

Article

Not peer-reviewed version

---

# A Multi-Parameter Persistence Algorithm for the Automatic Energy Calibration of Scintillating Radiation Sensors

---

Guglielmo Ferranti , [Chiara Rita Failla](#) , [Paolo Finocchiaro](#) <sup>\*</sup> , [Alessandro Pluchino](#) , [Andrea Rapisarda](#) , [Salvatore Tudisco](#) , [Gianfranco Vecchio](#)

Posted Date: 6 June 2025

doi: 10.20944/preprints202506.0487.v1

Keywords: Peak Detection; Automatic Energy Calibration; Topological data analysis; Multi-parameter Persistence; Persistent Homology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

## Article

# A Multi-Parameter Persistence Algorithm for the Automatic Energy Calibration of Scintillating Radiation Sensors

Guglielmo Ferranti <sup>1,2,\*</sup> , Chiara Rita Failla <sup>1,3</sup> , Paolo Finocchiaro <sup>3,\*</sup> ,  
Alessandro Pluchino <sup>1,2</sup> , Andrea Rapisarda <sup>1,2,4</sup> , Salvatore Tudisco <sup>3</sup>   
and Gianfranco Vecchio <sup>3</sup> 

<sup>1</sup> Department of Physics and Astronomy, University of Catania, Via S. Sofia 64, 95123 Catania, Italy

<sup>2</sup> Sezione di Catania, INFN, Via S. Sofia 64, 95123 Catania, Italy

<sup>3</sup> Laboratori Nazionali del Sud, INFN, Via S. Sofia 62, 95123 Catania, Italy

<sup>4</sup> Complexity Science Hub, Metternichgasse 8, 1030 Wien, Austria

\* Correspondence: finocchiaro@lns.infn.it (P.F.); guglielmo.ferranti@phd.unict.it (G.F.)

**Abstract:** Peak detection is a fundamental task in spectral and time-series data analysis across diverse scientific and engineering disciplines, yet traditional approaches are highly sensitive to the choice of algorithm parameters, complicating reliable and consistent interpretation. Triggered by the requirement of the energy calibration for the 128 detectors of the PI3SO gamma ray scanner, we introduce a versatile methodology inspired by concepts from persistent homology, extending the traditional notion of persistence to a multi-parameter setting. Our approach systematically explores the space defined by multiple detection parameters and quantifies peak robustness through the hyper-volume in parameter space where each peak is consistently identified. This volumetric multi-parameter persistence (VM-PP) measure enables robust peak ranking and significantly reduces the sensitivity of detection outcomes to individual parameter selection, demonstrating utility across simulated and experimental spectral datasets. Extensive validation reveals that this method reliably differentiates genuine peaks from noise-induced fluctuations under diverse noise conditions, proving effective in practical spectroscopic calibration scenarios. This framework, general by design, can be readily adapted to diverse signal-processing applications, enhancing interpretability and reliability in complex feature-detection tasks.

**Keywords:** peak detection; automatic energy calibration; topological data analysis; multi-parameter persistence; persistent homology

## 1. Introduction

Peak detection is a fundamental task in signal processing, underpinning applications in spectroscopy, mass spectrometry, astrophysics, and biomedical signal analysis [1,2]. This study was specifically motivated by the challenge of performing energy calibration for the 128 scintillation detectors of the PI3SO gamma-ray scanner [3]. In this context, spectra are acquired from standardized gamma-ray sources—typically  $^{137}\text{Cs}$ ,  $^{60}\text{Co}$ , and  $^{22}\text{Na}$ —and a linear calibration is then applied to convert the known peak positions from channel units to energy units. However, identifying the peaks and determining their centroids is still largely a manual or semi-manual process: even when automated routines are used, close supervision and frequent corrections are often required. Traditional algorithms, including those based on wavelet transforms [4] or built-in functions from mathematical software packages, depend heavily on user-defined hyperparameters such as peak width and inter-peak distance. These methods are highly sensitive to small changes in parameter values, often resulting in inconsistent outcomes: a slight adjustment can lead to missing genuine peaks or, conversely, detecting spurious peaks caused by noise [5,6]. Recent advances have sought to mitigate these issues by either training Deep Neural Networks on raw Liquid Chromatography–Mass Spectrometry (LC–MS) [5],

or employing statistical modeling to reduce reliance on arbitrary thresholds validating a Bayesian framework for peak detection on challenging data coming from two-dimensional *Gas Chromatography coupled with Time-Of-Flight mass spectrometry* (GC×GC-TOF MS) [2]. Similarly, Kilgour et al. [7] introduced an unsupervised threshold optimization approach tailored to mass spectrometry, significantly enhancing detection reliability across instruments. While these methods mitigate noise and offer improvements in robustness, they either remain highly sensitive to individual hyper-parameter settings, or often obscure *why* a peak is retained or rejected, leaving room for approaches that more transparently quantify feature stability. Motivated by these challenges, our work draws inspiration from the field of topological data analysis (TDA) and the concept of persistent homology, which has been celebrated for its ability to capture the underlying shape of data through so called *barcode plots* that display the interval for which a feature is detected when varying a single *filtration parameter*: by plotting the ranges in which all features (e.g. peaks in a noisy signal) appear and vanish (the *birth-death* intervals), barcode plots highlight long-lived, and therefore significant, features [8]. Along this line, Scholkmann and colleagues [1] have presented the Automatic Multiscale Peak Detection (AMPD) algorithm, which considers a peak *robust* if it remains a local maximum at many scales on *periodic* data. Although effective in single-parameter scenarios, rigorously extending these ideas to multiple parameters encounters fundamental challenges as barcode plots fully characterize data shape for single-parameter filtrations only while, as shown by Carlsson and Zomorodian [9], no equally simple and complete representation exists for multiparameter settings. To address this, we introduce a practical framework that aggregates peak detection outcomes across a multidimensional hyperparameter space  $\Theta$ , deliberately simplifying the representation by disregarding detailed shape information. We quantify peak robustness as a scalar measure, defined as the accumulated hyper-volume within the hyperparameter space  $\Theta$  over which each peak is consistently identified. Concretely, each detected peak corresponds to a specific sub-region of  $\Theta$ , as shown in Figure 1: panels (b) and (c) in particular show that peaks which appear visually more significant in the raw spectra naturally occupy larger associated hyper-volumes, providing an intuitive basis for ranking stability using their persistence. While exploring an  $m^n$  hyperparameter space—where  $m$  represents the number of discretized values per parameter and  $n$  the number of parameters—can potentially lead to significant computational costs, we effectively mitigate this by employing computationally efficient peak detection routines and extensive parallelization techniques, keeping the runtime of the energy calibration conducted in Section 3.2 at around 40 spectra (each consisting of 2048 channels) per second running the code locally on a laptop. Additionally, our framework is well-suited for future enhancements such as adaptive sampling strategies on the parameter space  $\Theta$ , which can significantly reduce computational overhead by concentrating computational resources in regions of  $\Theta$  where peak detection outcomes exhibit greater variability.

The remainder of the paper is organized as follows. Section 2 describes our methodology in detail, Section 3.1 presents validation on simulated data, and Section 3.2 demonstrates the application of our method to the real case of the PI3SO spectroscopic data.

## 2. Materials and Methods

The identification of peaks in spectral or time-series data typically requires careful tuning of various hyperparameters. Traditional methods evaluate these parameters individually or in limited combinations, resulting in peak detection outcomes that are highly sensitive to specific settings. Our methodology, inspired by the field of Topological Data Analysis and *persistent homology* techniques [8–11], addresses this by introducing the concept of **Volumetric Multi-Parameter Persistence** or **VM-PP**. Here, the robustness of a detected peak is quantified by systematically exploring and aggregating detection results across a multi-dimensional parameter space.

Formally, we define the hyperparameter space  $\Theta$  as a discrete grid spanned by the following parameters, each defined as:

- **Smoothing window size:** number of neighboring data points averaged to smooth short-term fluctuations.
- **Bin aggregation factor:** number of adjacent spectral channels combined into one, reducing spectral resolution and noise.
- **Prominence:** the minimum vertical distance between the peak and the lowest point one must descend to reach a higher peak. It quantifies how much a peak stands out from its surrounding landscape—not just its immediate neighbors—in the context of the whole signal.
- **Threshold:** the minimum vertical difference between the peak and its *immediate neighbors*. Unlike prominence, this is a strictly local criterion: it filters out small fluctuations or noise spikes that do not sufficiently rise above their direct surroundings.
- **Width:** required minimum width of detected peaks at half-prominence.
- **Peak-to-peak distance:** minimum separation allowed between neighboring peaks.

The grid cells in this parameter space correspond to unique parameter combinations, each having equal volume  $\Delta V$ . While cell volume could be defined proportional to parameter step sizes, such an approach may disproportionately emphasize parameters explored more coarsely. Thus, for simplicity and fairness, we assign each cell unitary volume, making the persistence of a peak simply equal to the count of parameter cells where it is detected:

$$P(f) = \sum_{k \in K(f)} \Delta V_k \quad (1)$$

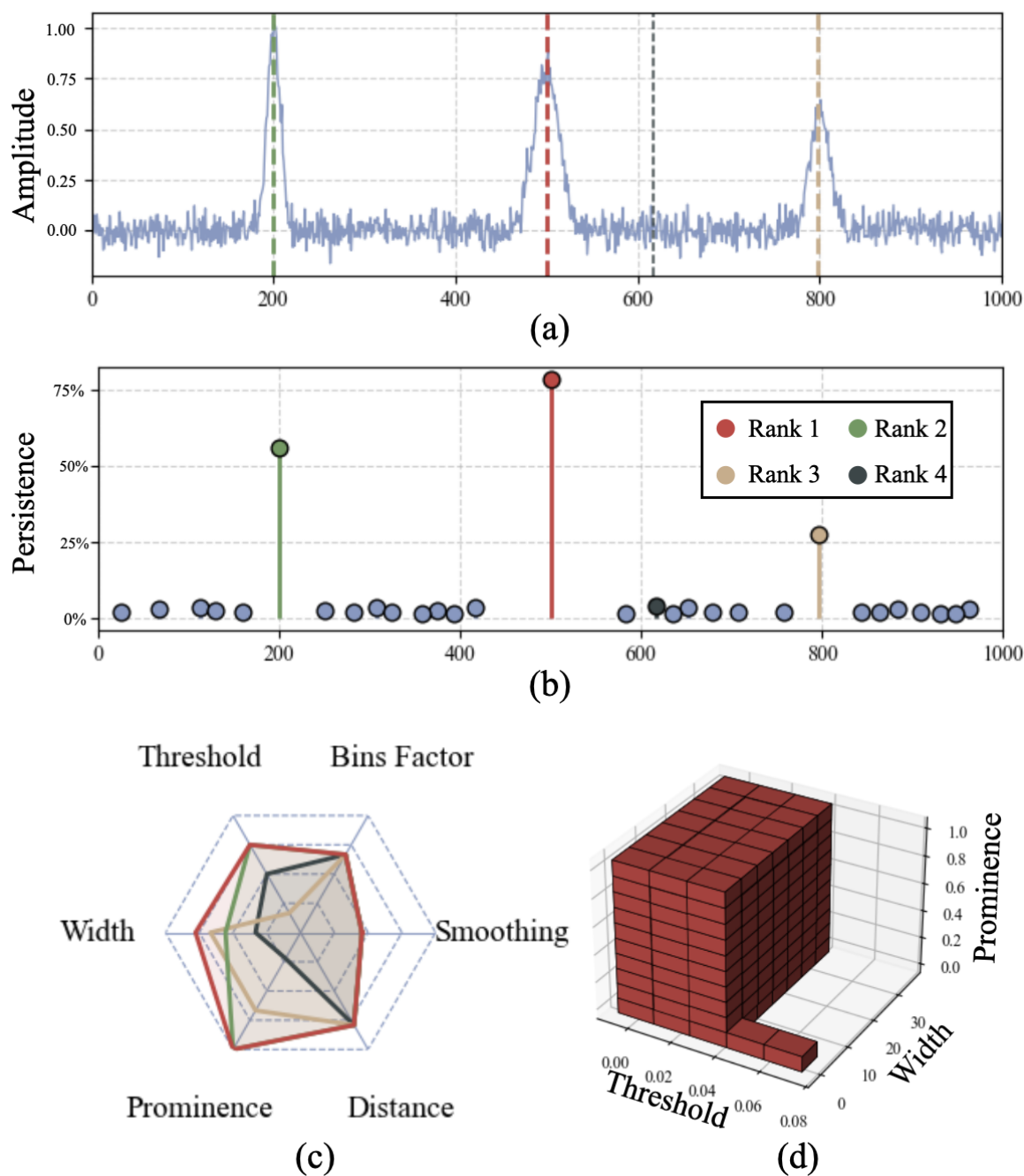
where  $K(f)$  denotes the set of cells in parameter space where feature  $f$  is identified. In other words, under this simplification, the method simply measures the fraction of parameter combinations that result in a peak being detected, using this value as a metric proportional to the peak significance. While this may seem a trivial approach, our tests (exemplified visually by Figures 1 and 2) show a remarkable ability of the method to robustly separate genuine peaks from noise induced fluctuations, minimizing spurious detections when evaluated on simulated signals (Figure 3). Even though in the following we will always apply this simplification, the method remains flexible for future tasks that may benefit from setting specific weights to each explored dimension in order to tune each cell's volume based on the impact of individual parameters on peak detection.

Peak detection within each cell utilizes Python's `scipy.signal.find_peaks` function. To enhance computational efficiency, the parameter exploration is structured in two nested grids: an outer grid iterates over preprocessing parameters (smoothing and bin aggregation), while an inner grid extensively covers detection parameters (prominence, width, threshold, and distance). For each outer grid configuration, data is preprocessed once, and peak detection is executed in parallel across all inner grid configurations. Detected peaks in closely adjacent positions (within a predefined spatial merging tolerance) are merged and mapped back to their original spectral channels when bin aggregation is applied. Each peak thus acquires a persistence value quantifying its detection robustness across the parameter space. Peaks are ranked according to their persistence scores, which correlate directly with physical significance and detection stability.

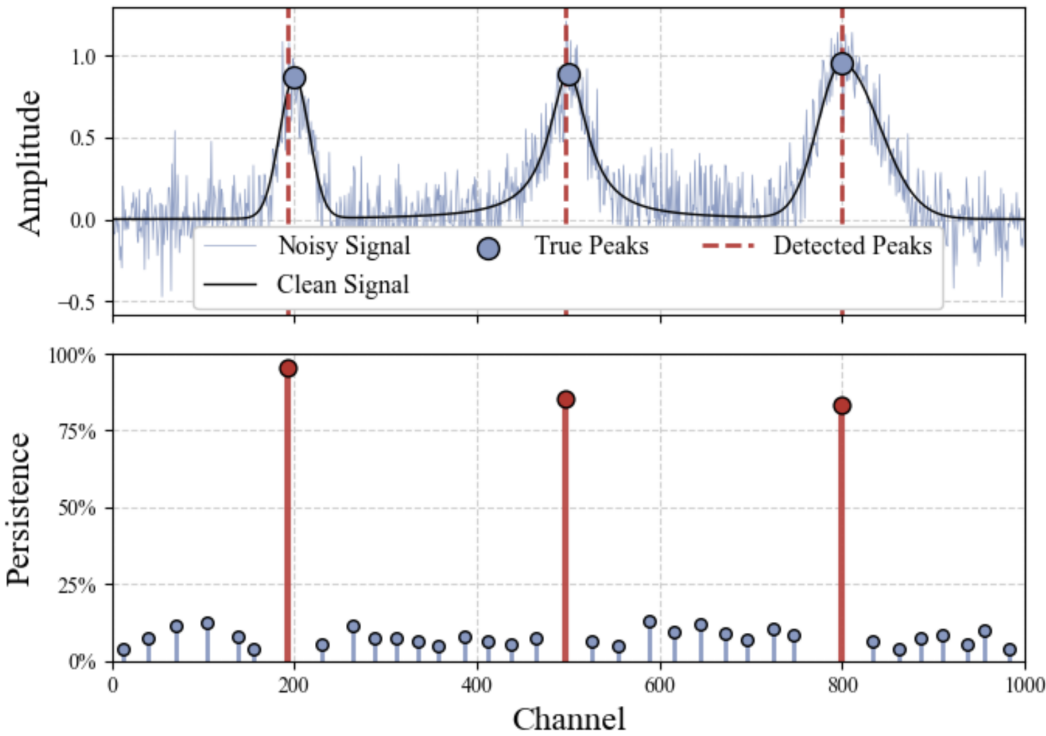
For illustrative clarity, Figure 1 panels (c) and (d) visualize this concept explicitly, showing the hyper-volumes occupied by the top ranked peaks in the parameter space. Specifically, panel (c) shows a linear projection of all 6 explored dimensions that is useful for tuning the ranges of parameters to explore, while panel (d) shows the 3D Width-Threshold-Prominence parameter sub-space, with the region occupied by the most persistent peak of the signal in panel (a) colored in red. This approach is able to effectively separate genuine peaks from noise induced fluctuations even in intense noise conditions, as showcased in Figure 2, where the amplitude of fluctuations has a comparable scale to the clean signal. When the number  $k$  of expected peaks in a spectrum is known a-priori, as is the case for the spectra employed for energy calibration detailed in Section 3.2, one can simply rank all detected peaks by their persistence and take the *top  $k$  most persistent* peaks, discarding all others without the need of additional hyper-parameters. For more agnostic exploratory tasks where the



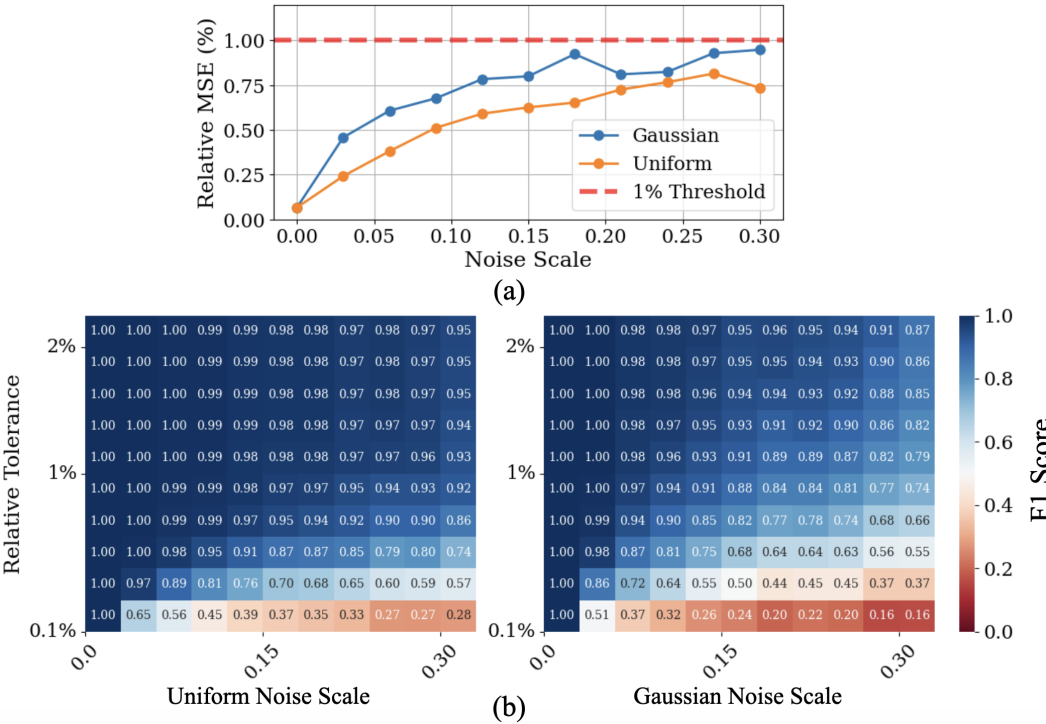
number of expected peaks is not known, we can select a persistence threshold below which peaks are discarded: the bottom panel of Figure 2 shows that the method effectively separates genuine peaks from noise induced detections, making the appropriate range for this threshold quite wide even for very noisy spectra.



**Figure 1.** (a) Simulated signal, obtained by adding a bounded-range noise ( $\sigma = 0.05$ ) to a clean spectrum consisting of 3 gaussian peaks, overlaid with the top 4 most persistent detected peaks: because only 3 true peaks exist, rank 4 necessarily corresponds to noise. (b) True peaks (ranks 1–3) exhibit markedly higher persistence than any noise-induced candidate; the rank 4 point (dark circle) lies well below the true peaks, demonstrating easy separability. (c) Radar chart projecting the six-dimensional persistence hyper-volumes to 2D. This is useful for tuning the parameter ranges to explore but, because of linear projection, the axis magnitudes do not preserve the true volume ratios seen in panel b. (d) 3D slice of the threshold–width–prominence subspace, with the region in which the rank 1 peak is consistently detected shaded in red.



**Figure 2.** *Top panel:* simulated test spectra composed of Gaussian, Lorentzian, and asymmetric peaks under intense Gaussian noise ( $\sigma = 0.15$ ). Vertical lines indicate detected peak channels. *Bottom panel:* VM-PP of peaks expressed in percentage of the total parameter space, allowing to discard peaks that fall below the 25% persistence threshold.



**Figure 3.** Performance of the multi-parameter persistence approach under varying noise conditions. Panel (a) plots the Relative Mean-Square-Error (%) against Noise-Scale for Gaussian (blue) and Uniform bounded-range (orange) noise, with a horizontal red line marking the 1% threshold; Panel (b) displays maps of detection rate as a function of noise scale (x-axis) and relative tolerance (y-axis) for Uniform (left) and Gaussian (right) noise types. Cooler colors in the heatmaps indicate higher detection rates. It is worth underscoring that the ranges explored represent quite heavy noise conditions, as shown in Figure 2 which reports an example spectrum under gaussian noise with  $\sigma = 0.15$ .

To contextualize computational performance, our implementation processes approximately 30 to 50 spectra per second (depending on the resolution and size of the parameter space) using parallel computation on a standard laptop (Python version 3.11.x, tested primarily on macOS Sequoia 15.4 but portable to any platform supporting Python), for spectra consisting of 2048 channels. Despite the exponential scaling inherent to a full  $m^n$  exploration of the parameter space (with  $m$  steps per  $n$  parameters), computational costs can be mitigated: specifically, the use of `scipy.signal.find_peaks` ensures minimal per-cell runtime, while independent parameter grid cells are evaluated in parallel using multiprocessing or distributed frameworks. Moreover, because the detection function is piecewise continuous (Appendix A) over the parameter space, adaptive sampling strategies can be employed in future implementations to reduce the number of necessary evaluations. The methodology described has been fully implemented in Python, leveraging parallelization libraries to ensure efficiency. To foster reproducibility and support the scientific community, our implementation, along with an interactive demonstration, is publicly accessible via GitHub [18].

### 3. Results

The following section is divided into two parts: in the first, we benchmark our method on synthetic data and evaluate its performance upon varying noise conditions; in the second, we showcase an application of our method to real spectroscopy data.

#### 3.1. Synthetic Data Benchmark

To evaluate the robustness of our multi-parameter persistence method under controlled conditions, we generated a synthetic dataset consisting of 5000 test spectra. Each test spectrum contained three distinct peaks: one Gaussian, one Lorentzian, and one asymmetric, carefully positioned to avoid overlap. The amplitudes of these peaks were normalized so that each clean (noise-free) spectrum remained within the range  $(0, 1)$ , ensuring that no individual peak exceeded an amplitude of 1.

To simulate realistic measurement conditions, we introduced two types of noise separately into each test spectrum:

- **Uniform noise**, where each data point was randomly perturbed by an amount between  $-a$  and  $+a$ , with the noise amplitude  $a$  varied from 0 to 0.3.
- **Gaussian noise** (Additive White Gaussian Noise, AWGN), where each data point was perturbed by a value drawn from a Gaussian distribution with mean 0 and standard deviation  $\sigma$  varied from 0 to 0.3.

An example synthetic spectrum with moderate Gaussian noise ( $\sigma = 0.15$ ) is shown in Figure 2. Detected peaks are marked by vertical dashed lines. The bottom panel illustrates peak persistence scores, clearly showing how the three true peaks stand out significantly from random noise fluctuations.

The performance of our detection method across varying noise levels is summarized in Figure 3. To quantify detection accuracy clearly, we defined two separate metrics:

- **Relative Mean Detection Error**: For each detected peak, we measured the absolute difference between the predicted channel (detected peak position) and the actual known peak position. This error was normalized relative to the full range of the spectrum to produce a percentage error. We then averaged this error over all three peaks in each spectrum and over all 5000 test spectra for each noise level. Panel (a) displays this averaged error metric, showing how detection accuracy degrades with increasing noise.
- **Detection Rate (F1 Score)**: To assess how reliably peaks were found, we used a common classification metric known as the F1 score. This metric combines two important aspects into a single value:
  1. **Recall**: The fraction of actual peaks successfully detected.
  2. **Precision**: The fraction of detected peaks that corresponded to actual (rather than noise-induced) peaks.

The F1 score is defined as the *harmonic mean* of Recall and Precision and reaches its maximum value of 1.0 (100%) when all true peaks are detected correctly without false positives. For instance, an F1 score of 0.98 indicates that overall, about 2% of detections were incorrect—either missed true peaks or included spurious noise peaks.

Panels (b) and (c) of Figure 3 show the F1 score for varying noise levels and different tolerance values. The **relative tolerance** refers to how close a detected peak must be to the true peak position to be considered a correct detection, again expressed as a percentage of the full spectral range. Darker blue areas in the heatmaps indicate excellent detection performance (close to 1), whereas lighter (warmer) colors indicate lower reliability.

The heatmaps demonstrate that our approach maintains strong performance under substantial noise. The uniform noise scenario (panel b) consistently showed excellent performance, while the Gaussian noise scenario (panel c) exhibited slightly lower but still robust detection rates, reflecting the more challenging nature of unbounded Gaussian fluctuations.

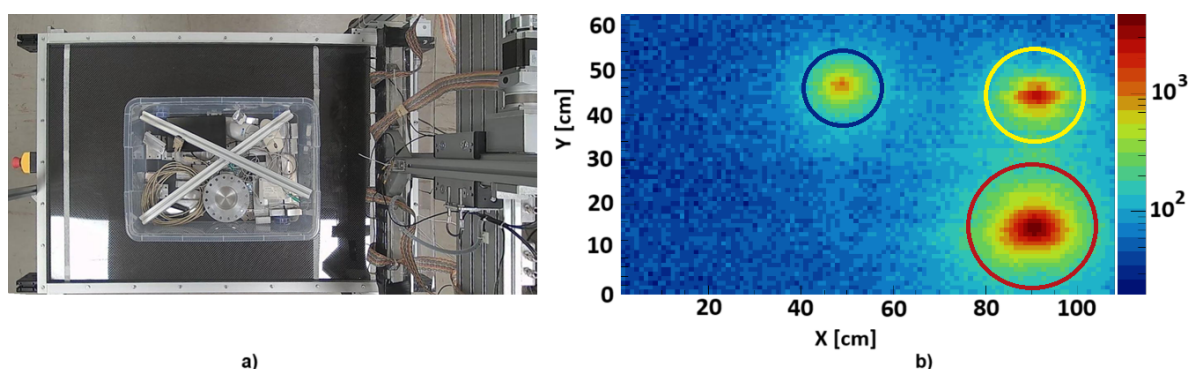
In summary, the synthetic tests clearly demonstrate that our multi-parameter persistence approach reliably detects and accurately locates peaks even under high noise conditions. The results underscore both the method's ability to consistently identify all peaks (high recall) and avoid false detections (high precision), qualities that directly translate into practical advantages when applied to experimental spectroscopic data.

### 3.2. Validation on Spectroscopy Data

To validate the effectiveness of our peak detection and classification methodology, we utilized experimental gamma-ray spectroscopy data obtained using the Proximity Imaging System for Sort and Segregate Operations (PI3SO) [3]. This advanced spectroscopic system automates the identification of gamma-ray emitting sources within radioactive waste, specifically targeting Intermediate Level Waste (ILW), Low Level Waste (LLW), and Very Low Level Waste (VLLW) [12,13]. The PI3SO system can improve radwaste (re)conditioning and management by rapidly identifying radioactive hotspots with minimum human intervention.

#### 3.2.1. PI3SO Spectroscopic System

The PI3SO project primarily addresses two interconnected tasks: *Hot Spot Search* and *Gamma Spectrometry*. Hot spot search involves identifying areas containing radioactive objects (see Figure 4-Panel (a)). Once identified, these hotspots become regions of interest (ROI) for subsequent detailed gamma spectroscopic analyses, as illustrated in Figure 4-Panel (b).

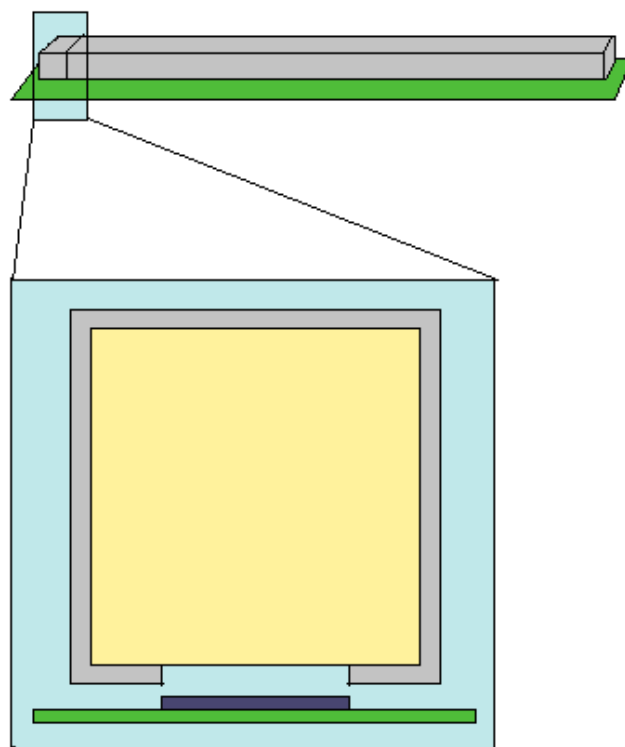


**Figure 4.** (a): System operation diagram: Radioactive waste is distributed over the scanning table to identify radioactive hotspots. (b): Visualization of the result of a scan of three different radioactive point-like sources; color scale on the right report the activity in cps. Red circle: source of <sup>137</sup>Cs with an activity of 1.4000 MBq; yellow circle: source of <sup>60</sup>Co with an activity of 0.0560 MBq; blue circle: source of <sup>22</sup>Na with an activity of 0.0154 MBq.

The PI3SO instrumentation consists of a robust table coupled with a mechanical sliding bridge. Two linear arrays of 64 gamma-ray detectors, are arranged above and beneath the table. Each array



contains four modules of 16 detectors, each detector coupling a cubic ( $1\text{ cm}^3$ ) CsI(Tl) scintillator with a Silicon Photomultiplier (SiPM). The detectors are encapsulated in reflective casings enhancing photon collection efficiency. The structural layout and module configuration are detailed in Figure 5.



**Figure 5.** *Top:* Sketch of a single component of one of the two linear arrays, which consists of 4 modules like the one shown, placed next to each other in a row; each module houses a total of 16 CsI(Tl) embedded in a reflective resin case and coupled to as many SiPMs that are housed on an electronic board. *Bottom:* Sketch of a single detector: in gray is the reflective mask in which the  $1\times1\times1\text{ cm}^3$  crystal of CsI(Tl) (in yellow) is embedded; the latter is optically coupled through its free face to a  $6\times6\text{ mm}^2$  SiPM (blue), which is housed in the electronic board (green).

CsI(Tl) crystals were chosen for their optimal properties, including high-density ( $4.51\text{ g/cm}^3$ ), superior energy resolution, and significant light yield ( $60,000\text{ photons/MeV}$ ) with a peak emission around  $550\text{ nm}$  [14,15]. Each scintillator is coupled to an MPPC (Multi-Pixel Photon Counter) SiPM from Hamamatsu, featuring an active area of  $6\times6\text{ mm}^2$  and 14,331 cells [16]. Previous tests confirmed an energy resolution (FWHM) around 5-6% for the entire detection system [3]. Signals from each detector are digitized using two VX2745 digitizers [17], each one capable of simultaneously sampling signals from 64 SiPM channels at  $125\text{ MSamples/s}$  with a 16-bit resolution. Data acquisition is synchronized via timestamps, allowing precise temporal alignment of detected events. The acquisition system, interfaced via Ethernet, supports real-time data analysis, spectral generation, and comprehensive event logging. During typical calibration or operational measurements, nearly all 128 detectors are actively engaged, generating extensive datasets. Each dataset typically comprises multiple gamma-ray spectra requiring energy calibration and isotope identification. Given the complexity and volume of data, manual analysis is impractical. Consequently, an automated, robust peak identification system becomes essential.

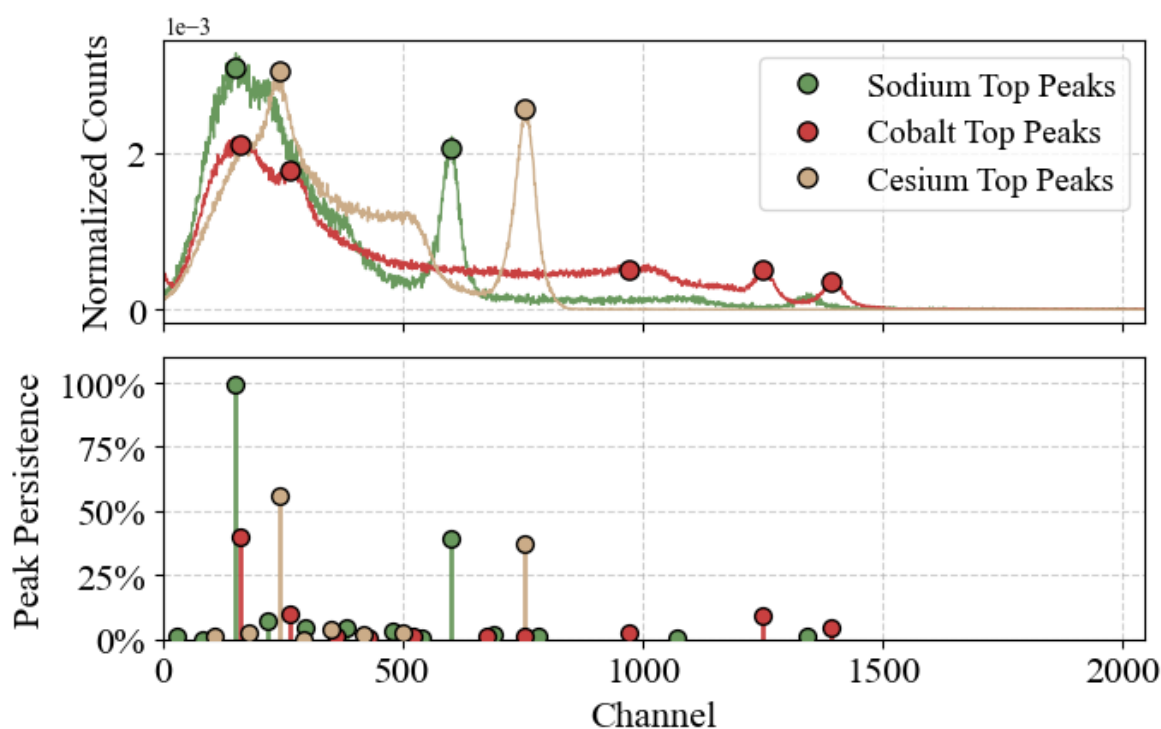
Our validation procedure leveraged detector calibration tasks involving three standardized radioactive sources:  $^{137}\text{Cs}$  ( $1.4000\text{ MBq}$ ,  $0.662\text{ MeV}$ ),  $^{60}\text{Co}$  ( $0.0560\text{ MBq}$ ,  $1.173\text{ MeV}$  and  $1.331\text{ MeV}$ ), and  $^{22}\text{Na}$  ( $0.0154\text{ MBq}$ ,  $0.511\text{ MeV}$  and  $1.274\text{ MeV}$ ). Each of the 128 detectors was exposed to these sources individually to validate automated peak detection and classification without manual intervention.

### 3.2.2. Peak Ranking and Automatic Selection

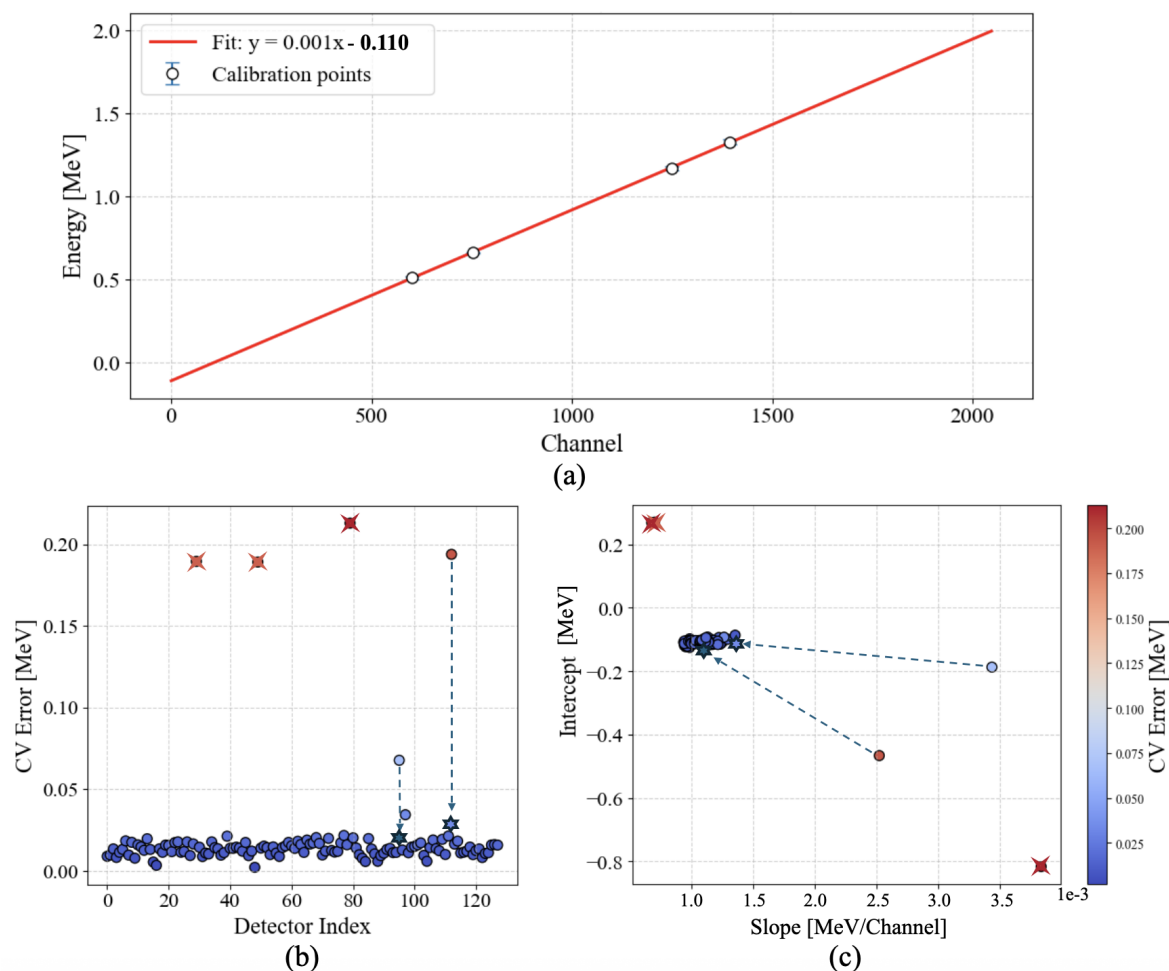
Figure 6 illustrates the application of our VM-PP method to typical emission spectra from Sodium, Cesium, and Cobalt radioactive sources, for which the noise scale was estimated between  $\sigma = 0.05$  and  $\sigma = 0.10$  (lower noise scales could be achieved at the cost of increasing the exposure time), highlighting the robustness of the peak-ranking process across spectra with varying complexities. Using our multi-parameter persistence ranking, we select from each detector the top  $k$  most persistent peaks. The choice of  $k$  depends on the source's spectral shape, but requires almost no additional filtering:

- For **Sodium**, we set  $k = 2$  to catch both the backscattering peak and the main emission peak at higher energy (0.511 MeV).
- For **Cesium**,  $k = 2$  again recovers the main peak (0.662 MeV) plus a secondary region at lower energy.
- For **Cobalt**,  $k = 5$  covers the overall complexity of the spectrum well; the two highest-energy peaks of those five correspond to the known 1.17 and 1.33 MeV lines.

Initially, three detectors were excluded from the analysis due to severe hardware damage, clearly identified by significant deviations in their calibration fits (Figure 7 (b-c)). Among the remaining 125 detectors, the correct calibration peaks were identified directly within the top- $k$  sets for 123 cases without any additional processing. However, in two detectors, the peaks of interest were ranked just outside the top- $k$  set, appearing as the  $(k + 1)$ -th most persistent peaks. We found that minimal additional filtering—such as applying a broad energy/channel cut to eliminate clear backscattering peaks—was sufficient to recover these peaks accurately. Ultimately, using the persistence-based ranking approach combined with this minor post-processing step in the two challenging instances, we successfully identified *all* four targeted emission peaks across the full set of 125 operational detectors.



**Figure 6.** VM-PP applied to peak detection on 3 sample sources: *top panel* shows the emission spectra of Sodium, Cobalt and Cesium sources, normalized by area under curve and annotated with the top  $k$  most persistent peaks for each source; *bottom panel* shows the *persistence* of detected peaks, reported as the percentage of volume in parameter space where the peak is detected. For all 3 spectra, despite significant differences in peak shapes, the same portion of parameters space is explored, showcasing the versatility and robustness of our method.



**Figure 7.** Calibration results: (a) An example linear best fit on a single detector's data, using 4 emission peaks of known energy from Sodium (0.511 MeV), Cesium (0.662 MeV) and Cobalt (1.17 and 1.33 MeV). (b) The leave-one-out (LOO) cross-validation (CV) error of each detector's calibration, highlighting 3 outlier crossed points resulting from faulty detectors. The 2 remaining points with poor performance are easily recovered with minimal post processing steps, the results of which are shown as stars. (c) Distribution of best-fit slope and intercept parameters across all detectors, colored by the LOO-CV error. Negative intercept values arise from well known non-linearities near the extremes of detectors' ranges.

### 3.2.3. Calibration Outcomes

Once the correct peak positions are identified in channel space, a simple linear fit converts channel number to energy. The fit uses the known energies of each peak as the  $x$  values and the measured channel positions as the  $y$  values. In Figure 7, we illustrate the overall calibration performance:

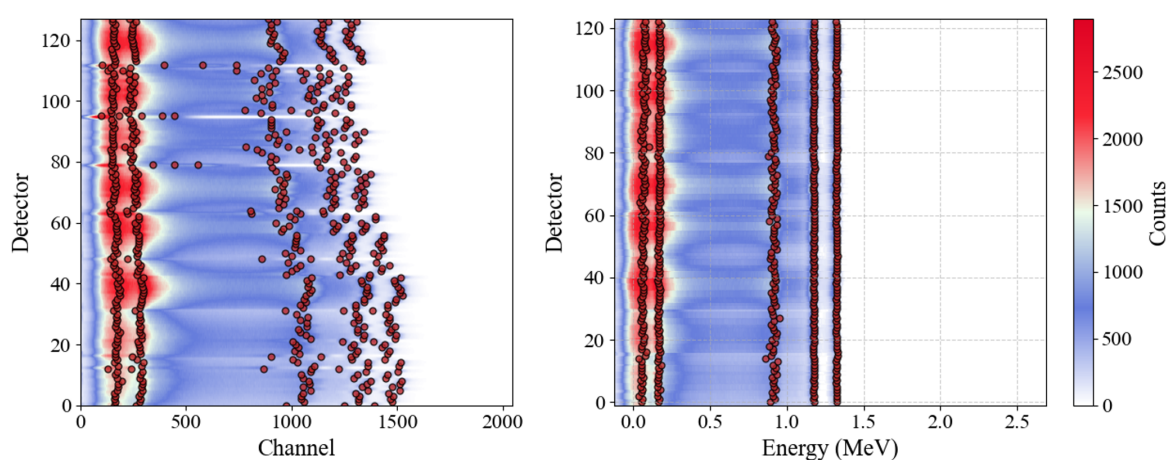
Panel (a) of Figure 7 shows a typical example of a linear energy calibration fit for a single detector, using the four known gamma-ray emission peaks. To evaluate the reliability of these calibrations across all detectors, we employed a standard technique known as **leave-one-out cross-validation (LOO-CV)**. In this procedure, the calibration is repeated four times per detector, each time leaving out one of the four reference peaks, fitting a line to the remaining three, and then computing the prediction error on the omitted point. Panels (b) and (c) summarize the results of this analysis across the full detector array. For each detector, the individual prediction errors from the four LOO-CV iterations are averaged to yield a single scalar error value. These average errors are reported in Panel (b) as a measure of calibration reliability and are also used to color-code the corresponding slope and intercept values in Panel (c). Lower average error indicates that all four calibration peaks lie close to a consistent linear fit, while higher error suggests issues such as misidentified peaks or detector faults. In particular, the three marked detectors with extremely high errors were identified as irreparably damaged and removed from subsequent analyses. The remaining two elevated-error

points arise from the two detectors in which the primary Cobalt peaks initially ranked as  $(k + 1)$ -th. Because the correct peaks are nonetheless detected (just slightly lower in the persistence order), a mild filter on backscattering peaks fixes the issue.

Panel (c) in Figure 7 summarizes how fit parameters vary across the detectors. Most calibrations cluster around consistent slope/intercept values, with small variations attributed to manufacturing differences, gain settings, or mild hardware non-linearities. Negative intercepts persist in all detectors, reflecting known edge effects and electronic offsets but still yielding an accurate mapping over the region of interest.

### 3.2.4. Channel-to-Energy Translations

To visualize the overall outcome of this multi-detector calibration, we compare raw (channel-space) spectra to their energy-space equivalents. Figure 8 shows, for a Cobalt source, how the large spread of peak positions among the 128 detectors (left) collapses into well-aligned peaks once the linear transformations are applied and the 3 faulty detectors removed (right).



**Figure 8.** Example Cobalt emission spectra from all detectors, annotated with the top 5 most persistent peaks. Left: raw data in channel space reveal large variability in apparent peak locations. Right: the same spectra after linear calibration and removal of 3 faulty detectors, translating channel to energy. The highest two (energy-wise) of these 5 persistent peaks correspond precisely to the known 1.17 MeV and 1.33 MeV emissions. The triangular shaped shadows visible in the region around 0.3 MeV with a periodicity of 16 detectors depends on the different counting rates due to different center-to-side solid angle when placing the pointlike source in front of each 16-unit detector module.

By ranking peaks via multiparameter persistence, each detector automatically yields the relevant peaks. Even small deviations in detection (as for the 2 problematic detectors mentioned above) are trivially corrected by discarding the low-energy backscattering regions. Hence, without ever manually labeling any peak, we successfully identify all calibration peaks for 125 detectors, requiring minimal filtering to recover even the two initially mislabeled peaks. This amounts to detecting (and correctly matching)  $4 \times 125 = 500$  emission peaks across all spectra, confirming the robustness of the proposed persistence-based approach for automating detector calibration.

## 4. Discussion

The proposed multi-parameter persistence framework demonstrates substantial advantages in robustness, interpretability, and practical utility for peak detection across both simulated and real-world datasets. The methodology generalizes traditional single-parameter persistence concepts, quantifying the stability of detected features by integrating across multiple influential detection parameters such as prominence, distance, width, and threshold. This comprehensive approach significantly mitigates the sensitivity to arbitrary parameter selection, an inherent limitation of conventional peak detection algorithms. Validation on test datasets (Section 3.1) underscores the robustness of this approach.



Our experiments revealed consistently low detection errors—under 1%—across a wide range of noise intensities, including challenging scenarios with both Gaussian and uniform noise distributions. Heatmap analyses further confirmed the method's resilience, highlighting expansive parameter regions where near-perfect detection rates were consistently achieved, thus establishing its reliability under various practical conditions. When applied to real-world spectroscopic data for detector calibration (Section 3.2), the method's effectiveness was clearly demonstrated. Using automated, persistence-based peak ranking without manual labeling, all critical calibration peaks across a large detector array were successfully identified. The framework proved not only capable of accurately calibrating functional detectors but also adept at diagnosing and isolating hardware faults, streamlining the instrument maintenance process. In cases where the automated method initially ranked correct peaks slightly lower in persistence, minimal post-processing sufficed for full recovery, highlighting the practicality and flexibility of the approach.

An important strength of this method lies in its intuitive interpretability. By assigning each detected feature a clear scalar metric—the accumulated hyper-volume across parameter space—practitioners gain a straightforward and quantitative measure of peak robustness. This metric inherently simplifies the feature-ranking process, distinguishing meaningful peaks from noise-induced artifacts effectively and transparently. From a theoretical standpoint, this work bridges traditional signal processing techniques with concepts from multi-parameter persistent homology (MPH). While formal MPH can offer meaningful insights by analyzing data through intricate topological invariants, our simpler volumetric persistence measure provides an accessible yet powerful alternative that is both computationally efficient and easy to interpret. Future research could further explore adaptive or probabilistic sampling strategies to optimize computational efficiency or integrate formal MPH invariants to uncover additional structural insights in complex multi-dimensional datasets.

In conclusion, the multi-parameter persistence method introduced here represents a robust, versatile, and intuitive solution to feature detection tasks across a variety of applications. Its successful application in simulated and real-world spectroscopic scenarios demonstrates its broad potential, offering enhanced reliability, reduced parameter sensitivity, and clear interpretability that are essential for modern scientific and engineering analyses.

**Author Contributions:** Conceptualization, G. F.; methodology, G. F. ; software, G. F. and G. V. ; validation, G. F. ; formal analysis, G. F.; investigation, G. F.; resources, G. F. and C.R. F.; data curation, C.R. F., G. F. and P. F.; writing—original draft preparation, G. F. and C.R. F.; writing—review and editing, G. F., C.R. F. and P. F.; visualization, G. F. C.R. F.; supervision, A. R., A. P., P. F. and S. T.; project administration, A. R., A. P., P. F. and S. T.; funding acquisition, A. R. and P. F.; All authors have read and agreed to the published version of the manuscript.

**Funding:** PNRR, Regione Sicilia

**Data Availability Statement:** The spectroscopy data analyzed in this study were provided under license by the *INFN-Laboratori Nazionali del Sud* and are not publicly available. The complete Python implementation of the multi-parameter persistence peak detection method introduced in this study is openly available at <https://github.com/gullo97/Volumetric-MPP>. An interactive online demonstration of the method, allowing users to test it on custom synthetic data, as well as user-provided signals, will also be publicly accessible at the same repository upon publication.

**Acknowledgments:** G.F., A.P. and A.R. acknowledge support by the University of Catania with the project “Linea di Intervento 1. Progetti di Ricerca Collaborativa” of Department of Physics and Astronomy “E.Majorana”. The authors also thank PhD. Enrico Martello for the feedback regarding some images in this paper. During the preparation of this manuscript, the author(s) used OpenAI's o3 LLM for the purposes of translating and streamlining text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
LOOCV	Leave-One-Out Cross-Validation
MPP	Multi-Parameter Persistence
LC-MS	Liquid Chromatography–Mass Spectrometry
GC×GC-TOF MS	2D Gas Chromatography with Time-Of-Flight mass spectrometry
VM-PP	Volumetric Multi-Parameter Persistence

## Appendix A

### *Piecewise Continuity of the Detection Function*

In this section we establish that, under reasonable assumptions, the detection function used in our methodology is piecewise continuous. Let  $\Theta \subset \mathbb{R}^n$  denote the hyperparameter space and consider the function

$$f : \Theta \rightarrow \{0, 1\},$$

where  $f(\theta) = 1$  indicates that a particular feature (e.g., a peak) is detected for the hyperparameter setting  $\theta$ , and  $f(\theta) = 0$  otherwise.

We assume that the underlying data function  $D(x)$  is continuous and that the detection of a feature at a candidate location  $x_0$  is determined by inequalities of the form

$$D(x_0) \geq \tau(\theta) \quad \text{and} \quad D(x_0) - D(x) \geq \delta(\theta)$$

for  $x$  in a neighborhood of  $x_0$ , where  $\tau : \Theta \rightarrow \mathbb{R}$  and  $\delta : \Theta \rightarrow \mathbb{R}$  are continuous functions representing, respectively, a detection threshold and a margin requirement.

Suppose first that for some  $\theta_0 \in \Theta$  the detection conditions are satisfied (i.e.,  $f(\theta_0) = 1$ ). By the continuity of  $\tau$  and  $\delta$ , there exists an open neighborhood  $U \subset \Theta$  containing  $\theta_0$  such that the inequalities

$$D(x_0) \geq \tau(\theta) \quad \text{and} \quad D(x_0) - D(x) \geq \delta(\theta)$$

continue to hold for all  $\theta \in U$ . Hence,  $f(\theta) = 1$  for all  $\theta \in U$ . Conversely, if  $f(\theta_0) = 0$ , then the failure of the detection condition persists in an open neighborhood  $V \subset \Theta$  about  $\theta_0$ , so that  $f(\theta) = 0$  for all  $\theta \in V$ .

Thus, aside from the boundaries where the detection outcome changes, the function  $f$  is constant—and therefore continuous—on open subsets of  $\Theta$ . The discontinuities, if any, are confined to the boundaries between these open regions, which are typically of lower dimension. In this sense, the detection function  $f$  is piecewise continuous.

This property is essential for justifying adaptive sampling strategies. In regions where  $f$  is continuous, a coarse sampling suffices to capture its behavior, while finer resolution can be concentrated near the transition boundaries, thereby mitigating the computational burden of a uniform grid search.

### *Connection to Multiparameter Persistent Homology*

From a TDA perspective, detecting and tracking features across an  $n$ -dimensional parameter space is akin to *multiparameter persistence*. Classical persistence homology often focuses on a single parameter (e.g., a filtration value), but  $\Theta$  here involves multiple dimensions—leading to more complex invariants in multiparameter persistent homology (MPH) [11]. In principle, one could model feature detections themselves (e.g., peaks) as *0D topological features* that appear or vanish across the hyperparameter lattice.

Our extension of “persistence” to a higher-dimensional hyperparameter domain is conceptually aligned with MPH, yet we opt for a simpler, domain-specific *aggregator*: the **accumulated hyper-**

**volume** in which a feature is detected. This single scalar score captures the same spirit of “robustness under parameter variation” without the overhead of constructing a full multiparameter barcode. It is:

- **Straightforward to compute:** A grid search (or analogous sampling) across the parameter space directly yields the measure of persistence for each feature.
- **Intuitive to interpret:** It quantifies how “hard” it is to destroy a feature by shifting the algorithm’s hyperparameters.
- **Avoids complexity:** Formal MPH approaches may require identifying complex topological invariants which can lead to intricate representations of the data and high computational overhead.

In the TDA literature, advanced tools like bigraded Betti numbers or rank functions can become unwieldy in higher dimensions, because multiparameter persistence is known to be “wild” from a representation-theoretic perspective [8–11]. By contrast, our aggregated volume measure is easy to compute and well-suited to routine signal-processing tasks. Nevertheless, an interesting future direction is to compare our hyper-volume aggregator with more formal MPH tools, to see whether further insights or additional topological structures might be revealed. We view this approach as a simple, application-driven extension of persistence ideas into robust feature detection while still connecting to the broader multiparameter TDA framework [9].

## References

1. Scholkmann, F.; Boss, J.; Wolf, M. An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals. *Algorithms* **2012**, *5*, 588–603. <https://doi.org/10.3390/a5040588>.
2. Kim, S.; Ouyang, M.; Jeong, J.; Shen, C.; Zhang, X. A New Method of Peak Detection for Analysis of Comprehensive Two-Dimensional GC×GC-TOF Mass Spectrometry Data. *Ann. Appl. Stat.* **2014**, *8*, 1209–1231. <https://www.jstor.org/stable/24522093>.
3. Poma, G. E.; Failla, C. R.; Amaducci, S.; Cosentino, L.; Longhitano, F.; Vecchio, G.; Finocchiaro, P. PI3SO: A Spectroscopic  $\gamma$ -Ray Scanner Table for Sort and Segregate Radwaste Analysis. *Inventions* **2024**, *9*, 85. <https://doi.org/10.3390/inventions9040085>.
4. Guo, Tiantian and Zhang, Tongpo and Lim, Enggee and López-Benítez, Miguel and Ma, Fei and Yu, Limin; IEEE Access; A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities; DOI: 10.1109/ACCESS.2022.3179517
5. Deep Learning for the Precise Peak Detection in High-Resolution LC–MS Data. Arseny D. Melnikov, Yuri P. Tsentalovich, and Vadim V. Yanshole Analytical Chemistry 2020 92 (1), 588-592 DOI: 10.1021/acs.analchem.9b04811
6. Zhang, J.; Gonzalez, E.; Hestilow, T.; Haskins, W.; Huang, Y. Review of Peak Detection Algorithms in Liquid-Chromatography–Mass Spectrometry. *Curr. Genomics* **2009**, *10*, 388–401. DOI: 10.2174/138920209789177638.
7. Kilgour, D.P.A., Hughes, S., Kilgour, S.L. et al. Autopiquer - a Robust and Reliable Peak Detection Algorithm for Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 253–262 (2017). <https://doi.org/10.1007/s13361-016-1549-z>
8. Ghrist, R. Barcodes: The Persistent Topology of Data. *Bull. Amer. Math. Soc.* **2008**, *45*, 61–75. DOI:10.1090/S0273-0979-07-01191-3
9. Carlsson, G., Zomorodian, A. The Theory of Multidimensional Persistence. *Discrete Comput Geom* **42**, 71–93 (2009). <https://doi.org/10.1007/s00454-009-9176-0>
10. Lesnick, M.; Wright, M. Interactive Visualization of 2-D Persistence Modules. *arXiv* **2015**, 1512.00180. Available online: <https://doi.org/10.48550/arXiv.1512.00180>
11. Botnan, M. B.; Lesnick, M. An Introduction to Multiparameter Persistence. *arXiv* **2022**, 2203.14289. Available online: <https://doi.org/10.48550/arXiv.2203.14289>
12. Bernardini, G. Definizione Normativa e Classificazione dei Rifiuti Radioattivi. *Cammino Diritto*, 22 October 2022. Available online: [https://rivista.camminodiritto.it/public/pdfarticoli/8838\\_10-2022.pdf](https://rivista.camminodiritto.it/public/pdfarticoli/8838_10-2022.pdf) (accessed 22 October 2022).
13. IAEA. Classification of Radioactive Waste. IAEA Safety Standards Series No. GSG-1; IAEA: Vienna, 2009. Available online: [https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1419\\_web.pdf](https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1419_web.pdf).
14. Rossi, F.; Cosentino, L.; Longhitano, F.; Minutoli, S.; Musico, P.; Osipenko, M.; Poma, G. E.; Ripani, M.; Finocchiaro, P. The Gamma and Neutron Sensor System for Rapid Dose Rate Mapping in the CLEANDEM Project. *Sensors* **2023**, *23*, 4210. <https://doi.org/10.3390/s23094210>.

15. Longhitano, F.; Poma, G. E.; Cosentino, L.; Finocchiaro, P. A Scintillator Array Table with Spectroscopic Features. *Sensors* **2022**, *22*, 4754. <https://doi.org/10.3390/s22134754>.
16. Hamamatsu Photonics. MPPC S14160-6050HS. Available online: <https://www.hamamatsu.com/eu/en/product/optical-sensors/mppc/mppc-array/S14160-6050HS.html> (accessed 2023).
17. CAEN Digitizer VX2745. Available online: <https://www.caen.it/products/vx2745/> (accessed on 31 May 2025).
18. <https://github.com/gullo97/Volumetric-MPP>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.