Improving the *Fermi* LAT Source Catalog

Jean Ballet[*]

Laboratoire AIM, CEA-IRFU/CNRS/Universite Paris Diderot, Service d'Astrophysique, CEA Saclay, F-91191 Gif sur Yvette, France

E-mail: jballet@cea.fr

on behalf of the *Fermi*-LAT collaboration

The *Fermi* Large Area Telescope has been routinely gathering science data since August 2008, surveying the full sky every three hours. The current source catalog (3FGL) is based on four years of data. Besides a longer time interval, the next source catalog will be based on the new Pass 8 data, which introduces a number of improvements at all energies, and in particular enhanced response at low energy. This opens the possibility of broadening the energy range to reach below 100 MeV, but it also introduces new challenges due to the larger number of low-energy photons in a range where the point-spread function is several degrees wide. In that regime the source-to-background ratio is small (at the percent level) so systematic errors on the background model are critical, and the Earth limb contamination is important. The energy dispersion must also be accounted for. We describe here how we plan to address those challenges in order to keep the catalog robust.

The 34th International Cosmic Ray Conference,
30 July- 6 August, 2015
The Hague, The Netherlands

---

[*]Speaker.

1. Introduction

Since *Fermi* was launched in 2008, the LAT collaboration has strived to keep pace with the data and provide ever deeper source catalogs. The current version (3FGL, [2]) is based on four years of Pass 7 Reprocessed data. In parallel to working on 3FGL, the LAT collaboration has prepared and made public a new event reconstruction (Pass 8, [4]) which significantly extends the capabilities of the LAT. The main challenge of the next version of the LAT source catalog will be robustness, in particular with respect to the systematic uncertainties. The modelling of the Galactic diffuse background [5] will undoubtedly improve but will necessarily remain imperfect, and the large additional effective area at low energy (+ 60% at 100 MeV) means that second-order effects now need to be accounted for. This contribution describes four important aspects which need to be addressed for the next-generation source catalog. They are considered independently, in order to quantify each of them.

Other improvements, not discussed further here, may enter in 4FGL:

Extend the energy range upward to 1 TeV, and downward to 30 MeV.

Include additional extended sources and improve the description of existing ones.

For faint pulsars, fix the superexponential index to a value less than 1, as indicated by the average obtained for the brightest ones [1].

Implement phase selection around bright pulsars to facilitate detection of faint sources in their vicinity.

2. Likelihood weights

In 3FGL we introduced a first attempt at accounting for systematic errors in the maximum likelihood process itself, at the source detection level. It was not used in the source characterization, however, for lack of a suitable framework. The standard way to account for systematic errors (for example in *XSPEC*[1]) is to define them as a fraction ε of the signal and add them to the statistical errors in quadrature, in a χ^2 formalism. This can be adapted to the maximum likelihood framework by introducing weights and using the fact that the statistical variance in Poisson statistics is the signal itself, so that (M_i and n_i are the model and observed counts in each bin)

$$\log \mathscr{L} = \sum_i w_i (n_i \log M_i - M_i) \;\; \text{with} \;\; w_i = \frac{M_i}{M_i + (\varepsilon M_i)^2} = \frac{1}{1 + \varepsilon^2 M_i} \tag{2.1}$$

where the sum runs over all bins in space and energy. The source significance can then be quantified in the same way, via the Test Statistic $TS = 2 \log(\mathscr{L}/\mathscr{L}_0)$ in which $\mathscr{L}$ and $\mathscr{L}_0$ are the (weighted) likelihood with and without the source of interest, respectively.

However, that formulation has a serious flaw, which is that it is not stable to rebinning. If one splits the bins in half, then M_i is split in half while ε stays the same (it is defined externally). In the limit of very small bins, obviously the weights will all tend to 1 and the $\log \mathscr{L}$ formula will tend to the unweighted one, even though nothing has changed in the underlying data or the model.

The solution we propose is to define a suitable integral over energy (E) and space ($\mathbf{r}$) $N(\mathbf{r}, E)$ which will not depend on binning. M_i in the weight formula will then be replaced by $N(\mathbf{r}_i, E_i)$

[1]https://heasarc.gsfc.nasa.gov/xanadu/xspec/
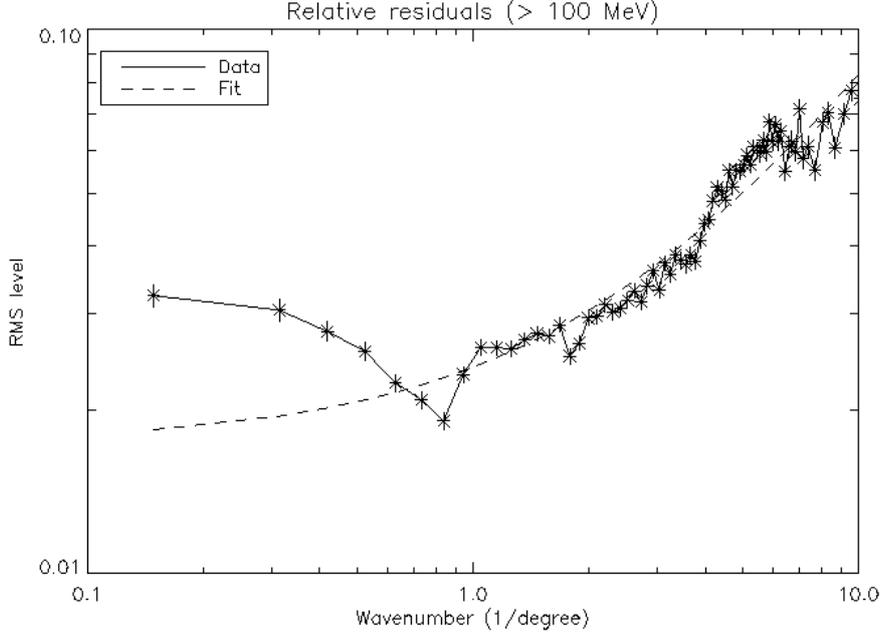
Figure 1: Relative residuals between the data and the model (diffuse background + sources) in Fourier space, as a function of wavenumber $k = 2\pi/\lambda$. This is obtained for the 3FGL data set (4 years of Pass 7 Reprocessed data, 100 MeV to 300 GeV). The dashed line is a simple linear fit: $RMS = 0.017 + 6.5 \times 10^{-3}\,k$.

taken at the event's coordinates. For the integral over space, since the catalog mostly deals with point sources, the logical solution is to convolve the model with the PSF $P(\mathbf{r}, E)$, normalized to 1 at the peak (this is equivalent, for a flat diffuse model, to multiplying by the PSF solid angle). Note that the model already contains the PSF, so this amounts to applying a double convolution to the sky model.

For the energy integral the choice is less obvious. The source spectrum is not a narrow line, so convolving with the energy dispersion (similar to what is done for space) is not justified. Since the spectrum is broadband and TS is global, an integral over the full energy range could look more appropriate. But this would give the same weight to all energies, which is clearly not what we want (there is no reason to downplay the few high-energy events). The option we are considering at present is to integrate over all energies above the current one. That integral is dominated by the lowest energies because of the PSF convolution, so the exact upper bound is not important. Another change is that, since the origin of the systematic error is the diffuse model, we would compute the integral counts $N(\mathbf{r}, E)$ from the diffuse model B only (there are also systematic uncertainties on sources due to calibration, but they do not affect the significance).

$$w_i = \frac{1}{1 + \varepsilon^2 N(\mathbf{r}_i, E_i)} \text{ with } N(\mathbf{r}_i, E_i) = \int_{E_i}^{E_{\max}} S(\mathbf{r}_i, E)\, dE \text{ and } S(\mathbf{r}, E) = \frac{dB}{dE}(\mathbf{r}, E) * \frac{P(\mathbf{r}, E)}{P(0, E)} \quad (2.2)$$

It remains to be defined what ε should be. To that end we have built the 3FGL model of the full Galactic plane up to $\pm 30°$ and compared it to the Pass 7 Reprocessed data. We show in Fig.1 the RMS of the residuals (*i.e.* the square root of the power spectrum of the residuals divided by that of the full model). The result shows residuals at the level of 3% (consistent with what we used
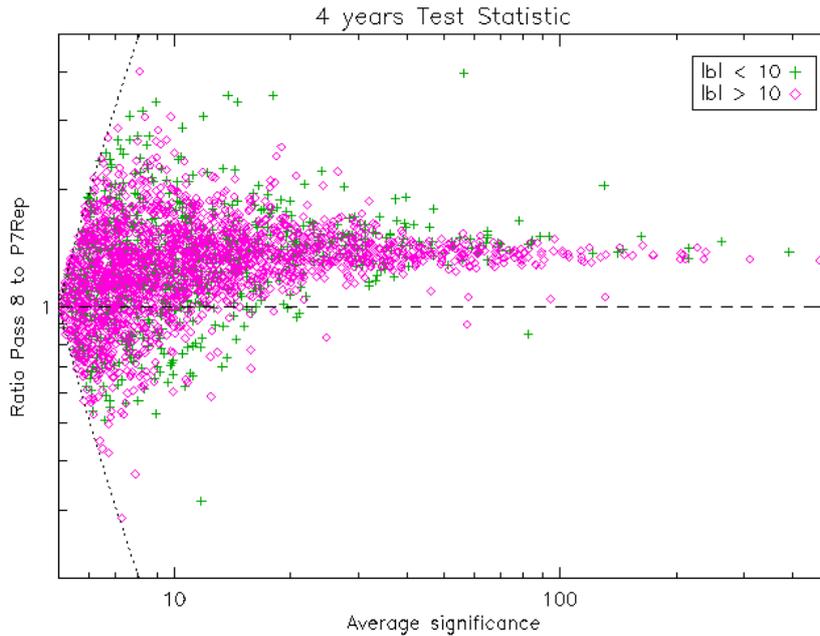
Figure 2: Ratio of the Test Statistic values between the Pass 8 data and the original 3FGL (Pass 7 Reprocessed) analysis, for the same time interval of 4 years, as a function of the approximate source significance $(\sqrt{TS_{\text{Pass8}}} + \sqrt{TS_{\text{P7Rep}}})/2$. Sources within $10°$ of the Galactic plane are drawn in green, higher-latitude sources appear in pink. The dotted lines mark the cuts at $TS > 25$.

in 3FGL) with a slight dependence on wavenumber. Eq.2.2 could be generalized to include scale dependence.

3. Switching to Pass 8 data

The main feature of Pass 8 data from the point of view of the source catalog is that it increases the effective area by about 20% between 1 and 10 GeV where most of the significance of LAT sources comes from. The improvement is larger at both low and high energy. This helps broaden the useful energy range for the catalog analysis. In particular, a large number of sources can be observed directly above 50 GeV [3]. Pass 8 introduces also PSF event types which allow splitting the data into four subsets on the basis of the PSF quality. This will be discussed in Sect.4.

In order to test the impact of this improvement, we repeated the 3FGL source characterization but using Pass 8 data. Apart from that, everything was exactly the same (same energy range and time interval, same zenith angle cut, same use of *Front* and *Back* data, same use of unbinned likelihood above 3 GeV, same set of seeds to start from). The Galactic diffuse model was rescaled to Pass 8 data and the isotropic template recomputed, as explained in the FSSC page [2].

The result is illustrated in Fig.2, which compares the resulting Test Statistic. The average ratio is 1.27, quantifying the improvement. The number of resulting sources at $TS > 25$ is not actually larger in this test, because it is limited by the seeds which still come from Pass 7.

---

[2]http://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html
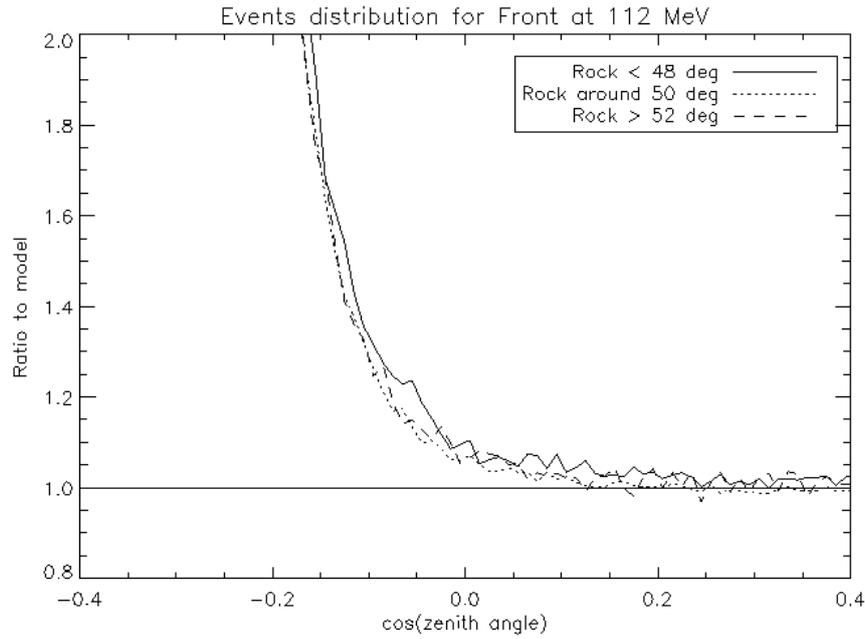
Figure 3: Count rate increase due to the Earth limb as a function of zenith angle, for three rocking angle selections, based on 6 years of Pass 8 data. $50°$ is the standard rocking angle in normal survey mode. This particular plot is for *Front* events between 100 and 126 MeV.

Besides that, we observe a few small biases: the energy flux is lower with Pass 8 by 0.5 σ on average, while scatter is introduced at the 1 σ level (Pass 8 is a full new event reconstruction, in which many events are different from Pass 7); faint sources appear softer by about 0.1 in the Galactic Ridge. Both effects are probably related to the Galactic and isotropic diffuse models, and the fact that low-energy events are more numerous in Pass 8 and so effectively have more weight in the likelihood analysis.

4. Handling the Earth limb

For the 3FGL analysis (which was above 100 MeV) the contamination from Earth limb γ-rays was reduced by excluding events with zenith angle (angle between the current zenith and the reconstructed event direction) larger than $100°$ as well as time intervals when the rocking angle (angle between current zenith and pointing direction of the LAT) was larger than $52°$, and modeling the residuals with a template. That does not work as well with Pass 8 data which has greater effective area at low energy. In addition, *Fermi* spent most of 2014 in a special combination of pointing and survey modes designed to favor the Galactic center, which exceeded rocking angle $52°$ regularly. So we reviewed that issue.

Starting from 6 years of Pass 8 data, we built an empirical model of the full sky from low zenith angles ($< 66°$, the minimum possible while still covering the full sky for *Fermi*'s orbit inclined at $25.6°$). Then we computed the total emission (relative to that expected from the empirical model) in a number of zenith angle bins and for three rocking angle selections. An example is shown on Fig.3. The conclusion was that large rocking angles do not adversely affect the data. This was
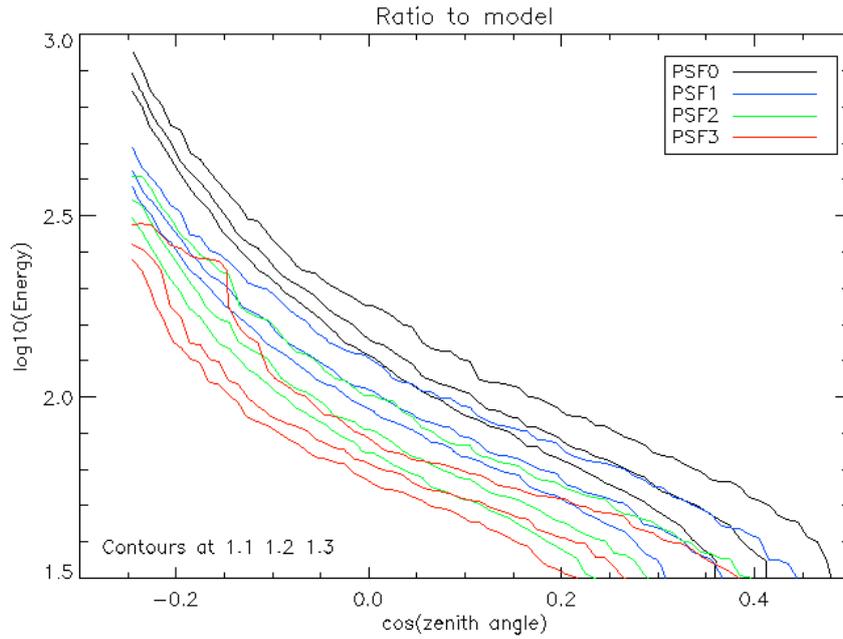
Figure 4: Contours of equal relative contamination of the Earth limb in zenith angle and energy (in MeV), based on 6 years of Pass 8 data. Three contours are drawn for each PSF type, at levels of 10, 20 and 30%.

not obvious, because it depends on a subtle balance between the effective area and the PSF wings, as a function of off-axis angle (the main effect of increasing the rocking angle is that areas of the sky further from the zenith are observed closer to the telescope axis). The same conclusion was qualitatively true at all energies and for *Back* events.

Retaining all rocking angles means that we cannot easily construct a spatial template for the residual as in 3FGL, because Earth limb events can appear (nearly) anywhere in the sky. On the other hand, because events are plentiful at low energy where the Earth limb is the brightest, cutting out the Earth limb entirely (to the point where it is not necessary to include it in the model at all) by introducing an energy-dependent zenith angle cut is an alternative that we have pursued.

The way to reach that objective is illustrated in Fig.4. To build that figure we have used the Earth limb excess as in Fig.3 but at all energies. From that it is possible to define a region in energy vs. zenith angle space where the contamination is less than 10% (at that level, it has no measurable effect). Fig.4 was obtained for the new Pass 8 PSF types. As expected, the Earth limb is less of a contaminant for the 'good' event types (large PSF number). From that plot, we could for example use PSF3 alone below 100 MeV with zenith angle cut $z_{max} = 70°$, PSF2+3 between 100 and 300 MeV with $z_{max} = 90°$, PSF1+2+3 between 300 MeV and 1 GeV with $z_{max} = 100°$, and all events above 1 GeV with $z_{max} = 105°$.

This has a side effect, however, because a zenith angle cut amounts to a time selection. Using a different cut at different energies means that the time coverage is not the same. This is not innocuous for variable sources (such as most blazars) so it would not be recommended for a scientific study of particular sources. For a catalog, however, all it means is that different parts of the spectrum have different weights, so the best-fit spectrum will reflect those weights, but because
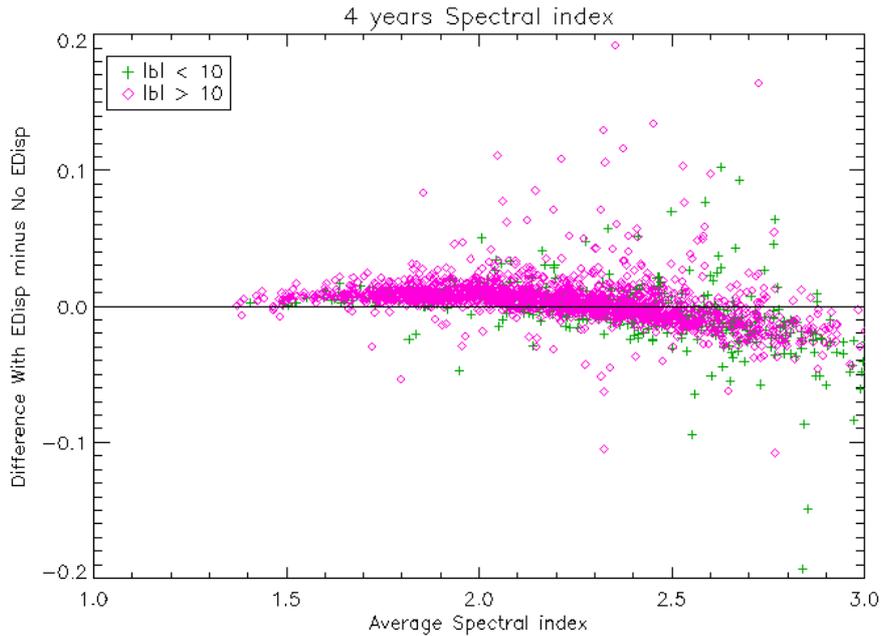
Figure 5: Difference of the spectral index values between the analysis with and without energy dispersion, as a function of spectral index itself. Same color convention as in Fig.1.

that selection is independent of the source state it should not bias population studies one way or another. We prefer using the same zenith angle selection for different event types at a given energy, in order not to introduce tension in the fit between those event types in cases of strongly variable sources.

5. Energy dispersion

The energy reconstruction in *Fermi* is relatively good, better than 10% over most of the energy range, and no sharp feature is expected in the source spectra or the acceptance, so energy dispersion was not explicitly accounted for in the LAT source catalogs up to 3FGL. For the next generation we wish to broaden the energy range to below 100 MeV. At those energies, the steeply rising effective area combined with energy dispersion around 20% do not allow neglecting it.

Before moving to low energies, a difficult challenge because of the large confusion incurred by the broad PSF (even for PSF3 events the 68% containment radius is 3° at 100 MeV), we quantified the effect of energy dispersion inside the 3FGL energy range by enabling evaluation of its effect on the likelihood function in the Science Tools for all sources (but not for the components of the diffuse model, the spectra of which were fitted neglecting energy dispersion). The reference was the Pass 8 run described in Sect.3, and everything else was left identical.

The Test Statistic was found to be similar. The energy flux was lower by 2.4% on average with energy dispersion, with relatively little scatter (0.27 σ) out of the Galactic plane. That bias depends on spectral shape, and goes to 0 for hard sources. The spectral shape is also affected at a low level, even though there is no average bias. Fig.5 shows that hard sources ($\Gamma \leq 2$) become a little softer, while soft sources ($\Gamma \geq 2.5$) become a little harder. Both effects are very small (at the level of
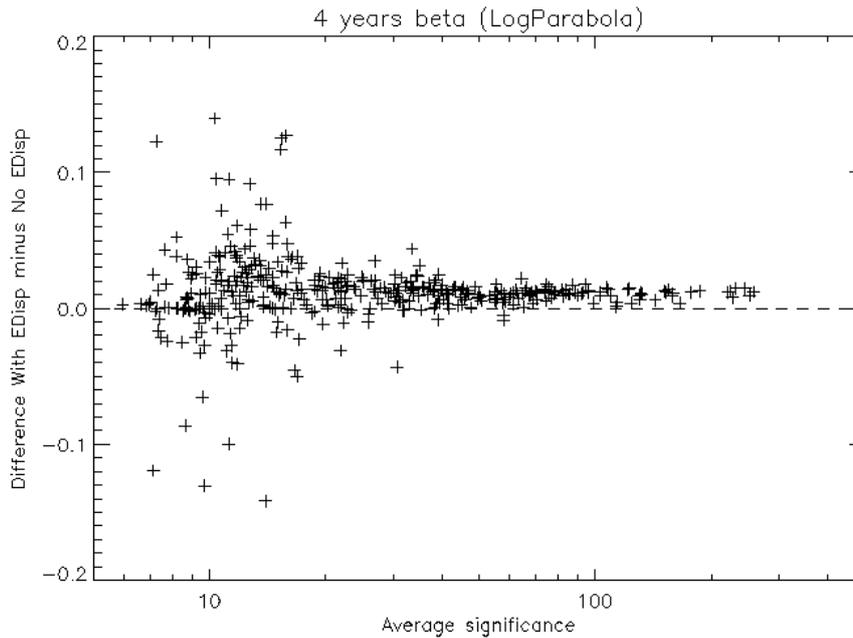
Figure 6: Difference of the curvature values for LogParabola sources between the analysis with and without energy dispersion, as a function of the source significance as in Fig.1. Same data selection as above.

0.01). In the same vein, Fig.6 shows that sources appear a little more curved when accounting for energy dispersion, by 0.014 on average.

Acknowledgments

References

[1] A. A. Abdo *et al.* 2013, *The Second Fermi Large Area Telescope Catalog of Gamma-Ray Pulsars*, *ApJS* **208**, 17 [astro-ph/1305.4385]

[2] F. Acero *et al.* 2015, *Fermi Large Area Telescope Third Source Catalog*, *ApJS* **218**, 23 [astro-ph/1501.02003]

[3] M. Ajello *et al.* 2015, *2FHL, the second Catalog of Hard Fermi-LAT Sources*, this conference

[4] W. Atwood *et al.* 2013, *Pass 8: Toward the Full Realization of the Fermi-LAT Scientific Potential*, *2012 Fermi Symposium proceedings*, eConf C121028, astro-ph/1303.3514

[5] J.-M. Casandjian *et al.* 2015, *The Fermi-LAT model of interstellar emission for standard point source analysis*, *2014 Fermi symposium proceedings*, astro-ph/1502.07210