

Bayesian Mixture Modeling of the High-Energy Photon Counts collected by the Fermi Large Area Telescope

Denise Costantin^{1 2 3}, **Andrea Sottosanti**³, **Alessandra
R. Brazzale**³, **Denis Bastieri**^{1 4 5} and **JunHui Fan**^{1 2}

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, Padova, Italy

⁴ Department of Physics and Astronomy “G. Galilei”, University of Padova, via F. Marzolo 8, Padova, Italy

⁵ National Institute for Nuclear Physics, Padova Division, Padova, Italy

Address for correspondence: Denise Costantin, Center for Astrophysics, Guangzhou University, Guangzhou 510006, China.

E-mail: denise.costantin@gmail.com.

Phone: .

Fax: .

Abstract: Identifying as yet undetected high-energy sources in the γ -ray sky is one

of the declared objectives of the *Fermi* LAT Collaboration. We develop a Bayesian mixture model which is capable of disentangling the high-energy extra-galactic sources present in a given sky region from the pervasive background radiation. We achieve this by combining two model components. The first component models the emission activity of the single sources and incorporates the instrument response function of the *Fermi* γ -ray space telescope. The second component reliably reflects the current knowledge of the physical phenomena which underly the γ -ray background. The model parameters are estimated using a reversible jump MCMC algorithm, which simultaneously returns the number of detected sources, their locations and relative intensities, and the background component. Our proposal is illustrated using a sample of the *Fermi* LAT data. In the analyzed sky region, our model correctly identifies 116 sources out of the 132 present. The detection rate and the estimated directions and intensities of the identified sources are largely unaffected by the number of detected sources.

Key words: bivariate exponential distribution; γ -ray photon, instrument response function; mixture model, reversible jump MCMC

1 Introduction

The Large Area Telescope (LAT) is the primary instrument onboard the *Fermi* γ -ray space telescope (Atwood et al., 2009). The main objective of the *Fermi* LAT Collaboration¹ is to provide a catalog of sources present in the γ -ray sky. Since the *Fermi*

¹<http://fermi.gsfc.nasa.gov/>

LAT began its survey on August 11th 2008, the Collaboration has released several catalogs which are accessible on-line. The latest version of the catalog, the Third High-Energy Source Catalog (3FHL), lists the 3,034 γ -ray sources which have been detected until now (Ajello et al., 2017). The data collected by the *Fermi* LAT are so-called *events* in the high-energy band. Each event corresponds to a high-energy photon, or γ -ray, which interacts with the space telescope. At every hit several measurements are taken. We will use the sky directions of the incoming photons, their energy and the quality of the reconstructed events, the so-called event type, to pinpoint high-energy extra-galactic emitting sources.

The main difficulty of identifying high-energy sources in astrophysical count maps is to separate the signal stemming from the sources from the background component which spreads over the whole sky. The method used to detect the sources listed in the 3FHL catalog consists of two steps. The first step uses an image analysis technique based on wavelet transforms. Broadly speaking, two wavelet transforms (see, e.g., Damiani et al., 1997; Starck and Pierre, 1998; Ciprini et al., 2007) translate the incoming signal into a shape which is suitable for detecting a large variety of sources. These can range from point sources to highly extended sources and their signal may be smeared by the detector of the LAT. The second step quantifies the statistical significance of a possible source detection using the likelihood ratio statistic (Mattox et al., 1996).

In this article we abandon the commonly used two-step technique described above and consider a Bayesian mixture model instead. We follow previous work by Jones et al. (2015) on the identification of X-ray sources and extend their model into two directions. First, the uncertainty of the measurements is modeled using the instrument

response functions provided by the *Fermi* LAT Collaboration. Second, we translate the simulation-based background model developed by the same Collaboration into a workable parametric formulation. As in practice we do not have any knowledge about the origin of a single photon, these two model components are combined into a finite mixture model with an unknown number of components. The number of sources, their locations and relative intensities, and the background component are estimated using a tailored reversible jump MCMC algorithm.

Our newly developed model was tested on a collection of γ -ray photons recorded by the *Fermi* LAT during a period of around 7.2 years of operation with promising results. We were able to identify almost 90% of the sources present in the analyzed sky region and to reconstruct their directions with high precision. A careful sensitivity analysis revealed that the parameter estimates of the identified sources remain stable once a sufficiently large number of components is specified. Although the purpose of the analysis is to disentangle high-energy emitting sources from the prominent background radiation, our model can easily be adapted to other types of highly contaminated data in the 2-dimensional space where the set of locations at which a spatial process is observed are of interest. As such, our model belongs to the wider class of models known in spatial statistics as *Spatial Point Processes* (SPPs). The approach we follow differs from the better-known branch of spatial statistics called geostatistics (or point-referenced data) which tries and reconstructs a spatial process that varies continuously in space but is observed only at a few points. A general reference for SPPs is [Gaetan and Guyon \(2010, Chap. 3\)](#), which also well explains the difference between the two approaches.

The paper is organized as follows. Section 2 gives a general overview of the *Fermi*

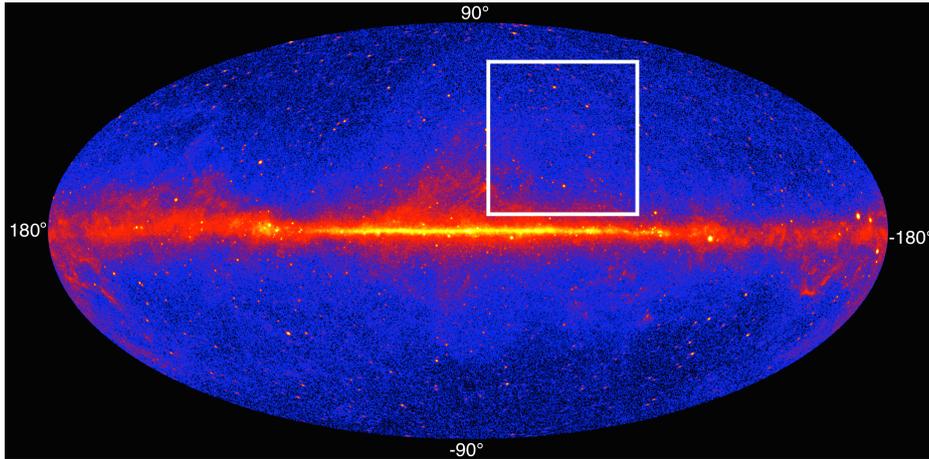


Figure 1: Whole sky map at γ -ray wavelengths and energies larger than 1 GeV based on data accumulated by the LAT over a period of five years of operation (Image Credit: NASA/DOE/Fermi LAT Collaboration). The region framed in white represents the area analyzed in this paper.

Large Area Telescope and the datasets considered in our work. The description of the proposed statistical model and of how we fit it are presented in Sections 3 and 4. Section 5 outlines the main results of our analysis, which are then discussed in Section 6.

2 The Fermi LAT data

The *Fermi* LAT is a wide field-of-view pair-conversion telescope which covers the energy range from 20 MeV to more than 300 GeV (Atwood et al., 2009). Figure 1 reproduces the whole γ -ray sky in galactic coordinates and at energies larger than 1 GeV according to the data collected by the *Fermi* Large Area Telescope over a 5 years period. We can clearly identify extra-galactic point and small-area γ -ray sources. The brighter the color, the brighter are the emitting sources. The brilliant

horizontal stripe in the center of the figure corresponds to the Galactic plane which is mainly formed by the γ -ray photons emitted by our galaxy, the Milky Way, whose center hosts a supermassive black hole. Indeed, the count map shown in Figure 1 is a mixture of photons which are emitted from both, known and as yet undetected high-energy sources and the diffuse γ -ray background.

Our primary dataset consists of the 26, 208 events recorded by the *Fermi* LAT during a period of around 7.2 years of operation and which fall in the subregion framed in white in Figure 1. This region covers the portion of the sky which lies between $[-10^\circ, -80^\circ]$ of galactic longitude and $[10^\circ, 80^\circ]$ of galactic latitude, respectively. The detected γ -ray photons cover an energy range between 10 GeV and 1, 000 GeV. Figure 2 plots the nonparametric kernel estimate of the photon density distribution. The various spikes most likely correspond to known and as yet unrevealed high-energy emitting sources. The irregularly shaped γ -ray background which spreads over the entire area is clearly visible.

Several measurements are taken by the γ -ray detector onboard the *Fermi* spacecraft for every collected event. Our purpose is to model the directions, expressed in galactic coordinates, of the incoming γ -ray photons and to use the gained insight to identify known and possibly new extra-galactic high-energy emitting sources. We will also consider the energy content of the incoming photons and the associated event type. The latter variable characterizes the quality of the detection, that is, it specifies how accurately the detector identified an incoming photon. There exist four different classes of event types which, in increasing order of reliability, are called PSF0, PSF1, PSF2 and PSF3. We will integrate both pieces of information, energy content and event type, into the model which reconstructs the direction of the high-energy

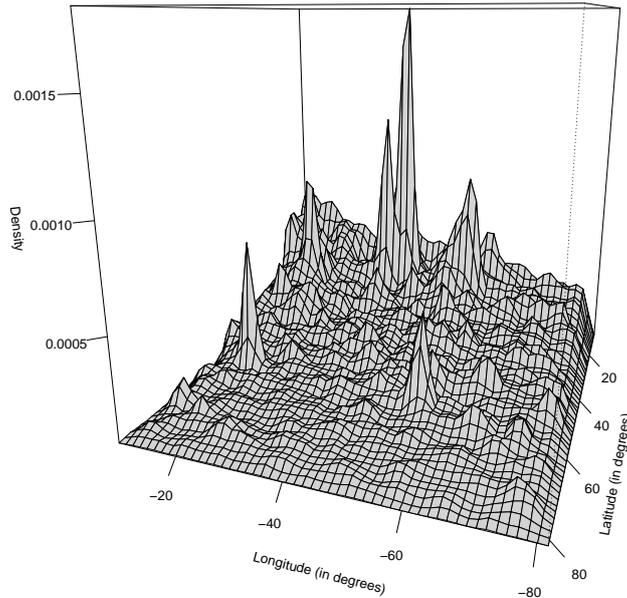


Figure 2: Nonparametric kernel estimate of the photon density based on the γ -ray count map accumulated by the *Fermi* LAT over a 7.2 years period. The map is expressed in galactic coordinates and refers to the sky region $[-10^\circ, -80^\circ] \times [10^\circ, 80^\circ]$, that is, to the area framed in white in Figure 1. The spikes represent potential high-energy extra-galactic emitting sources. The irregularly shaped γ -ray background is clearly visible.

photons. The details are given in Section 3.1.

As mentioned in the introduction, the detection of high-energy emitting sources is hampered by the presence of the photon emitting background which spreads over the whole sky. Jones et al. (2015) model it with a uniform distribution, which is fine for the X-ray photons considered in their work, but inadequate to capture the rather irregular background structure of γ -ray photons. The contaminating effect in particular dominates the galactic part of the sky map and involves non negligible difficulties when we try to discover new sources whose directions lie in this area. This is why in this paper we restrict our attention to subregions of the sky whose latitudes lie above 10° . In Section 3.2 we will present a suitable model to describe the behaviour of the γ -ray background. This model is based on a careful analysis of the intensity maps provided by the *Fermi* LAT Collaboration. The details are given in Appendix A.

3 Model definition

Denote by $x_i \in (180^\circ, -180^\circ)$ and $y_i \in (-90^\circ, 90^\circ)$ the observed galactic longitude and galactic latitude of photon i with $i = 1, \dots, n$. Let furthermore $E_i \in \mathbb{R}^+$ be the energy content of an observed photon. Each photon was either emitted by an extra-galactic source or is part of the diffuse background radiation which spreads over the entire observed area. We will hence model the directions of the photons using two different models depending on which situation occurs. These are the single source model $f_j(x_i, y_i \mid E_i, \boldsymbol{\mu}_j)$, described in Section 3.1, which characterizes how photons scatter around source j , and the background model $g_b(x_i, y_i \mid \sigma_x, \sigma_y, \delta)$ of Section 3.2.

As in practice we do not have any knowledge about the origin of a single photon, these two model components are combined into a finite mixture model of the form

$$f(x_i, y_i | E_i, \boldsymbol{\theta}_K, K) = \sum_{j=1}^K \omega_j f_j(x_i, y_i | E_i, \boldsymbol{\mu}_j) + \omega_0 g_b(x_i, y_i | \sigma_x, \sigma_y, \delta). \quad (3.1)$$

The vector $\boldsymbol{\omega} = (\omega_0, \omega_1, \dots, \omega_K)$ contains the mixing proportions $\omega_j \in (0, 1)$, with $\sum_{j=0}^K \omega_j = 1$. We can interpret these as the relative intensities of the various identified sources ω_i , for $i = 1, \dots, K$, and as the relative intensity ω_0 of the background. That is, they determine the proportion of photons emitted from an identified source and the proportion of photons which are to be assigned to the background. The vector $\boldsymbol{\mu}_j = (\mu_{jx}, \mu_{jy}) \in (180^\circ, -180^\circ) \times (-90^\circ, 90^\circ)$ contains the direction of the j th source, while $(\sigma_x, \sigma_y, \delta)$ parametrize the background distribution, with σ_x and σ_y positive values and $\delta \in (0, 1)$. Further details of the two components of model (3.1) follow in Sections 3.1 and 3.2.

Our model is hence characterized by a set $\boldsymbol{\theta}_K = \{\boldsymbol{\omega}, \{\boldsymbol{\mu}_j\}_{j=1}^K, \sigma_x, \sigma_y, \delta\}$ of $3(K+1)$ parameters which need to be estimated. Recall, furthermore, that the number $K \in \mathbb{N}$ of undetected sources is itself supposed to be unknown and needs to be estimated. We will make inference on $(\boldsymbol{\theta}_K, K)$ using Bayes' rule. The a priori available information on these parameters is summarized in Section 3.3.

3.1 The single source model

Suppose that photon i was emitted by the j th high-energy source with sky direction $\boldsymbol{\mu}_j$. From the physical point of view, the emission activity of the source is point-like. However, we cannot express it by a point mass in $\boldsymbol{\mu}_j$ as this neglects all instrumental aspects which need to be accounted for; see Appendix A for the details. We will model the direction (x_i, y_i) of photon i which is emitted by source j by a so-called *point*

spread function (PSF). This is a class of functions which describe the instrumental response of optical measurement devices. In case of the *Fermi* LAT, the PSF describes how the incoming photons spatially scatter around $\boldsymbol{\mu}_j$, the direction of the emitting source, as a function of their energy content E_i and of the geometry of the detector. In particular, we will adopt the point spread function developed by the *Fermi* LAT Collaboration ([Ackermann et al., 2013](#)).

Let $t_d(\cdot \mid \mathbf{b}, \mathbf{C})$ be the density of a bivariate Student t distribution with location parameter \mathbf{b} , scale matrix \mathbf{C} and d degrees of freedom. The density $f_j(\cdot \mid \cdot)$ in Formula (3.1) becomes

$$f_j(x_i, y_i \mid E_i, \boldsymbol{\mu}_j) = p_i t_{\nu_i, core}(x_i, y_i \mid \boldsymbol{\mu}_j, \tilde{\sigma}_{i, core}^2 I_2) + (1 - p_i) t_{\nu_i, tail}(x_i, y_i \mid \boldsymbol{\mu}_j, \tilde{\sigma}_{i, tail}^2 I_2), \quad (3.2)$$

where I_2 is the 2×2 identity matrix. That is, the *Fermi* LAT PSF is a mixture of two bivariate Student t densities $t_{\nu_i, core}(x_i, y_i \mid \boldsymbol{\mu}_j, \tilde{\sigma}_{i, core}^2 I_2)$ and $t_{\nu_i, tail}(x_i, y_i \mid \boldsymbol{\mu}_j, \tilde{\sigma}_{i, tail}^2 I_2)$, respectively. Both distributions are centered at $\boldsymbol{\mu}_j$, but differ in their scales $\tilde{\sigma}_i^2$ and their degrees of freedom ν_i to characterize the scattering for high and low quality detections. The scale parameters and degrees of freedom depend on the energy content E_i of the photon and on its event type. They are calculated as given below using the information provided by the *Fermi* LAT Collaboration for each single detection.

For each photon i define the scale factor

$$S_p(E_i) = \sqrt{\left\{ c_{0,i} \left(\frac{E_i}{100 \text{ MeV}} \right)^{-0.8} \right\}^2 + c_{1,i}^2},$$

where the constants $c_{0,i}$ and $c_{1,i}$ depend on the event type of the photon and are listed in Table 1. Each photon is furthermore characterized by two additional sets of constants, $(\sigma_{i, core}, \gamma_{i, core}, N_{i, core})$ and $(\sigma_{i, tail}, \gamma_{i, tail}, N_{i, tail})$, whose values depend again

Event Type	c_0 (radians)	c_1 (radians)
PSF0	1.53×10^{-1}	5.70×10^{-3}
PSF1	9.64×10^{-2}	1.78×10^{-3}
PSF2	7.02×10^{-2}	1.07×10^{-3}
PSF3	4.97×10^{-2}	6.13×10^{-4}

Table 1: Values for c_0 and c_1 for the four types of event classes detected by the LAT.

on the event type, but also on the energy content E_i of the photon and on how steeply the photon hits the surface of the LAT. These constants can be retrieved from the Science Tools repositories² of the *Fermi* LAT project. We define the parameters $(\tilde{\sigma}_{i,core}^2, \nu_{i,core})$ as follows

$$\tilde{\sigma}_{i,core}^2 = \sigma_{i,core}^2 S_p^2(E_i) \frac{\gamma_{i,core}}{\gamma_{i,core} - 1}, \quad \nu_{i,core} = 2(\gamma_{i,core} - 1).$$

The same expression holds for $(\tilde{\sigma}_{i,tail}^2, \nu_{i,tail})$ only that $(\sigma_{i,core}, \gamma_{i,core})$ is replaced by $(\sigma_{i,tail}, \gamma_{i,tail})$. Finally, we define the mixture weights³ as $p_i = \sigma_{i,core}^2 / (\sigma_{i,core}^2 + N_{i,tail} \sigma_{i,tail}^2)$.

To sum up, the single source model (3.2) depends on the unknown source direction $\boldsymbol{\mu}_j$ and on a number of event specific constants, provided by the *Fermi* LAT for each single photon, which are directly linked to the most complex detection mechanism of the space telescope.

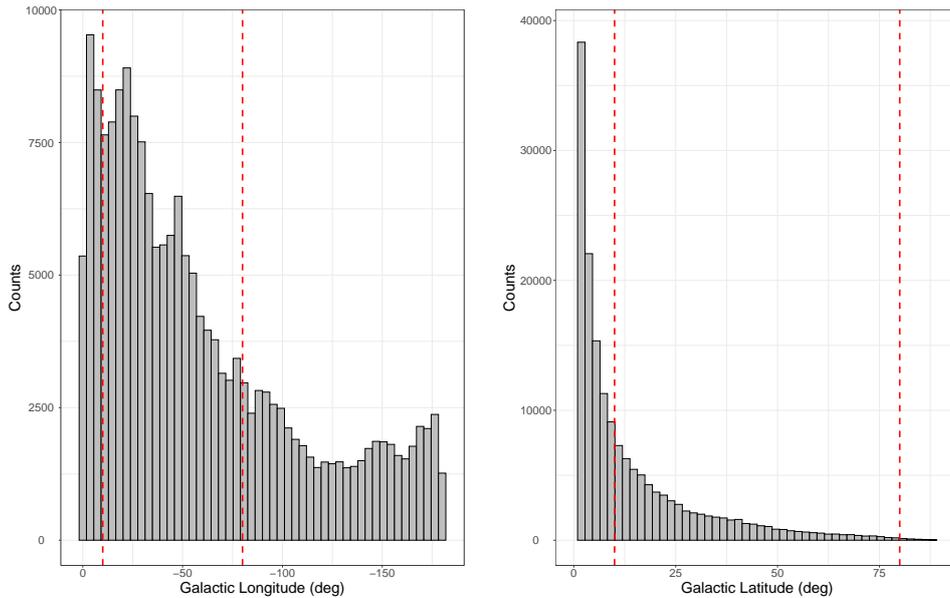


Figure 3: Histograms for the marginal distributions of the sky directions for the photons emitted by the γ -ray background as simulated by the *Fermi* LAT Science Tools for the second quadrant spanned by $[0^\circ, -180^\circ] \times [0^\circ, 90^\circ]$ and an observation period of 7.2 years. Left: galactic longitude. Right: galactic latitude. The vertical dashed lines delimit the sky region analyzed in Section 5.

3.2 The background model

The emission activity of the γ -ray background is characterized by the density function $g_b(\cdot | \cdot)$ which appears in the second part of Model (3.1). Figure 3 plots the marginal distributions of the galactic coordinates of the γ -ray photons emitted by the background as simulated using the *Fermi* LAT Science Tools; see Appendix A for the details. The figure considers the second quadrant spanned by $[0^\circ, -180^\circ] \times [0^\circ, 90^\circ]$ and an observation period of 7.2 years. The vertical dashed lines delimit the sky re-

²<https://fermi.gsfc.nasa.gov/ssc/data/analysis/software/>

³https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_PSF.html

gion analyzed in Section 5. The values of both the longitude and the latitude exhibit an exponential decay when getting away from the galaxy centre. Simply using two independent exponential distributions does not properly fit the background photon emission. We have to account for the dependence between the longitude and the latitude which is clearly visible in the contour plot of Figure 4.

In this article we propose to use the bivariate exponential distribution defined by Gumbel (1960) to model the complex structure of the background component in a simple though theoretically coherent way. We define $g_b(\cdot | \cdot)$ as

$$\begin{aligned}
 g_b(x_i, y_i | \sigma_x, \sigma_y, \delta) &= \sigma_x \sigma_y \exp\{-\sigma_x(-x_i) - \sigma_y y_i - \delta \sigma_x \sigma_y (-x_i) y_i\} \times \\
 &\quad \times [\{1 + \delta \sigma_x(-x_i)\} \{1 + \delta \sigma_y y_i\} - \delta],
 \end{aligned} \tag{3.3}$$

where $\sigma_x > 0$ and $\sigma_y > 0$ are two scale parameters, while $\delta \in (0, 1)$ expresses the correlation between the longitude and the latitude. It is straightforward to show that, for $\delta = 0$, the model reduces to the product of two independent exponential distributions with rates σ_x and σ_y . The motivations behind this model choice are given in Appendix A.

Differently to what holds for the PSF, where all scale and shape parameters are fixed and provided by the LAT, here the background parameters need to be estimated.

3.3 Prior belief

In Section 4 we will estimate Model (3.1) using Bayes' rule. The prior distributions for the model parameters $\boldsymbol{\theta}_K = \{\boldsymbol{\omega}, \{\boldsymbol{\mu}_j\}_{j=1}^K, \sigma_x, \sigma_y, \delta\}$ and K are set using the available information on the various physical processes and the current knowledge on the already catalogued sources.

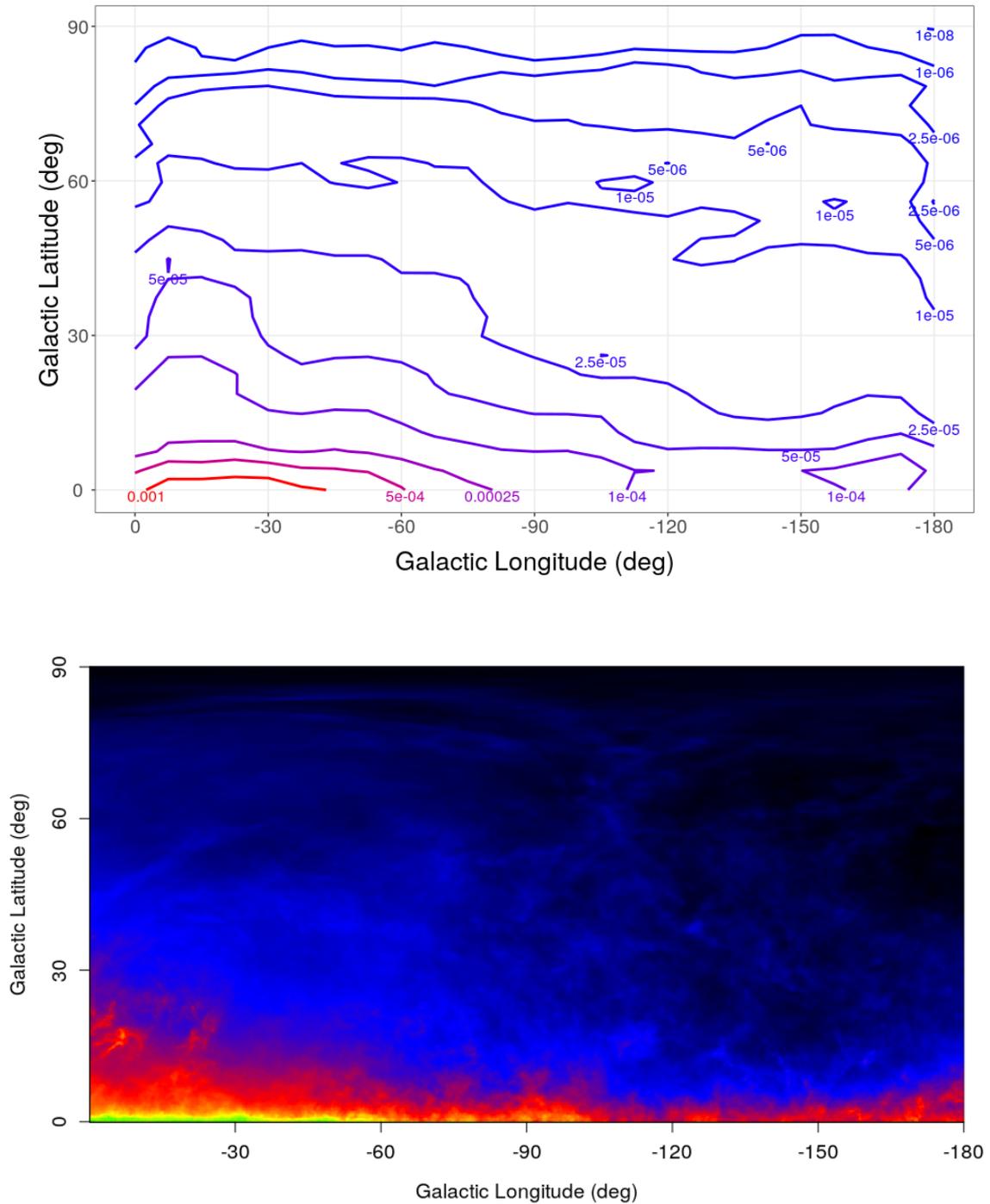


Figure 4: Top: Contour plot of the simulated background photon density for the second sky quadrant spanned by $[0^\circ, -180^\circ] \times [0^\circ, 90^\circ]$. Bottom: Image plot of the approximate count map reconstructed using the 30 GIE intensity maps provided by the LAT.

The first element in $(\boldsymbol{\theta}_K, K)$ are the mixing proportions $\boldsymbol{\omega}$. Conditional on the number K of sources present in the map, we choose a Dirichlet prior

$$\boldsymbol{\omega} \mid K \sim \text{Dir}(1, \dots, 1)$$

with concentration parameters equal to 1. This corresponds to assuming a flat distribution over the simplex of dimension $K + 1$. Though some prior belief on the intensity of the background radiation ω_0 is in principle available, we let the model free to specify the single contributions of the sources and of the background.

The only unknown parameters of the point spread functions are the galactic coordinates $\boldsymbol{\mu}_j$ of the putative sources, for $j = 1, \dots, K$. Since the universe contains a potentially infinite number of high-energy emitting phenomena at any possible direction, we use as prior for the $\boldsymbol{\mu}_j$'s a uniform distribution

$$\mu_{jx} \sim U(180, -180), \quad \mu_{jy} \sim U(-90, 90)$$

which extends over the whole map. Two exponential distributions with rate 1 are assigned as priors to the two scale parameters σ_x and σ_y of the background model

$$\sigma_x \sim \text{Exp}(1), \quad \sigma_y \sim \text{Exp}(1), \quad \delta \sim U(0, 1), \quad (3.4)$$

while a uniform distribution over $(0, 1)$ will be adopted for the correlation parameter δ . For the ease of computation, these three parameters were further transformed to the real domain as $\log \sigma_x$, $\log \sigma_y$ and $\text{logit}(\delta)$.

Finally, our model requires to set a prior distribution for the unknown number of sources K which we may detect in the map. This is the most tricky choice. We assume a truncated Poisson distribution

$$K \sim tP(\kappa \mid \kappa_{min}, \kappa_{max}),$$

where the hyper-parameters κ , κ_{min} and κ_{max} are set according to the specific sky area which is being analyzed using as scientific input the number of catalog sources in the region of interest (Ajello et al., 2017). As we will discuss later on, the directions and relative intensities of the detected sources remain stable if a sufficiently large number K of sources is considered.

4 Bayesian model fitting

We will follow the Bayesian paradigm and conduct inference on the model parameters $\boldsymbol{\theta}_K = \{\boldsymbol{\omega}, \{\boldsymbol{\mu}_j\}_{j=1}^K, \sigma_x, \sigma_y, \delta\}$ and K by exploring their posterior distribution. This distribution is derived in Section 4.1 up to a normalizing constant as the product of the likelihood function of Model (3.1) and of the prior densities and probabilities specified in Section 3.3. Section 4.2 outlines the reversible jump MCMC algorithm used to fit our model.

4.1 The posterior distribution

A common strategy to write down the likelihood function of a finite mixture model is to highlight the latent group variable which pinpoints the component of the mixture which generated a specific observation (Richardson and Green, 1997). With respect to Model (3.1), write $Z_i = 0$ if the generic photon i comes from the background with probability ω_0 , and $Z_i = j$ if the photon was emitted by the j th source with probability ω_j , for $j \in \{1, \dots, K\}$. By construction we have that $\sum_{i=1}^n \mathbb{1}(Z_i = j) = n_j$, where $\mathbb{1}(\cdot)$ is the indicator function and n_j the number of photons emitted by source j . It

follows that, conditional on the number K of sources,

$$(n_0, \dots, n_K) \mid K \sim \text{Mult}(n, \boldsymbol{\omega})$$

is a multinomial distribution indexed by the intensity vector $\boldsymbol{\omega}$.

Let now $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ be the vectors of observed longitudes and latitudes and $\mathbf{z} = (z_1, \dots, z_n)$ represent the unobserved latent class values. The full data likelihood function

$$\begin{aligned} L(\boldsymbol{\theta}_K, K \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) &= p(\mathbf{x}, \mathbf{y} \mid \mathbf{z}, \boldsymbol{\theta}_K, K) p(\mathbf{z} \mid \boldsymbol{\theta}_K, K) \\ &= \left[\prod_{i: z_i=0} g_b(x_i, y_i \mid \sigma_x, \sigma_y, \delta) \prod_{j=1}^K \left\{ \prod_{i: z_i=j} f_j(x_i, y_i \mid \boldsymbol{\mu}_j) \right\} \right] \prod_{j=0}^K \omega_j^{n_j} \end{aligned} \quad (4.1)$$

is given by the joint density function of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. By Bayes' theorem, the posterior distribution of the model parameters $(\boldsymbol{\theta}_K, K)$ is

$$\pi(\boldsymbol{\theta}_K, K \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) \propto L(\boldsymbol{\theta}_K, K \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) \pi(\boldsymbol{\theta}_K, K). \quad (4.2)$$

Here,

$$\pi(\boldsymbol{\theta}_K, K) = \pi(\boldsymbol{\theta}_K \mid K) \pi(K),$$

represents the joint prior distribution obtained by assuming

$$\pi(\boldsymbol{\theta}_K \mid K) = \pi(\boldsymbol{\omega} \mid K) \pi(\{\boldsymbol{\mu}_j\}_{j=1}^K \mid K) \pi(\sigma_x, \sigma_y, \delta), \quad (4.3)$$

that is, that the parameters $\boldsymbol{\omega}$, $\{\boldsymbol{\mu}_j\}_{j=1}^K$ and $\{\sigma_x, \sigma_y, \delta\}$ of $\boldsymbol{\theta}_K$ are conditionally independent given K . See Section 3.3 for the elicitation of the individual densities in (4.3) and for the prior $\pi(K)$.

4.2 The reversible jump MCMC algorithm

We implemented a two-step procedure for estimating the parameters $(\boldsymbol{\theta}_K, K)$ which combines Gibbs sampling with a Metropolis step. In turn, values are generated from

the full conditional distributions with density $\pi(\boldsymbol{\theta}_K \mid K, \mathbf{x}, \mathbf{y}, \mathbf{z})$ and probability $\pi(K \mid \boldsymbol{\theta}_K, \mathbf{x}, \mathbf{y}, \mathbf{z})$. At Step 1 the number of sources K is kept fixed, while at Step 2 the value of $\boldsymbol{\theta}_K$ is given. That is, we alternately estimate $\boldsymbol{\theta}_K$ for the given number of sources and then update the number of sources for a given set of parameters.

The following two paragraphs describe a single iteration of the algorithm; these need be repeated for a suitably large number T .

4.2.1 *Estimate model parameters for given K (Step 1)*

Assign the n observed γ -ray photons to the $K+1$ mixture components with allocation probabilities proportional to $\omega_j f_j(x_i, y_i \mid \boldsymbol{\mu}_j)$ for source $j = 1, \dots, K$, and $\omega_0 g_b(x_i, y_i \mid \sigma_x, \sigma_y, \delta)$ for the background, for $i = 1, \dots, n$. The model parameters $\boldsymbol{\theta}_K$ are updated following the order in which they are listed in the joint prior density (4.3).

- i) Update the vector of mixture weights $\boldsymbol{\omega}$ by drawing from the Dirichlet distribution $Dir(n_0 + 1, \dots, n_K + 1)$ whose hyper-parameters depend on the numbers (n_0, n_1, \dots, n_K) of photons assigned to the K sources and to the background. This step is straightforward given that the Dirichlet distribution is conjugate to the Multinomial.
- ii) Update the location parameters $\boldsymbol{\mu}_j$ by applying a double data augmentation strategy to the PSF of the single source model. This strategy allows us both, to avoid the mixture representation of Model (3.2) and to rewrite the Student t distribution as a product of a bivariate Gaussian and a gamma distribution (Gelman et al., 2013, Section 18.3).

Let $S_i \sim Be(p_i)$ be the random indicator variable which determines whether

photon i comes from the *core* density $t_{\nu_{i,core}}(x_i, y_i \mid \boldsymbol{\mu}_j, \tilde{\sigma}_{i,core}^2 I_2)$ or from the *tail* density $t_{\nu_{i,tail}}(x_i, y_i \mid \boldsymbol{\mu}_j, \tilde{\sigma}_{i,tail}^2 I_2)$ of the two-component PSF mixture. The corresponding contribution to the full data likelihood (4.1) of photon i given that $Z_i = j$ is

$$\left\{ \phi_2 \left(x_i, y_i \mid \boldsymbol{\mu}_j, \frac{\tilde{\sigma}_{i,core}^2}{w_i} I_2 \right) g \left(w_i \mid \frac{\nu_{i,core}}{2}, \frac{\nu_{i,core}}{2} \right) \right\}^{\mathbb{1}(S_i=1)} \times \\ \left\{ \phi_2 \left(x_i, y_i \mid \boldsymbol{\mu}_j, \frac{\tilde{\sigma}_{i,tail}^2}{w_i} I_2 \right) g \left(w_i \mid \frac{\nu_{i,tail}}{2}, \frac{\nu_{i,tail}}{2} \right) \right\}^{1-\mathbb{1}(S_i=1)},$$

where $\phi_2(\cdot \mid \cdot)$ and $g(\cdot \mid \cdot)$ denote, respectively, the densities of a bivariate Gaussian and of a gamma random variable, and I_2 is again the identity matrix of order 2. Thus,

$$x_i, y_i \mid S_i = 1, Z_i = j, w_i \sim N \left(\boldsymbol{\mu}_j, \frac{\tilde{\sigma}_{i,core}^2}{w_i} I_2 \right),$$

$$w_i \mid S_i = 1 \sim Ga(\nu_{i,core}/2, \nu_{i,core}/2)$$

and

$$x_i, y_i \mid S_i = 0, Z_i = j, w_i \sim N \left(\boldsymbol{\mu}_j, \frac{\tilde{\sigma}_{i,tail}^2}{w_i} I_2 \right),$$

$$w_i \mid S_i = 0 \sim Ga(\nu_{i,tail}/2, \nu_{i,tail}/2).$$

A direct Gibbs sampler for $\boldsymbol{\mu}_j$ is achieved in three steps:

(a) for every photon i assigned to the j th source, draw from

$$S_i \mid w_i, \boldsymbol{\mu}_j, x_i, y_i \sim Be(\tilde{p}_i), \quad \text{with}$$

$$\tilde{p}_i \propto \phi_2 \left(x_i, y_i \mid \boldsymbol{\mu}_j, \frac{\tilde{\sigma}_{i,core}^2}{w_i} I_2 \right) g \left(w_i \mid \frac{\nu_{i,core}}{2}, \frac{\nu_{i,core}}{2} \right) p_i;$$

(b) for every photon i assigned to the j th source draw from

$$w_i \mid S_i = 1, \boldsymbol{\mu}_j, x_i, y_i \sim Ga \left(\frac{\nu_{i,core}}{2} + 1, \frac{1}{2} \left\{ \nu_{i,core} + \frac{\|\mathbf{r}_i^\top - \boldsymbol{\mu}_j^\top\|^2}{\tilde{\sigma}_{i,core}^2} \right\} \right)$$

or

$$w_i \mid S_i = 0, \boldsymbol{\mu}_j, x_i, y_i \sim Ga \left(\frac{\nu_{i,tail}}{2} + 1, \frac{1}{2} \left\{ \nu_{i,tail} + \frac{\|\mathbf{r}_i^\top - \boldsymbol{\mu}_j^\top\|^2}{\tilde{\sigma}_{i,tail}^2} \right\} \right),$$

depending on the value of S_i at the previous step. Here, $\|\mathbf{r}_i^\top - \boldsymbol{\mu}_j^\top\|$ represents the squared Euclidean distance between the observed direction $\mathbf{r}_i = (x_i, y_i)$ and the direction $\boldsymbol{\mu}_j$ of source j ;

- (c) write $\mathbf{x}_j = (x_1, \dots, x_{n_j})$, $\mathbf{y}_j = (y_1, \dots, y_{n_j})$, $\mathbf{s}_j = (s_1, \dots, s_{n_j})$ and $\mathbf{w}_j = (w_1, \dots, w_{n_j})$. Draw from

$$\boldsymbol{\mu}_j^\top \mid \mathbf{s}_j, \mathbf{w}_j, \mathbf{x}_j, \mathbf{y}_j \sim \mathcal{N}_2(\boldsymbol{\psi}, \boldsymbol{\xi}),$$

where

$$\boldsymbol{\psi} = \left(\sum_{i=1}^{n_j} \frac{w_i}{\tilde{\sigma}_{i,s_i}^2} \right)^{-1} \left(\sum_{i=1}^{n_j} \frac{w_i}{\tilde{\sigma}_{i,s_i}^2} \mathbf{r}_i^\top \right), \quad \boldsymbol{\xi} = \left(\sum_{i=1}^{n_j} \frac{w_i}{\tilde{\sigma}_{i,s_i}^2} \right)^{-1}$$

with $\tilde{\sigma}_{i,S_i=1}^2 = \tilde{\sigma}_{i,core}^2$ and $\tilde{\sigma}_{i,S_i=0}^2 = \tilde{\sigma}_{i,tail}^2$, guarantees that the values are bounded into the map limits.

- iii) Jointly update the transformed background parameters $\{\log \sigma_x, \log \sigma_y, \text{logit}(\delta)\}$ with a Metropolis-Hastings step. The proposed values are sampled from a 3-dimensional Student t distribution with scale matrix $c_{prop} \cdot \Sigma_{prop}$ and degrees of freedom ν_{prop} . Set Σ_{prop} equal to minus the inverse observed Fisher information matrix obtained by fitting the background model (3.3) to the γ -ray photon counts simulated using the *Fermi* LAT Science Tools (see Appendix A.). Set the tuning parameters c_{prop} and ν_{prop} to 0.6 and 3 to guarantee an acceptance rate of about 60% and a moderate autocorrelation among the draws.

4.2.2 Update K for given model parameters (Step 2)

We update the size of the finite mixture model (3.1) by drawing from the posterior probability $\pi(K \mid \boldsymbol{\theta}_K, \mathbf{x}, \mathbf{y}, \mathbf{z})$ of K . This step changes the dimension of the parameter space and is carried out using the reversible jump MCMC algorithm described in Green (1995). According to Richardson and Green (1997) and Jones et al. (2015), we set four different types of possible moves called *split*, *birth*, *combine* and *death*. The first two increase the dimension of the parameter space by adding a new source to the model; the latter two reduce the size K of the finite mixture by one. The four moves are chosen at random with equal probability. Note that adding a new source to the model or removing a source doesn't change the background component $g_b(x_i, y_i \mid \sigma_x, \sigma_y, \delta)$.

In a *split* move, the algorithm attempts to augment the parameter space passing from a mixture of size K to size $K' = K + 1$. Given a source j , chosen at random, with location $\boldsymbol{\mu}_j$ and relative weight ω_j , a *split* implies the original parameters to be separated into two components $(\boldsymbol{\mu}_{j_1}, \boldsymbol{\mu}_{j_2})$ and $(\omega_{j_1}, \omega_{j_2})$. This is achieved by first generating three independent draws $\mathbf{u} = (u_1, u_2, u_3)$ from a Beta(2,2) distribution, as outlined on page 739 of Richardson and Green (1997, Section 3.2), and then transforming $(\boldsymbol{\mu}_j, \omega_j, \mathbf{u})$ as given there. The new entire set of parameters is called $\boldsymbol{\theta}_{K'}$. The *split* move is accepted with probability

$$\alpha_K^{split} = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}_{K'}, K' \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) q(K' \rightarrow K) \mathcal{J}(\boldsymbol{\theta}_K, \mathbf{u})}{\pi(\boldsymbol{\theta}_K, K \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) q(K \rightarrow K') b(\mathbf{u})} \right\}.$$

Here, $q(K \rightarrow K') = 1/(4K)$ is the probability of augmenting the parameter space with the selected *split* move, while $q(K' \rightarrow K) = 1/\{4(K + 1)\}$ is the probability of the reverse move. The expression $b(\mathbf{u})$ represents the product of three Beta(2,2)

densities evaluated at \mathbf{u} and $\mathcal{J}(\theta_K, \mathbf{u}) = \omega_1/\{u_1(1 - u_1)\}$ is the Jacobian of the transformation entailed by the *split* move.

A *combine* move reduces the size of the mixture from K to $K' = K - 1$. The corresponding acceptance probability is just the inverse of α_{K-1}^{split} when we augment the size of the mixture from $K - 1$ to K . That is, $\alpha_K^{comb} = (\alpha_{K-1}^{split})^{-1}$.

The *birth* and *death* moves are simpler. For a *birth* move we generate a random direction $\boldsymbol{\mu}^*$ from the prior density $\pi(\boldsymbol{\mu}_j)$ with corresponding intensity $\omega^* \sim \text{Beta}(1, K)$. Note that the existing intensities need be rescaled by $\omega_j = \omega_j(1 - \omega^*)$, for $j = 0, \dots, K$ so that all weights of the augmented model sum to 1. In a *death* move, an existing component is randomly chosen and deleted. Again, the remaining weights are rescaled so as to sum up to 1. Note that all existent/remaining directions are maintained in a *birth/death* move.

5 A real data example

We fitted the model presented in Section 3 to the high-energy photons recorded by the *Fermi* LAT in the sky region framed in white in Figure 1. As already mentioned, this region is of size $[-10^\circ, -80^\circ] \times [10^\circ, 80^\circ]$ and contains 26,208 events whose energy contents range from 10 GeV to 1,000 GeV. The 3FHL catalog lists 132 confirmed sources for this same region. Section 5.1 presents the parameter estimates, while in Section 5.2 we validate and benchmark our model by cross-checking the identified sources with the ones listed in the 3FHL catalog.

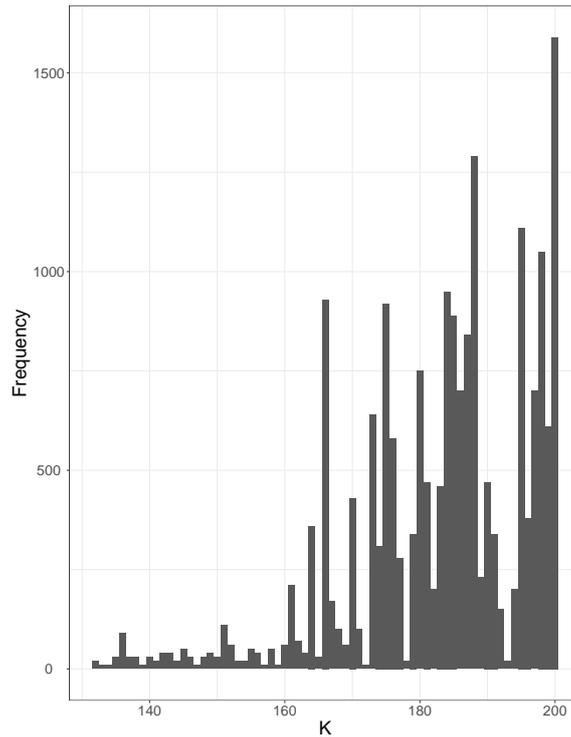


Figure 5: Posterior distribution of the number of mixture components K for the analyzed sky region. The modal value $K = 200$ appears 1,590 times. The minor mode $K = 188$ was visited 1,290 times.

5.1 Parameter estimates

Figure 5 plots the posterior distribution of the number K of identified sources obtained from 20,000 runs of the algorithm described in Section 4. All components of the chain stabilized very quickly. The posterior mode agrees with the upper bound $\kappa_{max} = 200$ of the support of K and was visited 1,590 times after a burn-in of suitable length was discarded. Here, we will illustrate the results obtained by focusing on the minor mode $K = 188$ which was visited 1,290 times. Indeed, because of as yet unclear reasons, our algorithm tends to explore the upper bound of the support of K a larger number of times as expected. A careful sensitivity analysis on the number of sources K (results

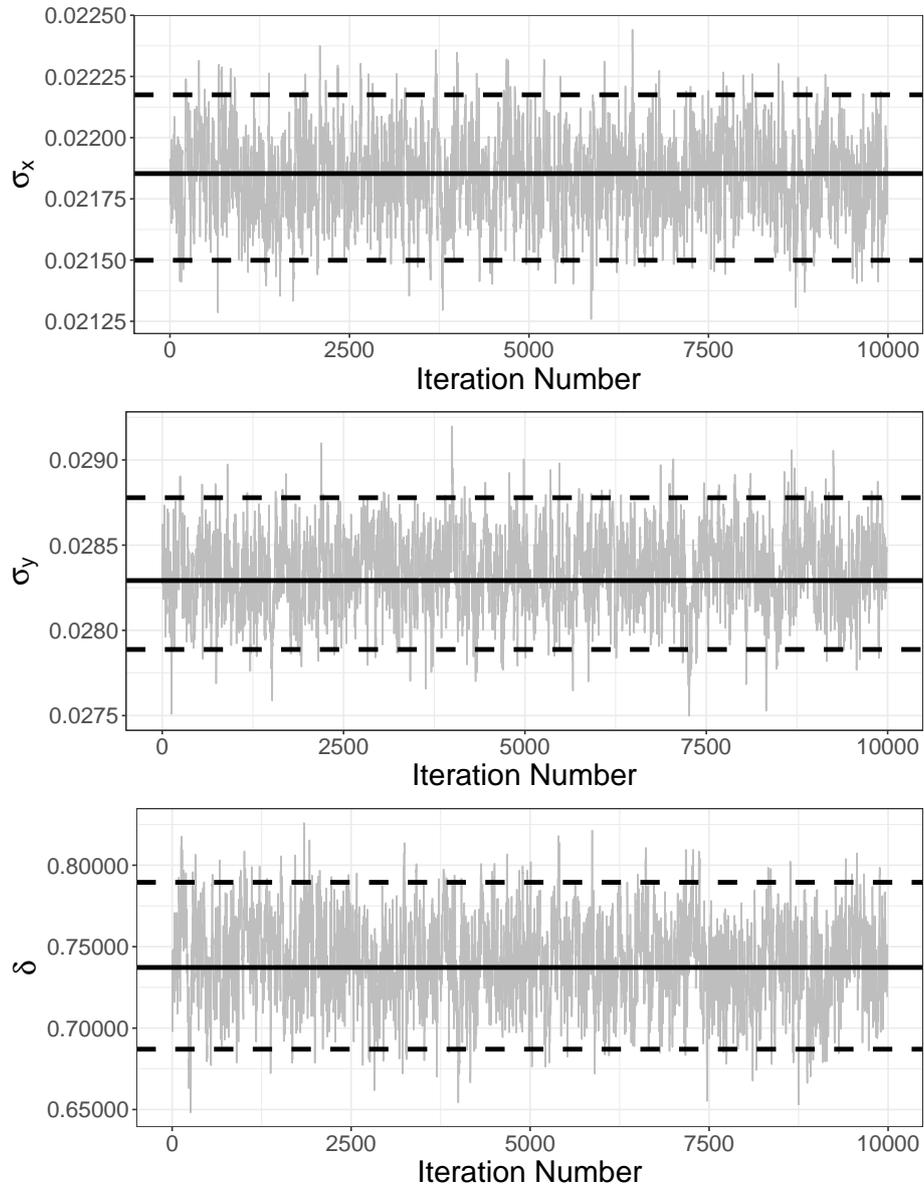


Figure 6: Trace plots of the 10,000 simulated values of the background model parameters obtained for the modal value $K = 188$: σ_x (top), σ_y (middle), δ (bottom). The solid horizontal line at the center represents the posterior mode; the two dashed lines delimit the 95% highest posterior density (HPD) interval.

not reported here) revealed, however, that the directions and relative intensities of the detected sources are largely unaffected by the final value chosen for K . That is, for $K = 200$ we re-detect the sources of $K = 188$, with almost no change in their parameter estimates, plus additional 12 ones.

We hence rerun the algorithm for 10,000 iterations with the number of components fixed at $K = 188$. Figure 6 plots the traces of the 10,000 simulated values for the background parameters σ_x , σ_y and δ obtained for the model value $K = 188$. The posterior modes and corresponding 95% highest posterior density (HPD) intervals are: $\tilde{\sigma}_x = 0.0218$ [0.0215 – 0.0222], $\tilde{\sigma}_y = 0.0282$ [0.0279 – 0.0288] and $\tilde{\delta} = 0.7372$ [0.6871 – 0.7895]. The marginal distribution of the longitude for γ -ray photons generated by the diffuse background is more dispersed than for the latitude as we may have expected from Figure 3. The Bayesian estimate of the correlation parameter δ is quite high. There is a rather strong linear dependence between the two exponential distributions which marginally model the background contamination along its longitude and its latitude.

Table 2 lists the first 10 sources detected by our model with the corresponding estimated longitudes (l_{alg}) and latitudes (b_{alg}) in Columns 2 and 3. Column 4 reports the estimated mixing proportions ω . All detected sources contribute a total of 6.27% of the detected γ -ray photons with an average intensity of 0.000293, while the intensity of the diffuse background is $\omega_0 = 0.9373$ (HPD interval [0.9340 – 0.9399]).

5.2 Model validation

Our model detected a total of 188 high-energy emitting sources. We were able to match 116 of these with one of the 132 confirmed sources listed in the 3FHL catalog

Detected source				Confirmed source			
ID_{alg}	l_{alg}	b_{alg}	ω	ID_{3FHL}	l_{3FHL}	b_{3FHL}	distance($^{\circ}$)
1	-79.213	17.062	1.27e-04	1	-79.214	17.071	0.010
2	-79.431	21.486	1.76e-04	2	-79.424	21.503	0.019
3	-79.339	28.077	1.65e-04	3	-79.333	28.090	0.014
4	-75.963	23.756	2.08e-04	4	-75.954	23.754	0.009
5	-74.738	21.444	1.17e-04	5	-74.759	21.463	0.028
6	-76.516	28.177	3.01e-04	6	-76.512	28.177	0.003
7	-70.647	13.130	4.38e-04	7	-70.624	13.160	0.037
8	-70.732	22.939	3.21e-04	8	-70.744	22.951	0.016
9	-78.489	46.552	2.43e-04	9	-78.463	46.580	0.034
10	-73.005	35.191	2.40e-04	10	-72.991	35.182	0.014

Table 2: Sample list of the sources detected by our model in the analyzed sky region. Columns 2 and 4 report the estimated longitudes (l_{alg}) and latitudes (b_{alg}); Column 4 records the estimated mixing proportions (ω). These sources were associated with confirmed sources present in the 3FHL catalog within an angular distance of 1° (Column 8). The longitudes (l_{3FHL}) and latitudes (b_{3FHL}) listed in the catalog are given in Columns 6 and 7. The complete list can be found in the Supplementary Material.

within and angular distance of 1° . An excerpt is given in Table 2. Columns 6 and 7 of the table report the longitude (l_{3FHL}) and latitude (b_{3FHL}) of the matched confirmed source. Column 8 reports the angular distance between the two sources, the detected and the confirmed one. The complete list of all possible matches is given in the Supplementary Material. This list consists of 127 rows, 11 of which are struck through. These are for the 9 detected sources with two possible matches, and the 2 catalog sources associated with two detected sources. For these multiple matches, the finally selected matched pair corresponds to the two sources whose angular distance is smaller. All observed angular distances are below 0.5° and most are within the typical values of $0.2 - 0.3^\circ$ used in Astrophysics.

In all, we were able to identify 116 confirmed sources out of 132 which corresponds to a success rate of 88%; 16 sources remained unidentified. There are 72 detected sources with no correspondence in the 3FHL catalog. We used the a posteriori available information on their intensities to further discriminate whether these correspond to real γ -ray emitting sources. A threshold value was defined as the median of the estimated mixing proportions ω_j for the 116 identified sources. Out of the 72 detected but unmatched sources, 20 sources exhibited posterior modes for ω_j which exceeded this threshold value. We qualified them as possible unidentified sources. They are currently under examination as foreseen by the *Fermi* LAT Collaboration. The 52 discarded sources may be false positives or they may correspond to as yet unidentified sources which are characterized by too faint signals to be detected with sufficiently high power.

6 Discussion

The model we developed in this paper represents a promising resource for identifying extra-galactic high-energy emitting sources using data from the *Fermi* LAT Collaboration. Indeed, we were able to identify 116 of the 132 γ -ray emitting sources confirmed for the analyzed sky region, which corresponds to a detection rate of almost 90%. Of the 72 sources detected by our algorithm, which we were not able to match with a confirmed source, 20 are currently under investigation. However, the same model can easily be adapted to other types of highly contaminated data in the 2-dimensional space where the set of locations at which a spatial process is observed are of interest.

Nonetheless, there are still a number of issues with the model and with the algorithm that need be settled.

A first question is why the posterior distribution of K tends to concentrate on the upper bound of its support κ_{max} . To explore the implications we carried out a sensitivity analysis by inspecting the posterior distributions of the remaining model parameters for a set of values of K between the minor mode $K = 188$ and $K = 200$. The reconstructed directions and relative intensities of the detected sources remained stable for all these scenarios. The high number of detected sources which found no confirmed counterpart may be a side effect of the parametric background model which is not flexible enough to capture local irregularities and thus generates a large amount of false positives. Some of them, however, may correspond to sources whose signal is too faint to be detected at a sufficiently high significance. Beware that the commonly used two-step procedure which led to the 3FHL catalog described in the introduction

classifies a new discovery using the 5 sigma standard, that is, for p -values smaller than 2.87×10^{-7} (Dorigo, 2016). The heuristic approach used to qualify the newly detected sources based on the estimated intensities needs to be replaced by a formal procedure which accounts also for the further available information on the photons such as the energy content. The ultimate solution would be to directly model the *Fermi* LAT intensity maps, especially if we wanted to work on a sky portion of the Galactic plane where it is more complicated to disentangle sources from the background.

A further issue is how to incorporate the possible fish eye effect into the current PSF. Indeed, as we used a mixture of two bivariate Student t distributions with equal spread on both directions (longitude and latitude), the scattering of the γ -ray photons around their emitting source is necessarily symmetric. Asymmetric model formulations are available, but the corresponding densities are no longer available in closed form. Future developments must also consider the fine tuning of our algorithm so as to be able to analyze larger portions of the sky.

This paper includes a substantial amount of improvements of the model and of the fitting routine with respect to a very first and highly incomplete attempt which was presented at a Italian national conference (Sottosanti et al., 2019). In particular, we replaced the simplistic approximation of the instrument response function based on King's PSF (King, 1962) with the more elaborate point spread function proposed in Ackermann et al. (2013), as discussed in Section 3.1. This allowed us to take into account both the energy content of the photon and its event type, which determine the quality of each event. We furthermore carefully studied the GIE component of the γ -ray emitting background. This allowed us to formulate the simple though theoretically coherent parametric model of Section 3.2 for the diffuse background component

which incorporates all information provided on it by the *Fermi* LAT project. On the computational side, we heavily improved the efficiency of the reversible jump MCMC algorithm of [Sottosanti et al. \(2019\)](#) both in terms of methodological innovations and coding; see Section 4. Last but not least, we carried out a sensitivity analysis aimed at assessing the extent to which the parameter estimates of the single source and background models may depend on the number K of detected sources.

7 Supplementary material

Supplementary materials for this paper are available from <http://www.statmod.org/smij/archive.html>.

Acknowledgements

We would like to thank the Associate Editor and an anonymous Referee for their most useful comments on a previous version of the paper. 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. We furthermore acknowledge the financial support by Prof. Junhui Fan (grants n. NSFC11733001 and n. NSFCU1531245).

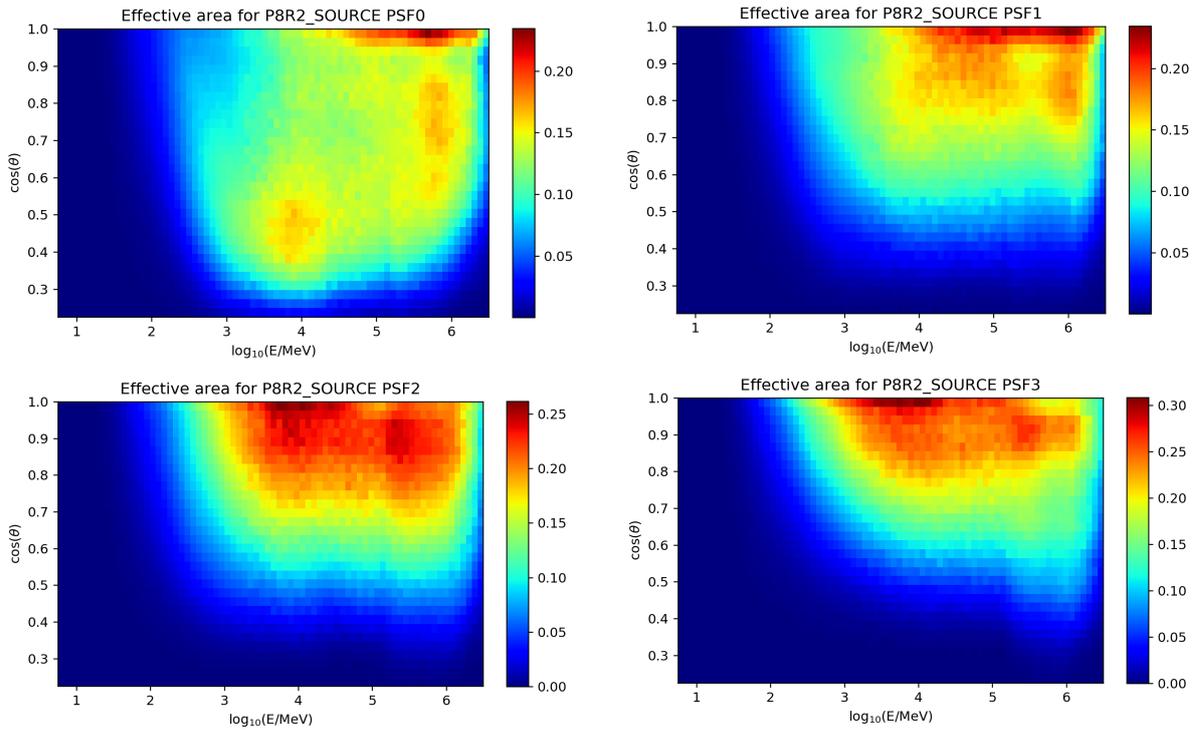


Figure 7: The effective area (in m^2) for the Pass8 release of the *Fermi* LAT software. The quality of the measurements, as expressed by the four event types PSF0, PSF1, PSF2 and PSF3, increases from the top left to the bottom right panel. The x-axis reports the energy content of the photon (in units of MeVs) on the \log_{10} scale; the y-axis reports the cosine of the incident angle (θ).

A Appendix

Despite its simple form, the background model (3.3) wraps up the insight gained from a detailed study of the available information on the diffuse γ -ray background. Indeed, the background emission itself has two origins: the isotropic γ -ray background which uniformly spreads over the entire map and the so-called Galactic Interstellar Emission (GIE) which dominates the galactic part of the sky map.

The *Fermi* LAT Collaboration provides data on the GIE component in the form of 30 so-called intensity maps measured in $cm^{-2}s^{-1}sr^{-1}MeV^{-1}$. These maps are based on an elaborate theoretical model which incorporates the current knowledge of the physical phenomena which determine galactic interstellar emission (Acero et al., 2016). Informally, we can interpret these intensities as expected counts which were normalized with respect to four features of the space telescope. These are the so-called *effective area* which broadly speaking corresponds to the size (in cm^2) of the surface of the LAT, the duration of observation (in seconds, s), the volume of the sky (in steradians, sr) covered by the space telescope and the energy contents of the γ -ray photon (in MeV). The 30 intensity maps correspond to 30 energy bins which range from 58 to 513,056 MeV with a step size of 0.136 measured on the \log_{10} scale. Each map divides the whole sky into a spatial grid of $2,880 \times 1,441$ pixels at which the intensities are provided (Gorski et al., 2005).

In Section 3.2 we simulated a count map of the GIE using the intensities of these 30 GIE maps thanks to the support provided by the *Fermi* LAT Science Tools⁴. Here, we will present the different steps of an empirical analysis which allowed us to

⁴<https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/>

drop the dependence on the features of the space telescope when modeling the diffuse background.

We first converted the 30 intensity maps into a unique whole sky count map, that is, into the same form our primary dataset is delivered. We can easily account for the energy layer by multiplying each intensity by the central value of the energy bin it refers to. These intermediate results are then multiplied by $4\pi/(2.800 \times 1.441)$, which is the solid angle covered by the *Fermi* LAT at every pixel, and by the total observation time of 7.2 years (expressed in seconds). An issue remains with the effective area component. This area depends on the geometrical conformation of the space telescope as well as on the capacity of the detector to convert and correctly identify incoming γ -rays. The behaviour of the effective area is described on the *Fermi* LAT performance page⁵ as a function of the energy content of the photon and of its direction expressed in polar coordinates. The four panels of Figure 7 plot the effective area (in m^2) which characterizes the latest release of the *Fermi* LAT software (Pass8) for the four event types PSF0, PSF1, PSF2, PSF3. Recall that these correspond to increasing quality of the measurements. Independently of the event type, the higher the energy content of the photon is and the steeper it hits the surface of the LAT, the larger is the effective area. It seems, however, that with the exception of the less accurate event type PSF0 and for incident angles which are steep enough we may approximate the effective area by a constant value for energy levels larger than 10 GeV (that is, 4 on the \log_{10} scale), which corresponds to the energy range we are interested in. This motivated us to discard in our analysis all events of type PSF0.

⁵http://www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm

A rough approximation of the expected number of counts is hence provided by summing up the rescaled intensities over the 30 maps on a pixel by pixel basis. The bottom panel of Figure 4 reports the approximate count map we reconstructed according to the steps outlined in the previous paragraphs. This map is in close agreement with the simulated count map of Section 3.2 and shown in the top panel of Figure 4. On this basis we decided to neglect the influence of the effective area of the space telescope in the formulation of our background model (3.3).

As far as the prior elicitation on the parameters of Model (3.3) goes, an alternative could have been to fit the bivariate exponential model (3.3) by maximum likelihood to the reconstructed count maps and to center the prior distributions (3.4) at the corresponding maximum likelihood estimates. We deliberately didn't choose this solution because of two reasons. First, Model (3.3) embraces both diffuse background components, that is, the GIE and the isotropic one, while the reconstructed count maps are for the GIE. Second, the reconstructed count map depends heavily on how we rescaled it with respect to the effective area. In any case, the fraction of γ -ray photons which are emitted by the background is so high that their likelihood contribution overweights any prior elicitation.

References

Acero, F., Ackermann, M., Ajello, M., Albert, A., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bellazzini, R., Bissaldi, E., et al. (2016). Development of the model of galactic interstellar emission for standard point-source analysis of Fermi large area telescope data. *The Astrophysical Journal Supplement Series*, **223**(2), 26.

Ackermann, M., Ajello, M., Allafort, A., Asano, K., Atwood, W., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bechtol, K., et al. (2013). Determination of the point-spread function for the Fermi large area telescope from on-orbit data and limits on pair halos of active galactic nuclei. *The Astrophysical Journal*, **765**(1), 54.

Ajello, M., Atwood, W., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bellazzini, R., Bissaldi, E., Blandford, R., Bloom, E., et al. (2017). 3FHL: the third catalog of hard Fermi-LAT sources. *The Astrophysical Journal Supplement Series*, **232**(2), 18.

Atwood, W., Abdo, A. A., Ackermann, M., Althouse, W., Anderson, B., Axelsson, M., Baldini, L., Ballet, J., Band, D., Barbiellini, G., et al. (2009). The large area telescope on the Fermi gamma-ray space telescope mission. *The Astrophysical Journal*, **697**(2), 1071.

Ciprini, S., Tosti, G., Maruccci, F., et al. (2007). The First GLAST Symposium. In Ritz, S., Michelson, P., and Meegan, C. A., editors, *American Institute of Physics Conference Series*, volume 921, pages 546–547.

Damiani, F., Maggio, A., Micela, G., and Sciortino, S. (1997). A method based on wavelet transforms for source detection in photon-counting detector images. I. Theory and general properties. *The Astrophysical Journal*, **483**(1), 350.

Dorigo, T. (2016). *Anomaly! Collider Physics and the Quest for New Phenomena at Fermilab*. World Scientific.

Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling*. Springer-Verlag.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. ISBN 9781439840955.

Gorski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, **622**(2), 759.

Green, P. J. (1995). Reversible jump markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, **55**(292), 698–707.

Jones, D. E., Kashyap, V. L., and Van Dyk, D. A. (2015). Disentangling overlapping astronomical sources using spatial and spectral information. *The Astrophysical Journal*, **808**(2), 137.

King, I. (1962). The structure of star clusters. I. An empirical density law. *The Astronomical Journal*, **67**, 471.

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

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with Discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59**(4), 731–792.

- Sottosanti, A., Costantin, D., Bastieri, D., and Brazzale, A. R. (2019). Discovering and Locating High-Energy Extra-Galactic Sources by Bayesian Mixture Modelling. In Petrucci, A., Racioppi, F., and Verde, R., editors, *New Statistical Developments in Data Science*, pages 135–148. Springer International Publishing.
- Starck, J.-L. and Pierre, M. (1998). Structure detection in low intensity x-ray images. *Astronomy and Astrophysics Supplement Series*, **128**(2), 397–407.