of eight survey-scale spectra, with additional external datasets from published Zenodo archives.

Dataset	Source	Role
example_spectrum.csv	In-house benchmark	Primary detailed comparison
AS Sample 0–AS Sample 6	Suliga et al. [12]	Primary multi-spectrum validation
XPS results by Marx Patrick.xls (+ MA94/MA309 C 1s)	Marx et al. [8]	External oxide thin-film check
OAD_XPS_survey (+ FOIL/OAD Ni 2p as-received)	Longo [7]	External OER/HAXPES-related check

The in-house spectrum contains Si 2p, Si 2s, C 1s, N 1s, O 1s, and F 1s features on a background from ~1,280 counts (high BE) to >125,000 counts (low BE). High-BE noise σ ≈ 974 counts (bulk SNR ~3.9). The primary corpus is modest in size (eight spectra) but spans independent survey data suitable for a software-methods paper. Marx and Longo spectra probe generalization beyond the Suliga survey domain; they are reported separately because default survey settings are not expected to be optimal for every instrument class or narrow-window region spectrum. Larger corpora with expert-annotated peak lists remain future work (the Discussion section).

Configuration

CG-DPF defaults: σ = 8 eV, k = 1.0, minimum prominence 0.1% of max intensity, minimum spacing 0.5 eV, configured FWHM bounds 0.2–5.5 eV (UI default 0.2–5.0 eV).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Background subtraction: All tabulated benchmark deconvolutions used multi-step sigmoid background subtraction (linear baseline plus two logistic steps), which better models survey-scale step continua than a single straight line. Shirley [11] and Tougaard [13] models are available in the tool but were not used for the tabulated runs.

Baselines: (1) unfiltered first-derivative crossings with $\theta = 0$ at $\sigma = 8$ (computed via test/k_sensitivity_example.py), (2) adaptive multi-scale detection (Medium and Research presets) [5] and (3) CG-DPF at default and relaxed k.

Metrics and Evaluation Rationale

Absence of ground truth. XPS survey spectra generally lack an authoritative, peer-reviewed peak-count reference. Quantitative precision, recall, and F1 against a gold standard are therefore not reported; surrogate metrics are used instead:

- Phys. Ratio: fraction of fitted peaks whose FWHM lies within the configured range. This measures compliance with a width prior, not chemical correctness or assignment accuracy.
- Corr. RMSE / Corr. R^2: reconstruction quality on background-corrected intensities after sequential fitting.
- Peak count: number of initialized and fitted components.
- Timing: wall-clock detection and process durations from export.

On the in-house benchmark, six major core levels are identifiable by binding energy [6]. The default CG-DPF recovers five (missing weak N 1s), providing informal support for conservative recall without constituting a formal detection benchmark.

FWHM bounds. The lower bound 0.2 eV is a permissive limit for narrow, well-resolved photoemission lines on modern instruments; the upper bound 5.5 eV (batch script with UI default 5.0 eV) accommodates broadened or poorly resolved survey features while excluding grossly out-of-range widths. Typical condensed-phase core-level FWHM values fall in the ~0.5–3 eV range for many elements [2,4]. The configured window is intentionally wide so that screening flags widths outside typical instrument/core-level ranges rather than enforcing a tight chemical-state prior.

Table 2. An example ($\sigma = 8$, $k = 1$), multi-step sigmoid background, 5 peaks, all within the FWHM window with Phys. Ratio = 1.00.

Peak Number	Center (eV)	Amplitude	FWHM (eV)	Physically Reasonable (0.2–5.5 eV)	Width σ (eV)
1	532.5256999	103312.2951	2.4879	Yes	1.056493376
2	687.9106958	22978.4212	5.3262	Yes	2.261844828
3	285.6015587	29624.95374	2.91	Yes	1.235755566
4	154.1615011	17439.17379	3.1539	Yes	1.33932264
5	103.2546256	21190.11936	2.4928	Yes	1.058582832

Exported workbooks record smoothing σ, standard deviation of d^2I/dE^2, θ, multiplier $k = (\mu - \theta)/\sigma I''$, peak parameters, metrics, and embedded chart PNGs (Input Parameters, Peak Parameters, Deconvolution Results, Charts sheets).

Computing Environment

Timing values were recorded during batch export on a single workstation: Windows 10 (build 19041), AMD64 processor (R7 5700X, 16 logical cores), Python 3.11, NumPy 1.26.2, SciPy 1.11.4. Detection times are wall-clock values from the exported Peak Detection Time (s) field (single-process batch run). Absolute timing will vary with hardware and load. The relative comparison between CG-DPF and Adaptive Detection under identical conditions is the intended use.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Statistical Analysis

For paired comparisons across the seven Suliga spectra (CG-DPF default vs. Adaptive Medium), paired t-tests and Wilcoxon signed-rank tests are reported on per-spectrum Phys. Ratio, peak count, and Corr. R^2 [3]. Tests are exploratory given $n = 7$ and $p < 0.05$ is noted where applicable.

RESULTS

Results are organized as a primary evaluation (in-house + seven Suliga surveys under identical multi-step sigmoid / Gaussian sequential fitting) followed by external transfer checks (Marx, Longo). Primary claims rest on Tables 3–5 and Figures 1–2; external datasets test whether survey defaults transfer without retuning (Tables 6–7 and Figure 3).

Detection Performance (In-House Benchmark)

Table 3. Detection and fitting comparison on example_spectrum.csv (FWHM range 0.2–5.5 eV, multi-step sigmoid background; values from Deconvolution Results sheets in data/results/). Precision, recall, and F1 are omitted because no reference peak list exists for this spectrum (the Metrics section).

Method	Peaks	FWHM-compliant	Phys. Ratio	Corr. RMSE	Corr. R^2
CG-DPF, unfiltered (σ=8, θ=0) †	47	9	0.19	708.64	0.986
CG-DPF, relaxed (σ=8, k≈0.32)	7	7	1.00	740.54	0.985
CG-DPF, default (σ=8, k=1)	5	5	1.00	816.42	0.981
Adaptive, Medium	10	6	0.60	814.89	0.981
Adaptive, Research	120	12	0.10	1138.07	0.963

† Unfiltered curvature at $\sigma = 8$ ($\theta = 0$), from test/k_sensitivity_example.py. Detection times are wall-clock Peak Detection Time from exported workbooks.

CG-DPF optimizes conservative initialization rather than maximum residual reduction. Adaptive Medium achieves comparable R^2 (0.981 vs. 0.981) but only 60% FWHM compliance, whereas the CG-DPF default retains five major core-level features with Phys. Ratio = 1.0 (Figure 1a). Unfiltered curvature ($\theta = 0$) detects 47 peaks and only 9 (19%) satisfy the FWHM window despite $R^2 = 0.986$, which is indicative of over-parameterization. Adaptive Research detects 120 peaks (12 FWHM-compliant, 10%) with substantially higher fitting time. Figure 1b and Figure 4 show the curvature gate θ on d^2I/dE^2 with retained centers marked.

At default $k = 1$, all five detected peaks (Si 2p, Si 2s, C 1s, O 1s, F 1s) fall within the FWHM window and N 1s (~399 eV) is excluded. Relaxing to $k \approx 0.32$ ($\theta \approx -248$ eV^2) recovers N 1s at 399.67 eV (FWHM = 3.89 eV) while maintaining Phys. Ratio = 1.0 and $R^2 = 0.985$.

Multi-Spectrum Validation (Suliga Dataset)

Table 4. Per-spectrum summary for CG-DPF default (σ = 8, k = 1) versus Adaptive Medium (multi-step sigmoid background).

Spectrum	CG-DPF Peaks	CG-DPF Phys. Ratio	CG-DPF R^2	Adaptive Peaks	Adaptive Phys. Ratio	Adaptive R^2
AS Sample 0	4	0.75	0.903	14	0.29	0.944
AS Sample 1	4	0.75	0.894	19	0.26	0.959
AS Sample 2	6	0.50	0.888	23	0.26	0.933
AS Sample 3	5	0.80	0.882	20	0.25	0.874
AS Sample 4	7	0.71	0.909	22	0.41	0.902
AS Sample 5	4	0.75	0.896	21	0.29	0.951
AS Sample 6	6	0.33	0.871	20	0.35	0.878
Mean ± SD	5.1 ± 1.1	0.66 ± 0.16	0.89 ± 0.01	19.9 ± 2.7	0.30 ± 0.05	0.92 ± 0.03

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Across seven independent survey spectra, CG-DPF consistently proposes fewer peaks with higher mean FWHM compliance than Adaptive Medium (Figure 2), at the cost of lower mean R^2 on several samples (notably AS Sample 6, where conservative filtering yields Phys. Ratio = 0.33 despite R^2 = 0.871). This trade-off is expected: CG-DPF is designed for trustworthy initialization, not exhaustive residual minimization.

Statistical significance (n = 7 paired spectra). CG-DPF yielded higher Phys. Ratio than Adaptive Medium (Wilcoxon W = 1, p = 0.031 and paired t-test p = 0.003) and fewer detected peaks (Wilcoxon p = 0.016 and paired t-test $p < 0.001$), although the sample size is small. Mean Corr. R^2 was lower for CG-DPF (paired t-test p = 0.051 and Wilcoxon p = 0.156), consistent with the distinct optimization objectives. These tests are therefore supportive rather than definitive [3].

Table 5. Aggregate comparison including high-recall configurations (Suliga samples 0–6, mean ± SD).

Method	Peaks	Phys. Ratio	Corr. R^2
CG-DPF, default (σ=8, k=1)	5.1 ± 1.1	0.66 ± 0.16	0.89 ± 0.01
CG-DPF, no smoothing (σ=0, k=1)	43.3 ± 13.4	0.14 ± 0.03	0.90 ± 0.03
Adaptive, Medium	19.9 ± 2.7	0.30 ± 0.05	0.92 ± 0.03
Adaptive, Research	249.7 ± 32.4	0.04 ± 0.01	0.81 ± 0.06

Relative to Adaptive Research, CG-DPF default reduces mean peak count by ~50× while raising mean Phys. Ratio from 0.04 to 0.66, reinforcing that high-recall multi-scale presets over-parameterize survey initialization under the same FWHM window.

External Validation (Marx Patrick Dataset)

Table 6. Marx Patrick oxide thin-film spectra [8] under the same survey-default pipeline (multi-step sigmoid). CG-DPF σ = 8 was unavailable for MA94/MA309 C 1s windows (no peaks detected or run missing); MA309 uses σ = 4 where noted.

Spectrum	Method	Peaks	Phys. Ratio	Corr. R^2
Marx survey	CG-DPF σ=8	6	0.00	0.438
Marx survey	Adaptive Medium	30	0.30	−0.235
MA309 C 1s	CG-DPF σ=4	28	0.00	−0.710
MA309 C 1s	Adaptive Medium	52	0.08	−4.454
MA94 C 1s	Adaptive Medium	53	0.04	−3.001

On the Marx survey, CG-DPF still returns fewer candidates than Adaptive Medium (6 vs. 30), but Phys. Ratio collapses to 0 at σ = 8 and reconstruction quality is poor for both methods under Gaussian sequential fitting with survey-default FWHM bounds (Figure 3b). Narrow-window C 1s regions are likewise poorly served by this pipeline: adaptive presets over-detect while CG-DPF either fails at σ = 8 or yields many non-compliant widths. These results are reported as negative external evidence that default survey initialization does not transfer automatically to every published oxide thin-film or region spectrum without retuning σ, k, background model, and line shape.

External Validation (Longo Filippo Dataset)

Table 7. Longo Filippo XPS spectra [7] (OAD survey and as-received Ni 2p regions) under the same multi-step sigmoid pipeline.

Spectrum	Method	Peaks	Phys. Ratio	Corr. R^2	Notes
OAD XPS survey	CG-DPF σ=8	7	0.00	0.272	Survey default
OAD XPS survey	Adaptive Medium	27	0.11	0.702	Higher R^2, low Phys. Ratio
OAD XPS survey	CG-DPF σ=2	8	0.25	0.693	Best CG-DPF Phys. Ratio among σ

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Spectrum	Method	Peaks	Phys. Ratio	Corr. R^2	Notes
					grid
FOIL Ni 2p (ar)	CG-DPF σ=4	1	0.00	0.398	σ=8: no peaks detected
FOIL Ni 2p (ar)	Adaptive Medium	1	1.00	0.539	Single-component region
OAD Ni 2p (ar)	CG-DPF σ=8	1	0.00	0.409	Under-parameterized at default σ
OAD Ni 2p (ar)	Adaptive Research	4	1.00	0.714	Medium preset: no peaks

As with Marx, Longo survey and Ni 2p region spectra expose the limits of a single survey-tuned ($\sigma = 8, k = 1$) setting: CG-DPF remains more conservative in peak count on the OAD survey (7 vs. 27) but FWHM compliance and Corr. R^2 degrade relative to the Suliga primary corpus. Lower σ (e.g., $\sigma = 2$ on OAD survey) improves Phys. Ratio and R^2, reinforcing that external datasets require analyst-guided retuning rather than blind reuse of the in-house default.

Parameter Sensitivity

Table 8. Curvature threshold (k), σ = 8 eV (example_spectrum.csv).

Configuration	k	θ (eV^2)	Peaks	Phys. Ratio	Corr. R^2
Unfiltered	0.00	0	47	0.19	0.986
Relaxed	0.32	−248	7	1.00	0.985
Default	1.00	−776	5	1.00	0.981

Smoothing σ (k = 1): σ = 0 yields 129 candidates (Phys. Ratio = 0.09), σ = 1 yields 27 (Phys. Ratio = 0.19), σ = 2 yields four peaks (Phys. Ratio = 1.0), and σ = 4, 8, and 16 each yield five peaks at Phys. Ratio = 1.0 on the in-house benchmark. σ = 8 was retained as the survey-scale default.

CG-DPF peak detection averaged 2 ms per spectrum versus 18 ms for adaptive methods across the eight primary benchmark spectra (the Computing Environment section).

Peak Assignments

Table 9. CG-DPF centers with default (σ = 8, k = 1) on example_spectrum.csv. Binding-energy labels are illustrative comparisons to tabulated core-level values [6]. CG-DPF does not perform chemical-state assignments.

Peak (eV)	Assignment (illustrative)	FWHM (eV)	Within FWHM window
103.25	Si 2p	2.49	Yes
154.16	Si 2s	3.15	Yes
285.60	C 1s	2.91	Yes
532.53	O 1s	2.49	Yes
687.91	F 1s	5.33	Yes

With relaxed threshold ($\theta \approx -248$ eV^2), N 1s at 399.67 eV (FWHM = 3.89 eV) is recovered with Phys. Ratio = 1.0.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Visual Results

Publication figures regenerated from exported Spectrum Data / Peak Parameters (Times New Roman, 300 dpi):

Figure 1. In-house CG-DPF (σ = 8 eV, k = 1): (a) original, fitted, residual, and peak components; (b) second derivative with curvature threshold θ and retained peak centers.

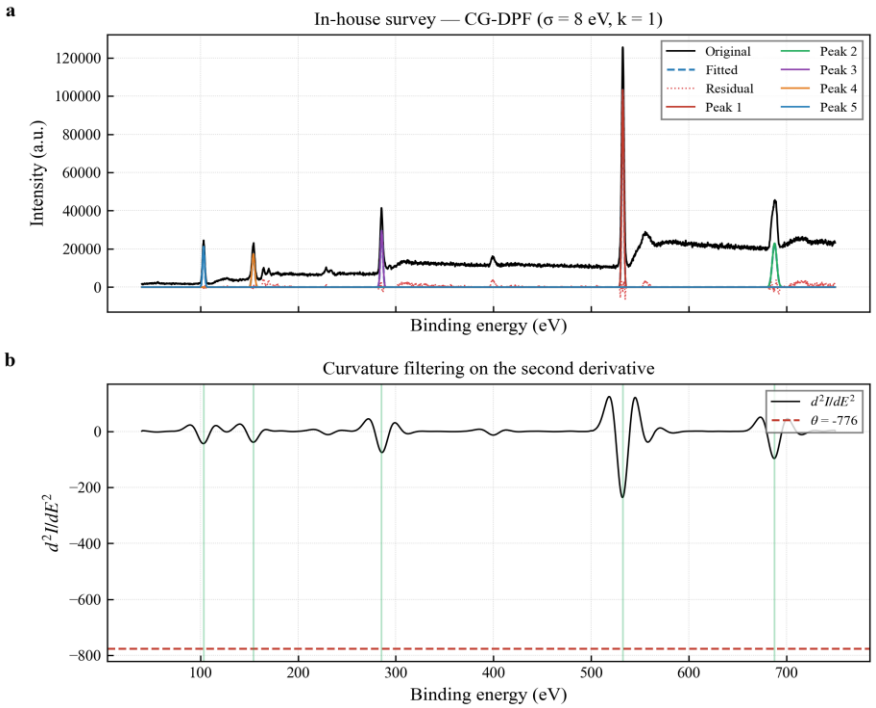
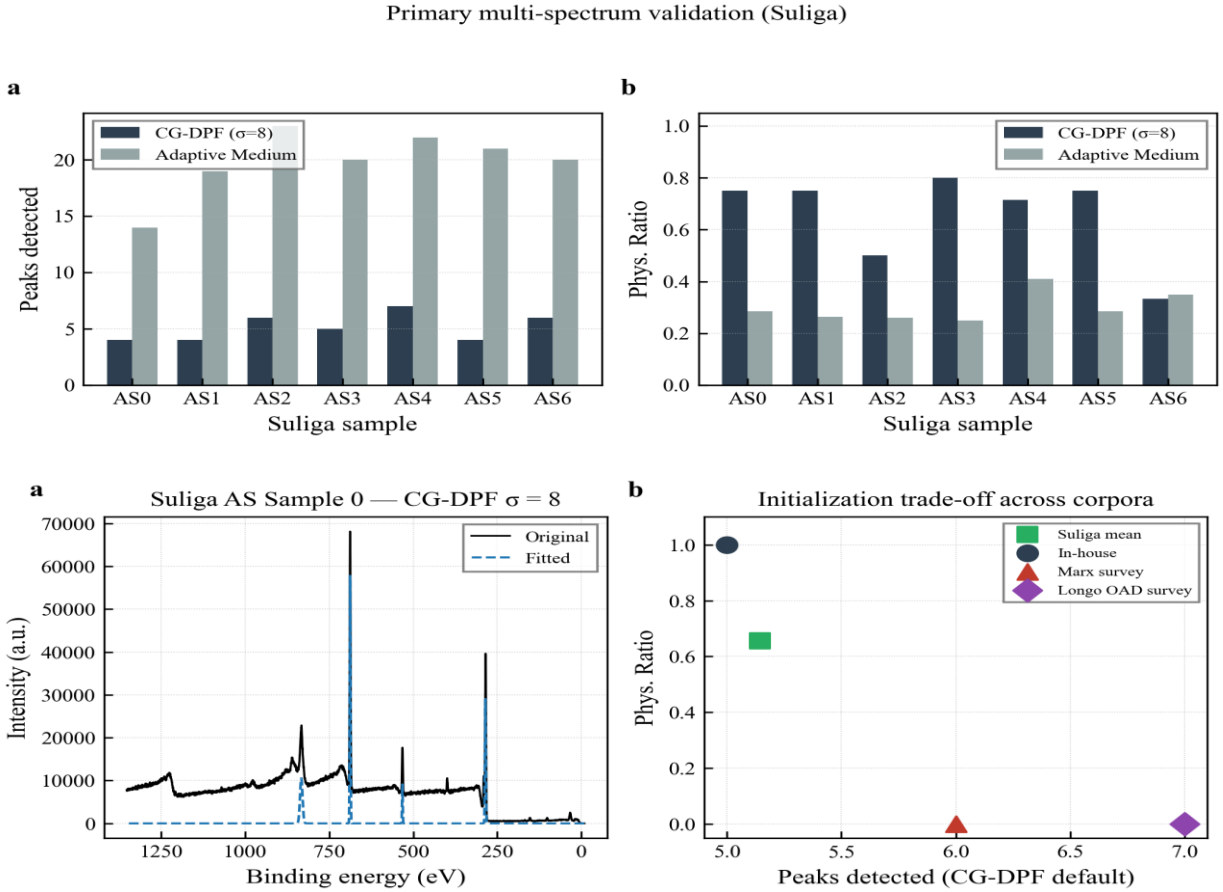


Figure 2. Suliga primary validation: per-sample (a) peak count and (b) Phys. Ratio for CG-DPF σ = 8 versus Adaptive Medium.



ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Figure 3. (a) Suliga AS Sample 0 CG-DPF fit; (b) peaks vs. Phys. Ratio across in-house, Suliga mean, Marx survey, and Longo OAD survey under CG-DPF defaults.

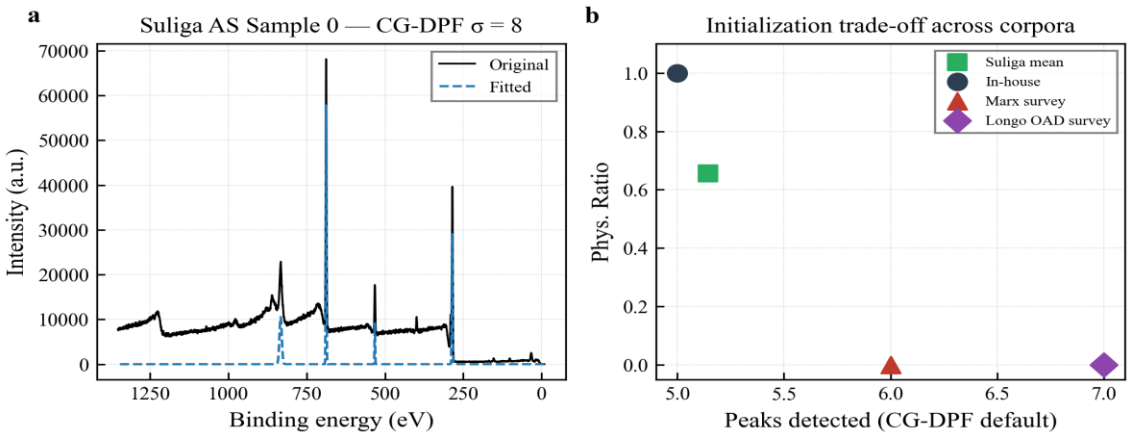
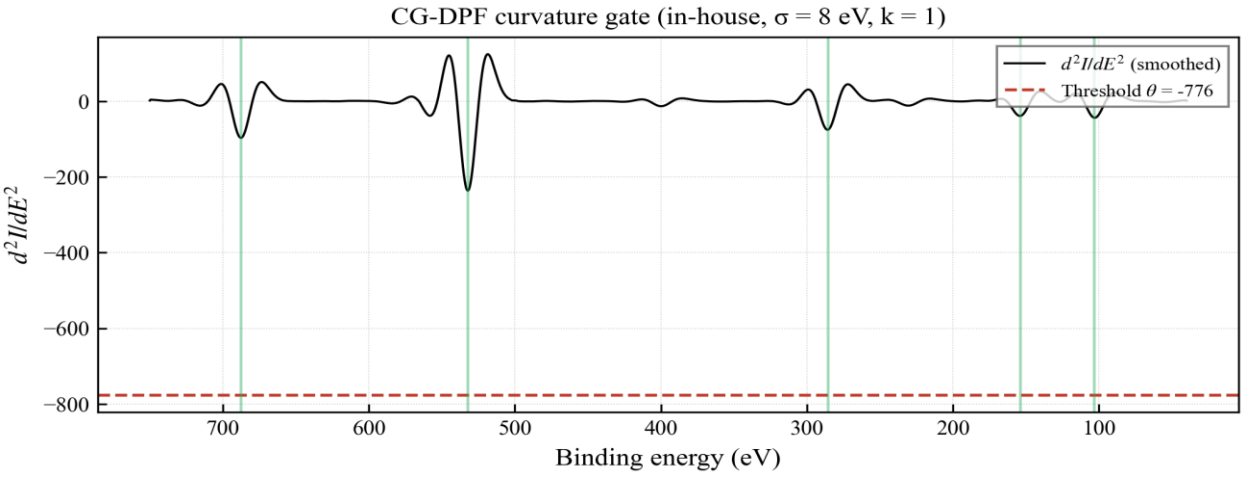
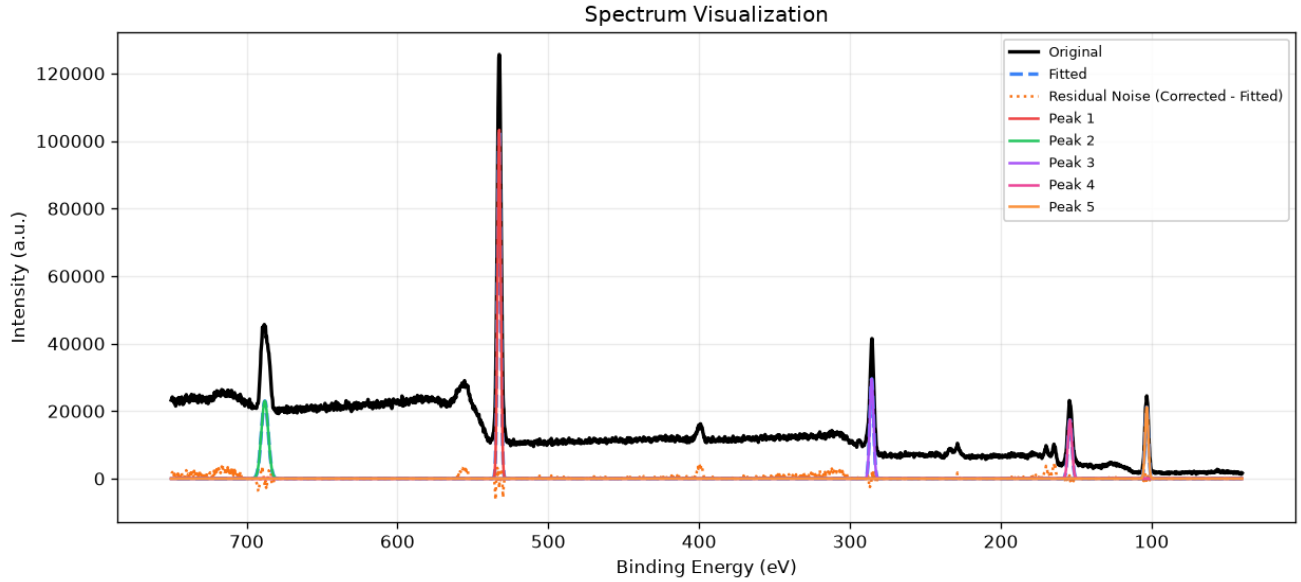


Figure 4. Curvature gate on d^2I/dE^2 for the in-house benchmark (σ = 8 eV, k = 1); vertical markers indicate fitted centers.



Figures extracted from the exported workbook Charts sheets:

Figure 5. Spectrum: original, background-corrected, fitted overlay, and peak components (σ = 8, k = 1) on the in-house benchmark.



ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Figure 6. First derivative dI/dE with zero-crossing positions (in-house, CG-DPF σ = 8).

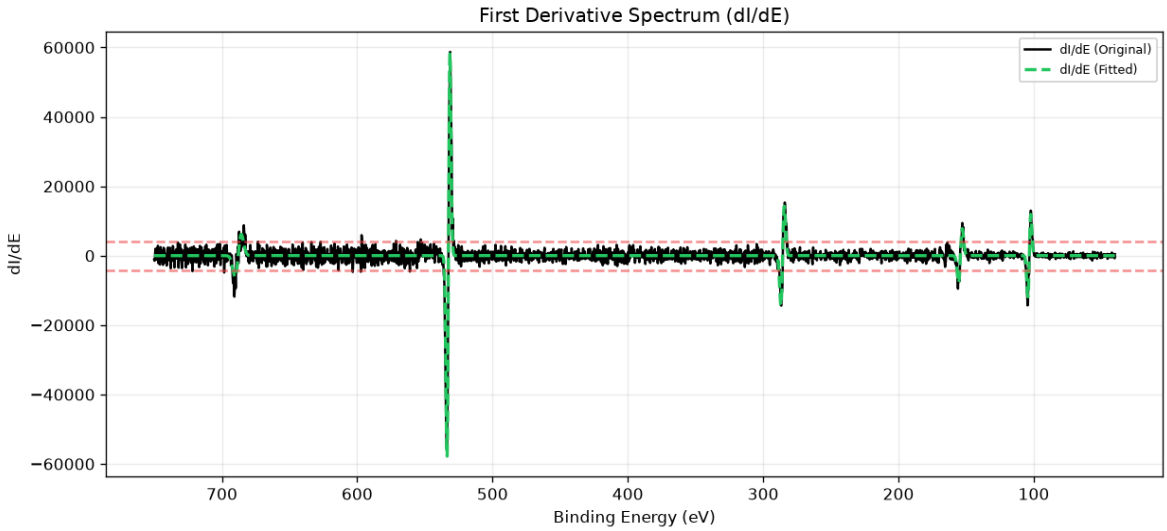


Figure 7. Second derivative d²I/dE² with curvature threshold (in-house, CG-DPF σ = 8, k = 1). Candidates below the threshold line (negative curvature exceeding θ) are retained. The green/red markers denote FWHM-compliant / out-of-range widths after fitting.

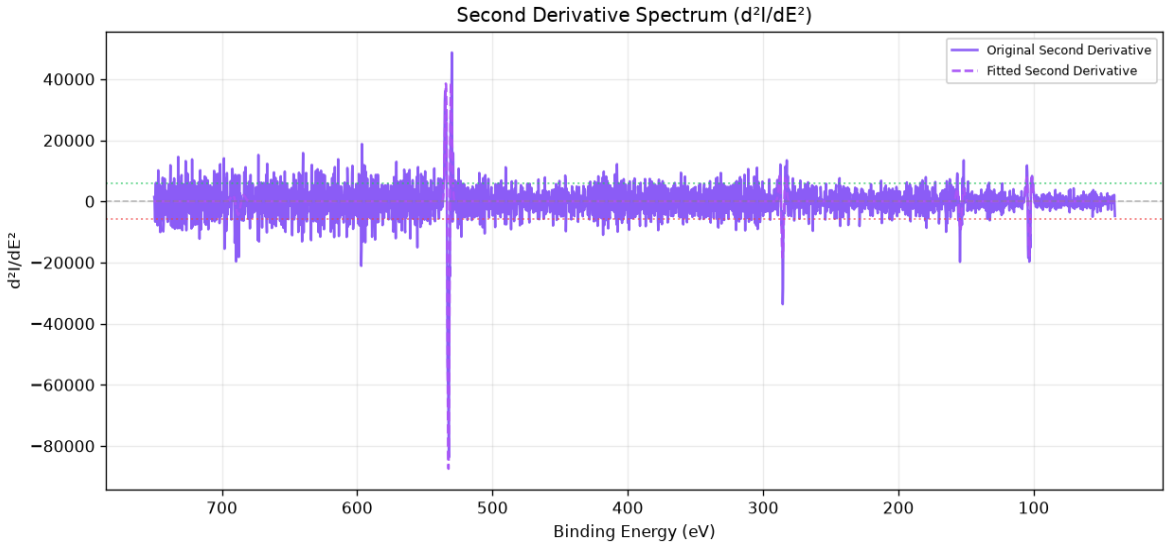
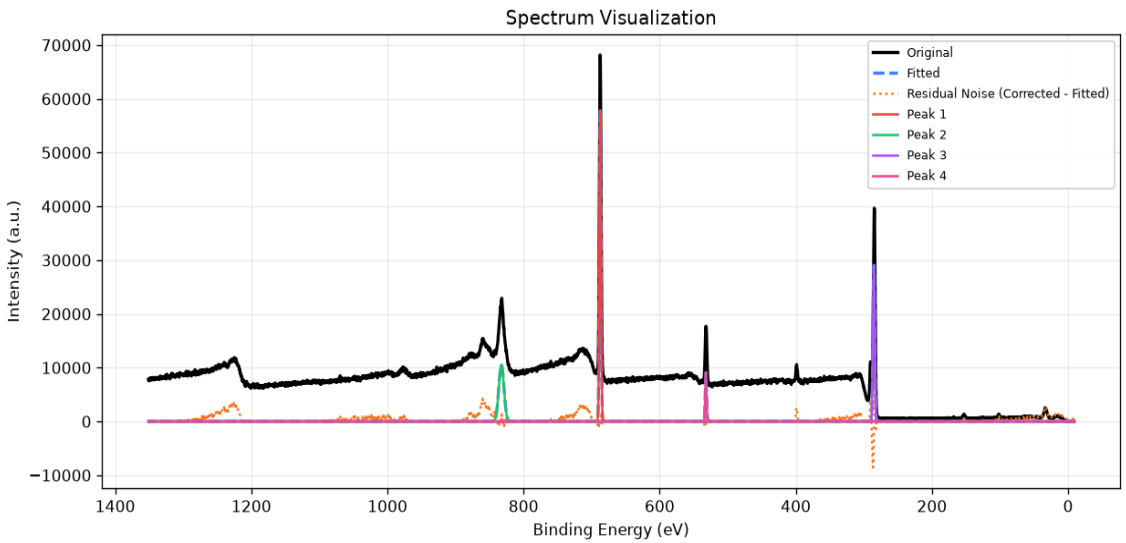


Figure 8. Suliga AS Sample 0, CG-DPF σ = 8 spectrum overlay.



ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Figure 9. Marx Patrick survey, CG-DPF σ = 8 spectrum overlay.

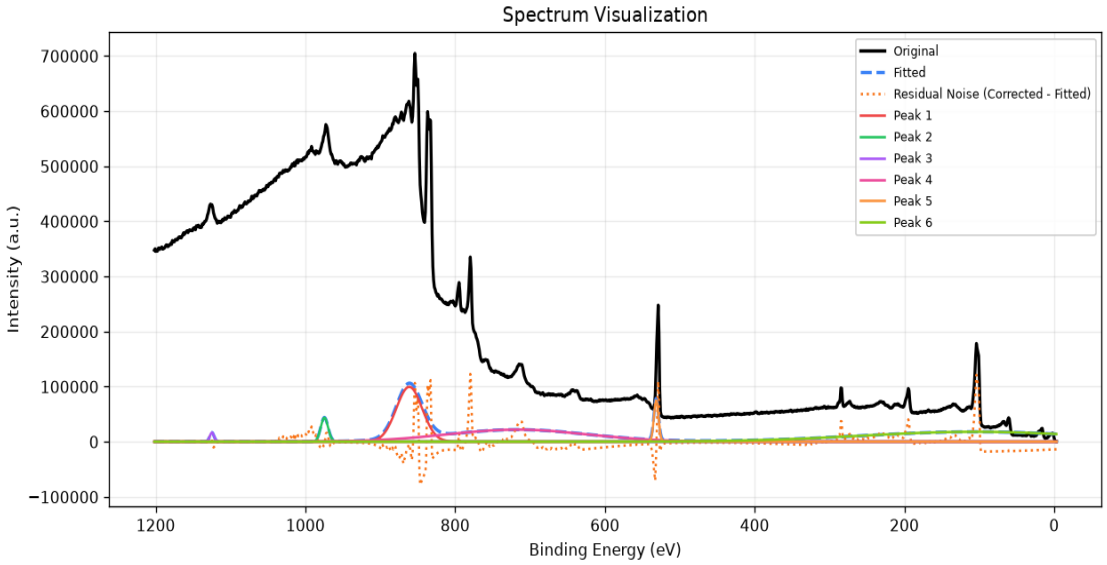
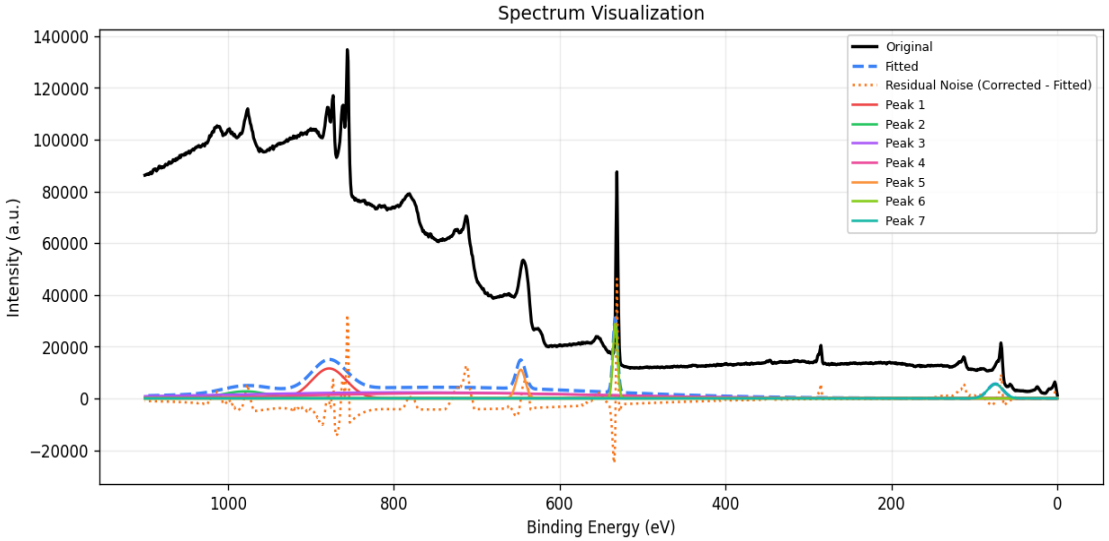


Figure 10. Longo OAD XPS survey, CG-DPF σ = 8 spectrum overlay.



DISCUSSION

CG-DPF does not replace expert interpretation and does not determine chemical-state assignments; it supplies reproducible initial centers and width screening only. FWHM compliance does not imply chemical correctness.

Evaluation limits. The primary benchmark comprises eight survey spectra without expert-annotated ground-truth peak lists; reported metrics are surrogate measures. Precision, recall, and F1 require reference annotations not available here. The small corpus limits generalization; bootstrap or cross-validation on larger annotated datasets remains future work [3]. External Marx [8] and Longo [7] check further show that survey-default σ and k can yield low Phys. Ratio and poor Corr. R^2 on oxide thin-film surveys and narrow Ni 2p windows, so external validation here is diagnostic of transfer limits rather than confirmatory of universal defaults.

Future evaluation. CG-DPF should next be assessed on larger and more diverse XPS collections that span different materials, chemical environments, and experimental conditions (survey vs. region scans, laboratory XPS vs. HAXPES, varied SNR and energy sampling). In parallel, the field would also benefit from standardized benchmark datasets with expert-annotated peak assignments, enabling objective comparison of initialization methods using precision, recall, F1 score, and related metrics rather than surrogate FWHM compliance and reconstruction quality.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VI June 2026

Baseline coverage. Quantitative comparison is limited to in-tool Adaptive Detection presets under identical preprocessing. Benchmark against Savitzky-Golay derivative peak picking [10], scipy.signal.find_peaks (implemented in SciPy [14]), MATLAB findpeaks, or wavelet detectors, was not conducted as such comparisons would strengthen claims of practical advantage but are outside the present software-unified study.

Method limits. CG-DPF requires expert threshold tuning for difficult spectra, derivative quality depends on energy sampling, low-SNR features may remain undetectable at conservative k, single-scale σ cannot optimally resolve peaks of very different widths, and peak-type classification (core-level vs. satellite vs. plasmon) remains the analyst's responsibility. Asymmetric line shapes, minimum area, and binding-energy priors are not enforced. Full reproducibility additionally requires reporting TRF stopping tolerances and any non-default background model.

Adaptive threshold recommendation. Further development could investigate adaptive methods that recommend curvature thresholds (θ) and smoothing parameters (σ) directly from spectrum statistics while preserving analyst override, transparent parameter logging, and interpretability on the d^2I/dE^2 plot. Such approaches should complement, rather than replace, the human-in-the-loop philosophy of CG-DPF.

CG-DPF should therefore be regarded as an optimizer-independent preprocessing and initialization strategy. It can be integrated with conventional Gaussian, Lorentzian, Voigt, or pseudo-Voigt fitting workflows using Levenberg-Marquardt, Trust Region Reflective, or other nonlinear least-squares optimizers.

CONCLUSION

This work aimed to improve the robustness and reproducibility of initial peak selection for sequential XPS spectral deconvolution by introducing an interpretable curvature-guided filtering framework. CG-DPF augments first-derivative zero-crossing detection with tunable second-derivative curvature filtering, reducing false positives (47 to 5 at $\sigma = 8$, $k = 1$ on the in-house benchmark) while retaining major core-level features. The scientific contribution is a filtering strategy and reproducible workflow curvature as a post-crossing acceptance criterion with logged thresholds rather than a new differentiation formula. Validation on seven independent Suliga survey spectra confirms that CG-DPF consistently favors FWHM-compliant initialization over peak count relative to adaptive baselines (mean Phys. Ratio 0.66 vs. 0.30 for Adaptive Medium, Wilcoxon $p = 0.031$, although the sample size is small). External Marx and Longo datasets demonstrate that the same defaults do not automatically generalize to all published oxide and region spectra, underscoring the need for analyst-guided σ/k returning. The default setting ($k = 1$) supports conservative initialization on survey-like data; operational analysis may prefer $k \approx 0.3–0.5$ for improved recall ($R^2 = 0.985$ at $k \approx 0.32$ on the in-house benchmark). Transparent threshold control, integrated width screening, sequential fitting support, and reproducible workbook export provide an auditable initialization path between manual peak placement and fully automatic detection. The method is implemented in XPS Spectra Deconvolution Tool v2.3 [5]. Future work should evaluate CG-DPF on substantially larger and more diverse XPS datasets spanning different materials, chemical environments, and acquisition conditions. Additional research should establish annotated benchmark datasets for objective comparison and investigate adaptive curvature-threshold recommendation while preserving analyst control and interpretability.

This work aimed to improve the robustness and reproducibility of initial peak selection for sequential XPS spectral deconvolution by introducing an interpretable curvature-guided filtering framework. CG-DPF augments first-derivative zero-crossing detection with tunable second-derivative curvature filtering, reducing false positives (47 to 5 at $\sigma = 8$, $k = 1$ on the in-house benchmark) while retaining major core-level features. The scientific contribution is a filtering strategy and reproducible workflow curvature as a post-crossing acceptance criterion with logged thresholds, rather than a new differentiation formula. Validation on seven independent Suliga survey spectra confirms that CG-DPF consistently favors FWHM-compliant initialization over peak count relative to adaptive baselines (mean Phys. Ratio 0.66 vs. 0.30 for Adaptive Medium, Wilcoxon $p = 0.031$, although the sample size is small). External Marx and Longo datasets demonstrate that the same defaults do not automatically generalize to all published oxide and region spectra, underscoring the need for analyst-guided σ/k returning. The default setting ($k = 1$) supports conservative initialization on survey-like data; operational analysis may prefer $k \approx 0.3–0.5$ for improved recall ($R^2 = 0.985$ at $k \approx 0.32$ on the in-house

benchmark). Transparent threshold control, integrated width screening, sequential fitting support, and reproducible workbook export provide an auditable initialization path between manual peak placement and fully automatic detection. The method is implemented in XPS Spectra Deconvolution Tool v2.3 [5]. Future work will evaluate CG-DPF on larger annotated XPS datasets and extend the framework with adaptive threshold selection and chemically informed constraints.

ACKNOWLEDGEMENTS

This work was developed as part of the XPS Spectra Deconvolution Tool project. Thanks to the open-source Python ecosystem (NumPy, SciPy, Pandas, Flask) for the computational foundation. The Suliga survey spectra are used under their published Figshare license for primary external validation. Additional published datasets from Longo and Marx et al. (Zenodo) were used for external transfer checks.

REFERENCES

1. Bajrai, S., "XPS_Spectra_Deconvolution," v2.4, Zenodo, 2026. https://doi.org/10.5281/zenodo.21143077
2. Briggs, D. and Seah, M. P. (Eds.), *Practical Surface Analysis, Volume 1: Auger and X-ray Photoelectron Spectroscopy*, 2nd ed. Chichester, UK: John Wiley & Sons, 1990.
3. Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap*. New York, NY, USA: Chapman & Hall/CRC, 1993.
4. Fairley, N., Fernandez, V., Richard-Plouet, M., Guillot-Deudon, C., Walton, M., Smith, E., Flahaut, D., Greiner, S., Biesinger, M., Tougaard, M., and Morgan, S., "Systematic and Collaborative Approach to Problem Solving Using X-ray Photoelectron Spectroscopy," *Applied Surface Science Advances*, vol. 5, Article 100112, 2021. https://doi.org/10.1016/j.apsadv.2021.100112
5. Hiro, *XPS Spectra Deconvolution Tool v2.3: Curvature-Guided Derivative Peak Filtering for Sequential XPS Spectral Deconvolution* (software), Version 2.3, 2026. Source code and benchmark datasets available from the project repository.
6. Lee, A. Y., Powell, C. J., Gorham, J. M., Morey, A., Scott, J. H. J., and Hanisch, R. J., "Development of the NIST X-ray Photoelectron Spectroscopy (XPS) Database, Version 5," *Data Science Journal*, vol. 23, Article 45, 2024. https://doi.org/10.5334/dsj-2024-045
7. Longo, F. (2024). Hard X-ray photoelectron spectroscopy probing Fe segregation during oxygen evolution reaction [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.13908770
8. Marx, P., Shukla, S., Pérez Mendoza, A. E., Lourens, F., Andronescu, C. & Ludwig, A. (2025). Dataset - Structural and Electrocatalytic Properties of La-Co-Ni Oxide Thin Films [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.15754300
9. O'Haver, T. C., *Introduction to Signal Processing in Chemical Spectroscopy*. University of Maryland. Available: https://www.terpconnect.umd.edu/~toh/spectrum/Differentiation.html (accessed July 2026).
10. Savitzky, A. and Golay, M. J. E., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627-1639, 1964. https://doi.org/10.1021/ac60214a047
11. Shirley, D. A., "High-Resolution X-ray Photoemission Spectrum of the Valence Bands of Gold," *Physical Review B*, vol. 5, no. 12, pp. 4709-4714, 1972. https://doi.org/10.1103/PhysRevB.5.4709
12. Suliga, A., "XPS data," Figshare, 2018. https://doi.org/10.6084/m9.figshare.5671645
13. Tougaard, S., "Quantitative Analysis of the Inelastic Background in Surface Electron Spectroscopy," *Surface and Interface Analysis*, vol. 11, no. 9, pp. 453-472, 1988. https://doi.org/10.1002/sia.740110902
14. Virtanen, P. et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261-272, 2020. https://doi.org/10.1038/s41592-019-0686-2
15. Workman Jr., J. and Weyer, L., "Derivative Spectroscopy," in L. Weyer and J. Workman Jr. (Eds.), *Handbook of Signal Processing in Spectroscopy*, Academic Press, 2012, ch. 15, pp. 1-28.