A novel approach for pre-filtering event sources using the von Mises–Fisher distribution

D. Costantin[1,2,3] • G. Menardi[3] • A.R. Brazzale[3] • D. Bastieri[1,4,5] • J.H. Fan[1,2]

Abstract Searching for as yet undetected γ-ray sources is one of the main stated goals of the Fermi Large Area Telescope Collaboration. In this paper, we explore the capability of a filtering method based on a finite mixture of von Mises–Fisher distributions. The proposed procedure is specifically designed to handle data with support on the unit sphere. The assumption of a parametric model for each high energy emitting source allows us to derive an explicit expression for both the direction of the sources and their angular resolutions. The corresponding measures are based on the directional mean and the quantiles of the single mixture components. Sound criteria of model selection can provide an automatic way to determine the number of detected sources. Additionally, a likelihood-ratio test is developed to evaluate their significance. The procedure is tested on simulated data sets of photon emissions from high energy sources within the energy range $[10 - 1,000]$ GeV. A real data example consisting of a sample of the Fermi LAT data collected over a period of about 7.2 years within the energy range $[10 - 1,000]$ GeV, in a subregion of the γ-ray sky, is furthermore provided.

Keywords astrostatistics · finite mixture · high energy photon · von Mises-Fisher distribution

D. Costantin[1]

G. Menardi

A.R. Brazzale

D. Bastieri

J.H. Fan

[1]Center for Astrophysics, Guangzhou University, Guangzhou 510006, China.

[2]Department of Education, Astronomical Science and Technology Research Laboratory, Guangdong Province PRC.

[3]Department of Statistical Sciences, University of Padova, via Cesare Battisti 241, Padua, Italy.

[4]Department of Physics and Astronomy "G. Galilei", University of Padua, via F. Marzolo 8, Padua, Italy.

[5]National Institute for Nuclear Physics, Padua Division, Padua, Italy.

mail corresponding author: denise.costantin@gmail.com

1 Introduction

The Large Area Telescope (LAT) onboard the Fermi spacecraft is a powerful instrument for exploring the high energy γ-ray sky (Atwood et al. 2009). It is a wide field-of-view pair-conversion telescope which covers the energy range from 20 MeV to more than 300 GeV (Ackermann et al. 2012). The telescope aims at resolving the high energy sky for as yet unidentified photon emitting sources and diffuse emission, leading to a better understanding of the mechanisms that both create and accelerate particles emitted by celestial objects. In more than 10 years of activity, the Fermi LAT Collaboration released various catalogs of multiple kinds of sources which cover different energy ranges.

The standard procedure of the Fermi LAT for source detection is targeted at obtaining the spectral parameters of the sources and their significance. It relies on a single source model which implements two working steps. The former step estimates the locations of the sources and provides a preliminary estimate of their spectral parameters using the pointlike method developed by Kerr (2011). This method requires the sky map to be partitioned into multiple small regions, called partition or pixelization, which is provided by the HEALPix software (Górski et al. 1998). It creates a series of tiles of equal area whose size may vary depending on the needs of the analyzer. The second step quantifies how much a source emerges from the background. The significance of the source is derived from a Poisson regression which models the number of photons

at each pixel. A first example of application is Mattox et al. (1996). Further details are provided by (Hobson et al. 2010, Section 7.4).

A different strategy for detecting high energy sources relies on variable number source models (Hobson et al. 2010, Section 7.3). The number of sources is identified by examining the whole sky map in one go, without the need of splitting it into smaller regions. A recent application to the analysis of X-ray data can be found in Jones et al. (2015), while a first attempt to extend it to the γ-ray sky is developed by Sottosanti et al. (2019). Both contributions develop a suitable Bayesian mixture model which represents the diffuse background component and identifies the directions of the photon emitting sources. A further interesting approach, which uses Bayesian techniques, is described in Selig and Enßlin (2015) and applied to the Fermi LAT data by Selig et al. (2015). Here, the γ-ray sky map is denoised, deconvolved and decomposed to obtain an individual reconstruction of the point-like photon fluxes as well as the diffuse background component.

The aim of our work is to develop a model for the high energy emitting sources present in the whole γ-ray sky. In particular, we assume that background processes were pre-filtered out, so that we can explicitly focus on the detection of the emitting sources. Our data consist in the directions of the photons observed by the Fermi LAT which we model using a finite mixture of parametric distributions with support on the unit sphere. More specifically, we will use the von Mises–Fisher distribution to represent the photon emission of each single source. The model is estimated by maximum likelihood using the expectation-maximization (EM) algorithm. Explicit expression for both, the directions of the sources and their angular resolutions, will be derived using the directional mean of the von Mises–Fisher distribution and its quantiles. The determination of the number of potential sources is an integrated part of the proposed procedure and is achieved via the application of a likelihood-based criterion for model selection. Additionally, a parametric bootstrap likelihood-ratio test is developed to evaluate the significance of the identified sources.

The paper is organized as follows. Section 2.1 provides a description of the reference data, and specifies the model. Its estimation, the determination of the number of sources and the test for evaluating their significance are detailed in Sections 2.2-2.4. The procedure is illustrated in practice and validated in Section 3. Some concluding remarks close the paper in Section 4.

2 Material and methods

2.1 Data and model specification

This contribution focuses on the photon emission of high energy γ-ray sources, as observed by the Large Area Telescope (LAT) onboard the Fermi spacecraft. Each photon, also called event, is described by a pair of Galactic coordinates which represent the direction of the emission. Additional information is available, such as the energy of the recorded photon, the time of the detection and the event type which characterizes the quality of the detection. These pieces of information will be recalled and considered later in our analysis.

The goal of our analysis is to identify the positions $\boldsymbol{\mu}$ of the unobserved sources, and to provide a measure δ of their angular resolution. To this aim, the Galactic coordinates (l, b) of each photon are suitably transformed into cartesian coordinates on the unit sphere $\boldsymbol{x} = (x, y, z) \in \mathbb{S}^2$. We furthermore assume that the photon emission of each source can be described by the von Mises–Fisher (vMF) distribution with probability density function

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}, k) = c_3(k) e^{k\boldsymbol{\mu}^T \boldsymbol{x}}. \tag{1}$$

The parameter $\boldsymbol{\mu}$ is directly linked to the mean direction of the distribution and has unit norm; $k > 0$ is a concentration parameter which characterizes how widely the photon emission spreads around the mean direction of the source. Figure 1 shows three examples of vMF distributions projected on the unit circle. All three instances share the same direction $\boldsymbol{\mu}$ but differ in their concentration, which takes on the values $k = 1$, $k = 10$ and $k = 100$. When the concentration is small, the vMF distribution reduces to the uniform distribution on the unit sphere, while, on the contrary, it collapses into a point mass centered at $\boldsymbol{\mu}$ if the concentration $k \to \infty$. The normalizing constant $c_3(k)$ is defined as

$$c_3(k) = \frac{k^{1/2}}{(2\pi)^{3/2} I_{1/2}(k)},$$

where $I_r(\cdot)$ represents the modified Bessel function of the first kind and order r.

Model (1) is particularly suitable to model high energy photon emission, since it can be regarded as akin to an isotropic Gaussian distribution on the unit sphere. That is, the vMF distribution describes observations on a sphere which are elliptically symmetrically distributed around their mean direction.

The assumption of a parametric model for each γ-ray emitting source allows us to derive an explicit expression for both the direction of the sources and their angular resolutions. Specifically, the parameter $\boldsymbol{\mu}$ of the

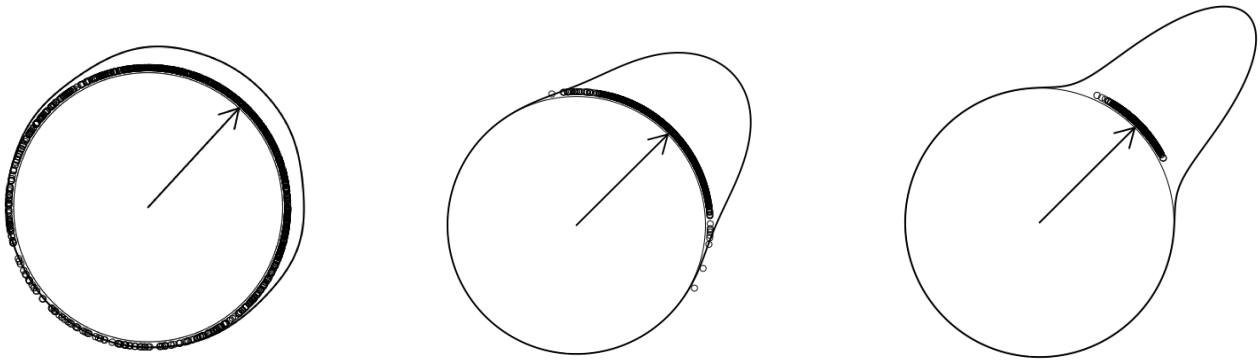


Fig. 1 Three examples of von Mises–Fisher distribution which share the same direction $\boldsymbol{\mu}$ (arrow), but differ in their concentration k, projected on the unit circle. The concentration increases from left to right: $k = 1$, $k = 10$, $k = 100$.

vMF distribution represents a natural measure of the center of the source, while the quantiles of the same distribution provide a robust measure of its angular resolution. Further distributions to model directional data are described in Sra (2016) which, however, focus on different characteristics of the data. Just to name one, the Watson distribution is a bipolar distribution which more suitably describes events which accumulated with respect to a given axis.

Suppose there are N sources in the observed sky region. We model the photon emission of these multiple sources using a mixture of parametric densities (see, e.g. McLachlan et al. 2000)

$$f(\boldsymbol{x} \mid \Theta) = \sum_{j=1}^{N} \alpha_j f_j(\boldsymbol{x} \mid \boldsymbol{\theta_j}), \tag{2}$$

where f_j is the density of a vMF distribution with parameters $\boldsymbol{\theta}_j = (\boldsymbol{\mu}_j, k_j)$, and α_j represents the probability that the j-th vMF distribution stands in the model.

2.2 Model estimation

Estimation of model (2) can be obtained by maximizing the likelihood of the observed events $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, $\boldsymbol{x}_i \in \mathbb{R}^3$, $i = (1, \ldots, n)$ with respect to all model parameters $\Theta = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, \alpha_1, \ldots, \alpha_N)$. To do so, we regard the mixture model (2) as an incomplete data structure model, where we do not observe the component membership of each observation $\boldsymbol{x}_i$. This membership can be represented by a binary vector $\boldsymbol{z_i} = (z_{i1}, \ldots, z_{iN})$ with $z_{ij} = 1$ if $\boldsymbol{x}_i$ arises from component j, and 0 otherwise. The maximum likelihood estimates of the model parameters are usually obtained by maximizing the complete data log-likelihood

$$\ln P(\boldsymbol{X}, \boldsymbol{Z} \mid \Theta) = \sum_{j=1}^{N} \sum_{i=1}^{n} \ln(\alpha_{z_{ij}} f_{z_{ij}}(\boldsymbol{x}_i \mid \boldsymbol{\theta_{z_{ij}}})), \tag{3}$$

where, however, we have to face the difficulty that the $\boldsymbol{z_i}$ are not observed. The Expectation-Maximization (EM, Dempster et al. 1977) algorithm overcomes this problem.

The EM algorithm is an iterative method which is heavily used in the statistical literature for maximizing the log-likelihood of observed data in the presence of missing values, as the $\boldsymbol{z_i}$ are. A complete formulation of the EM algorithm to estimate the parameters of a mixture of vMF distribution can be found in (Banerjee et al. 2005, Appendix A.2). Hereinafter we provide a summary of the main steps. The pseudo-code for its implementation is listed in Algorithm 1. Given an initial, possibly random, guess about the parameter vector $\Theta = (\boldsymbol{\theta_j}, \alpha_j)_{j=1,\ldots,n}$, the algorithm computes the probabilities τ_{ij} that observation x_i arises from component j

$$\tau_{ij} = \frac{\alpha_j f_j(\boldsymbol{x}_i \mid \boldsymbol{\theta_j})}{\sum_{j=1^N} \alpha_j f_j(\boldsymbol{x}_i \mid \boldsymbol{\theta_j})}.$$

The Expectation (E) step is used to determine the expectation of the complete data likelihood 3 given the probabilities τ_{ij}

$$\mathbb{E}[\ln P(\boldsymbol{X}, \boldsymbol{Z} \mid \Theta)] = \sum_{j=1}^{N} \sum_{i=1}^{n} \ln(\alpha_j)\tau_{ij} + \sum_{j=1}^{N} \sum_{i=1}^{n} \ln(f_j(\boldsymbol{x}_i \mid \boldsymbol{\theta_j}))\tau_{ij}. \tag{4}$$

The subsequent Maximization (M) step produces a novel estimation of the model parameters Θ by maximizing expression 4 under the constraints that $\sum_{j=1}^{n} \alpha_j = 1$, $\boldsymbol{\mu}_j^T \boldsymbol{\mu}_j = 1$ and $k_j \geq 0$. To this aim, the Lagrangian

$$L(\{\alpha_j\}_{j=1}^{N}, \lambda) = \sum_{j=1}^{N} \sum_{i=1}^{n} \ln(\alpha_j)\tau_{ij} + \lambda(1 - \sum_{j=1}^{N} \alpha_j) \tag{5}$$

Algorithm 1 Expectation-Maximization algorithm for a mixture of vMF distributions in $\mathbb{S}^2$

Input: A set of cartesian coordinate $\boldsymbol{x}_i$ on $\mathbb{S}^2$
Initialize all $(\alpha_j, \boldsymbol{\mu}_j, k_j)$, $j = 1, \ldots, N$
repeat
  {The E-step}
  for $i = 1$ to n **do**
    for $j = 1$ to N **do**
      $f_j(\boldsymbol{x}_i \mid \theta_j) \leftarrow c_3(k_j) e^{k_j \boldsymbol{\mu}_j^T \boldsymbol{x}_i}$
    end for
    for $j = 1$ to N **do**
      $\tau_{ij} \leftarrow \frac{\alpha_j f_j(\boldsymbol{x}_i|\theta_j)}{\sum_{l=1}^{j} \alpha_l f_l(\boldsymbol{x}_i|\theta_l)}$
    end for
  end for
  {The M-step}
  for $j = 1$ to N **do**
    $\alpha_j \leftarrow \frac{1}{n} \sum_{i=1}^{n} \tau_{ij}$
    $\boldsymbol{\mu}_j \leftarrow \sum_{i=1}^{n} p(j \mid \boldsymbol{x}_i, \Theta)$
    $\bar{r} \leftarrow ||\boldsymbol{\mu}_j||/(n\alpha_j)$
    $\boldsymbol{\mu}_j \leftarrow \boldsymbol{\mu}_j / ||\boldsymbol{\mu}_j||$
  end for
until convergence

is maximized with respect to α_j, which leads to

$$\hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} \tau_{ij}.$$

Similarly, the Lagrangian

$$L(\{\boldsymbol{\mu}_j, k_j, \lambda_j\}_{j=1}^{N}) = \sum_{j=1}^{N} \sum_{i=1}^{n} \ln(f_j(\boldsymbol{x}_i \mid \boldsymbol{\theta_j}))\tau_{ij} + \sum_{j=1}^{N} \lambda_j (1 - \boldsymbol{\mu}_j^T \boldsymbol{\mu}_j)$$

serves to maximize the expected likelihood with respect to the vMF parameters $(\boldsymbol{\mu}_j, k_j)$, while assuring that the constraint $k > 0$ is respected. After some algebra, it follows that

$$\hat{\boldsymbol{\mu}}_j = \frac{\boldsymbol{r}_j}{||\boldsymbol{r}_j||} \quad \text{and}$$
$$\frac{I_{3/2}(\hat{k}_j)}{I_{1/2}(\hat{k}_j)} = \frac{||\boldsymbol{r}_j||}{\sum_{i=1}^{n} \tau_{ij}}, \tag{6}$$

where $\boldsymbol{r}_j = \sum_{i=1}^{n} \boldsymbol{x}_i \tau_{ij}$. Since estimating k in Equation 6 involves a ratio of Bessel functions, numerical and asymptotic approximations are available. As detailed in Banerjee et al. (2005, Section 4.1), using an asymptotic approximation combined with an empirically estimated correction terml eads to the following estimate

$$\hat{k}_j = \frac{\bar{r}d - \bar{r}^3}{1 - \bar{r}^2} \tag{7}$$

where $\bar{r} = I_{3/2}(k_j)/I_{1/2}(k_j)$, which is more efficient in the presence of high dimensional data.

Steps E and M are alternated until convergence of the algorithm. Once model (2) is estimated, we can associate each event to the corresponding potential source via the maximum a posteriori rule. That is, $\mathbf{x}_i$ is assigned to component j if

$$\hat{\tau}_{ij} = \frac{\alpha_j f_j(x_i \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{j}})}{\sum_{\ell=1^N} \hat{\alpha}_\ell f_\ell(x_i \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}})} \tag{8}$$

is maximum among the set $\{\hat{\tau}_{ij}\}_{j=\ldots N}$.

2.3 Model choice

So far, model estimation has been performed by assuming that the number N of mixture components is known. In fact, this is far from being a realistic assumption, hence some criterion to select the number of components is required. To this aim, the Bayesian Information Criterion (BIC) Schwarz et al. (1978) is unarguably the most common approach, as it is shown to consistently estimate the number of mixture components under some regularity conditions (see, e.g., McLachlan et al. 2000, Chapter 6). In general, the more complex the model, the higher the data likelihood. However, to avoid overfitting, choosing the maximum likelihood model, among competitors characterized by a different number of parameters is not suggested. In this direction, the BIC provides a measure of goodness of fit

by penalizing the model likelihood with respect to the number of parameters. Formally,

$$BIC(N) = \sum_{i=1}^{n} \log \sum_{j=1}^{N} f(\boldsymbol{x}_i \mid \hat{\Theta}) - \frac{1}{2}\nu_N \log(n), \tag{9}$$

where ν_N is the number of parameters of a model with N components. According to this criterion, we choose the model which maximizes the BIC. A further possibility is the Integrated Classification Likelihood (ICL) criterion (Biernacki et al. 2000). With respect to the BIC, this criterion uses a penalty which favours mixture models with well-separated components.

Indeed, selecting the best model according to the BIC, or any other selection criterion, does not exclude the possibility of false positives. In other words, there is no guarantee that all identified mixture components correspond to real sources. Taking advantage of the parametric nature of model (2), we hence developed a measure of significance for the potentially detected sources based on a likelihood ratio test (LRT) which we specifically tailored the underlying context. Likelihood ratio tests are well studied in statistics and widely used in applications, because of their appealing asymptotic properties. Given two nested models, the LRT compares the likelihood of the global model with the likelihood of a reduced (null) model. The larger the former is with respect to the latter, the more evidence do the data provide against the null model.

However, spherical data are often much concentrated around their location $\boldsymbol{\mu}$, while the classical asymptotic results apply in cases where the concentration parameter k is neither too small nor too large (Paindaveine and Verdebout 2019). In addition, the common χ^2 limiting null distribution of the LRT statistics requires a set of regularity conditions to hold, which are not satisfied by general mixture models (Hartigan 1977, 1985). Because of these two difficulties, we will compute the null distribution of the likelihood ratio test using a modified parametric bootstrap procedure.

Consider model (2) and suppose we want to assess the significance of the j_0-th component, that is,

$$H_0 : \alpha_{j_0} = 0 \quad \text{against} \quad H_1 : \alpha_{j_0} > 0. \tag{10}$$

We define the null model as

$$\begin{aligned} f_0(\boldsymbol{x} \mid \hat{\Theta} \setminus \{\hat{\alpha}_{j_0}, \hat{\boldsymbol{\theta}}_{\boldsymbol{j_0}}\}) \\ = \hat{\alpha}_1^* f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{1}}) + \cdots + \hat{\alpha}_{j-1}^* f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{j-1}}) + \\ + \hat{\alpha}_{j+1}^* f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}_{j+1}) + \cdots + \hat{\alpha}_N^* f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{N}}), \end{aligned} \tag{11}$$

where $\hat{\alpha}_j^* = \hat{\alpha}_j/(1 - \hat{\alpha}_{j_0})$. Note that the values of the vMF parameters $\boldsymbol{\theta_j}$ and of the mixing proportions α_j are kept fixed to their estimated values $\hat{\boldsymbol{\theta}}_{\boldsymbol{j}}$ and $\hat{\alpha}_j$. That is, we assume that the directions $\boldsymbol{\mu_j}$ of the "confirmed" sources were correctly identified. Furthermore, the corresponding mixing proportions $\hat{\alpha}_j$ are renormalized so as to sum up to one, and all events are re-allocated to the $N-1$ model components via the maximum a posteriori rule (8).

The alternative hypothesis is represented by the embedding model

$$f(\boldsymbol{x} \mid \hat{\Theta}) = \pi_0 f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}_{\boldsymbol{j_0}}) + (1-\pi_0) f_0(\boldsymbol{x} \mid \hat{\Theta} \setminus \{\hat{\alpha}_{j_0}, \hat{\boldsymbol{\theta}}_{\boldsymbol{j_0}}\}),$$

where the mixing proportion $\pi_0 \in [0,1]$ is the only unknown parameter which we estimate by maximum likelihood. Note that, if the omitted model component corresponds to a true source, we expect the estimate of π_0 to be close to $\hat{\alpha}_{j_0}$, and almost zero otherwise. The two hypothesis in (12) can be rewritten as

$$H_0 : \pi_0 = 0 \quad \text{against} \quad H_1 : \pi_0 > 0. \tag{12}$$

We define the likelihood ratio test as in Davison and Hinkley (1997, §4.3.2)

$$T = \frac{1}{n} \sum_{i=1}^{n} \log \frac{f(\boldsymbol{x} \mid \hat{\Theta})}{f_0(\boldsymbol{x} \mid \hat{\Theta} \setminus \{\hat{\alpha}_{j_0}, \boldsymbol{\theta}_{\boldsymbol{j_0}}\})} \tag{13}$$

and compute its null distribution via parametric bootstrapping. That is, we generate B samples of size n from the mixture of vMF densities $f_0(\boldsymbol{x} \mid \hat{\Theta} \setminus \{\hat{\alpha}_{j_0}, \hat{\boldsymbol{\theta}}_{\boldsymbol{j_0}}\})$ and define the corresponding p-value as

$$p = \frac{1 + \sum_{b=1}^{B} (T_b > T)}{1 + B}. \tag{14}$$

Here, T_b denotes the test statistic (13) computed on the b-th bootstrap sample and T the observed value. Once that one non-significant component is removed from the model, the mixing proportions α_j can be renormalized to guarantee they sum up to one, and the events reallocated to the remaning components via the maximum a posteriori rule (8).

2.4 Implementaion details

Event reconstruction is subject to several sources of uncertainty. These include among others the quality with which the photon direction is reconstructed by the LAT, as outlined by the Fermi Science Support Center (FSSC). In addition, high energy photon emitting sources are typically highly concentrated. A possible side effect is that we encounter "overpowered" p-values, which are detrimental if the aim of our LRT is to pinpoint spurious sources. Indeed, two sources may be qualified as different based on their p-values,

when a simple eye inspection of the map highlights that they actually represent the same cluster. We hence re-designed the classical procedure for likelihood ratio testing to overcome these difficulties incorporating two amendments.

First, we account for the uncertainty of the reconstructed events by including into our analysis the 95% containment angle (CA95) of the corresponding event class (Ackermann et al. 2013).

More precisely, instead of using the original set of observations, $\{(l_i, b_i)\}_{i=1,\dots,n}$, we rely upon a new set of Galactic coordinates $\{(l_i^*, b_i^*)\}_{i=1,\dots,n}$, where

$$l_i^* = l_i + \epsilon_i$$
$$b_i^* = b_i + \epsilon_i$$

and $\epsilon_i \sim N(0, CA95_i^2)$. This blurred set of observations incorporates the uncertainty on the data by noising the reconstructed events via the addition of a normal error term with standard deviation set to the reconstructed 95% containment angle. Model (2) is then estimated using the cartesian transformation of the pairs $\{(l_i^*, b_i^*)\}_{i=1,\dots,n}$.

The second adaptation modifies the bootstrap procedure by including an additional component of variability which avoids overpowered p-values in the presence of highly concentrated clusters. This is achieved by reducing all concentration parameters k of the null model (11) by a fixed amount.

3 Model Validation

3.1 Empirical setting

This section presents the results of the application of the proposed procedure on some sets of sources simulated from one of the catalogues released from the Fermi LAT collaboration and spread on the whole sky map, and on a set of real data selected from a circumscribed region.

For each of the considered data sets, a set of models has been estimated with different number of mixture components. Then the best one according to the BIC has been selected. It is worth to note, however, that similar, not reported, results hold by selecting the best model according to the ICL criterion. The 5% significance of each component has been evaluated via the likelihood ratio test described in Section 2.3, run with 500 bootstrap samples. Due to the limited number of bootstrap samples, we have not provided any correction for multiple testing; however, it is worth highlighting that the computed p-values cluster homogeneously around values that are either very close to zero or one.

In both the simulated and the real data, the true source of photon emission is known, as well as its positional error. Once that model (2) has been estimated on the data, and cleaned out from the not significant components via the test (13), each component is associated to the source which reports the minimum distance. The association is judged to be successful if the distance between the true and the detected source is below the sum of the positional error of the true source and the positional error of the detected one. The latter value is computed in all analyses which follow as the distance between the estimated centroid and the 95% quantile of the estimated component.

As a benchmark, we have also applied to all data sets the single linkage clustering method, associated with the Minimum Spanning Tree (MST) of the data, based on the angular distance between photons (see, Campana et al. 2007, for a comprehensive development of this procedure for the detection of γ-ray sources). Since, in this case, there is no obvious method to select the number of detected sources, for the sake of comparison, we set this number to the actual number of sources, as reported by the catalogues released from the the Fermi LAT.

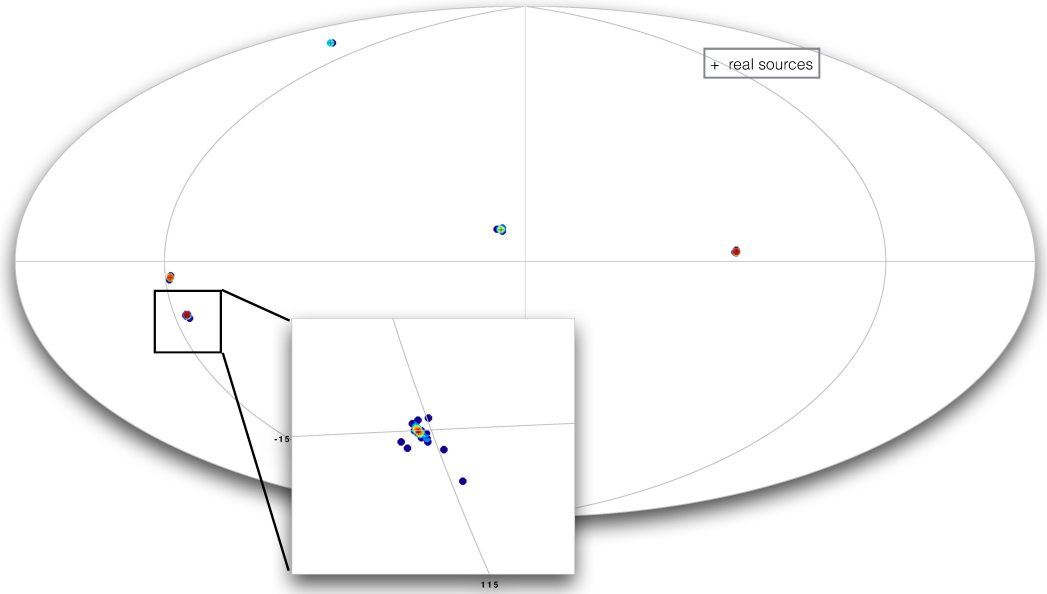
All analyses have been performed within the R computing environment (R Core Team 2019) with the aid of the `movMF` package (Hornik and Grün 2014) as well as some ad-hoc built routines.

3.2 A simple illustrative example

As a first step of our empirical work, we illustrate the proposed procedure on the simulated toy example displayed in Figure 2. Five sources were simulated from the 3FHL catalog (Ajello et al. 2017), using the Fermi LAT tool `gtobssim` and the same instrument response functions (IRFs) employed to produce the 3FHL catalog, that is, `Pass8`. Each source emission has been generated within an energy range [10 – 1,000] GeV assuming a power law (PL) spectrum. The simulated sample is rather small, amounting to $n = 88$ photons.

The BIC leads to the choice of a mixture components with $N = 10$ sources, whose estimated parameters are listed in Table 1, along with all the results of the analysis. The LRT test confirms the 5% significance of the 5 sources.

The performance of the proposed method was evaluated according to two criteria. As a first criterion, we associate the estimated sources with the true ones which present minimum distance, and confirm the association if the distance between the true and the detected source is below the sum of their positional errors. As a summarizing measure of the quality of the association,



ID	L (deg)	B (deg)	n_j
3FHL J0007.9+4711	115.31	-15.03	30
3FHL J0014.7-5801	118.08	-4.50	19
3FHL J1125.0-5806	-68.29	2.87	26
3FHL J1726.2-1710	7.63	10.03	9
3FHL J1153.2+4929	145.62	64.99	4

Fig. 2 Aitoff projection of photons simulated from 5 sources. The crosses identify the real source direction. Red colours indicate a higher concentration of photons than those tending towards blue. Below, the 3FHL name of the sources, their Galactic coordinates (L,B), and the number of simulated photons n_j.

we compute the True Positive Rate (TPR) and the False Positive Rate (FPR). The former index is defined as the proportion of true sources correctly detected, while the latter one corresponds to the proportion of estimated components which are not associated with any source. In all cases, the distance between the estimated and the detected source does not exceed the tolerance threshold given by the sum of the positional errors of the true and the detected sources. The overall TPR takes its maximum value one and the FPR gets the optimal value zero.

As a second evaluation criterion, in the case of simulated data, we take advantage of the knowledge of the pertaining source of each simulated photon emission. This allows us to discuss the use of the proposed model not only with respect to the detected source location, but also with respect to the ability of the estimated model to associate the events to the pertaining detected source. To this aim, a common measure of quality of the classification is known as the Adjusted Rand Index (ARI, Hubert and Arabie 1985). The ARI has expected value 0 under random allocation while the higher its value (with maximum set to 1), the better the agreement between two partitions; negative values are also possible, indicating a classification worse than what would be expected under a random allocation. The ARI of the classification of the photons simulated in the toy example is equal to 0.67 before removing the not significant components, and it grows to one when the only significant components are considered, that is, every simulated photon has been correctly allocated.

The performance of the competitor MST is also excellent in the considered setting. Note, however, that the number of sources to detect has been set to the true value 5.

3.3 Simulated data

The procedure has been afterwards validated on two further more challenging examples, built by simulating all the sources with more than 3 photons registered by the 3FHL catalogue. We consider photon emissions from two time spans, the former covering an observational time of 2 years, from 247000129 MET to 309831982 MET (Figure 3, left), the latter within the same observational time of the LAT, that is, of about 7.2 years, from 247000129 MET to 474175306 MET (Figure 3, right). In the 2 years setting the number

Table 1 Empirical results on the toy example of Figure 2. The Galactic coordinates of the estimated μ_j are provided, along with the estimated k, number of allocated photons n_j, the p-value of the test on each component p; n'_j represents the number of photons assigned to the estimated source after removing the non significant components, ID is the label ID of the true sources, whose coordinates are reported in Figure 2. Additionally, the distance $d(\mu_j, \hat{\mu}_j)$ is provided, the angular resolution of the estimated sources δ_j, and the sum of the angular resolutions of the estimated and the true sources ERR_j. The last column displays a Y for the confirmed components, that is, the ones for which the distance from the associated true source is less than ERR. The lower panel displays the TPR, the FPR, the ARI before applying the test (ARI_0), the final ARI and the ARI of the MST competitor.

	l (deg)	b (deg)	k	n_j	p	n'_j	ID	$d(\mu_j, \hat{\mu}_j)$	δ_j	ERR_j	confirmed
1	115.30	-15.01	178623.48	17	0.002	30	3FHL J0007.9+4711	0.02	0.3	0.35	Y
2	114.91	-15.51	3595.84	13	0.988	-	-	-	-	-	-
3	118.02	-4.55	54404.91	16	0.002	19	3FHL J0014.7+5801	0.08	0.6	0.63	Y
4	118.09	-4.56	4624.85	3	1	-	-	-	-	-	-
5	-68.16	2.97	137635.91	19	0.002	26	3FHL J1125.0-5806	0.2	0.4	0.42	Y
6	-66.70	4.50	121550.16	3	1	-	-	-	-	-	-
7	-68.57	2.53	617003.95	4	1	-	-	-	-	-	-
8	7.63	10.05	5482.15	9	0.002	9	3FHL J1726.2-1710	0.02	1.9	1.94	Y
9	145.00	64.56	17709.63	4	0.002	4	3FHL J1153.2+4929	0.51	1.1	1.11	Y
	TPR = 1										
	FPR = 0										
	$\text{ARI}_0 = 0.67$										
	ARI = 1										
	ARI(MST) = 1										

of simulated sources which emit more than 2 photons amounts to 889 while in the 7.2 one it grows to 1,481.

The sky distribution of the data of both the time range simulations is complex. On one hand there is a strong overlapping between sources emissions in the Galactic region. On the other hand the photons are less concentrated in the extragalactic sky map. In both cases, the number of photons emitted by the true sources cover a wide range of values. The number of photons emitted by a source ranges from 3 to 3,572. The source that emits more photons, 3572, is the well known Vela Pulsar (or PSR J0835-4510), lying on the Galactic plane.

In principle, the proposed procedure would allow its direct application to the whole data set. In fact, a mixture model with 1,481 components (as it would be if we detected all the sources) would be characterized by more than 6 thousands parameters, a large number which is expected to jeopardize the accuracy of the estimates and the computational feasibility. For this reason, we follow an alternative route, consisting in splitting our main goal in multiple minor sub-goals. Specifically, the sphere surface is cut into 36 regions, obtained by the cross-product of the following partitions of the latitude and longitude values:

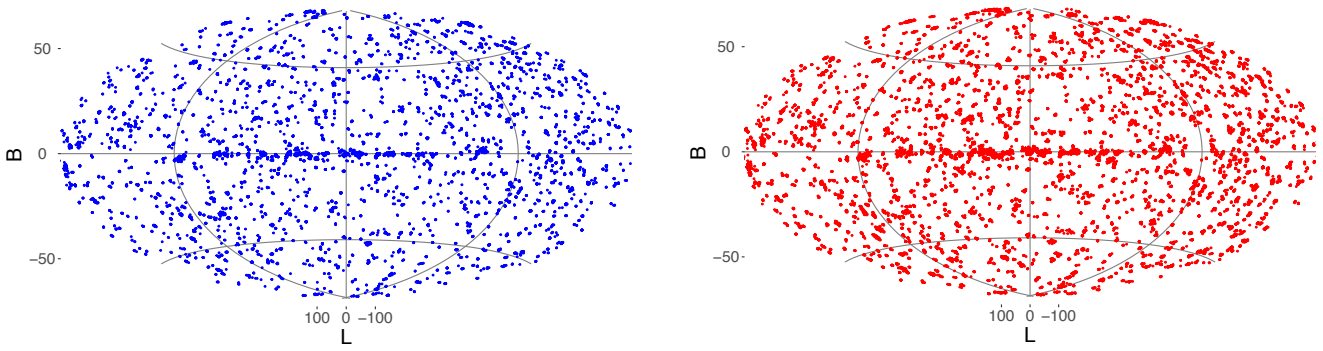
$$l \in [(0, 60), (60, 120), (120, 180), (180, 240), (240, 300), (300, 360)]$$

$$b \in [(-90, -45), (-45, -22.5), (0, 22.5), (22.5, 45), (45, 90)]$$

Instead of estimating the model on these separate regions, which would possibly make difficult the identification of sources which lie between two adjacent cuts, we enlarge each of the regions by 10 degrees along both dimensions. For instance, the region $(120, 180) \times (22.5, 45)$ is enlarged to $(115, 185) \times (17.5, 50)$ and the model is estimated on this enlarged region. Afterwards, we restrict the model to the only detected (and confirmed) sources with directional mean lying in the smaller cut.

Other criteria of splitting the data and analyzing them separately could be considered. For instance, we also evaluated the use of random subsets sampled from the whole sky map. However, as expected, this choice has resulted less effective than the above illustrated space cut approach, because it does not account for the spatial correlation of the data.

Consistently with the toy example, for each cut we have evaluated the performance of the proposed procedure in terms of TPR, FPR, and ARI. Results of the single cuts are listed in Tables 2 and 3. In terms of classification, the increase of the ARI after removing the not significant components confirms the effectiveness of the proposed test. The high values of the ARI show an excellent ability of the method to correctly allocate the photons to the pertaining source. The only few exceptions of lower values of the ARI refer to the cut regions where the information provided by the Fermi LAT is



pdg

Fig. 3 Aitoff projection of all photons generated from the 3FHL catalog sources. (Left) In a 2 years simulation, the number of simulated photons is 20,474. Out of 1,556 sources, 63 sources (4%) do not emit within the selected time and energy range. (Right) In ~ 7.2 years simulation, the number of simulated photons is 73,389. Simulated sources 1,556, of these 1,481 sources emit a number of photons > 3.

itself less reliable, due to a consistent presence of the background.

The classification of the proposed procedure is overall comparable withe one of the MST method, whose results are biased by the head start provided to the procedure by suggesting the true number of sources to detect.

With respect to the association of the estimated and the true sources based on their distance, very low values of the FPR indicate that the procedure does not detect nonexistent sources. The TPRs, on the other hand, are not completely satisfactory, as the proposed procedure tends to underestimate the true number of sources. In fact, the disagreement between low values of the TPR and high values of the ARI, in some cut, suggest that the undetected sources are the one with fewer photons. Results are consistent for the 2 years and the 7.2 years settings, with slightly worse results in the latter case where the number of data and sources is much higher.

Once that a model has been estimated and the spurious components removed at each sub-region, we are able to combine all the models together by taking advantage of its mixture structure. Specifically, given the estimated cut models $\{f^{(l)}(\cdot \mid \hat{\Theta}_l)\}_{l=1,\dots,36}$, each with structure of the (2), the overall merged model will take the form

$$f(\boldsymbol{x} \mid \hat{\Theta}) = \frac{1}{36} \sum_{l=1}^{36} f^{(l)}(\boldsymbol{x} \mid \hat{\Theta}_l)$$

Results deriving from the model merging unavoidably worsen the partial results, especially in the 7.2 years setting, where more events are observed, possibly also lying between the region cuts, which may lead to an incorrect allocation of some photons.

3.4 Real data analysis

Although our method was designed to be applied on data after a pre-processing step expected to filter out the background, we tentatively apply the developed strategy to a set of real data, collected by the Fermi LAT over the period 247000129 − 474175306 (MET), that is, of about 7.2 years. The data present a rather complex structure, exhibiting not only contributions from source emissions but also from two different background components. The first component includes a diffuse γ-ray emission, given by the interactions of Galactic cosmic rays with gas and radiation fields, which mainly affect the Galactic plane; the second component is the residual all-sky emission, characterized by an isotropic distribution in the sky, referred to as the extragalactic γ-ray background (Ackermann et al. 2015). The photons coming from all these astrophysical processes cannot be distinguished by the observer.

The analysis has been restricted to photons with energy range from 10 to 1,000 GeV, where both background components are expected to be limited, and to a sky map portion identified by the range defined by the latitude, and respectively, logitude extremes $(120, 180) \times (45, 90)$. In this region, the diffuse γ-ray emission is expected to be neglectable. Besides the favouring setting, the background, in the considered data, amounts to about the 90%. At the selected energy range, a further advantage is that also the 95% containment angle of the direction of the reconstructed photons, by normal incidence (and a certain classification of the event), is expected to be smaller than at lower energy.

Figure 4 illustrates the whole sky map from which the real data were drawn. Results are reported in Table 4. While they are remarkably worsened, with respect to the simulated setting, the overall performance

of the procedure are overall quite satisfactory, given the consistent presence of background data. In the considered cut, the model identifies 26 out of the 55 sources. Among these, the 50% has distance less than 0.27 degrees from the associated true source. The TPR amounts to 0.47 and the FPR to 0.19.

4 Discussion

In this paper we propose a pre-filtering method based on a finite mixture of von Mises–Fisher distributions which is specifically designed to handle observations with support on the unit sphere. The intent is to model and cluster the γ-ray photons recorded by the Fermi LAT as they occur in the 3-dimensional space. Every mixture component is supposed to represent a high energy emitting source. The model is estimated by maximum likelihood using the EM algorithm. The number of mixture components is determined according to the BIC, though the ICL criterion could equally well be used. We furthermore developed a parametric bootstrap procedure to assess the significance of the identified sources.

We tested and benchmarked our algorithm on a number of simulated data sets of increasing complexity. The proposed procedure performed optimally on the toy example of Section 3.2, where it identified all five sources present in the analyzed sky region and correctly assigned all detected photons to their corresponding emitting source. The algorithm furthermore performed well on the two additional, more challenging examples of Section 3.3, which consider whole-sky photon emission over 2 and about 7.2 years of observation, respectively. Identifying the 889 and 1,481 sources present in these two data sets is challenging because of both, the overlapping of the sources and the high variability of the number of photons emitted by the sources. We were capable of identifying around 60% of the sources present in both analyzed sky regions within a median distance of 0.15 and 0.09, respectively, and a FPR of only 4%. The ARIs amount to 0.79 and 0.65, respectively, which reveals that a large portion of the γ-ray photons get correctly assigned to their emitting source.

The procedure was also tested on the real data set presented in Section 3.4 which considers the photons collected by the Fermi LAT over the period 247000129 – 474175306 (MET) with energy ranging from 10 to 1,000 GeV in the sky region defined by $(120, 180) \times (45, 90)$. Despite we did not filter out the background, as required by our procedure, our finite mixture model performed remarkably well. We identified almost half of the sources present within an angular distance of less than 0.27 degrees from the associated true sources and a FPR of 0.19.

Though our procedure has proved to perform more than satisfactorily on both, simulated and real, data, there are some margins of improvement.

First, our procedure requires that the diffuse and isotropic γ-ray background be filtered out. Future research will focus on how to introduce an additional component into the mixture model to represent the background. A possibility is to mimic the flat though locally irregularly shaped background with a mixture of vMF distributions all characterized by a very low values of k. A further aspect which needs be considered is that the vMF distribution is a good approximation for point-like sources, but it may behave less satisfactorily when diffuse sources need be modelled. In addition, the emissions of the sources could scatter in a non circular way around their center. An alternative to the vMF is the Fisher-Bingham distribution, which accounts for different shapes around the estimated direction (Mardia and Jupp 2009, p. 174). The major drawback of this distribution is that there is no integral representation of the normalizing constant and a total o 8 parameters need be estimated.

From the statistical point of view, we want to gain more insight into the ad hoc bootstrap procedure. Preliminary investigation (results not shown here) revealed that the amount of variability we introduce in the bootstrap little influences the significance of the sources provided it is larger than a factor of 10. Although all FPRs encountered in our analysis are acceptable, a suitable form of correction for the look-elsewhere-effect needs be introduced especially if we aim at data sets with a very high number of sources.

A last point of improvement is on the computational side. Currently, we are not capable of analyzing the whole sky map in a go. Indeed, the computational burden for the two simulated whole-sky data sets had to be distributed over a suitable number of adjacent cuts.

Acknowledgements

We would like to thank the Associate Editor and an anonymous Referee for their most useful comments which largely helped us improving a previous version of this manuscript. We acknowledge the financial support by Prof. Junhui Fan (grants n. NSFC11733001 and n. NSFCU1531245). Furthermore, this research was supported by SID 2018 grant "Advanced statistical modeling for indexing celestial objects" (BIRD185983) awarded by the Department of Statistical Sciences of the University of Padova.

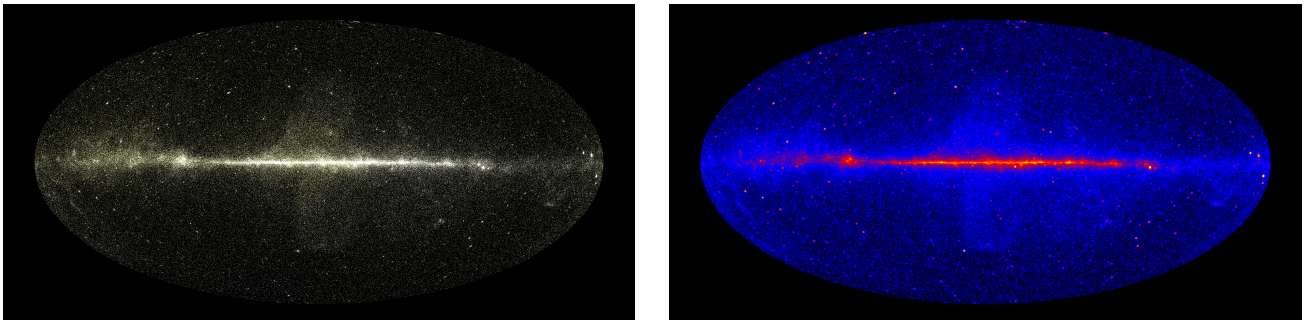


Fig. 4 Image maps of all photons in the real data, within the energy range $10 - 1,000$ GeV, collected over a period of 7.2 years. Image scale of 0.1 degrees/pixel and in logarithm scale. Brighter colors indicates brighter sources.

Table 2 Summarizing results for the single space regions: N is the number of true sources in the region, $\hat{N}$ the number of estimated sources, $d(\mu_j, \hat{\mu}_j)$ is the median distance between the true and the detected sources, δ is the median positional error of the estimated sources in each cut, ERR is the median sum of the positional error of the true and estimated sources, TPR is the True Positive Rate in the cut, FPR is the False Positive Rate in the cut, ARI_0 and ARI are the ARI values before and after removing the not significant components while ARI(MST) is the ARI resulting from the application of the MST clustering method. Results refer to the 2 years data.

	cut	N	$\hat{N}$	$d(\mu_j, \hat{\mu}_j)$	δ	ERR	TPR	FPR	ARI_0	ARI	ARI (MST)
1	$(0, 60) \times (-90, -45)$	19	11	0.13	0.59	0.63	0.58	0.08	0.27	0.5	0.99
2	$(60, 120) \times (-90, -45)$	16	11	0.19	1.24	1.27	0.69	0.08	0.83	0.9	0.99
3	$(120, 180) \times (-90, -45)$	22	15	0.22	0.98	1.01	0.68	0	0.74	0.81	0.77
4	$(180, 240) \times (-90, -45)$	16	13	0.13	0.54	0.59	0.81	0	0.59	0.92	0.68
5	$(240, 300) \times (-90, -45)$	24	15	0.22	0.93	0.96	0.62	0	0.53	0.8	0.7
6	$(300, 360) \times (-90, -45)$	16	13	0.11	0.61	0.62	0.81	0	0.74	0.88	0.72
7	$(0, 60) \times (-45, -22.5)$	16	12	0.39	0.85	0.89	0.75	0	0.77	0.58	0.66
8	$(60, 120) \times (-45, -22.5)$	20	16	0.12	0.61	0.64	0.8	0	0.71	0.97	0.62
9	$(120, 180) \times (-45, -22.5)$	24	19	0.15	0.55	0.58	0.79	0	0.71	0.97	0.92
10	$(180, 240) \times (-45, -22.5)$	18	12	0.48	1.27	1.33	0.67	0	0.7	0.79	0.91
11	$(240, 300) \times (-45, -22.5)$	28	14	0.14	0.51	0.55	0.5	0	0.6	0.94	0.99
12	$(300, 360) \times (-45, -22.5)$	18	14	0.17	1.01	1.04	0.78	0	0.75	0.92	0.71
13	$(0, 60) \times (-22.5, 0)$	43	23	0.13	0.59	0.64	0.53	0	0.55	0.57	0.71
14	$(60, 120) \times (-22.5, 0)$	34	23	0.19	0.78	0.81	0.68	0	0.74	0.87	0.88
15	$(120, 180) \times (-22.5, 0)$	33	10	0.18	0.64	0.66	0.3	0	0.26	0.83	0.88
16	$(180, 240) \times (-22.5, 0)$	24	10	0.18	0.98	1	0.42	0.17	0.16	0.41	0.75
17	$(240, 300) \times (-22.5, 0)$	30	16	0.21	1.14	1.16	0.53	0	0.14	0.87	0.99
18	$(300, 360) \times (-22.5, 0)$	42	19	0.13	0.66	0.68	0.45	0.05	0.48	0.67	0.93
19	$(0, 60) \times (0, 22.5)$	36	13	0.31	1.32	1.35	0.36	0	0.5	0.35	0.75
20	$(60, 120) \times (0, 22.5)$	45	22	0.17	0.73	0.77	0.49	0.08	0.57	0.7	0.92
21	$(120, 180) \times (0, 22.5)$	48	23	0.17	0.68	0.71	0.48	0	0.49	0.74	0.86
22	$(180, 240) \times (0, 22.5)$	38	19	0.26	0.75	0.78	0.5	0	0.41	0.92	0.89
23	$(240, 300) \times (0, 22.5)$	20	11	0.2	0.82	0.84	0.55	0.08	0.1	0.67	0.86
24	$(300, 360) \times (0, 22.5)$	44	16	0.18	0.9	0.92	0.36	0	0.54	0.68	0.61
25	$(0, 60) \times (22.5, 45)$	14	9	0.36	0.72	0.76	0.64	0.18	0.4	0.59	0.92
26	$(60, 120) \times (22.5, 45)$	33	13	0.14	0.86	0.9	0.39	0.13	0.44	0.67	0.84
27	$(120, 180) \times (22.5, 45)$	40	21	0.3	1.45	1.48	0.52	0	0.54	0.88	0.93
28	$(180, 240) \times (22.5, 45)$	16	5	0.36	1.56	1.59	0.31	0	0.67	0.4	0.74
29	$(240, 300) \times (22.5, 45)$	19	12	0.42	2.07	2.12	0.63	0	0.79	0.88	0.8
30	$(300, 360) \times (22.5, 45)$	21	13	0.17	1.02	1.06	0.62	0	0.73	0.89	0.95
31	$(0, 60) \times (45, 90)$	19	13	0.14	0.74	0.79	0.68	0	0.38	0.92	0.93
32	$(60, 120) \times (45, 90)$	22	16	0.12	0.81	0.83	0.73	0	0.81	0.94	0.94
33	$(120, 180) \times (45, 90)$	31	18	0.07	0.49	0.51	0.58	0.05	0.3	0.65	0.94
34	$(180, 240) \times (45, 90)$	20	10	0.3	0.97	0.98	0.5	0	0.14	0.6	0.43
35	$(240, 300) \times (45, 90)$	18	12	0.21	0.91	0.94	0.67	0	0.73	0.91	0.96
36	$(300, 360) \times (45, 90)$	16	11	0.12	0.69	0.71	0.69	0.08	0.67	0.91	0.89
		889	516	0.15	0.77	0.80	0.58	0.04		0.79	0.41

Table 3 Cf. Table 2. Results refer to the 7.2 years data.

	cut	N	$\hat{N}$	$d(\mu_j, \hat{\mu}_j)$	δ	ERR	TPR	FPR	ARI_0	ARI	ARI (MST)
1	$(0, 60) \times (-90, -45)$	36	18	0.11	0.6	0.65	0.5	0.1	0.22	0.53	0.96
2	$(60, 120) \times (-90, -45)$	29	21	0.08	0.44	0.5	0.72	0	0.56	0.88	0.74
3	$(120, 180) \times (-90, -45)$	32	26	0.07	0.57	0.62	0.81	0	0.74	0.91	0.64
4	$(180, 240) \times (-90, -45)$	35	20	0.11	0.49	0.53	0.57	0	0.35	0.85	0.79
5	$(240, 300) \times (-90, -45)$	41	28	0.12	0.61	0.64	0.68	0.03	0.36	0.82	0.84
6	$(300, 360) \times (-90, -45)$	30	24	0.08	0.45	0.5	0.8	0	0.64	0.87	0.68
7	$(0, 60) \times (-45, -22.5)$	31	19	0.14	0.72	0.77	0.61	0	0.72	0.79	0.61
8	$(60, 120) \times (-45, -22.5)$	36	21	0.1	0.54	0.57	0.58	0	0.55	0.91	0.6
9	$(120, 180) \times (-45, -22.5)$	44	28	0.09	0.56	0.59	0.64	0	0.53	0.95	0.89
10	$(180, 240) \times (-45, -22.5)$	31	18	0.17	0.66	0.7	0.58	0	0.47	0.85	0.96
11	$(240, 300) \times (-45, -22.5)$	49	26	0.13	0.43	0.48	0.53	0.07	0.43	0.77	0.94
12	$(300, 360) \times (-45, -22.5)$	38	27	0.08	0.53	0.56	0.71	0	0.66	0.91	0.68
13	$(0, 60) \times (-22.5, 0)$	65	37	0.11	0.51	0.53	0.57	0.08	0.48	0.7	0.31
14	$(60, 120) \times (-22.5, 0)$	55	37	0.1	0.51	0.55	0.67	0.05	0.57	0.82	0.71
15	$(120, 180) \times (-22.5, 0)$	54	24	0.07	0.54	0.56	0.44	0.04	0.18	0.93	0.89
16	$(180, 240) \times (-22.5, 0)$	39	20	0.09	0.65	0.67	0.51	0.05	0.15	0.48	0.75
17	$(240, 300) \times (-22.5, 0)$	44	22	0.05	0.54	0.56	0.5	0.04	0.1	0.45	0.98
18	$(300, 360) \times (-22.5, 0)$	68	30	0.09	0.5	0.53	0.44	0	0.37	0.94	0.5
19	$(0, 60) \times (0, 22.5)$	52	28	0.11	0.65	0.69	0.54	0	0.45	0.73	0.86
20	$(60, 120) \times (0, 22.5)$	56	31	0.08	0.59	0.63	0.55	0.06	0.44	0.78	0.82
21	$(120, 180) \times (0, 22.5)$	61	27	0.09	0.57	0.6	0.44	0	0.43	0.78	0.7
22	$(180, 240) \times (0, 22.5)$	57	29	0.07	0.6	0.63	0.51	0.03	0.22	0.68	0.8
23	$(240, 300) \times (0, 22.5)$	46	19	0.1	0.71	0.77	0.41	0.1	0.07	0.78	0.65
24	$(300, 360) \times (0, 22.5)$	66	39	0.09	0.53	0.56	0.59	0	0.39	0.83	0.66
25	$(0, 60) \times (22.5, 45)$	30	15	0.09	0.69	0.74	0.5	0.06	0.24	0.72	0.88
26	$(60, 120) \times (22.5, 45)$	59	28	0.14	0.48	0.5	0.47	0.03	0.39	0.61	0.81
27	$(120, 180) \times (22.5, 45)$	57	33	0.1	0.77	0.81	0.58	0.03	0.44	0.86	0.72
28	$(180, 240) \times (22.5, 45)$	40	19	0.15	0.58	0.65	0.48	0	0.61	0.77	0.74
29	$(240, 300) \times (22.5, 45)$	37	22	0.07	0.58	0.62	0.59	0	0.68	0.81	0.75
30	$(300, 360) \times (22.5, 45)$	49	29	0.1	0.58	0.62	0.59	0	0.52	0.91	0.75
31	$(0, 60) \times (45, 90)$	34	20	0.09	0.51	0.53	0.59	0.05	0.26	0.61	0.93
32	$(60, 120) \times (45, 90)$	47	28	0.07	0.48	0.53	0.6	0.03	0.61	0.86	0.85
33	$(120, 180) \times (45, 90)$	57	28	0.04	0.36	0.38	0.49	0.18	0.18	0.58	0.95
34	$(180, 240) \times (45, 90)$	33	13	0.06	0.56	0.6	0.39	0.07	0.11	0.84	0.42
35	$(240, 300) \times (45, 90)$	36	25	0.09	0.54	0.57	0.69	0	0.52	0.9	0.92
36	$(300, 360) \times (45, 90)$	29	17	0.09	0.75	0.79	0.59	0	0.49	0.9	0.94
		1481	890	0.09	0.54	0.57	0.60	0.04		0.65	0.41

Table 4 Cf. Table 2. Results refer to the real data. For this reason, the pertaining source of each photon is not known, hence the ARI cannot be computed.

cut	N	$\hat{N}$	$d(\mu_j, \hat{\mu}_j)$	δ	ERR	TPR	FPR
$(120, 180) \times (45, 90)$	55	26	0.27	1.35	1.38	0.47	0.19

References

Acero, F., Ackermann, M., Ajello, M., Albert, A., Atwood, W., Axelsson, M., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., *et al.*: The Astrophysical Journal Supplement Series **218**(2), 23 (2015)

Ackermann, M., Ajello, M., Albert, A., Allafort, A., Atwood, W.B., Axelsson, M., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., et al.: The Astrophysical Journal Supplement Series **203**(1), 4 (2012)

Ackermann, M., Ajello, M., Allafort, A., Asano, K., Atwood, W., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bechtol, K., *et al.*: The Astrophysical Journal **765**(1), 54 (2013)

Ackermann, M., Ajello, M., Albert, A., Atwood, W., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bechtol, K., Bellazzini, R., *et al.*: The Astrophysical Journal **799**(1), 86 (2015)

Ajello, M., Atwood, W., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bellazzini, R., Bissaldi, E., Blandford, R., Bloom, E., *et al.*: The Astrophysical Journal Supplement Series **232**(2), 18 (2017)

Atwood, W., Albert, A., Baldini, L., Tinivella, M., Bregeon, J., Pesce-Rollins, M., Sgrò, C., Bruel, P., Charles, E., Drlica-Wagner, A., et al.: arXiv preprint arXiv:1303.3514 (2013)

Atwood, W., Abdo, A.A., Ackermann, M., Althouse, W., Anderson, B., Axelsson, M., Baldini, L., Ballet, J., Band, D., Barbiellini, G., *et al.*: The Astrophysical Journal **697**(2), 1071 (2009)

Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S.: Journal of Machine Learning Research **6**(Sep), 1345 (2005)

Baudry, J.-P., Cardoso, M., Celeux, G., Amorim, M.J., Ferreira, A.S.: Advances in Data Analysis and Classification **9**(2), 177 (2015)

Biernacki, C., Celeux, G., Govaert, G.: IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(7), 719 (2000)

Campana, R., Massaro, E., Gasparrini, D., Cutini, S., Tramacere, A.: A MST algorithm for source detection in γ-ray images **921**(1), 536 (2007). American Institute of Physics

Celeux, G.: Advances in Data Analysis, 3 (2007)

Damiani, F., Maggio, A., Micela, G., Sciortino, S.: The Astrophysical Journal **483**(1), 350 (1997)

Davison, A.C., Hinkley, D.V.: Bootstrap Methods and Their Application. Cambridge University Press, Cambridge, UK (1997)

Dempster, A.P., Laird, N.M., Rubin, D.B.: Journal of the Royal Statistical Society: Series B (Methodological) **39**(1), 1 (1977)

Górski, K.M., Hivon, E., Wandelt, B.D.: arXiv preprint astro-ph/9812350 (1998)

Gorski, K.M., Hivon, E., Banday, A., Wandelt, B.D., Hansen, F.K., Reinecke, M., Bartelmann, M.: The Astrophysical Journal **622**(2), 759 (2005)

Hartigan, J.: Classification and Clustering, 45 (1977)

Hartigan, J.: Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer **II**, 807 (1985)

Hillen, T., Painter, K.J., Swan, A.C., Murtha, A.D.: Mathematical Biosciences & Engineering **14**(3), 673 (2017)

Hobson, M.P., Jaffe, A.H., Liddle, A.R., Mukherjee, P., Parkinson, D.: Bayesian Methods in Cosmology (2010)

Hornik, K., Grün, B.: Journal of Statistical Software **58**(10), 1 (2014)

Hornik, K., Feinerer, I., Kober, M., Buchta, C.: Journal of Statistical Software **50**(10), 1 (2012)

Hubert, L., Arabie, P.: Journal of Classification **2**(1), 193 (1985)

Jones, D.E., Kashyap, V.L., Van Dyk, D.A.: The Astrophysical Journal **808**(2), 137 (2015)

Kerr, M.: arXiv preprint arXiv:1101.6072 (2011)

Kume, A., Wood, A.T.: Biometrika **92**(2), 465 (2005)

Mardia, K.V., Jupp, P.E.: Directional Statistics **494** (2009)

Mattox, J.R., Bertsch, D., Chiang, J., Dingus, B., Digel, S., Esposito, J., Fierro, J., Hartman, R., Hunter, S., Kanbach, G., *et al.*: The Astrophysical Journal **461**, 396 (1996)

McLachlan, G.J., Peel, D.: John Wiley & Sons (2000). DOI:10.1002/0471721182

Paindaveine, D., Verdebout, T.: arXiv preprint arXiv:1901.00359 (2019)

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2019). R Foundation for Statistical Computing. https://www.R-project.org/

Schwarz, G., *et al.*: The Annals of Statistics **6**(2), 461 (1978)

Selig, M., Enßlin, T.A.: Astronomy & Astrophysics **574**, 74 (2015)

Selig, M., Vacca, V., Oppermann, N., Enßlin, T.A.: Astronomy & Astrophysics **581**, 126 (2015)

Sottosanti, A., Costantin, D., Bastieri, D., R., B.A.: New Statistical Developments in Data Science, SIS 2017 **288**, 135 (2019)

Sra, S.: arXiv preprint arXiv:1605.00316 (2016)

Sra, S., Karp, D.: Journal of Multivariate Analysis **114**, 256 (2013)

Wilks, S.S.: The Annals of Mathematical Statistics **9**(1), 60 (1938)

This manuscript was prepared with the AAS LaTeX macros v5.2.