Technical Note
radphi-2002-601

Monte Carlo study of the shower shape in the LGD

Mihajlo Kornicer
Richard T. Jones
University of Connecticut, Storrs, CT

June 7, 2002

Abstract

We present an approach of improving the LGD clustering scheme by using the Point Spread Function (PSF). As the first approximation, the shower shape of isolated photons in the LGD, generated by the Gradphi, is analysed using a single Gaussian as a PSF. We found that the shower energy distribution over the LGD blocks is strongly peaked around the shower center. As a result, the shower radial width weakly depends on energy. Fitted shower energies are systematically lower than generated photon energies by $\approx 7 - 10\%$. This confirms that the shower shape is not a single Gaussian. In addition, it is found important to integrate a PSF function over the block area, instead of just sampling a function at the center of the block. The difference between these two cases can be significant when the PSF radial width is smaller than the size of the LGD block. The next step in this approach is to probe the PSF on real data.

1 Introduction

Current mechanism for reconstructing photon showers, their energy, position, and multiplicity in the LGD is based on the IU clusterizer. One of the factors that limits the clusterizer efficiency in reconstructing showers from the hits in the LGD is the finite size of the LGD blocks (4x4 cm). Showers separated by less than $\approx 6-8$ *cm* are very hard to distinguish. In addition, low energy photons can escape from the eye of the clusterizer, and can be merged into a high-energy neighbor. From previous study [1] it is realized that accurate photon multiplicity reconstruction is crucial for eliminating background from the final states of interest to Radphi.

One approach to improve shower reconstruction is to exploit the shape of the shower in the LGD, depending on the photon energy and impact angle. The idea is based on the optical imaging technique, used for example in astrophysics, to reconstruct the point light source from the known response function in the detector, called the point spread function (PSF). The procedure to build a new algorithm should be as follows:

1 - use isolated showers to guess the shower shape and make a 1-cluster fitter based on the PSF that will depend only on the shower energy and position (E, x_0, y_0).

2 - obtain a good sample of isolated clusters from real data, and modify the PSF as necessary. Compare resolution and efficiency with the current clusterizer.

3 - design a multi-cluster fitter using MC data and refine on real data. Make decision regarding further analysis based on the comparison with the current clusterizer.

2 Single shower fitter

There are various functions that have been used to describe the shape of showers in EM calorimeters. Usually one uses a sum of 2 functions: a core part that describes the peak of the energy distribution and a second one that covers the tail. Among others, two exponential, exponential and Gaussian, and two Gaussian functions have been exploited [2]. The single shower fitter is built to use an arbitrary PSF function without changing the algorithm. The first step in fitting an event is to get initial values for the fitting parameters, (E, x_0, y_0), from the data. In the case of an isolated shower, E is set to the total energy in the LGD, E_{tot}, while (x_0, y_0) is set by the coordinates of the block with maximum energy yield. The fit is performed using MINIUT. In order to speed up fitting, we decided not to fit the whole LGD array, but rather a matrix (currently 9X9) around the maximum energy block. The fitter minimizes χ^2

$$\chi^2 = \Sigma_i \frac{(e_i - E_i)^2}{\sigma_i^2}, \tag{1}$$

where E_i is the observed block energy, and e_i is the PSF function taken at the center of the block. Errors σ_i in Eq. 1 are calculated from expression

$$\sigma_i = A\sqrt{E_i} + B, \tag{2}$$

with A=0.08 and B=0.01.

The first trial function to be used is a single Gaussian

$$G(x, y) = Ne^{(-\frac{1}{2}\frac{r^2}{R^2})}, \tag{3}$$

with r being the radial distance from the shower center $r = \sqrt{(x-x_0)^2 + (y-y_0)^2}$. The normalization N is determined by the the shower energy E, N = E/$(2\,\pi R^2)$. In order to get the proper overall normalization one has to multiply Eq. 3 with the LGD block area L^2. The radial Gaussian width R should be fixed, but can be a function of energy E and angle θ. In Fig. 1 is shown how R converges, when it is used as a free parameter, to the value $R \approx 1.5$ while shower energy increases. For $E > 0.5$ GeV, R is almost independent of energy. It seems this is a feature of MC showers, connected with the fact that core contribution, consisting of just two central blocks, is much more prominent than the tail. We will try to explain this later. An additional result from the first fit is that the fitted energy is biased by $\approx 7\%$ towards lower values with respect to the generated energy.

When R is fixed to its average value of 1.7, the negative energy bias in the fitted energy increases to $\approx 10\%$. Comparing data with function, block by block (Fig. 2), it is found that the negative bias comes

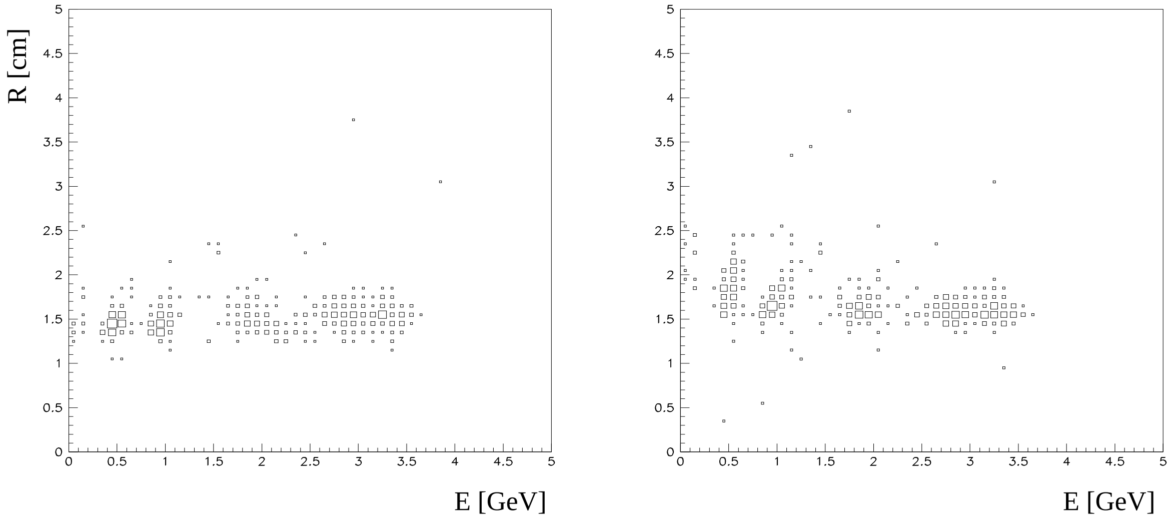

Figure 1: Radial width R vs. energy for 0.1, 0.5, 1.0, 2.0, 3.0, 3.5 GeV photons shot in the direction of $(\theta,\phi) = (10,5)^0$. Left: initial $R_0 = 1.5$, $B = 2.0$. Right: initial $R_0 = 2.5$, $B = 5.0$. Note: when R is allow to vary, B has to be large in order to stabilize fitter and get physically acceptable fits.

from an under-estimate of the two central blocks. There are two possible explanations: MC showers are not Gaussian, or we are misinterpreting the "true" shape by sampling a function at the center of the block. The second consideration is based on the following observation: for radial widths of order $R \approx L/2$, a Gaussian is changing rapidly across the block. While we sample the function at the center of the block, the actual function value can be quite different at the block edges. It might be necessary to integrate over a block to guess the shape of the PSF. Proper normalization of the function requires multiplication with the block area. This is equivalent to 0-th order expansion of the integral

$$I(x_c, y_c) = \int_{x_c-L/2}^{x_c+L/c} f(x)dx \approxeq f(x_c, y_c)L^2 = I_0(x_c, y_c), \tag{4}$$

around the center of the block (x_c, y_c). On the other hand, for any 2-dim function, a 2-nd order expansion can be obtained from

$$\begin{aligned} I_2(x_c, y_c) = & \int_{x_c-L/2}^{x_c+L/2} \int_{y_c-L/2}^{y_c+L/c} dxdy \\ & \left(f(x_c, y_c) + D_x(x-x_c) + D_y(y-y_c) + \frac{1}{2}D_{xx}(x-x_c)^2 + D_{xy}(x-x_c)(y-y_c) + \frac{1}{2}D_{yy}(y-y_c)^2 \right) \\ & = f(x_c, y_c)L^2 + \frac{1}{2}\left(D_{xx} + D_{yy}\right)|_{(x_c,y_c)} \frac{2L^4}{24}, \end{aligned} \tag{5}$$

where D stands for a partial derivative of f with respect to a particular index. In the case of the Gaussian, this is reduced to

$$I_2 = G(x_c, y_c)L^2 \left(1. + \frac{r^2 - 2R^2}{R^4}\frac{L^2}{24}\right). \tag{6}$$

The true value of the integral over the block area is given by

$$I = \frac{E}{2}\left[erf\left(\frac{\Delta x - L/2}{\sqrt{2}R}\right) - erf\left(\frac{\Delta x + L/2}{\sqrt{2}R}\right)\right]\left[erf\left(\frac{\Delta y - L/2}{\sqrt{2}R}\right) - erf\left(\frac{\Delta y + L/2}{\sqrt{2}R}\right)\right], \tag{7}$$

with $(\Delta x, \Delta y)$ being $((x_c - x_0), (y_c - y_0))$. Table 1 shows a block-by-block comparison of data with I_0, I_2, and I, from two showers with 1.0 GeV and 3.5 GeV photons. For the first shower, the I_0 approximation

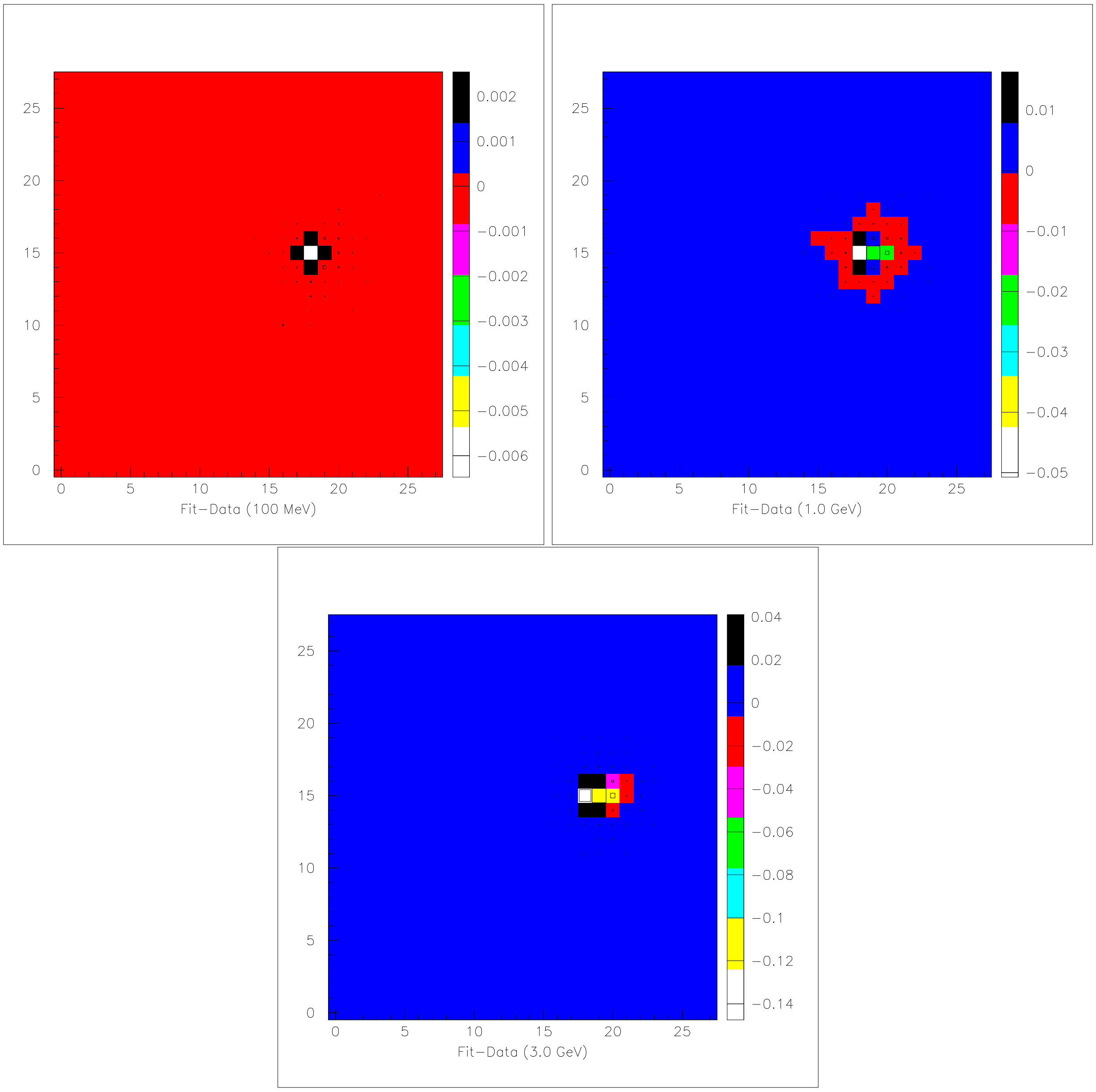

Figure 2: Fit-data difference, averaged over 100 shots, for 0.1, 1.0, and 3.5 GeV photons, (top-left, top-right and bottom colored plot respectively). Data are superimposed for comparison (black boxes).

Table 1: Comparison between I_0, I_2, and I from two fits. Note: not all blocks with non-zero data have been shown.

E_{fit}	X_0	Y_0	R		
0.891053	19.606381	5.840514	1.607639		
X_b	Y_b	Data	I_0	I_2	I
18.0	2.0	0.025080	0.030720	0.068005	0.065595
22.0	2.0	0.027360	0.016705	0.042231	0.044924
26.0	2.0	0.004560	0.000019	0.000112	0.000352
18.0	6.0	0.533520	0.530301	0.394643	0.408305
22.0	6.0	0.291840	0.288370	0.305232	0.279637
26.0	6.0	0.059280	0.000321	0.001467	0.002193
18.0	10.0	0.010260	0.018751	0.046285	0.046610
22.0	10.0	0.002280	0.010196	0.028374	0.031922
E_{fit}	X_0	Y_0	R		
2.701281	21.084819	6.086944	0.966082		
X_b	Y_b	Data	I_0	I_2	I
18.0	2.0	0.030780	0.000006	0.000115	0.005431
22.0	2.0	0.004560	0.000612	0.007948	0.036056
26.0	2.0	0.002280	0.000000	0.000000	0.000053
18.0	6.0	0.346560	0.044842	0.307623	0.339312
22.0	6.0	2.256060	4.686579	1.022599	2.252726
26.0	6.0	0.139080	0.000018	0.000317	0.003307
18.0	10.0	0.019380	0.000012	0.000229	0.008419
22.0	10.0	0.045600	0.001288	0.015373	0.055895

is used as the PSF and the radial width obtained is $\approx L/2$ if R is allowed to vary. One can see that I_0 significantly differs from I_2 and I. In the second case the integral I is used as the PSF, with $R \approx L/4$, where both I_0 and I_2 differ from I.

In Fig. 3 is shown the profile of the fitted energy as a function of the generated energy obtained using I. The energy bias is essentially unchanged, relative to the case where I_0 is used. The radial width again weakly depends on energy but it drops to $R = 1.12$ averaged over photon energies. Table 2 shows the average fitted R as a function of shower energy. Figs. 4, 5, and 6 show a comparison between MC data and the PSF (I function), for R fixed to the values from Table 2. Comparing fit-data differences for various shower energies (bottom right plots on Fig. 4, 5, and 6) with the similar ones obtained by I_0 (Fig. 2) one can see that central blocks are fitted better with I than with I_0.

Table 2: Average R and energies, and corresponding errors, for fits obtained by the I-function. Once R is fixed to the value from the left-most column, average fitted energy again decreases so that overall negative bias reappears.

E_γ $[GeV]$	R $[cm]$	ΔR $[cm]$	E_{fit} $[GeV]$	ΔE $[GeV]$	E_{fit} $[GeV]$(fixed R)
0.1	1.30	0.16	0.1	0.02	0.092
0.5	1.16	0.19	0.47	0.05	0.45
1.0	1.12	0.16	0.93	0.07	0.90
2.0	1.13	0.13	1.85	0.13	1.81
3.0	1.11	0.10	2.77	0.17	2.72
3.5	1.12	0.12	3.25	0.16	3.18

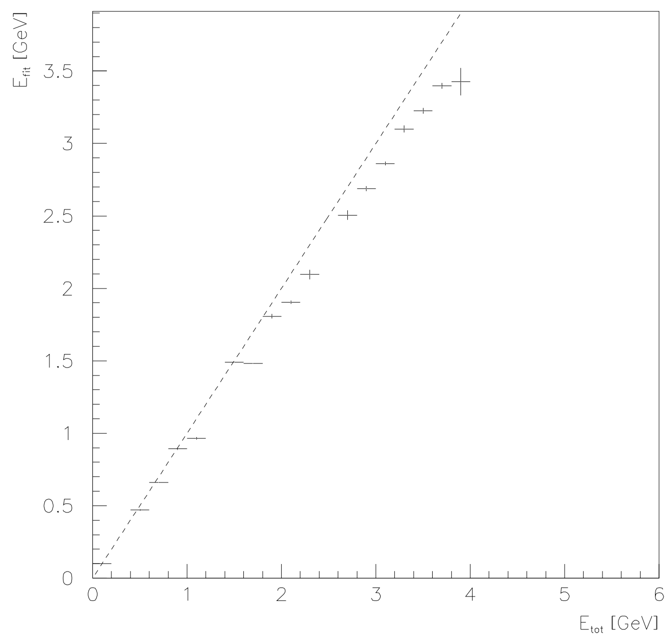

Figure 3: The profile of the fitted energy as a function of the total energy in the LGD, obtained using $I(x_c, y_c)$, and values of R from the Table 2. The straight dashed line would represent unbiased fit.

3 Discussion

Testing the single shower PSF fitter we realized that we are facing several difficulties in determining the correct shape of the shower:

- by nature, conversion of the EM energy in the calorimeter is a stochastic process: we can deduce only the average shower shape. Then we can compare fit-data difference with the "noise", and accept or reject the result.

- finite width of the block: we sample our function at the center of the block to compare it with the block energy. This is OK for a function that is changing slowly across the block, but might be misleading for rapidly changing functions. We can take care of this by replacing the value of the function at the center of the block by its integral over the block area. If the function is not analytically integrable we can use 2-nd, or a higher even-order expansion of the integral (in any integral expansion around the center of the bin only even terms survive). We found out that these corrections are very important for the Gaussian with small width compared to the block size. If a function is not analytically integrable, or if it is too complicated to find higher order corrections, numerical integration might be necessary.

- the shape of the shower is not radially symmetric with respect to the shower center, but rather distorted in the direction defined by impact angle (θ, ϕ). This can be taken into account by introducing an elliptical PSF function. However, this again complicates the fit if numerical integration is required.

This single Gaussian study shows that a single shower fitter is stable when R is fixed. The fitted shower energy is biased by more than $\approx 10\%$ for higher energy showers. The bias results from the fact that we fit a sharply peaked distribution with a single Gaussian. The bias remains even when we integrate the fitting function across the block area. The bias itself should not be a problem; it can be removed by proper calibration of the fitter. A bigger problem is the 2-block dominance in the shower energy distribution of MC showers. It might arise as a consequence of low light attenuation along the LGD blocks. We expect that real data will have a broader distribution because of higher attenuation. The next step is to look at real data before we fine-tune PSF.

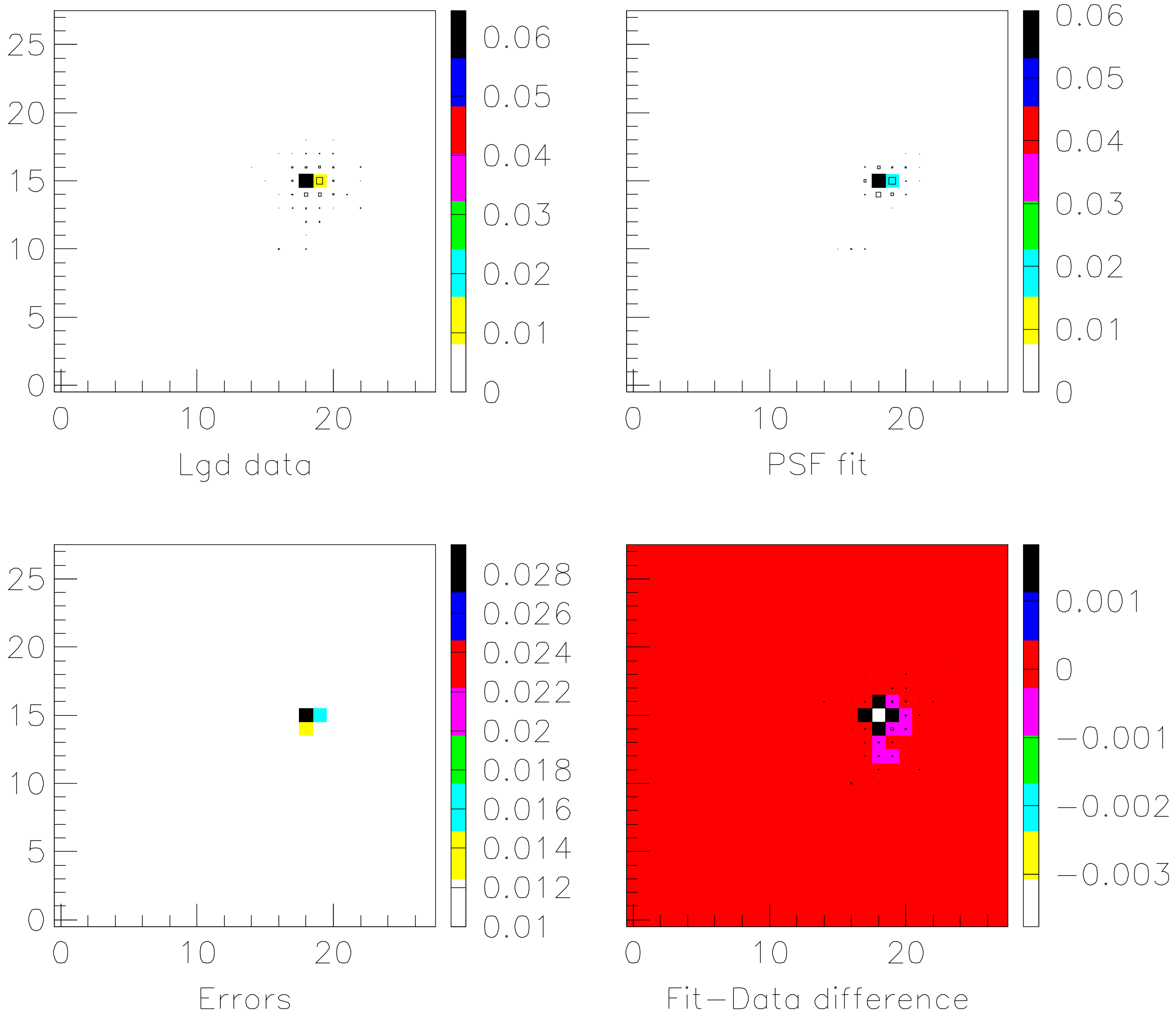

Figure 4: Comparison of average data with $I(x_c, y_c)$ and PSF (top-left and top-right plots respectively), for fixed R, in the case of 0.1 GeV photons. Bottom-left plot represent errors, while bottom-right shows fit-data difference. Box-plots are superimposed for comparison.

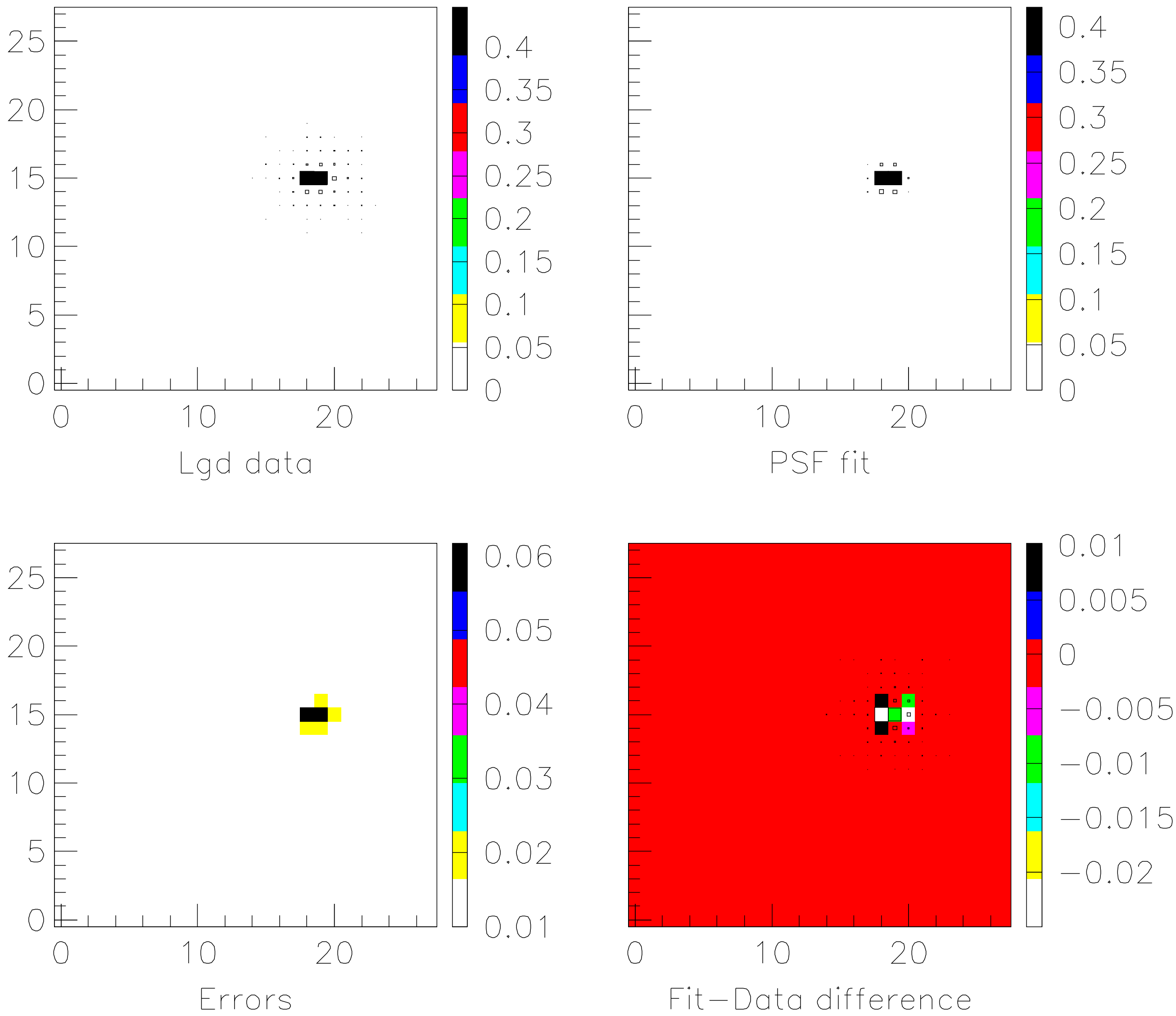

Figure 5: The same as Fig. 4 for 1.0 GeV photons.

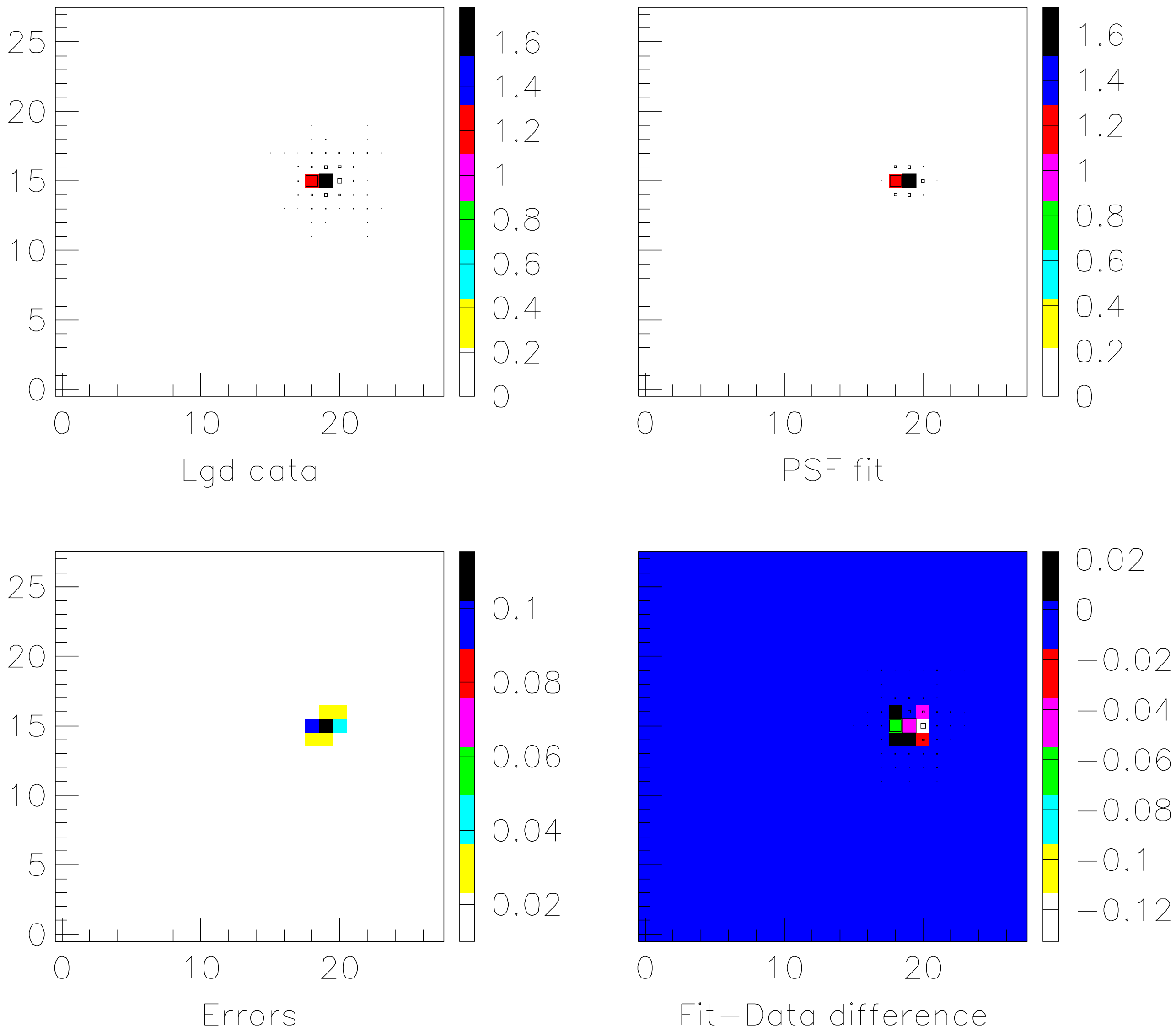

Figure 6: The same as Fig. 4 for 3.5 GeV photons.

References

[1] "MC study of background from b_1 reactions", M. Kornicer, R.T Jones, Radphi TechNote *2001-711* (November 2001).

[2] "Data Analysis Techniques for High-Energy Physics", R. Frühwirth, M. Regler, R. K. Bock, H. Grote, D. Notz, Cambridge University Press 2000.