Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **537,** 730–738 (2025) https://doi.org/10.1093/mnras/staf070
Advance Access publication 2025 January 14

Detection of new gamma-ray AGN candidates using SIMEFIC III and DBSCAN in *Fermi*-LAT data

M. Soor, F. Akhondi and H. Hedayati Kh ★

Department of Physics, K.N. Toosi University of Technology, P.O. Box 15875-4416, Tehran, Iran

Accepted 2025 January 7. Received 2025 January 4; in original form 2024 November 17

ABSTRACT

We present the detection of new high-energy gamma-ray active galactic nucleus candidates using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm on two-dimensional gamma-ray data from *Fermi* Large Area Telescope. Our approach involves iterative applications of the SIeving MEthod for FInding Core (SIMEFIC III) algorithm, designed to enhance point source detection. By integrating DBSCAN at each denoising step, we track and evaluate the significance of potential sources across iterations. Our findings indicate that source significance increases to a specific threshold, beyond which further denoising may remove genuine sources. At the optimal denoising stage, we identified 18 sources not listed in the *Fermi* catalogue, with several of these sources potentially matching entries in the the Candidate Gamma-Ray Blazar Survey (CRATES) and Roma-BZCAT (Roma-BZCAT Multi-Frequency Catalog of Blazars) catalogues.

Key words: methods: data analysis – gamma-rays: general.

1 INTRODUCTION

The *Fermi* Large Area Telescope (*Fermi*-LAT), aboard the *Fermi Gamma-ray Space Telescope* launched in 2008, is a highly sensitive instrument designed to detect gamma-rays across a wide energy range, from 20 MeV to over 1 TeV. The latest *Fermi*-LAT catalogue of gamma-ray sources, the Fourth Fermi Gamma-ray LAT Catalog (4FGL), has identified 7194 sources (Ballet et al. 2023), with over 3800 of these classified as active galaxies of the blazar type.

Blazars are a class of active galactic nuclei (AGNs) characterized by jets of charged particles directed nearly along our line of sight, causing them to appear extremely bright in gamma-ray observations. The gamma-rays detected from blazars are generated within these jets, where particles travel at relativistic speeds (Blandford & Rees 1978; Maraschi, Ghisellini & Celotti 1992). Despite progress in understanding these sources, several key questions remain unresolved: the mechanisms driving particle acceleration (Blandford, Meier & Readhead 2019), the exact location of gamma-ray emission within the jet (Arsioli & Chang 2018), the causes of AGN variability (Hawkins 2002), and the gamma-ray duty cycle of blazars (Meyer et al. 2011).

Distinguishing genuine sources from background noise is a substantial challenge, especially with the extensive data amassed over 16 yr. Advanced data analysis techniques – such as machine learning (e.g. Panes et al. 2021) and clustering algorithms (e.g. Campana et al. 2008; Tramacere & Vecchio 2013) – have been developed to enhance the detection of point sources. These methods enable the grouping of similar data points and pattern identification to better differentiate true sources from background fluctuations. Moreover, the integration of multifrequency seeds has proven to be a powerful approach for raising detection thresholds by utilizing cross-band information. This method, as demonstrated by Arsioli, Chang & Ighina (2024) and Arsioli, Chang & Musiimenta (2020), leverages data from multiple frequency bands to enhance sensitivity and reliability in source identification. However, processing such large data sets remains computationally demanding, potentially slowing down analysis and increasing the likelihood of misidentification.

Background noise frequently complicates point source detection, often causing clustering algorithms to misidentify spurious sources. This noise increases with longer observation durations in *Fermi*-LAT data. In 2013, Tramacere and Vecchio applied the DBSCAN algorithm to analyse approximately 11 000 photons (Tramacere & Vecchio 2013). In 2022, Campana et al., including Tramacere, expanded this analysis to a 9×12 deg^2 region of the Large Magellanic Cloud (Campana et al. 2022). Earlier, in 2007, Campana et al. conducted a minimum spanning tree (MST) analysis on a 500-point test field (Campana et al. 2008). However, their 2018 study encountered difficulties when attempting to process the full-sky MST catalogue containing around 290 000 photons (Campana, Massaro & Bernieri 2018). To reduce noise, Campana increased the photon energy threshold to 10 GeV and divided the sky into smaller regions, with the largest containing roughly 25 000 photons. In 2021, they worked with about 320 000 photons, segmenting the sky into areas with a maximum of 31 000 photons (Campana & Massaro 2021). Although these methods reduced noise by raising the energy threshold or focusing on smaller regions, challenges persisted, particularly near field boundaries. Furthermore, increasing the energy threshold risks excluding lower energy sources.

In our previous work (Hedayati Kh., Soor & Akhondi 2024), we introduced SIMEFIC III, an algorithm developed to enhance

★ E-mail: hedayati@kntu.ac.ir

source detection by reducing background noise, enabling more precise identification of genuine point sources while minimizing false positives. We successfully tested the DBSCAN algorithm on a field containing approximately 44 000 photons. SIMEFIC III operates by analysing the distances between data points and removing those with excessively large separations. Initially, SIMEFIC III used a thresholding technique based on the mean distance between data points, which yielded satisfactory results. However, to further refine photon removal, we experimented with alternative strategies, investigating iterative denoising processes and comparing their performance with the original method. As we will demonstrate, this iterative approach achieves improved accuracy over the initial method. Using this enhanced method, we are able to identify potential new sources that do not appear in the existing *Fermi* catalogue.

2 OUTLINE OF SIMEFIC III

In our previous work, we introduced an algorithm designed to reduce background noise and highlight point sources in a field of points, enabling more accurate source detection. Suppose we have N points distributed on a plane. The first step is to calculate the distance between every pair of points. Let d_{ij} represent the distance between points i and j.

(i) **Identify closest pair:** Identify the two points, i and j, with the smallest distance between them, denoted as d_{ij}. This step is crucial for locating pairs of points that are likely close to a genuine source rather than background noise.

(ii) **Find nearest point to mid-point:** Given the closest pair of points, i and j, calculate their mid-point $(x_\mathrm{m}, y_\mathrm{m})$ as follows:

$$(x_\mathrm{m}, y_\mathrm{m}) = \left(\frac{x_i + x_j}{2}, \frac{y_i + y_j}{2}\right).$$

Then, from the remaining points in the list, find the point k that has the smallest distance to this mid-point.

(iii) **Determine closest source candidate:** Next, we compare the distances d_{ik} and d_{jk}. If d_{ik} is smaller than d_{jk}, we select point i as closer to the source and assign it an index distance (ID) value, $D_i = d_{ij}$. Point i is then removed from the list of points. Otherwise, if d_{jk} is smaller, we select point j as closer to the source, assign its ID as $D_j = d_{ij}$, and remove point j from the list.

(iv) **Iterate until completion:** Repeat steps (i)–(iii) with the updated list of points until only two points remain. These final two points are removed without assigning IDs (or by setting their D values to ∞), thereby concluding the iterative process. At the end of this process, each retained point has an assigned distance value, D, which can then be used to filter out points with high distance values, treating them as noise.

For smaller data sets, we employ a brute-force approach to identify the closest pair of points, as this method is straightforward and efficient for a limited number of points. However, as the data set size increases, the quadratic time complexity $O(n^2)$ of the brute-force method renders it impractical. To overcome this limitation, we employ more efficient algorithms, such as the divide-and-conquer approach, which reduces the time complexity to $O(n \log n)$. This method operates by recursively partitioning the data set into smaller subsets, determining the closest pair within each subset, and then merging the results to find the closest pair across the entire data set. Comprehensive discussions and implementations of this algorithm are available in foundational computational geometry literature, such as De Berg (2000) and Preparata & Shamos (2012). These studies demonstrate how the divide-and-conquer strategy utilizes spatial properties and efficient merging techniques to significantly enhance computational performance, making it well suited for large-scale data sets.

3 EFFECT OF MULTIPLE DENOISING ITERATIONS ON NOISE REDUCTION

To further improve the denoising effectiveness of the SIMEFIC III method, we adopt an iterative approach that gradually refines source detection. In the initial iteration, points with an ID above the average ID are removed, effectively reducing background noise. After this initial pass, SIMEFIC III is reapplied to the remaining points, once again filtering out those with ID values exceeding the recalculated average. This iterative process continues until further iterations no longer yield meaningful noise reduction and instead start to impact genuine sources. The goal of this approach is to maximize noise reduction while preserving as many true sources as possible.

We determined the centre of each cluster using our algorithm, SIMEFIC III, by assigning a weight $w_i = 1/D_i$ to each point, where D_i is the ID calculated by SIMEFIC III. The weighted centre $(x_\mathrm{c}, y_\mathrm{c})$ of the cluster is then computed as follows:

$$x_\mathrm{c} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \qquad y_\mathrm{c} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}, \tag{1}$$

where n is the number of points in the cluster.

3.1 Data selection

Our study focused on high-energy gamma-rays within a specific region of the sky, defined by the coordinates $170^\circ < l < 240^\circ$ and $40^\circ < b < 60^\circ$, excluding the galactic plane and poles. This region aligns with the field analysed in our previous work (Hedayati Kh. et al. 2024).

For data collection, we downloaded weekly photon data covering a 16-yr period from 2008 August 4 to 2024 August 4 using Pass 8, Release 3 data. The *Fermi*-LAT data are publicly accessible via the LAT data server. We processed the data using FERMITOOLS 2.2.0 and FERMIPY 1.2.2, focusing on high-quality gamma-rays with energies above 3 GeV. Standard filtering criteria were applied, selecting source-class events (`evclass = 128`) and both front and back conversions (`evtype = 3`), with a maximum zenith angle of 90°. Following this filtration, the data set was narrowed to 44 499 high-quality photons for further analysis. Fig. 1 shows a two-dimensional (2D) scatter plot of this region.

3.2 DBSCAN clustering

Following our previous work, we applied the DBSCAN algorithm to detect clusters in *Fermi*-LAT data, first without noise reduction and subsequently after using the SIMEFIC III algorithm. For this analysis, we implemented a modified version of the DBSCAN algorithm, as introduced by Tran, Drab & Daszykowski (2013). To ensure consistency in comparing different denoising approaches on the same region, we used the same parameter values as in our prior study: a neighbourhood radius of $\varepsilon = 0.17$ and a minimum neighbourhood size of MinPts = 5, consistent with Tramacere & Vecchio (2013). As in previous analyses, the significance of each cluster was calculated using the method proposed by Li & Ma (1983):

$$S = \sqrt{2\left(N_\mathrm{on} \ln\left(\frac{1+\alpha}{\alpha}\frac{N_\mathrm{on}}{N_\mathrm{on}+N_\mathrm{off}}\right) + N_\mathrm{off} \ln\left((1+\alpha)\frac{N_\mathrm{off}}{N_\mathrm{on}+N_\mathrm{off}}\right)\right)}, \tag{2}$$

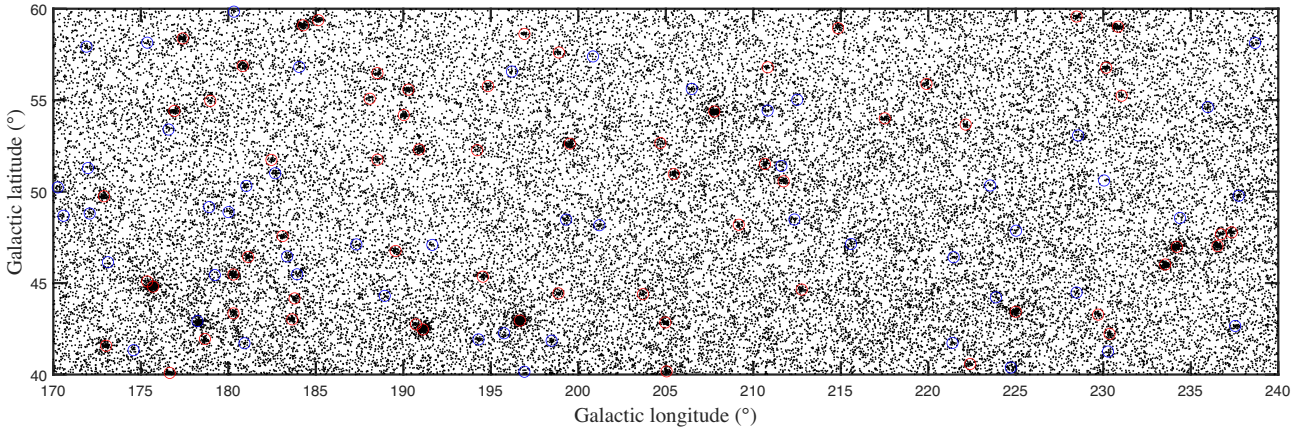


Figure 1. Black dots represent individual photons, while red circles mark sources with a TS above 25, and blue circles indicate sources with TS below 25. This field contains a total of 116 sources from the *Fermi* 14-yr catalogue (4FGL-DR4), with 64 sources showing TS $>$ 25 and 52 sources showing TS $<$ 25.

where $N_{\rm on}$ represents the number of points within the cluster area, $N_{\rm off}$ is the count of background points, and α denotes the ratio of exposure between the on-source and off-source regions. This equation provides a quantitative measure of the statistical significance of clusters, aiding in the effective differentiation of genuine sources from background noise.

In the 4FGL catalogue, the significance of a source is calculated using the square root of its test statistic (TS). The TS value is defined as

$$\mathrm{TS} = 2\log\left(\frac{L}{L_0}\right), \tag{3}$$

where L is the maximum likelihood with the source included in the model, and L_0 is the maximum likelihood with the source excluded. This calculation provides a quantitative measure of the source's presence, with higher TS values indicating a higher probability that the detected emission is due to an actual source rather than background fluctuations.

We set the significance threshold for the sources at $S = 2.5$, based on the relationship between the significance of the cluster and TS proposed by Tramacere & Vecchio (2013):

$$S \simeq 0.5\sqrt{\mathrm{TS}}. \tag{4}$$

This results in a TS value of TS $= 25$. Consequently, clusters with a significance level below 2.5 are excluded from the field.

We applied the SIMEFIC algorithm to the data set iteratively, from 1 up to 10 iterations. Initially, clustering was performed using DBSCAN without any denoising. In the subsequent steps, the SIMEFIC algorithm was applied to remove all points with an ID greater than the average, followed by reapplying DBSCAN. This process was repeated for up to 10 iterations, with points having an ID greater than the average being removed at each step. The results are summarized in Table 1, where each row corresponds to one iteration.

The first column of the table outlines the performance of DBSCAN across the denoising iterations. The second column indicates the number of photons remaining after each iteration. The third and fourth columns detail the identification of sources listed in the *Fermi* catalogue, categorized by their TS values (TS $>$ 25 and TS $<$ 25). The subsequent columns present the number of clusters not associated with the *Fermi* catalogue and the total clusters identified at each step. Differences between the total number of *Fermi* catalogue-listed sources and clusters not associated with the catalogue arise from overlapping clusters that persist across iterations.

As shown in Table 1, no sources with TS $>$ 25 are removed until the sixth denoising iteration. In the seventh iteration, the first source with TS $>$ 25 is lost, and by the tenth iteration, two more sources are removed. This indicates that overdenoising begins after the sixth iteration, leading to the unintended removal of genuine sources, which are misclassified as noise. Notably, after 10 iterations, no clusters unassociated with *Fermi* catalogue sources remain, as confirmed by the 4FGL Data Release 4 (4FGL-DR4) data.

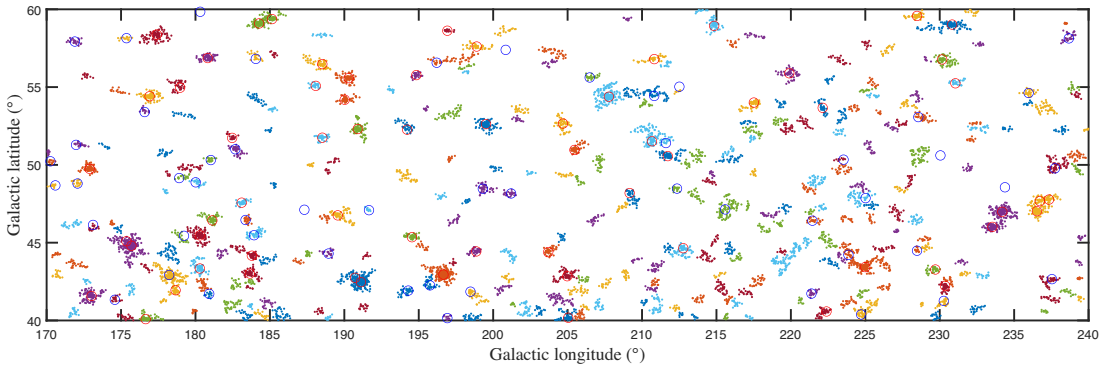
It should be noted that we are using 16 yr of *Fermi*-LAT data, while the 4FGL-DR4 corresponds to 14 yr of data. Therefore, the TS values for the sources used in this paper might be slightly different from the values mentioned currently. However, we consider them as a framework for our work.

Fig. 2 presents the DBSCAN clustering results at three stages of the denoising process: the first, sixth, and ninth iterations (corresponding to rows 2, 7, and 10 in Table 1). Fig. 2(a) shows the clustering after the first iteration, where points with an ID exceeding the average were removed. Despite the initial denoising, a significant number of clusters remain unassociated with the *Fermi* catalogue. Fig. 2(b) displays the results after 6 iterations, with the field significantly refined, leaving only 18 unassociated clusters, while retaining all sources with TS $>$ 25. In Fig. 2(c), the ninth iteration reveals only two unassociated clusters but results in the loss of one *Fermi* source with TS $>$ 25. Analysis of earlier denoising steps, along with the proximity of these clusters to *Fermi* catalogue sources, suggests that they are likely remnants of larger clusters originally associated with these sources. During the denoising process, these larger clusters may have fragmented, leading to the removal of their main components and leaving behind two residual clusters. This highlights the effectiveness of iterative denoising but suggests that denoising beyond the sixth iteration risks removing *Fermi* sources along with the noise.

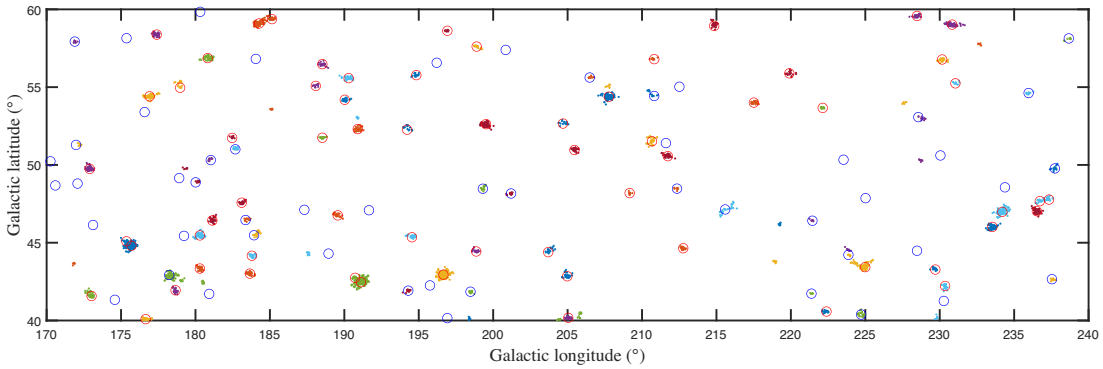
4 CLUSTER SIGNIFICANCE

In this study, as in our previous work, the sources are ordered by their TS in descending order [refer to part (a) of fig. 7 in Hedayati Kh. et al. 2024]. Given that the source properties have been updated in the latest *Fermi* catalogue, we have compiled an updated list of their features in Table 4.

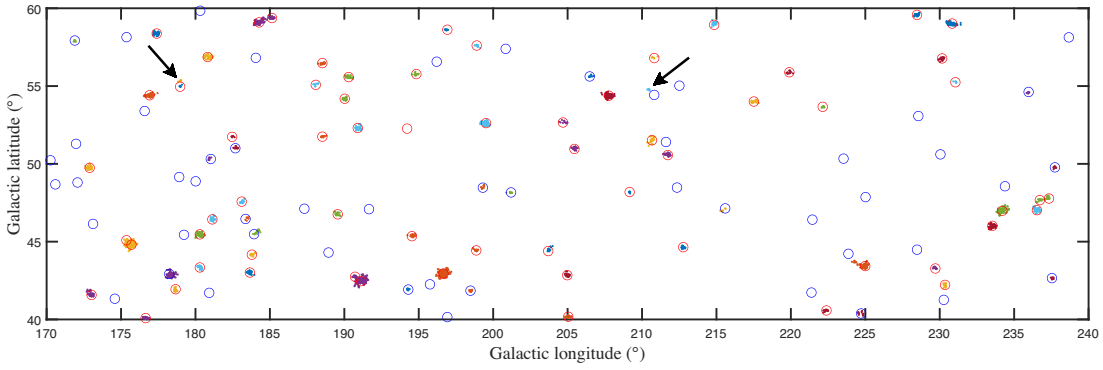
In the previous section, we discussed the relationship between significance and the TS as defined by equation (4). Now, we turn our attention to the impact of denoising on the significance of clusters.



(a) DBSCAN after the first denoising iteration, where points with ID greater than the average have been removed.



(b) DBSCAN after the sixth denoising iteration.



(c) DBSCAN after the ninth denoising iteration. Two clusters, not associated with any known source, remained after the ninth iteration and are marked with arrows for clarity.

Figure 2. DBSCAN clustering results after multiple denoising iterations using the SIMEFIC III algorithm. As in the previous figure, sources with a TS above 25 are marked in red, while those below 25 are marked in blue.

After each denoising step, we evaluated the significance of each cluster, and the results revealed noticeable changes. Specifically, we computed the significance at the first, sixth, and ninth denoising stages for comparison. Additionally, we examined the differences in significance between the sixth and first steps, as well as between the ninth and sixth steps.

Fig. 3 highlights the relationship between denoising iterations and cluster significance. While the initial denoising steps tend to increase the significance for many clusters, this trend reverses after the sixth iteration. By the ninth denoising iteration, the significance of all clusters has diminished. The key observation here is that significance continues to improve until overdenoising occurs, at which point valuable data are lost. Excessive denoising becomes evident when, instead of enhancing cluster significance, it causes a decline. After six iterations, not only are some clusters associated with *Fermi* catalogue sources removed, but portions of the remaining clusters are also affected. This highlights that overdenoising can lead to the loss of critical data points, compromising the integrity of the remaining clusters.

Notably, the seventh denoising resulted in the removal of a source from the field. Source number 64, 4FGL J0958.0+3222, was removed after the seventh denoising. This source had the lowest TS value among the sources, with $\sqrt{\mathrm{TS}} = 5.029$.

Four sources (9, 38, 49, and 50) exhibit notable variations in cluster significance during the denoising process. Source 50 initially shares a cluster with a source having a TS < 25. In subsequent steps,

Table 1. Clustering results after applying SIMEFIC and DBSCAN across multiple iterations. Each row represents a denoising iteration (with the first row indicating the data before denoising), followed by the corresponding counts of remaining photons, identified *Fermi*-listed sources (split into two columns for TS > 25 and TS < 25), clusters not associated with the *Fermi* catalogue, and the total number of clusters for that iteration. At each step, points with an ID exceeding the average are systematically removed, refining the data set iteratively.

Denoising iteration	Remaining photons	Sources (TS > 25)	Sources (TS < 25)	Clusters not associated with *Fermi* catalogue sources	Total clusters
No denoising	44 699	64	51	1534	1648
After 1st pass	30 809	64	46	210	329
After 2nd pass	21 296	64	43	145	245
After 3rd pass	14 917	64	38	76	171
After 4th pass	10 688	64	37	45	141
After 5th pass	7922	64	33	31	123
After 6th pass	6100	64	29	18	103
After 7th pass	4903	63	28	9	92
After 8th pass	4128	63	24	4	83
After 9th pass	3598	63	19	2	77
After 10th pass	3241	61	13	0	70

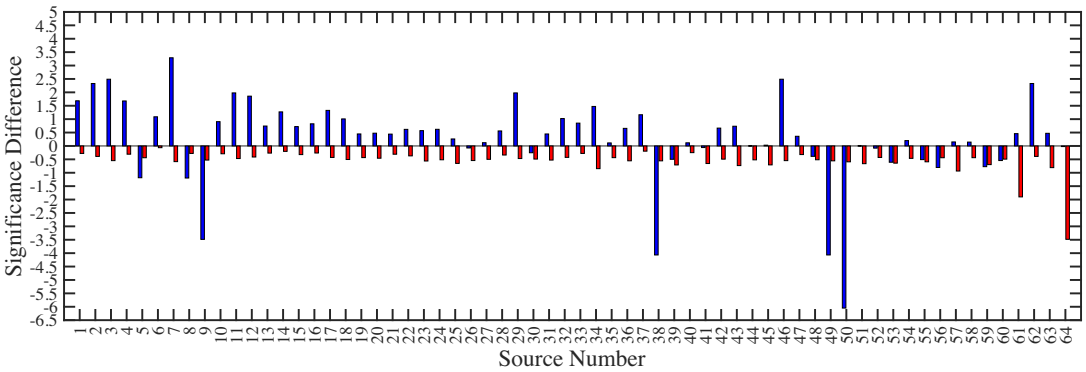


Figure 3. The significance differences between the first, sixth, and ninth denoising iterations are shown. Blue represents the change between the sixth and first stages, while red shows the change between the ninth and sixth stages. These colours correspond to the subtraction of significance values between the respective iterations.

the two sources separate, leading to a reduction in the significance of the cluster associated with source 50. For sources 38 and 49, the cluster significance diminishes as their shared cluster splits into subclusters during denoising (see Table 2 for a list of shared clusters). Initially, sources 9 and 5 are grouped due to their high $\sqrt{\mathrm{TS}}$ values and close proximity. By the sixth iteration, they form distinct clusters, illustrating the efficacy of the denoising process in distinguishing individual clusters. Therefore, the reduction in significance for clusters associated with these four sources, even if considerable, is entirely natural and to be expected. This decrease is a result of the denoising process, which often leads to the separation of overlapping clusters and improved accuracy in source identification.

5 THE PRECISION IN DETERMINING THE CENTROIDS OF THE CLUSTERS

Matching the cluster centres with the exact source locations from the *Fermi* catalogue is crucial. Fig. 4 illustrates the effect of denoising on the accuracy of cluster centre calculations. The angular distance between the calculated cluster centre (from equation 1) and the *Fermi* catalogue source location is averaged across all clusters. As denoising iterations increase, accuracy in determining cluster centres improves significantly. After six denoising steps, the average angular distance decreases from 0.370 to 0.088, indicating a notable enhancement in precision.

However, excessive denoising presents another drawback. While six iterations improve accuracy, beyond the seventh step, the results become inconsistent. The angular distance fluctuates – rising in the eighth step, dropping in the ninth, and increasing again in the tenth. This reinforces the conclusion that stopping denoising after the sixth iteration yields optimal results, with no need for further steps.

As shown in Table 2, we monitored the behaviour of various sources throughout the denoising process. Sources 2 and 62 consistently remained grouped, as did sources 3, 46, 11, and 29, indicating a strong relationship among these pairs.

Conversely, some sources separated from their initial clusters at different stages. For instance, sources 5 and 9 started together but split by the sixth iteration. Similarly, source 8, which was initially part of a common cluster with sources 38 and 49 and had the highest TS among them, separated from the cluster at this stage. This separation from shared clusters likely contributed to improved accuracy in determining the core location of the clusters.

6 DISCOVERY OF POTENTIAL ASTROPHYSICAL SOURCES THROUGH DENOISING

In our analysis, we identified 18 new clusters at the sixth denoising iteration, not previously catalogued in 4FGL, as shown in Table 3.

Table 2. This table summarizes key information for shared clusters identified in our analysis, including each cluster's source number, name, galactic coordinates (l, b), TS, photon count, and significance values for the first, sixth, and ninth denoising iterations. It also includes the differences in significance between the sixth and first, and between the ninth and sixth iterations.

Source Num	l(°)	b(°)	TS	1st iteration's photon count	1st iteration's Sig	6th iteration's photon count	6th iteration's Sig	9th iteration's photon count	9th iteration's Sig	Sig Diff between 6th and 1st iterations	Sig Diff between 9th and 6th iterations
2	175.7005	44.8052	33.0153	378	17.072	283	19.402	263	19.012	2.33	−0.39
62	175.3631	45.0912	5.1395	378	17.072	283	19.402	263	19.012	2.33	−0.39
3	191.1188	42.4638	28.5867	396	18.003	318	20.493	290	19.945	2.49	−0.548
46	190.7225	42.7437	6.8538	396	18.003	318	20.493	290	19.945	2.49	−0.548
5	234.2043	46.985	19.9019	330	14.754	142	13.568	126	13.128	−1.186	−0.44
9	233.5399	46.0162	16.2676	330	14.754	96	11.268	84	10.741	−3.486	−0.527
8	236.5262	47.0247	16.9685	216	11.822	85	10.625	78	10.346	−1.197	−0.279
38	237.3407	47.7723	7.9777	216	11.822	47	7.754	38	7.198	−4.068	−0.556
49	236.7238	47.6703	6.5115	216	11.822	47	7.754	38	7.198	−4.068	−0.556
11	184.2743	59.0923	15.3972	196	11.808	147	13.787	130	13.317	1.979	−0.47
29	185.1401	59.3715	9.4273	196	11.808	147	13.787	130	13.317	1.979	−0.47

Figure 4. The average angular distance between cluster centres (after denoising with SIMEFIC III) and their corresponding sources in the *Fermi* catalogue, shown across different denoising iterations. The first bar represents the average angular distance prior to any denoising.

These clusters exhibit high significance and photon counts, indicating that they could be candidate astrophysical sources.

To assess potential associations with known sources, we cross-referenced the positions of these clusters with multiple catalogues. These included the Roma-BZCAT (5th edition) by Massaro et al. (2015), listing coordinates for 3561 radio-detected sources; the CRATES catalogue by Healey et al. (2007), which provides high-precision positions; and the WIBRaLS (Wide-band IBL and Radio-Loud Sources) catalogue (Wright et al. 2010), an all-sky list of candidate gamma-ray blazars from WISE (Wide-field Infrared Survey Explorer) data.

Among the identified clusters, two showed positional correlations with Roma-BZCAT sources, at angular separations of 4.952 and 5.466 arcmin, respectively. One additional cluster aligned with a CRATES source within 4.727 arcmin. Given the average angular separation between the SIMEFIC III cluster centres and 4FGL sources after the sixth denoising iteration (see Fig. 5), these associations fall within a plausible range for positional accuracy.

Furthermore, we found eight additional clusters associated with catalogue sources at greater distances, from approximately 8 to 35 arcmin, as detailed in Table 3. We also analysed clusters near the 40° boundary, finding that clusters 7 and 4 connect to a source. This connection suggests that these clusters are remnants of larger clusters previously associated with 4FGL sources.

These findings underscore the effectiveness of SIMEFIC III's iterative denoising approach in enhancing source detection sensitivity and isolating potential new astrophysical sources. By systematically reducing background noise across each iteration, SIMEFIC III sharpens the resolution of genuine signal patterns, allowing clusters that may represent new sources to stand out more clearly against the reduced noise. This improvement not only aids in identifying previously undetected sources but also refines the characterization of existing ones by better separating them from background artefacts.

7 CONCLUSION

Our proposed method for denoising 2D gamma-ray images, utilizing iterative applications of the SIMEFIC III algorithm, demonstrates promising results in identifying point sources and offers substantial potential for further research.

We employed the DBSCAN clustering algorithm to detect point sources in *Fermi*-LAT data across each denoising iteration, calculat-

Table 3. DBSCAN clusters without 4FGL-DR4 associations (except for candidate source 17). Column *N* lists photon counts, 'Sig' represents significance via likelihood ratio test, as defined by Li and Ma (1983). Catalogues, source names, and distances (arcmin) are also shown.

Name	RA°(J2000)	Dec°(J2000)	$l°$	$b°$	N	Sig	Catalogue	Source name	Distance (arcmin)
Candidate source 1	150.3949	34.418 29	190.8841	53.066 89	5	2.587	CRATES	J100112+342455	4.7277
Candidate source 10	151.1648	37.821 91	185.1516	53.567 62	5	2.598	BZCAT	5BZB J1004+3752	4.9523
Candidate source 11	138.8614	47.598 54	171.7979	43.5893	5	2.576	BZCAT	5BZB J0915+4738	5.4669
Candidate source 12	136.1798	27.702 66	198.3855	40.1527	6	2.829	CRATES	J090504+274823	8.7297
Candidate source 13	136.9086	41.317 98	180.4707	42.478 58	7	3.038	CRATES	J090650+412426	11.284
Candidate source 14	138.054	22.286 95	205.8196	40.357 78	7	3.030	WIBRaLS	J091224.78+220506.2	16.272
Candidate source 15	153.8901	24.859 86	207.6771	55.037 67	5	2.552	WIBRaLS	J101353.43+244916.4	23.154
Candidate source 16	154.0324	11.326 42	228.6939	50.299 01	6	2.817	WIBRaLS	J101512.04+110219.5	24.197
Candidate source 17	147.3428	11.632 59	223.7469	44.648	6	2.826	4FGL	4FGL J0000.3−7355	26.714
Candidate source 18	137.1526	42.264 13	179.1882	42.669 33	5	2.569	BZCAT	5BZQ J0908+4150	33.712
Candidate source 2	137.4335	23.202 38	204.4538	40.078 96	5	2.581	BZCAT	5BZQ J0910+2248	35.001
Candidate source 3	144.9829	14.352 41	218.9444	43.736 49	7	3.037	No counterpart		
Candidate source 4	147.3802	15.111 38	219.2858	46.163 98	6	2.836	No counterpart		
Candidate source 5	146.758	41.831 74	179.2376	49.811 33	5	2.607	No counterpart		
Candidate source 6	145.6359	5.616 108	229.7712	40.251 82	5	2.582	No counterpart		
Candidate source 7	161.818	13.045 25	232.6163	57.755 79	5	2.573	No counterpart		
Candidate source 8	139.582	36.277 22	187.5339	44.331 58	5	2.570	No counterpart		
Candidate source 9	156.9721	13.686 71	227.5315	53.945 69	5	2.542	No counterpart		

Table 4. Specifications of sources with $\sqrt{\text{TS}}$ values, ordered by descending TS.

Source number	Source name	l	b	$\sqrt{\text{TS}}$
1	4FGL J0915.9+2933	196.6587	42.9494	37.0015
2	4FGL J0920.9+4441	175.7005	44.8052	33.0153
3	4FGL J0910.6+3329	191.1188	42.4638	28.5867
4	4FGL J1012.7+2439	207.772	54.3824	23.863
5	4FGL J1012.3+0629	234.2043	46.985	19.9019
6	4FGL J1033.1+4115	177.3956	58.3763	17.7376
7	4FGL J0946.6+1016	224.9955	43.4315	17.1349
8	4FGL J1016.0+0512	236.5262	47.0247	16.9685
9	4FGL J1008.0+0620	233.5399	46.0162	16.2676
10	4FGL J0923.5+4125	180.2933	45.4691	15.9832
11	4FGL J1032.6+3737	184.2743	59.0923	15.3972
12	4FGL J1049.8+1429	230.8346	59.0186	14.3561
13	4FGL J1023.1+3949	180.8222	56.8672	14.1664
14	4FGL J1001.1+2911	199.5183	52.6195	14.0459
15	4FGL J0950.2+4553	172.8934	49.7451	12.4915
16	4FGL J0854.3+4408	176.653	40.0876	12.4897
17	4FGL J1012.7+4228	176.917	54.4102	12.3498
18	4FGL J0957.8+3423	190.9019	52.3041	12.2519
19	4FGL J0921.7+2336	204.9769	42.8447	11.8331
20	4FGL J0910.8+3859	183.6529	43.0147	11.6153
21	4FGL J0903.1+4652	173.0175	41.5741	11.3386
22	4FGL J1049.5+1548	228.4733	59.5738	11.0453
23	4FGL J0912.2+4127	180.2943	43.3408	10.9904
24	4FGL J0928.5+4048	181.1333	46.4299	10.0703
25	4FGL J0925.7+3126	194.5429	45.352	9.9821
26	4FGL J1036.2+2202	214.8448	58.9344	9.6068
27	4FGL J0956.7+2516	205.4591	50.9595	9.5116
28	4FGL J0910.6+2247	205.0443	40.1737	9.4619
29	4FGL J1033.7+3708	185.1401	59.3715	9.4273
30	4FGL J0934.3+3926	183.105	47.5717	9.3577
31	4FGL J0936.5+1847	212.7692	44.6455	9.19
32	4FGL J0930.7+3502	189.5585	46.7562	9.0916
33	4FGL J1007.0+3455	190.0276	54.1875	8.8518
34	4FGL J1002.5+2215	210.6692	51.5171	8.5702
35	4FGL J0950.2+0615	230.3603	42.2209	8.3334
36	4FGL J1013.7+3444	190.2883	55.5775	8.3013
37	4FGL J0959.4+2120	211.7335	50.5642	8.1923
38	4FGL J1019.7+0511	237.3407	47.7723	7.9777
39	4FGL J0924.0+2816	198.8638	44.4369	7.7264
40	4FGL J1041.0+1342	230.1625	56.7762	7.7113

Table 4 – *continued*

Source number	Source name	l	b	$\sqrt{\text{TS}}$
41	4FGL J0952.8+0712	229.7059	43.2754	7.5128
42	4FGL J1018.4+3540	188.5222	56.477	7.4996
43	4FGL J1027.6+1828	219.8976	55.8881	7.1394
44	4FGL J0932.7+1041	222.392	40.5771	7.1179
45	4FGL J0916.7+3856	183.7876	44.1629	6.9947
46	4FGL J0911.7+3349	190.7225	42.7437	6.8538
47	4FGL J1018.1+1905	217.5125	54.0062	6.7711
48	4FGL J1028.3+3108	196.9208	58.6236	6.6315
49	4FGL J1018.4+0528	236.7238	47.6703	6.5115
50	4FGL J0904.6+4238	178.6639	41.9414	6.3528
51	4FGL J1036.5+1231	231.0529	55.2439	6.3357
52	4FGL J1014.7+3210	194.8385	55.7709	6.2446
53	4FGL J1024.8+2332	210.8217	56.7983	6.2125
54	4FGL J0955.1+3551	188.5231	51.7476	6.0052
55	4FGL J0956.0+3936	182.463	51.7362	5.9421
56	4FGL J1003.6+2605	204.689	52.6589	5.6364
57	4FGL J0927.2+2454	203.6964	44.3943	5.5717
58	4FGL J1011.6+3600	188.0816	55.0793	5.4419
59	4FGL J0947.6+2215	209.1688	48.1898	5.3551
60	4FGL J1021.1+1626	222.1377	53.6665	5.1915
61	4FGL J1014.3+4112	178.9692	54.9611	5.1407
62	4FGL J0922.6+4454	175.3631	45.0912	5.1395
63	4FGL J1023.8+3002	198.8926	57.6069	5.0955
64	4FGL J0958.0+3222	194.2182	52.2649	5.029

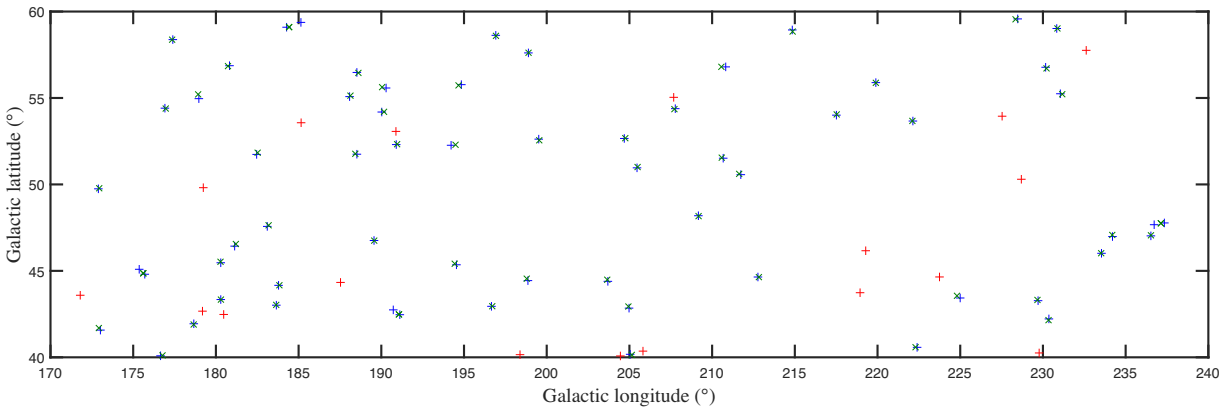


Figure 5. The exact locations of the sources from the *Fermi* catalogue are marked by blue '+' signs, while the green 'x' symbols represent the positions calculated in this analysis. The red '+' signs indicate the clusters that remain after the sixth denoising iteration.

ing the significance of each source at each step. The results indicate that the significance of sources varies across multiple denoising iterations.

A critical finding is that the significance of sources tends to increase until most relevant points are retained in the field. Beyond this point, further denoising iterations can remove real sources and critical points within remaining clusters, making additional iterations inadvisable. This trend can help establish an optimal limit for the number of denoising iterations.

An important observation is that even when we reach the optimal iteration limit, some remaining clusters may still exist that are not associated with any sources in the *Fermi* catalogue. Our investigation suggests that these clusters may not be entirely false but could instead be related to actual sources.

However, several questions remain unanswered. For instance, the strategy of removing points with an index higher than the average may not be the most optimal approach. Alternative thresholds could be more suitable, depending on either the number of iterations or specific regions of the sky. This issue remains open for further investigation and research.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to the anonymous referee for their thoughtful and constructive comments. Their time and effort in reviewing our manuscript are greatly appreciated.

DATA AVAILABILITY

The data underlying this study are publicly available. The gamma-ray data analysed in this work were obtained from the *Fermi*-LAT 4FGL-DR4 catalogue, which can be accessed at Fermi Science

Support Center. The positional data for cross-matched sources were derived from the Roma-BZCAT catalogue (5th edition), the CRATES catalogue, and the WIBRaLS catalogue, all of which are also publicly accessible. Derived data products generated during the study are available upon request from the corresponding author.

REFERENCES

Arsioli B., Chang Y., 2018, A&A, 616, A63
Arsioli B., Chang Y.-L., Musiimenta B., 2020, MNRAS, 493, 2438
Arsioli B., Chang Y.-L., Ighina L., 2024, preprint (arXiv:2411.18431)
Ballet J., Bruel P., Burnett T. H., Lott B., The Fermi-LAT collaboration, 2023, preprint (arXiv:2307.12546)
Blandford R. D., Rees M. J., 1978, in Wolfe A. M., ed., Proceedings of the Pittsburgh Conference on BL Lac Objects. Department of Physics and Astronomy, Pittsburgh, PA, p. 328
Blandford R., Meier D., Readhead A., 2019, ARA&A, 57, 467
Campana R., Massaro E., 2021, A&A, 652, A6
Campana R., Massaro E., Gasparrini D., Cutini S., Tramacere A., 2008, MNRAS, 383, 1166
Campana R., Massaro E., Bernieri E., 2018, A&A, 619, A23
Campana R., Massaro E., Bocchino F., Miceli M., Orlando S., Tramacere A., 2022, MNRAS, 515, 1676
De Berg M., 2000, Computational Geometry: Algorithms and Applications. Springer Science & Business Media, Berlin
Hawkins M., 2002, MNRAS, 329, 76
Healey S. E., Romani R. W., Taylor G. B., Sadler E. M., Ricci R., Murphy T., Ulvestad J. S., Winn J. N., 2007, ApJS, 171, 61
Hedayati Kh. H., Soor M., Akhondi F., 2024, ApJ, 968, 37
Li T., Ma Y., 1983, ApJ, 272, 317
Maraschi L., Ghisellini G., Celotti A., 1992, ApJ, 397, L5
Massaro E., Maselli A., Leto C., Marchegiani P., Perri M., Giommi P., Piranomonte S., 2015, Ap&SS, 357, 1
Meyer E. T., Fossati G., Georganopoulos M., Lister M. L., 2011, ApJ, 740, 98
Panes B., Eckner C., Hendriks L., Caron S., Dijkstra K., Jóhannesson G., de Austri R. R., Zaharijas G., 2021, A&A, 656, A62
Preparata F. P., Shamos M. I., 2012, Computational Geometry: An Introduction. Springer Science and Business Media, New York, NY
Tramacere A., Vecchio C., 2013, A&A, 549, A138
Tran T. N., Drab K., Daszykowski M., 2013, Chemometr. Intell. Lab. Syst., 120, 92
Wright E. L. et al., 2010, AJ, 140, 1868

This paper has been typeset from a TEX/LATEX file prepared by the author.