A new method to perform data-model comparison in Fermi-LAT analysis

P. Bruel

A&A 656, A81 (2021)
https://doi.org/10.1051/0004-6361/202141553

Astronomy & Astrophysics

A new method to perform data-model comparison in *Fermi*-LAT analysis

P. Bruel

Laboratoire Leprince-Ringuet, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France
e-mail: Philippe.Bruel@llr.in2p3.fr

Received 15 June 2021 / Accepted 15 September 2021

ABSTRACT

Context. The analysis of *Fermi*-Large Area Telescope (LAT) gamma-ray data in a given Region of Interest (RoI) usually consists of performing a binned log-likelihood fit in order to determine the sky model that after convolution with the instrument response best accounts for the distribution of observed counts.
Aims. While tools are available to perform such a fit, it is not easy to check the goodness-of-fit. The difficulty of the assessment of the data-model agreement is twofold. First of all, the observed and predicted counts are binned in three dimensions (two spatial dimensions and one energy dimension) and comparing two 3D maps is not straightforward. Secondly, gamma-ray source spectra generally decrease with energy as the inverse of the energy square. As a consequence, the number of counts above several GeV generally falls into the Poisson regime, which precludes performing a simple χ^2 test.
Methods. We propose a method that overcomes these two obstacles by producing and comparing, at each pixel of the analyzed RoI, spatially integrated count spectra for data and model. The comparison follows a log-likelihood approach that extends the χ^2 test to histograms with low statistics. This method can take into account likelihood weights that are used to account for systematic uncertainties.
Results. We optimize the new method so that it provides a fast and reliable tool to assess the goodness-of-fit of *Fermi*-LAT data and we use it to check the latest gamma-ray source catalog on 10 years of data.

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

1. Introduction

Since its launch in June 2008, the *Fermi* Large Area Telescope (LAT, Atwood et al. 2009) has been continuously observing the gamma-ray sky in the energy range between 30 MeV and 2 TeV. The *Fermi*-LAT data are public[1] and have been widely used by the gamma-ray community. The *Fermi*-LAT latest general catalog, 4FGL-DR2 (Abdollahi et al. 2020; Ballet et al. 2020), reports almost 5800 sources, among which are hundreds of pulsars, tens of supernovae remnants and pulsar wind nebulae, and thousands of blazars.

The usual way to analyze LAT data in a given Region of Interest (RoI) is to bin the data in a 3D map (two spatial dimensions and one energy dimension[2]) and then search for the sky model that best predicts the number of gamma rays observed in this RoI. The sky model is a list of gamma-ray sources, whose position, spatial nature (point-like or extended), as well as their energy spectrum are provided. The *Fermi*-LAT analysis package, `Fermitools`[3], is used to convolve the emission of each source with the instrument response functions (IRFs, Ackermann et al. 2012) in order to predict the number of observed counts. As a consequence, one of the main steps of the analysis is generally a log-likelihood fit that finds the source spectral parameters that give the best agreement between the observed and predicted 3D count maps.

After the fit has converged, it is important to check that the obtained agreement is satisfactory but the user faces the difficulty of comparing two 3D maps. The simplest way would be to define an energy band and produce the corresponding 2D residual (data–model) count map. However, the energy band definition and the analysis of the residual map is not straightforward because of the energy dependence of the point spread function (PSF) of the instrument: its 68% containment angle varies from 5° at 100 MeV to 0.1° at 30 GeV (Atwood et al. 2009, 2013)[4]. As a consequence, the binning of the residual map should depend on the energy range definition. Furthermore, since the spectral characteristics of a potential data-model mismatch are unknown, it is not obvious to choose an energy range a priori, often leading to look at several energy bands.

The `Fermitools` provide a method to quantify the data-model agreement which is frequently used. It consists of computing a test statistic (TS) map: at each pixel, the presence of an additional source is tested by computing twice the difference in log-likelihood obtained with and without the source. A TS = 25 corresponds to $\sim 4\sigma$ significance (Mattox et al. 1996). The drawbacks of this method are twofold: it is computationally intensive and, above all, it is not sensitive to negative deviations (data<model) because the flux of the additional source is bound to be positive. Another public analysis package, *Fermipy*

[1] https://fermi.gsfc.nasa.gov/ssc
[2] A typical example is a $12° \times 12°$ map with a pixel size of 0.1° and 10 logarithmically spaced bins per decade in energy.
[3] https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation
[4] https://www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm

(Wood et al. 2017), offers a simplified version of the TS maps that is much faster but is still blind to negative deviations.

In this paper we propose a method to perform data-model comparison that overcomes the aforementioned problems. As in the case of residual or TS maps, we want to provide spatial information. So the goal is to quantify the level of deviation between data and model at each pixel of the RoI. Since we go from 3D count maps to a 2D deviation map, it implies that the spectral information must be fully utilised in the process of the deviation assessment. So, for each pixel of the RoI, the method consists of building and comparing data and model count spectra. We present in Sect. 2 a simple way to build spatially-integrated count spectra that takes into account the PSF of the instrument. Since it is not always possible to perform a simple χ^2 test to compare these count spectra, we apply a log-likelihood approach that is presented in Sect. 3 and its extension to take into account systematic uncertainties is presented in Sect. 4. We optimize and verify this new method in Sect. 5 and the results obtained when using it to check the 4FGL-DR2 catalog are given in Sect. 6.

2. PSF-like integrated count spectra

Since we want to be as sensitive as possible to discrepancies between data and model, it is useful to list the possible causes for such discrepancies in order to find the optimal way to create the count spectra that we compare in Sect. 3. The two possibilities are that one source is missing in the model or that it is mismodeled. In both cases, the resulting discrepancy depends on the point-like or extended nature of the source. In the point-like case, the region of the discrepancy corresponds to the PSF and, as a consequence, its size follows the energy dependence of the PSF.

The LAT PSF 68% containment angle can be parameterized as $p_0(E/100\,\mathrm{MeV})^{-p_1} \oplus p_2$, the addition being in quadrature (Ackermann et al. 2013). For LAT PASS 8 SOURCE class events, p_0 and p_2 are about 5 and 0.1° and $p_1 = 0.8$.

If the source is slightly extended (e.g., 0.5°), the size of the discrepancy region can be modeled by the same parameterization, replacing p_2 by the sum of 0.1° and the extension of the source. For diffuse sources such as the Galactic diffuse emission, one can use an even larger p_2.

For these reasons, we propose to use an energy-dependent distance of the form $p_0(E/100\,\mathrm{MeV})^{-p_1} \oplus p_2$ to build for each pixel of the RoI the spatially integrated count spectra for both data and model. We start with the SOURCE class PSF 68% containment angle but the definition of the integration region is revisited and optimized in Sect. 5. Examples of PSF 68% integrated count spectra are shown in Fig. 1. For simplicity's sake, the spatial integration is performed directly on the data and model 3D count maps used in the fit (by summing over the pixels within the energy-dependent distance). This choice allows us to compare data and model using the closest information to the one used in the fit.

Regarding the energy binning of the count spectra, it is useful to consider the expected spectral features of potential data-model deviations. The known gamma-ray sources do not exhibit any narrow peak or dip in their spectra in the *Fermi*-LAT energy range and we thus do not expect particularly sharp data-model deviations from either a spectrally mismodeled source or a missing source. Any deviation would be smeared by the energy resolution of the LAT, which is 10% at 1 GeV, corresponding to a 68% containment interval of about 0.1 in $\log_{10} E$. As a result, it is not useful to use a bin smaller than 0.1. On the contrary, a larger binning may increase the deviation sensitivity by increasing the statistics in each bin. The choice of the optimal bin size is investigated in Sect. 5.

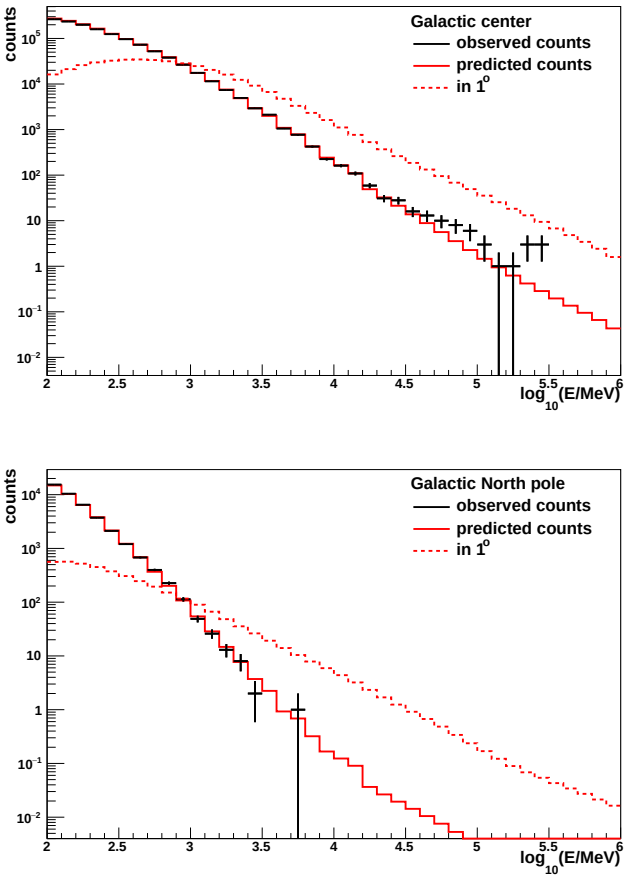

Fig. 1. PSF 68% integrated count spectra at the center pixel of the RoIs centered on the Galactic center (*top*) and the Galactic North pole (*bottom*) for data (black) and model (red). The dashed red line shows for comparison the model count spectra integrated over 1°. The average likelihood weights, introduced in Sect. 4, are also shown.

3. Estimation of the deviation probability

In order to assess the level of deviation between the data and model integrated count spectra, we want to compute the p-value, that is to say the probability that the statistical fluctuations can reach a level of deviation as large as the one observed in the data, under the assumption that the model represents the data.

If all the bins of the count spectra were in the Gaussian regime, we could simply perform a χ^2-test to estimate the p-value. However, because of the general $E^{-\gamma}$ power-law spectrum of gamma-ray sources (with photon indices γ between 1 and 3 for most of the sources as well as for the background) and the energy dependence of the PSF 68% selection, the integrated count spectra fall steeply with energy and the numbers of counts at high energy are generally in the Poisson regime, which precludes performing a χ^2-test.

In order to overcome this limitation, we adopt a log-likelihood approach and define the random variable

$$L = -\sum_k \log \mathcal{P}(x_k, m_k) \tag{1}$$

where $\mathcal{P}$ is the Poisson probability and x_k are independent random Poisson variables of mean m_k, the spatially integrated number of model counts in the spectral bin k. The p-value is the

integral of the probability distribution function (pdf) of L above $L_{\rm data} = -\sum_k \log \mathcal{P}(n_k, m_k)$, the value obtained with the data integrated count spectra n_k. In other words, the p-value is the L complementary cumulative distribution function (CCDF) at $L_{\rm data}$.

It is common to compute expected likelihood distributions by performing simulations but it can be very computationally intensive, especially when requiring the resulting statistical fluctuations to be below the percent level. Here we choose to compute the L pdf in an iterative way, as fully described in Appendix A. When the number of counts in part of the spectrum is large enough, it is possible to optimize the computation by taking advantage of the χ^2 approximation of the log-likelihood. For the spectral bins with Gaussian statistics, the Poisson probability of parameter m_k can be replaced with a Gaussian with mean and variance equal to m_k. Ignoring the constant term, we have

$$L_{\rm Gaus} = \frac{1}{2}\sum_k \frac{(x_k - m_k)^2}{m_k}. \tag{2}$$

As a consequence, the $L_{\rm Gaus}$ pdf can be simply derived from the χ^2 distribution with a number of degrees of freedom equal to the number of spectral bins with Gaussian statistics.

In order to compute the L pdf, we thus start by sorting the counts in decreasing order, then compute the $L_{\rm Gaus}$ pdf corresponding to all the bins with a number of counts greater than 100 and finally perform the iterative computation over the remaining bins. We note that this p-value computation provides a simple extension of the χ^2-test to histograms with counts in the Poisson regime.

We define PS[5], the data-model deviation estimator, as

$$|\mathrm{PS}| = -\log_{10}(p\text{-value}) \tag{3}$$

and give it the sign of the sum of the residuals in sigma units:

$$\mathrm{sign(PS)} = \mathrm{sign}\left(\sum_k (n_k - m_k)/\max\left(1, \sqrt{m_k}\right)\right) \tag{4}$$

which allows us to estimate whether the deviation is positive (data>model) or negative (data<model). To our knowledge, this is the first time that such an estimator is proposed.

The advantage of using the logarithm of the p-value rather than converting it into σ units is that, when considering the maximum value of a PS map, the correction for the number of trials is simply done by subtracting the logarithm of the number of pixels in the map. The $3, 4$ and 5σ thresholds correspond to PS = $2.57, 4.20$ and 6.24, respectively. For a typical 100×100 pixels map, assuming that the PS are independent, the $3, 4$ and 5σ thresholds correspond to uncorrected PS = $6.57, 8.20$ and 10.24, respectively.

4. Systematic uncertainty handling with weighted log-likelihood

Log-likelihood weights have been introduced in the *Fermi*-LAT general catalog analysis in order to account for systematic uncertainties, especially those coming from the modeling of the diffuse emission (Abdollahi et al. 2020). This is done by performing the spectral fit with the following definition of the log-likelihood (Hu & Zidek 2002):

$$\log \mathcal{L} = \sum_{i,j,k} W_{ijk}(N_{ijk} \log M_{ijk} - M_{ijk}) \tag{5}$$

where the indices i, j, k run over the 3D maps and W, N and M are the map pixel weight, number of observed and predicted counts, respectively. The *Fermi*-LAT catalog analysis uses data-based weights corresponding to a level of systematic uncertainty of 3% (see Appendix B of Abdollahi et al. 2020, for more details).

We note that in the Gaussian regime, the Poisson contributions in Eq. (5) can be replaced with Gaussian probabilities:

$$\frac{1}{2}W_{ijk}(N_{ijk} - M_{ijk})^2/M_{ijk} = \frac{1}{2}(N_{ijk} - M_{ijk})^2/(M_{ijk}/W_{ijk}) \tag{6}$$

which highlights that the effect of the weights is to increase the variance by $1/W_{ijk}$.

If log-likelihood weights are used in the spectral fit, they also have to be taken into account when assessing the goodness-of-fit. The first thing to do is thus to also introduce weights in the definition of the random variable L used to compute the p-value and Eq. (1) becomes

$$L = -\sum_k w_k \log \mathcal{P}(x_k, m_k). \tag{7}$$

Since m_k is the sum of predicted counts in the integration region, its variance is the sum of the pixel variances. Following Eq. (6), we interpret m_k/w_k as the expected variance of m_k and thus define w_k such that $m_k/w_k = \sum_{ij} M_{ijk}/W_{ijk}$, where the sum runs over the integration region corresponding to the spectral bin k. Since we adopt the data-based weights as in the 4FGL-DR2 analysis, the average is actually weighted with the data counts rather than the model counts: $n_k/w_k = \sum_{ij} N_{ijk}/W_{ijk}$. When $n_k = 0$, w_k is set to 1. Examples of such average weights are shown in Fig. 1.

Although Eq. (7) accounts for the relative importance of the systematic uncertainties between spectral bins, it does not take into account their absolute meaning. This is simply because the variances of the individual Poisson distributions are unchanged. A clear symptom of this problem is that the p-value is invariant with a global rescaling of the weights.

This serious limitation can be naturally overcome for the bins of the count spectrum that follow Gaussian statistics, for which the introduction of the weights modifies Eq. (2) as

$$L_{\rm Gaus} = \frac{1}{2}\sum_k \frac{(x_k - m_k)^2}{m_k/w_k}. \tag{8}$$

As a consequence, using the χ^2 approximation $L_{\rm Gaus}$ to compute the L pdf in the weighted log-likelihood case allows us to ensure that the absolute meaning of the systematic uncertainty is properly taken into account, as long as the uncertainties of the spectral bins are taken as $\sqrt{m_k/w_k}$ instead of $\sqrt{m_k}$.

In order to study the case of the spectral bins with Poisson statistics, we performed simulations for different levels of systematic uncertainties, whose results are described in Appendix C. The conclusion of this study is that introducing the weights in the definition of L is helpful for large systematic uncertainty and unimportant for small systematic uncertainty, where the needed correction is actually small. In this study considering only the bins with Poisson statistics, the resulting error on PS is within 3% for systematic uncertainties on the order of 3%, as currently used in LAT analyses. When computing the PS with all the bins of the count spectra, the true error on PS is actually smaller thanks to the bins in the Gaussian regime, which make up about 50% of the spectral bins in the analysis of 10 years of LAT data, as reported in the next section.

Because of the overall positive role of the weights to take systematic uncertainty into account, we decided to keep the

[5] The name PS was chosen because P can stand for both p-value and PSF and also because the output map name, PS map, sounds close to TS map.

weighted version of L to compute the PS. As a consequence, when computing the L pdf, we first compute the $L_{\rm Gaus}$ pdf corresponding to all the spectral bins with a number of counts greater than 100, using Eq. (8), and then perform the iterative computation over the remaining bins using Eq. (7). Compared to Eq. (4), the PS sign definition is modified as

$$\mathrm{sign(PS)} = \mathrm{sign}\left(\sum_k (n_k - m_k)/\max\left(1, \sqrt{m_k/w_k}\right)\right). \quad (9)$$

5. PS optimization and calibration

We use the PS method to assess how well the *Fermi*-LAT general catalog describes the whole sky in terms of predicted counts. This verification, named catXcheck, performed on the latest published catalog, that is 4FGL-DR2, gives us the opportunity to measure and optimize the PS sensitivity.

5.1. catXcheck framework

It consists of the analysis of 438 RoIs (120×120 pixels, with a pixel size of 0.1°) covering the whole sky. The RoI centers lie on Galactic parallels whose latitudes go from −90 to 90 with a 10° step. The longitude step is 10° for $|b| \leq 30$ and it is 12, 15, 20, 30 and 45° for $|b| = 40$, 50, 60, 70 and 80, respectively.

The analysis of each RoI is performed with the `Fermitools`. We use the same data as in the 4FGL-DR2 catalog analysis, namely Pass 8 SOURCE class data (Atwood et al. 2013; Bruel et al. 2018) collected during the first 10 years of the mission. We select data above 100 MeV whose zenith angle is less than 90° to avoid Earth's limb contamination. Using the PSF event-type partition improves on average the point source sensitivity but it is not particularly useful when assessing data-model agreement. We thus combine all events in the analysis. We use 10 bins per decade between 100 MeV and 1 TeV. On top of the Galactic diffuse emission and the isotropic template[6], as well as the Sun and Moon steady emission templates, the sky model comprises all point-like and extended sources from 4FGL-DR2, within $5 + 0.015(\sigma_{\rm src} - 4)$ degrees of the RoI border, where $\sigma_{\rm src}$ is the significance of the source as reported in the catalog.

The only free parameters of the spectral fit are the ones of the Galactic diffuse emission (power-law correction) and isotropic (normalization) templates. We use the `P8R3_SOURCE_V2` IRFs and energy dispersion is taken into account for the Galactic diffuse emission and the 4FGL sources.

5.2. PS optimization

The PS estimator is designed to detect data-model deviations. A way to quantify and optimize its sensitivity is to create artificial deviations. This is done within the catXcheck framework: for each 4FGL source inside an RoI and more than 1° away from its border, we removed the source from the model, recomputed the total predicted 3D count map and then computed the PS and TS around the position of the source. In order to find the spatial selection parameters as well as the energy binning of the count spectra that maximize the PS, we compared the PS to the TS over the whole set of 4FGL-DR2 sources. We stress that comparing the PS and TS does not imply that we suggest that the PS could replace the TS as a way to quantify the significance of a known gamma-ray source. The only goal of the PS estimator is to search for data-model deviations. As such, it can detect a potential missing source in the sky model of an RoI but it is not the optimal way to characterize a known gamma-ray source.

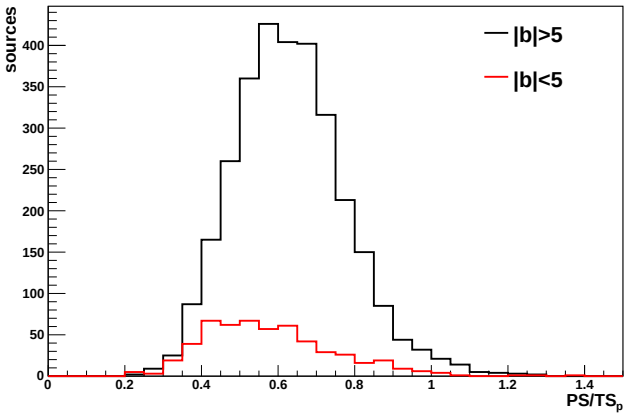

Fig. 2. Distribution of the ratio of PS over $\mathrm{TS_p}$ (TS expressed as $-\log_{10}(p\text{-value})$) for the 4FGL-DR2 sources closer (red) and further than 5° (black) from the Galactic plane.

For the integrated count spectra used to compute the PS, we first used a bin width of 0.1 in $\log_{10} E$. For the TS computation, the putative source was modeled with a power law with a photon index fixed to 2.3, which is the average photon index of the 4FGL-DR2 sources. So we assume that the TS follows a χ^2 distribution with one degree of freedom. The PS sensitivity can be compared to that of the TS by simply looking at the ratio of the maximum PS to $\mathrm{TS_p}$, the maximum TS expressed as $-\log_{10}(p\text{-value})$. This ratio (which, for simplicity's sake, we refer to as PS/TS) is shown in Fig. 2 for sources with $\mathrm{TS_p} > 7$ in ($|b| < 5°$) and outside ($|b| > 5°$) the Galactic plane. These distributions are relatively wide with a peak at around 0.6.

The PS is computed on the integrated count spectra and the energy-dependent spatial selection is defined with $p_0(E/100\,\mathrm{MeV})^{-p_1} \oplus p_2$. In order to find which parameters maximize the PS sensitivity, we computed the PS/TS ratio on a p_0, p_1 grid for $p_2 = 0.1, 0.15$ and 0.20°. Above 30 GeV, for a pixel size of 0.1°, these three values of p_2 correspond to an integration region of 5, 9 and 13 pixels, respectively.

For each configuration, we fitted the PS/TS distribution with a log normal in order to estimate the peak position. Figure 3 shows how this peak position varies with the parameters. There is no significant difference between the results obtained with $p_2 = 0.1$ and 0.15 but the results are on average worse with $p_2 = 0.2$. Over most of the p_0, p_1 grid, the variation is modest compared to the typical 0.15 width of the distributions. A precise choice of the parameters is thus not critical and we choose $p_0 = 4$ and $p_1 = 0.9$. We note the that the dependence of the ratio on the parameters p_0 and p_1 is opposite in the Galactic plane and away from it: a larger integration region at low energy is preferred in the former case. This is due to the fact that the spectrum of the Galactic diffuse emission is harder in the plane than away from it, whereas the 4FGL-DR2 sources are on average softer. Regarding p_2, we choose 0.1° in order to minimize the level of correlation between pixels, as discussed in the next section. These parameters, defining the optimized spatial selection, are the ones used to produce the PS/TS ratio distributions of Fig. 2.

All these results have been obtained with a TS computed as normally done in the computation of TS maps, that is to say assuming a power-law spectrum. This choice, along with the fact

[6] https://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html

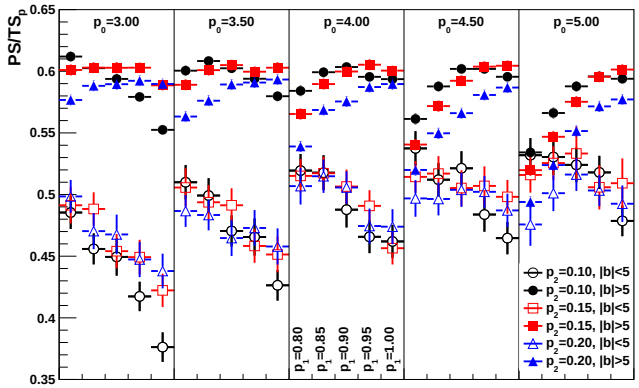

Fig. 3. Position of the maximum of the PS/TS distribution of low and high-latitude sources as a function of the spatial selection parameters p_0 and p_1 for three choices of $p_2 = 0.10, 0.15$ and $0.20°$.

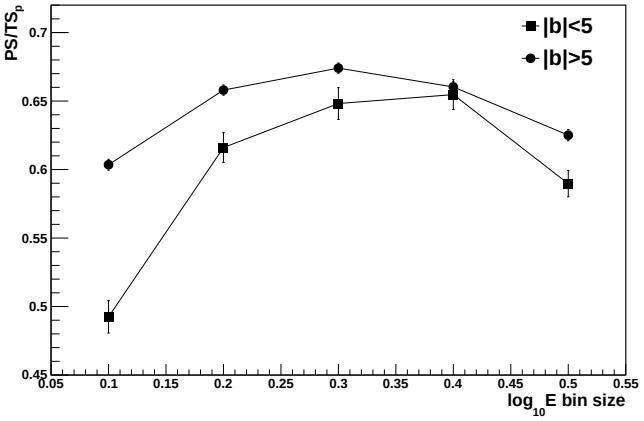

Fig. 4. Position of the maximum of the PS/TS distribution of low and high-latitude sources as a function of the energy binning.

that the PS and TS are computed using the same data and model 3D maps, ensures a fair PS/TS comparison. However ∼30% of the sources in 4FGL-DR2 are better modeled with a curved spectrum so the power-law assumption may bias the optimization of the spatial selection parameters. In order to check this possibility, we performed the same comparison but using the TS reported in the 4FGL-DR2 catalog, which is computed with the curved spectral shape (either a log normal or a subexponentially cutoff power law) when the source spectrum is found to be significantly curved. We find that, compared to the results presented in Fig. 3, the average value is decreased by ∼10% but the variation of the PS/TS ratio with the spatial selection parameters is the same, which shows that the optimization procedure was not biased by the power-law assumption.

So far we have used an energy binning of 0.1 in $\log_{10} E$. Using the optimized spatial selection, we computed the PS/TS ratio for a $\log_{10} E$ bin width of 0.2, 0.3, 0.4 and 0.5. Figure 4 shows that increasing the $\log_{10} E$ bin width reduces the difference between low and high Galactic latitude sources and that a maximum PS/TS ratio of 0.65 is reached for a bin width of about 0.3.

We conclude that the PS sensitivity is on average about 65% of the TS sensitivity. The sensitivity loss with respect to the TS is the cost of performing the energy-dependent integration. We note that the loss is rather modest, considering the pixel count information that is lost by the energy-dependent integration, and it is somewhat mitigated by the gain in computation speed.

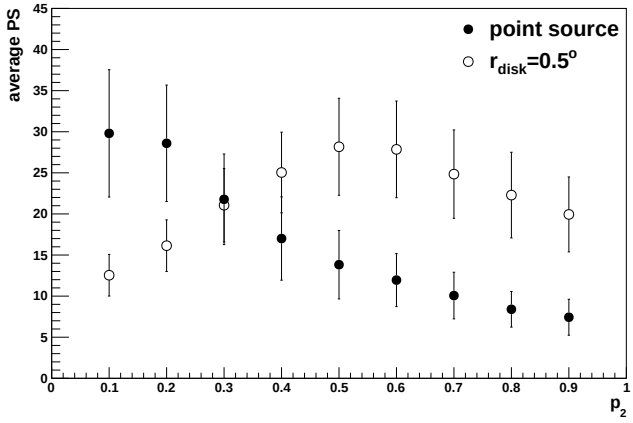

Fig. 5. Variation of the average PS with p_2 for a point-like source (filled circle) and an extended source (empty circle), whose spatial model is a uniform disk with a 0.5° radius. The average is computed over 20 simulations and the error bars correspond to the root mean square.

Using a larger energy binning has also the advantage of slightly increasing the fraction of the spectral bins that are in the Gaussian regime, and for which the absolute meaning of the systematic uncertainty is fully taken into account by the weights, as explained in Sect. 4. Since we use systematic uncertainties of about 3%, they do not have any impact on the bins with very few counts. The fraction is thus computed with respect to the number of spectral bins with at least one predicted count. For the RoI centered on the North Galactic pole, the average of this fraction over the pixels increases from 44% to 50% when the energy binning in $\log_{10} E$ is widened from 0.1 to 0.3. For the RoI with much more statistics centered on the Galactic center, it increases from 60% to 65%.

5.3. Extended deviation case

Since most of the 4FGL-DR2 sources are point-like sources, the optimized spatial selection, especially $p_2 = 0.1°$, is optimal for point-like sources. As noted in Sect. 2, we expect that a larger value of p_2 increases the PS sensitivity to extended deviations. In order to investigate this possibility, we computed the PS as a function of p_2 for a simulated source whose spatial extension is a uniform disk with a 0.5° radius and also for a point-like source. To perform this test, we used the catXcheck RoI centered on $(l = 0°, b = 20°)$, which has no 4FGL-DR2 source within 3° of its center. We placed the simulated source at the center of the RoI and set its spectrum to a power law with a photon index of 2. In both the point-like and extended cases, the flux has been chosen such that the resulting PS is about 30 on average.

We simulated 20 mock 3D count data maps with the simulated source in the model and computed the PS map with these data maps and the predicted count map when the simulated source is removed from the model. We then computed the average and the root mean square of the PS map maximum over the 20 simulations. Figure 5 shows the variation of the average PS with p_2. In the point-like case, the maximum is obtained with $p_2 = 0.1°$, as expected. In the case of the extended source, the maximum is reached for p_2 between 0.5 and 0.6°, which is on the order of the source extension, and is about twice the PS measured with $p_2 = 0.1°$, confirming that the PS sensitivity to extended deviations is significantly enhanced by increasing p_2.

Figure 6 shows some of the average PS maps obtained with these simulations. For the point-like source, the PS map with

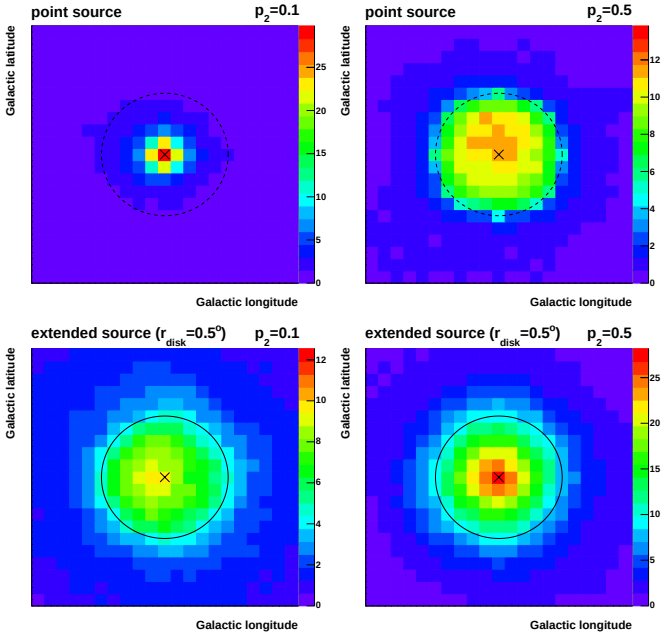

Fig. 6. 2°×2° region of the average PS map around the simulated source for a point-like source (*top*) and an extended source (*bottom*), whose spatial model is a uniform disk with a 0.5° radius. The PS is computed with $p_2 = 0.1$ (*left*) and 0.5 (*right*). The pixel size is 0.1° and the radius of the superimposed circle is 0.5°. The cross indicates the position of the simulated source.

$p_2 = 0.1°$ is clearly peaked at the position of the source, whereas with $p_2 = 0.5°$ it is almost flat over a disk of radius ∼0.5°. The results are reversed in the case of the extended source: the average PS map is flatter with $p_2 = 0.1°$ than with $p_2 = 0.5°$. But with the latter, the PS map is not as peaked as in the point-like source case with $p_2 = 0.1°$. We conclude that a flat PS peak with $p_2 = 0.1°$ likely corresponds to an extended deviation, which can be further confirmed by computing the PS with a larger p_2. We note however that the PS statistical fluctuations, visible in Fig. 5, are such that a precise extension measurement requires performing a standard TS-based analysis.

5.4. PS calibration

For each of the 438 RoIs, catXcheck produces a PS map and we look for $|\mathrm{PS}|_{\max}$, the maximum PS measured in the RoI. The asymptotic behavior of the expected CCDF of $|\mathrm{PS}|_{\max}$ can be easily derived from that of the PS if there is no correlation between pixels: it corresponds to a 10^{-x} function scaled by the trial correction factor 120×120 or, equivalently, shifted horizontally by $\log_{10}(120 \times 120) \sim 4.16$. In order to check the existence of correlations, we simulated 100 000 mock 3D count data maps and computed the corresponding PS maps, with the spatial selection parameters $p_0 = 4$ and $p_1 = 0.9$ and a $\log_{10} E$ bin width of 0.1. No systematic effect was included in this simulation and all weights were thus set to 1. We performed this test for two RoIs: the ones centered on the Galactic center and on the North Galactic pole. These two RoIs allow us to test very different situations in terms of statistics, as is clear from the integrated spectra at their respective centers shown in Fig. 1.

We first checked the |PS| CCDF derived from all the pixels, which is displayed in Fig. 7 for PS computed with $p_2 = 0.10°$. For the two RoIs, it closely follows the 10^{-x} expectation: the deviation from expectation expressed as an error on PS is within 3%. Figure 7 also shows the $|\mathrm{PS}|_{\max}$ CCDF. There is no significant deviation at large $|\mathrm{PS}|_{\max}$ values from the expectation after correction for the number of pixels. This demonstrates that the level of correlation between pixels is very low. When the PS is computed with $p_2 = 0.15°$, the |PS| CCDF is almost identical to $p_2 = 0.10$ but the $|\mathrm{PS}|_{\max}$ CCDF is systematically ∼3% below the expectation, as shown in Fig. 7. It actually follows the expectation with a trial correction corresponding to an effective number of pixels equal to $\sim 110^2$, smaller than 120^2, expected when there is no correlation. This shows that, for a pixel size of 0.1°, there is some correlation between adjacent pixels with $p_2 = 0.15°$.

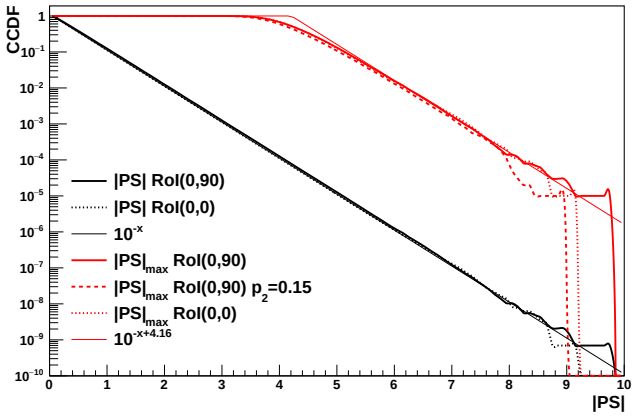

Fig. 7. Simulations of the Galactic North pole (solid) and the Galactic center (dotted) RoIs: |PS| (black) and $|\mathrm{PS}|_{\max}$ (red) CCDF as well as the corresponding expected distributions (thin solid).

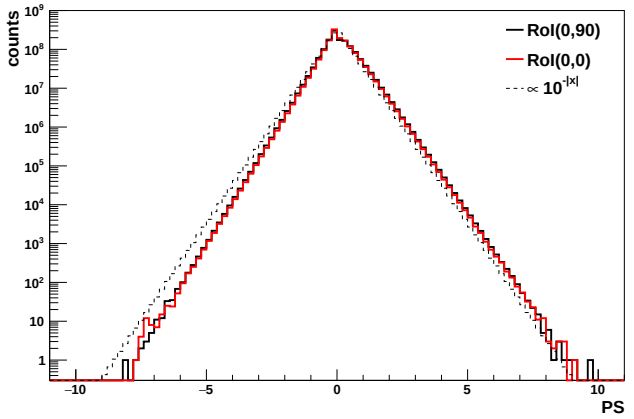

Fig. 8. PS distribution in simulations of the Galactic North pole (solid) and the Galactic center (dotted) RoIs, as well as the corresponding expected distributions (thin solid).

The distribution of the signed PS differs from the exponential expectation in two aspects, as can be seen in Fig. 8 for the two RoIs. The first difference is a systematic ∼0.1 horizontal shift with respect to the expectation. This is very likely due to the simple prescription we use to derive the PS sign. The second difference is at the peak around 0 and is the consequence of a lack of precision for p-values close to 1. This feature could be reduced by changing the L pdf computation parameters (especially n_{pdf}). However, the large p-value region is not critical when looking for significant data-model deviations so it does not seem useful to slow down the PS computation only to reach a better agreement with the expectation around 0. We note that, when PS is expressed in σ units, the notch at 0 is even more apparent.

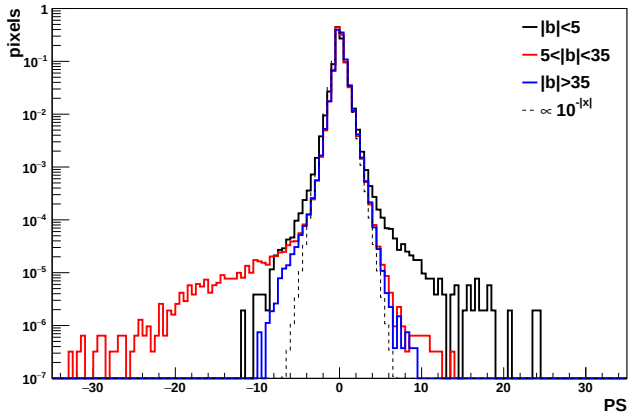

Fig. 9. PS distribution obtained in the verification of the 4FGL-DR2 catalog (10 years of data) for the three Galactic latitude selections. The thin solid histogram corresponds to the expected $10^{-|x|}$ distribution. All histograms are normalized such that their integral is 1.

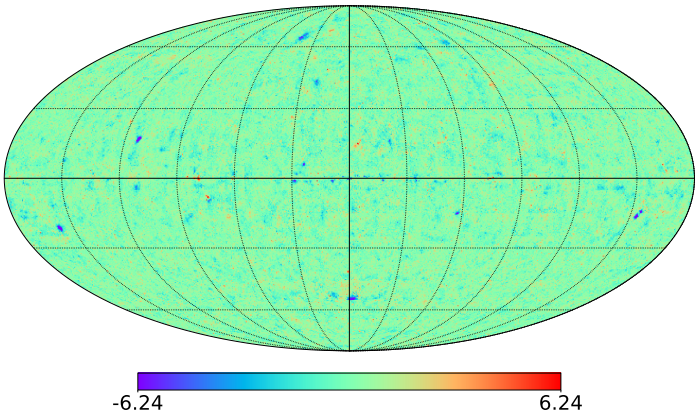

Fig. 10. All-sky PS map in Galactic coordinates (Mollweide projection).

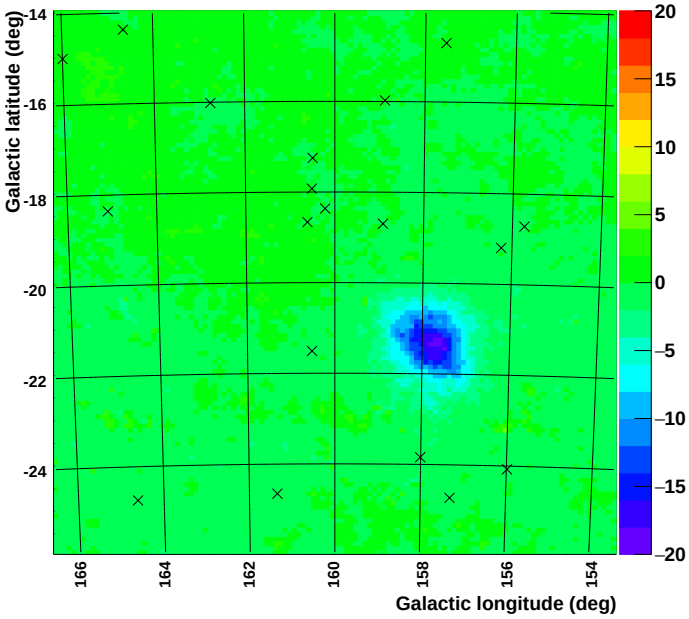

Fig. 11. PS map of the RoI centered on ($l = 160°, b = -20°$). The black crosses correspond to 4FGL-DR2 sources.

6. Results of the 4FGL-DR2 verification

In this section we present the results of the catXcheck analysis on the 4FGL-DR2 catalog. The PS map of each RoI was computed with the optimized spatial selection parameters and a $\log_{10} E$ bin width of 0.3 (see Appendix D for a detailed description of the PS map production). Since there are 120×120 pixels in each PS map and we consider 438 RoIs, the 3, 4 and 5σ thresholds correspond to PS = 9.4, 11 and 13, respectively. In the following we use PS > 11 to select significant deviations.

Figure 9 shows the PS distributions for three Galactic latitude samples: low ($|b| < 5°$), mid ($5° < |b| < 35°$) and high ($|b| > 35°$) latitudes. There is no significant deviation in the high-latitude selection whereas the low and mid-latitude distributions exhibit positive and negative broad tails, respectively.

From the PS maps of the 438 RoIs, it is possible to construct an all-sky PS map. We use a HEALPix (Górski et al. 2005) map in Galactic coordinates with $N_{\rm side} = 256$ and we set the PS of each pixel to the maximum of the PS found among the individual RoI pixels falling into that HEALPixel. The resulting all-sky PS map is shown in Fig. 10.

The most significant negative deviations (clusters of many pixels with PS < −11) correspond to five negative spots located at $(l, b) \sim (112.8°, 16.6°)$, $(157.6°, -21.2°)$, $(302.6°, -14.7°)$, $(206.9°, -16.7°)$ and $(205.0°, -14.1°)$. Since they are all close to large molecular clouds (the first one is associated with Cepheus, the second with Perseus, the third with Chamaeleon and the two last ones with Orion B), they are due to imperfections in the modeling of the Galactic diffuse emission[7]. A more detailed analysis of these imperfections will be included in the forthcoming 4FGL-DR3 catalog publication.

The PS map of the RoI centered on ($l = 160°, b = -20°$) containing the negative spot at ($l = 157.6°, b = -21.2°$) is shown in Fig. 11. The data and model integrated count spectra corresponding to the pixel with the lowest PS (−19.5) are shown in Fig. 12. One can see that the model overpredicts the data in the energy range between 0.3 and 10 GeV. Since the PS is relatively flat around the minimum, we performed a PS scan over the spatial selection parameter p_2. A minimum of −47.4, much lower than −19.5, is obtained around $p_2 = 1.3°$, indicating that the deviation is extended.

Both Figs. 9 and 10 show that there are several significant positive deviations. The one in the mid-latitude selection (PS = 12.3) is located in the RoI centered on ($l = 110°, b = -30°$) and it corresponds to a point-like excess which will be included in the next 4FGL-DR3 catalog (Fermi-LAT Collaboration, in prep.). It is also the case for most of the deviations in the low-latitude selection. The most significant one (PS = 24.25) is located at ($l = 286.55°, b = -1.15°$). The PS map of the corresponding RoI is shown in Fig. 13 and the data and model integrated count spectra at the maximum PS position are shown in Fig. 14. The excess of counts in data is visible above 1 GeV, which is typical of a missing source in the model. After investigation, it was found that this excess corresponds to the bright optical Nova ASASSN-18fv (Stanek et al. 2018) at ($l = 286.573°, b = -1.088°$), whose gamma-ray emission has been detected by *Fermi*-LAT (Jean et al. 2018; Aydi et al. 2020) around April 14, 2018. This source will be included in the 4FGL-DR3 catalog.

[7] https://fermi.gsfc.nasa.gov/ssc/data/analysis/software/aux/4fgl/Galactic_Diffuse_Emission_Model_for_the_4FGL_Catalog_Analysis.pdf

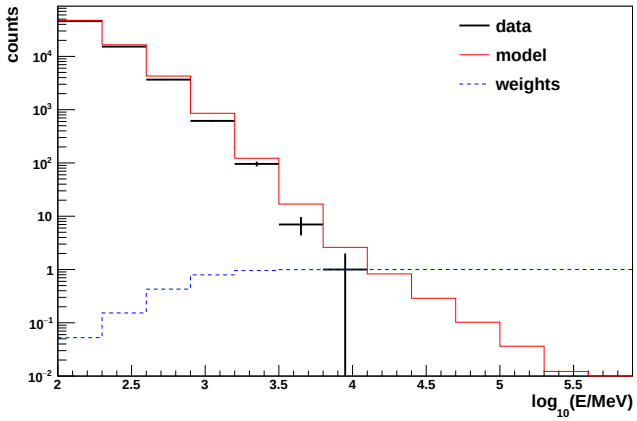

Fig. 12. Integrated count spectra corresponding to the maximum PS in the RoI centered on ($l = 160°, b = -20°$), for data (black) and model (red), as well as the average likelihood weights (dashed blue).

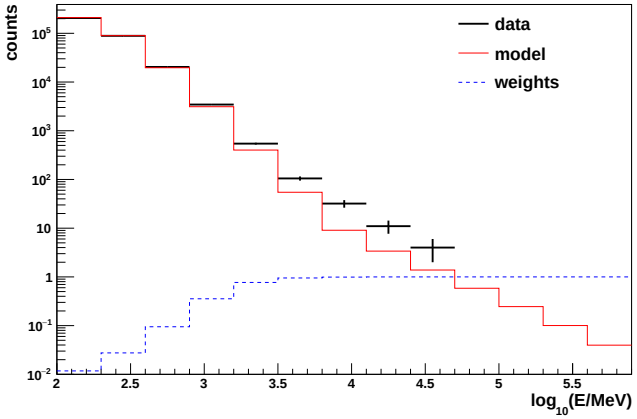

Fig. 14. Integrated count spectra corresponding to the maximum PS in the RoI centered on ($l = 290°, b = 0°$), for data (black) and model (red), as well as the average likelihood weights (dashed blue).

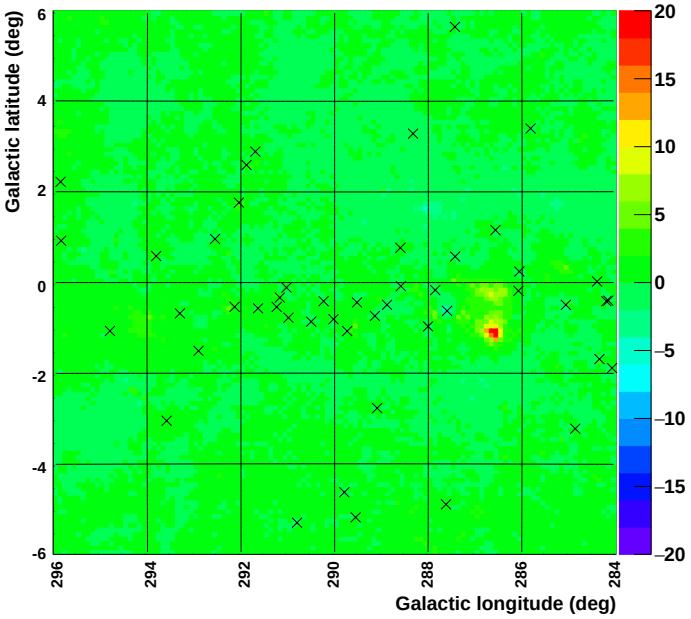

Fig. 13. PS map of the RoI centered on ($l = 290°, b = 0°$). The black crosses correspond to 4FGL-DR2 sources.

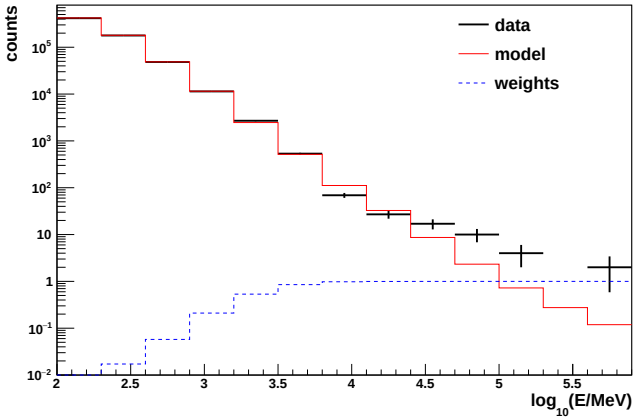

Fig. 15. Integrated count spectra at the position of the gamma-ray binary LS 5039, for data (black) and model (red), as well as the average likelihood weights (dashed blue).

The RoI centered on ($l = 20°, b = 0°$) provides an example of a source whose spectrum is mismodeled. The maximum of the PS map is 9.28 and its position corresponds to the gamma-ray binary LS 5039 (Abdo et al. 2009; Hadasch et al. 2012; Chang et al. 2016). The 4FGL-DR2 analysis considers three spectral models (power law, power law with subexponential cutoff and log normal) for each source and selects the best model. The LS 5039 spectrum is modeled with a log normal but this spectral shape is not able to reproduce the integrated count spectrum, as can be seen in Fig. 15. This is the consequence of the presence of a second spectral component, first reported by Hadasch et al. (2012) and recently confirmed by Yoneda et al. (2021).

7. Conclusion

Fermi-LAT analyses are generally based on a binned log-likelihood fit of a 3D count map but there is no fast, reliable and sensitive tool available to check the goodness-of-fit (including the case of negative residuals). In order to overcome the lack of such a tool, we have developed a new method that allows *Fermi*-LAT data users to quantify efficiently the data-model agreement after performing a fit of an RoI. The method is based on integrating the observed and predicted counts over an energy-dependent region around each pixel of the map and on computing a deviation estimator, named PS, between the integrated data and model count spectra. This method can incorporate the likelihood weights that are used in *Fermi*-LAT analyses to take into account some systematic uncertainty.

In order to minimize the computation time, the PS algorithm has been optimized while ensuring a PS precision of 3%. The PS statistical calibration has been checked with simulations and its average sensitivity to a point-like source deviation has been measured at 65% of the TS. This lower sensitivity is naturally explained by the energy-dependent spatial integration which dilutes the 3D information into 1D count spectra. However, this integration allows the PS map computation[8] to be

[8] A python script computing PS maps is available at the User contribution page of the *Fermi* Science Support Center website: https://fermi.gsfc.nasa.gov/ssc/data/analysis/user/. The links to the script and the documentation are https://fermi.gsfc.nasa.gov/ssc/data/analysis/user/gtpsmap/gtpsmap.py and https://fermi.gsfc.nasa.gov/ssc/data/analysis/user/gtpsmap/README.

much faster than for TS maps. This is very convenient while optimizing the sky model of an RoI. Another important advantage of the PS is that it is sensitive to both positive and negative deviations.

The use of PS maps to check 4FGL-DR2, the latest of the *Fermi*-LAT general catalog, has proven to be useful, reporting some positive deviations, actually corresponding to gamma-ray sources, and some negative deviations, related to imperfections in the modeling of the Galactic diffuse emission. The same verification is being performed in the preparation of 4FGL-DR3, the next catalog based on 12 years of data.

Acknowledgements. We thank our Fermi-LAT Collaborators Jean Ballet and Matthew Kerr for fruitful discussions. The *Fermi*-LAT Collaboration acknowledges generous ongoing support from a number of agencies and institutes that have supported both the development and the operation of the LAT as well as scientific data analysis. These include the National Aeronautics and Space Administration and the Department of Energy in the United States, the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique/Institut National de Physique Nucléaire et de Physique des Particules in France, the Agenzia Spaziale Italiana and the Istituto Nazionale di Fisica Nucleare in Italy, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), High Energy Accelerator Research Organization (KEK) and Japan Aerospace Exploration Agency (JAXA) in Japan, and the K. A. Wallenberg Foundation, the Swedish Research Council and the Swedish National Space Board in Sweden. Additional support for science analysis during the operations phase is gratefully acknowledged from the Istituto Nazionale di Astrofisica in Italy and the Centre National d'Études Spatiales in France. This work performed in part under DOE Contract DE-AC02-76SF00515.

References

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2009, ApJ, 706, L5
Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, ApJS, 247, 33
Ackermann, M., Ajello, M., Albert, A., et al. 2012, ApJS, 203, 4
Ackermann, M., Ajello, M., Allafort, A., et al. 2013, ApJ, 765, 1
Aydi, E., Sokolovsky, K. V., Chomiuk, L., et al. 2020, Nat. Astron., 4, 776
Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009, ApJ, 697, 1071
Atwood, W., Albert, A., Baldini, L., et al. 2013, ArXiv e-prints [arXiv:1303.3514]
Ballet, J., Burnett, T. H., Digel, S. W., et al. 2020, ArXiv e-prints [arXiv:2005.11208]
Bruel, P., Burnett, T. H., Digel, S. W., et al. 2018, ArXiv e-prints [arXiv:1810.11394]
Chang, Z., Zhang, S., Ji, L., et al. 2016, MNRAS, 463, 495
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Hadasch, D., Torres, D. F., Tanaka, T., et al. 2012, ApJ, 749, 54
Hu, F., & Zidek, J. V. 2002, Can. J. Stat., 30, 347
Jean, P., Cheung, C. C., Ojha, R., et al. 2018, ATel, 11546, 1
Mattox, J. R., Bertsch, D. L., Chiang, J., et al. 1996, ApJ, 461, 396
Stanek, K. Z., Holoie, T. W.-S., Kochanek, C. S., et al. 2018, ATel, 11454, 1
Wood, M., Caputo, R., Charles, E., et al. 2017, in 35th International Cosmic Ray Conference, 10–20 July, 2017, Bexco, Busan, Korea, 301
Yoneda, H., Khangulyan, D., Enoto, T., et al. 2021, ApJ, 917, 90

Appendix A: Iterative computation of the log-likelihood pdf

In this appendix, we describe the iterative procedure to compute the probability distribution function (pdf) of the random variable L defined in Equation 1 of Section 3. Let $N_{\rm bins}$ be the number of bins of the model integrated count spectrum and m_i the model integrated counts in bin i. Let pdf_i be the probability distribution function corresponding to the first i bins. pdf_1, corresponding to the first bin, is simply given by the Poisson distribution.

In order to build pdf_i from pdf_{i-1}, we must consider all the ways that a realization of the data could alter the values of the log-likelihood. Given the model integrated counts m_i, drawing a value of the counts k changes L by $-\log \mathcal{P}(k, m_i)$. Thus, the iteration consists of computing the probability weighted sum of all the possible values of k. In other words, each k contribution corresponds to L_{i-1} shifted by $-\log \mathcal{P}(k, m_i)$ and multiplied by $\mathcal{P}(k, m_i)$, and pdf_i is obtained by simply summing all these contributions:

$$\mathrm{pdf}_i(x) = \sum_k \mathcal{P}(k, m_i) \times \mathrm{pdf}_{i-1}\left(x + \log \mathcal{P}(k, m_i)\right). \tag{A.1}$$

In order to minimize the computing time, we apply the following prescriptions. Firstly, the L pdf is computed as an histogram with $n_{\rm pdf} \times N_{\rm bins}$ bins on the range $[0, 20N_{\rm bins}]$. Secondly, we introduce the precision parameter ϵ and, at each iteration, we only consider k in the interval $[k_0(m_i), k_1(m_i)]$ defined as the narrowest interval such that $\sum_k \mathcal{P}(k, m_i) \geq 1 - \epsilon$. The $k_0(m)$ and $k_1(m)$ parameterizations are given in Appendix B.

An example of such an iteration is shown in Figure A.1. In this example, we use $n_{\rm pdf} = 200$ and $\epsilon = 10^{-15}$ and the count spectrum corresponds to a 40 bins E^{-2} spectrum between 100 MeV and 1 TeV with 10000 counts in the first bin. Figure A.1 shows the 12th iteration, corresponding to the 12th spectral bin for which $m_i = 39.81$. The L pdf of the previous step is shown in black in the top panel. It is used to derive the k contributions to the L pdf for k in $[0, 100]$, as shown in the bottom panel after division by the maximum contribution (corresponding to $k = 39$). Summing all these contributions along the y-axis leads to the L pdf after the iteration that is shown in red in the top panel.

As explained in Section 3, we first compute the contribution of the spectral bins with Gaussian statistics from the χ^2 distribution with a number of degrees of freedom equal to the number of these bins, and then run the iterative computation over the spectral bins with Poisson statistics.

The algorithm to compute the L pdf thus depends on three parameters: ϵ, defining the interval $[k_0(m), k_1(m)]$, $n_{\rm pdf}$, defining the number of bins of the L pdf histogram ($n_{\rm pdf} \times N_{\rm bins}$) and N_g, the lower limit on the number of counts to decide whether a spectral bin is in the Gaussian regime.

The choice of N_g is not critical because of the steepness of the count spectra. For a $\log_{10} E$ bin size of 0.1, the count ratio of neighboring bins is typically 2. So increasing N_g from 100 to 200, for instance, moves at most one spectral bin from the Gaussian regime to the Poisson regime, which is not enough to change significantly the final L pdf.

In order to test the sensitivity to ϵ, we set $n_{\rm pdf} = 200$ (corresponding to a very fine binning of the L pdf histogram) and $N_g = 100$ and we compare the results obtained with ϵ ranging from 10^{-5} to 10^{-15}, the latter being used as the reference. For each ϵ value, we compute the L complementary cumulative distribution function (CCDF) in order to obtain the curve PS(L). Figure A.2 shows the ratio of these curves as a function of the reference PS. One can see that $\epsilon = 10^{-7}$ is small enough to get a 2% precision up to PS = 20, that is to say well above 10, which corresponds to $\sim 5\sigma$ for a typical 100×100 pixels map. Ensuring a 1% precision up to PS = 100 can be useful when using the PS map around a positive excess to estimate the position of

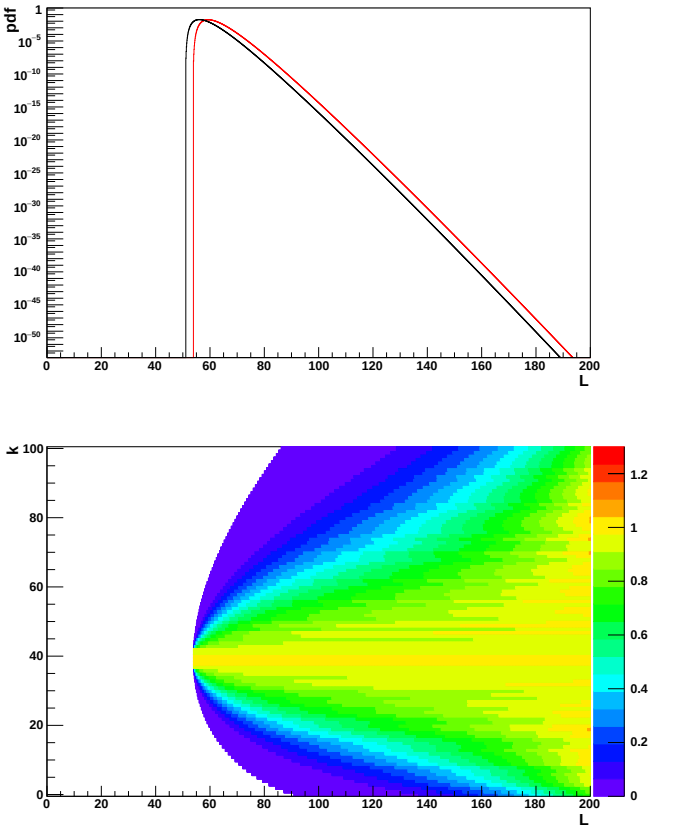

Fig. A.1. L distributions at the beginning (black) and the end (red) of an iteration of the L pdf computation algorithm (top) and the corresponding individual k contributions (bottom) after division by the maximum contribution (corresponding to $k = 39$). The number of predicted counts in this step is 39.81 and the y-axis range corresponds to the $[k_0, k_1]$ interval obtained with $\epsilon = 10^{-15}$.

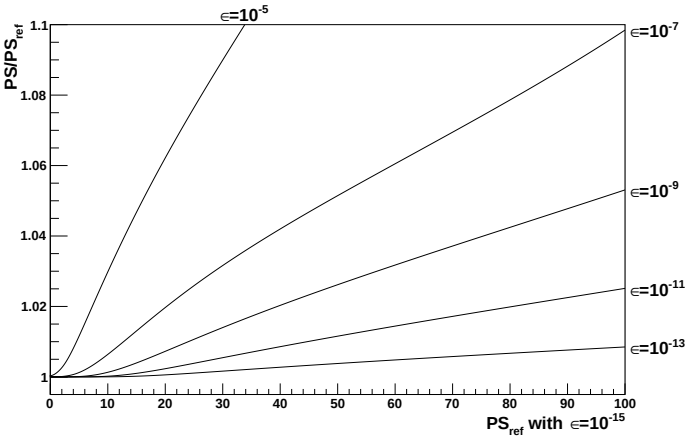

Fig. A.2. Ratio of PS to the reference PS as a function of the reference PS for several choices of ϵ, the precision parameter that defines the interval of k considered during the L pdf computation. The ratio is actually evaluated as a function of L and L is converted to the PS for $\epsilon = 10^{-15}$.

a possible point source at the origin of the deviation. In that case a smaller ϵ is needed, at the expense of computation time.

Regarding the choice of $n_{\rm pdf}$, we compare the results obtained with $n_{\rm pdf}$ = 10, 20, 50, 100 and 200, the latter being used as the reference. Differences larger than 1% are only seen for $n_{\rm pdf}$ = 10 and 20. In order to minimize computation time while ensuring a 2% precision, we use the following parameters: $\epsilon = 10^{-7}$, $n_{\rm pdf} = 50$ and $N_g = 100$.

Appendix B: $[k_{\rm min}, k_{\rm max}]$ interval parameterization

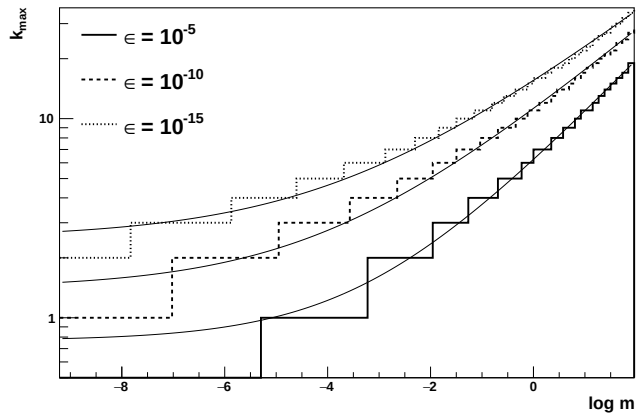

Fig. B.1. Fit of $k_{\rm max}$ as a function of $\log m$ for three choices of ϵ $(10^{-5}, 10^{-10}, 10^{-15})$.

For a given number of predicted counts m and a given precision parameter $\epsilon \ll 1$, let $k_{\rm min}$ and $k_{\rm max}$ be the boundaries of the narrowest k interval such that $\sum_k \mathcal{P}(k, m) \geq 1 - \epsilon$. In this appendix, we derive the parameterizations of $k_{\rm min}$ and $k_{\rm max}$ as functions of m and ϵ. We note that the goal of the introduction of ϵ is mostly to be able to vary the level of precision of the L pdf computation (described in Section 3) in order to find an optimal choice with respect to the computation speed. As a consequence, the fact that $\sum_k \mathcal{P}(k, m)$ monotonically increases when ϵ decreases is more important than ensuring a high level of accuracy for these parameterizations.

For $m < 5$, we set $k_{\rm min}(m, \epsilon) = 0$. In order to determine $k_{\rm max}(m, \epsilon)$, we first compute $k_{\rm max}$ as a function of $\log m$ for several values of ϵ, and fit these curves with the function $a + bm^c$, as shown in Figure B.1.

We then look at the variation of the parameters a, b and c as a function of $\log_{10} \epsilon$ between −15 and −5. We fit these curves with a simple quadratic function, as shown in Figure B.2. We find the following parametrizations:

$$a(\epsilon) = 0.715029 + 0.049825 \log_{10} \epsilon + 0.011768 (\log_{10} \epsilon)^2,$$

$$b(\epsilon) = -0.308206 - 1.309547 \log_{10} \epsilon - 0.028455 (\log_{10} \epsilon)^2,$$

$$c(\epsilon) = 0.817286 + 0.050841 \log_{10} \epsilon + 0.001828 (\log_{10} \epsilon)^2.$$

For $m \geq 5$, in order to find the narrowest $[k_{\rm min}, k_{\rm max}]$ interval such that $\sum_k \mathcal{P}(k, m) \geq 1 - \epsilon$, we sort the Poisson probabilities $\mathcal{P}(k, m)$ in decreasing order, find the smallest subset for which $\sum_k \mathcal{P}(k, m) \geq 1 - \epsilon$ and then find the lowest and highest k among this subset. Rather than trying to directly parameterize their variations with m and ϵ, we compare them to the Gaussian expectations $m \pm \sigma(\epsilon) \sqrt{m}$, with $\sigma(\epsilon) = \sqrt{2} \mathrm{erf}^{-1}(1 - \epsilon)$. Because of the positive skewness of the Poisson distribution, both $k_{\rm min}$ and $k_{\rm max}$

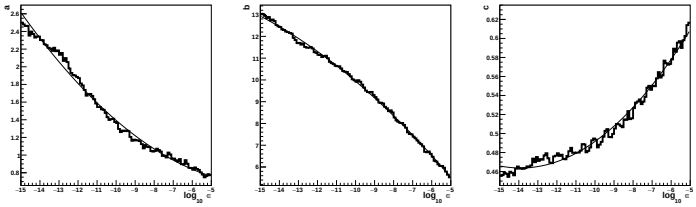

Fig. B.2. Variations with $\log_{10} \epsilon$ of the parameters of the function $a+bm^c$ used to fit the dependence of $k_{\rm max}$ with m.

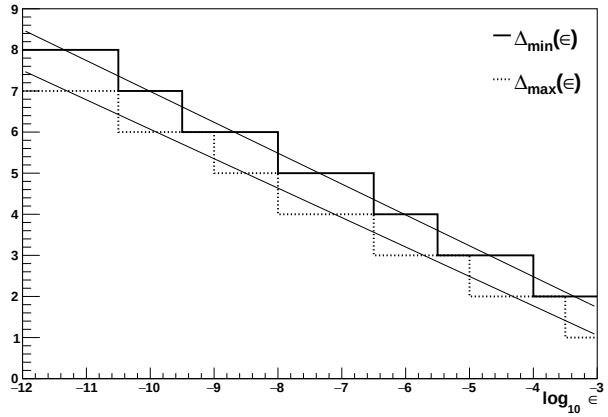

Fig. B.3. $\Delta_{\rm min}$ and $\Delta_{\rm max}$, the differences between $k_{\rm min}$ and $k_{\rm max}$ with respect to their Gaussian expectations, as a function of $\log_{10} \epsilon$.

are greater than their Gaussian expectations. We find that, for a given ϵ, these differences are almost constant (± 1 unit) for $m < 200$. So we define

$$\Delta_{\rm min}(\epsilon) = \min_{5<m<200} \left(k_{\rm min} - (m - \sigma(\epsilon) \sqrt{m}) \right),$$

$$\Delta_{\rm max}(\epsilon) = \max_{5<m<200} \left(k_{\rm max} - (m + \sigma(\epsilon) \sqrt{m}) \right).$$

Figure B.3 shows that the variation of $\Delta_{\rm min}$ and $\Delta_{\rm max}$ with $\log_{10} \epsilon$ can be approximated with a linear parameterization. We use

$$k_{\rm min}(m, \epsilon) = \max \left(0, m - \sigma(\epsilon) \sqrt{m} - 0.523564 - 0.75129 \log_{10} \epsilon \right),$$

$$k_{\rm max}(m, \epsilon) = m + \sigma(\epsilon) \sqrt{m} - 1.09374 - 0.716202 \log_{10} \epsilon.$$

Appendix C: Testing the weighted log-likelihood with simulations

When computing the PS, we use Equation 7 to take into account the likelihood weights in the log-likelihood function. These weights are associated to a certain level of systematic uncertainty. As explained in Section 4, the absolute meaning of the systematic uncertainty is ensured for the spectral bins with Gaussian statistics by using the χ^2 approximation to compute the corresponding L pdf.

In order to study the case of the spectral bins with Poisson statistics, we performed simulations in which the number of observed counts in bin k is drawn according to a Poisson distribution of mean μ_k, which is itself drawn, for each realization, according to a Gaussian distribution of mean m_k and standard deviation σm_k, where σ is the systematic uncertainty level. This simple simulation is not realistic because it ignores the likely correlations between spectral bins but it corresponds to a situation in which the correction brought by the χ^2 approximation works exactly for Gaussian statistics. Therefore this simulation

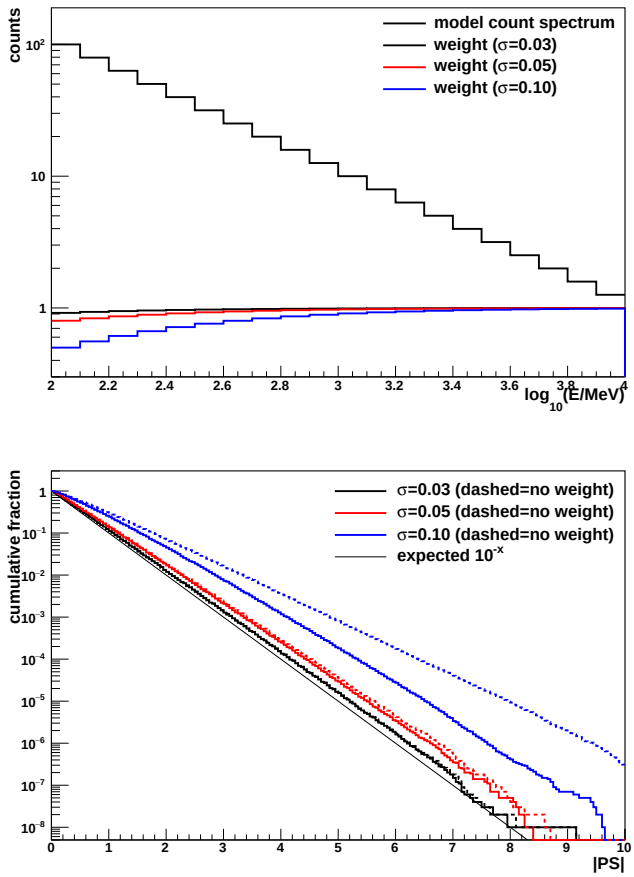

Fig. C.1. Simulations to test the use of likelihood weights to handle systematic uncertainty: count spectrum and average weights (top) and PS CCDF as a function of |PS| (bottom). The three levels of systematic uncertainty are 3, 5 and 10%. The solid and dashed histograms correspond to the weighted and unweighted PS, respectively. The expected 10^{-x} distribution is also shown (thin solid line).

allows us to compare how well the weighted log-likelihood correction of Equation 7 performs for Poisson statistics relatively to the corrected χ^2 approximation.

A simple analytical calculation shows that the mean and variance of the number of observed counts in each spectral bin are m_k and $m_k + (\sigma m_k)^2$, respectively. It means that systematic uncertainties lower than 10% have a negligible impact on spectral bins with $m_k \leq 1$.

We considered a 20 bins E^{-2} spectrum such that the number of counts decreases from 100 to 1 and performed simulations for three different values of σ: 3, 5 and 10%. We set the weights to $w_k = 1/(1 + m_k\sigma^2)$. The count spectrum and the weights are shown in the top panel of Figure C.1 and the resulting PS CCDF are shown in the bottom panel, as well as the ones obtained when no weight correction is applied.

If the weight correction were perfect, the PS CCDF should follow the 10^{-x} behavior. In the $\sigma = 10\%$ case, the weight correction has a significant effect but the error on PS is about 25%. For $\sigma = 3\%$, corresponding to the current LAT case, the weight correction has no effect but the error on PS is within 3%. In the $\sigma = 5\%$ intermediate case, the weight correction starts to play a role but the resulting error on PS is about 10%.

It appears from this study that, although the weighted version of the log-likelihood does not fully encompass the effect of systematic uncertainty, it allows the PS estimator to take it partly into account in the case of systematic uncertainty larger than 5%.

Appendix D: PS map production steps

In this appendix we recap the several steps that we go through to produce the PS map of a given RoI from the input data and model 3D count maps as well as the likelihood weight maps.

The data and model 3D count maps are produced with the `Fermitools`[9] `gtbin` and `gtmodel`, respectively. The spatial part of the maps is a $12° \times 12°$ map with a pixel size of 0.1°, while the energy part ranges from 100 MeV to 1 TeV with a $\log_{10} E$ bin size of 0.1. The likelihood weights are produced with `gtwtsmap` from a 3D data count map covering a larger spatial region ($22.6° \times 22.6°$ with a pixel size of 0.1°) in order to ensure a good estimation of the weights within the $12° \times 12°$ inner region that is used in the PS computation.

The three steps of the PS production, whose flowchart is shown in Figure D.1, are the spatial integration, the energy rebinning and the PS computation.

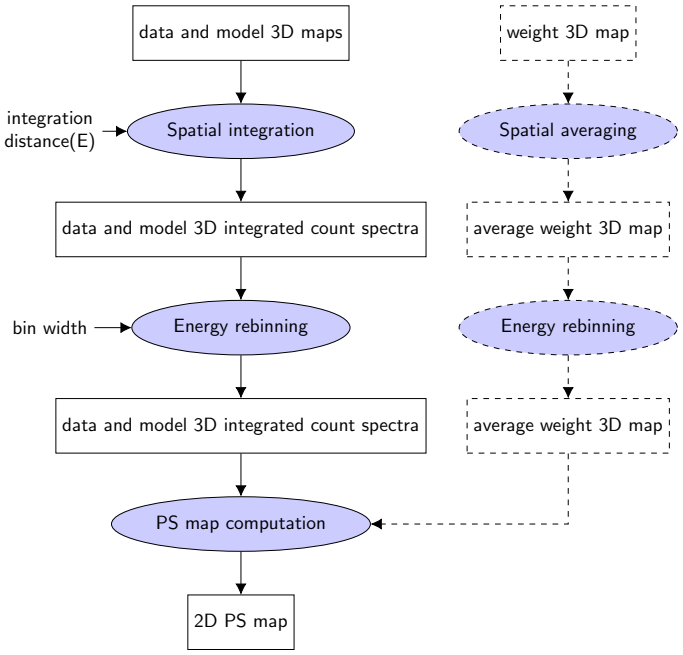

Fig. D.1. Flowchart of the PS map production.

D.1. Spatial integration

This step produces the 3D integrated count spectra. It is performed independently for each energy bin k, whose lower bound is noted E_k. For each (i, j) pixel, we sum all the counts in the pixels within a distance $d(E_k)$ from pixel (i, j), with $d(E_k) = 4°(E_k/100\,\mathrm{MeV})^{-0.9} \oplus 0.1°$. For the log-likelihood weight 3D map, we compute the average weight over the pixels within a distance $d(E_k)$ from pixel (i, j), as explained in Section 4.

D.2. Energy rebinning

This step produces 3D integrated count spectra with a $\log_{10} E$ bin size of 0.3 from the 3D maps produced in the previous step. For each (i, j) pixel, the energy bins of the data and model counts are summed by group of three, while for the log-likelihood weights we compute the average weight over the three merged bins (with the same prescription as explained in Section 4). The energy part

[9] https://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/references.html

of the output 3D maps ranges from 100 MeV to $10^{5.9}$ MeV with a $\log_{10} E$ bin size of 0.3.

D.3. PS computation

This step produces the 2D PS map. For each (i, j) pixel, we use the data and model integrated count spectra as well as the log-likelihood weights from the 3D maps produced in the previous step. The model integrated count spectra and the weights are used to compute the L pdf (starting with the pdf corresponding to all the spectral bins with Gaussian statistics and then performing the iterative computation described in Appendix A for the remaining bins), which allows us to compute the p-value corresponding to the likelihood obtained with the data integrated counts. The absolute value of the pixel PS is $-\log_{10}$(p-value) and its sign is given by Equation 9.