Algorithm of reconstruction for electromagnetic shower analysis in emulsion cloud chambers

Y. Caffari, F. Juget

Algorithm of reconstruction for electromagnetic shower analysis in emulsion cloud chambers

Y. Caffari, IPN Lyon, France
F. Juget, Institut de Physique de Neuchâtel, Switzerland

4th May 2006

Abstract

This note describes an algorithm of reconstruction for electromagnetic cascades in emulsion cloud chambers. This algorithm, performed under the ROOT framework, is tested over MC simulations and experimental data (6 GeV electrons, dry scan).

Contents

1 Introduction

An electron test beam has been performed at DESY in order to study electromagnetic showers and electron identification for the $\nu_\mu \rightarrow \nu_e$ oscillations and $\tau \rightarrow$ e decay channel. In this note we describe an algorithm of reconstruction specific to the electromagnetic cascades in emulsion cloud chambers. We will see that it is divided in two steps : the reconstruction of the primary track in the first five films and the reconstruction of the branches from the second film to the last one. In Section 2, we briefly remind the experimental set up. The third section is a presentation of the tools used for the data and Monte Carlo (MC) reconstruction. The section 4 is devoted to a precise description of the algorithm. The section 5 presents his performances with a comparison between experimental data and MC simulations.

2 Experimental Set Up

A test beam which had used the T24 electron beam line at DESY was performed in order to study electron identification and shower development in emulsion cloud chambers like in OPERA detector. This beam line allowed to have :

- a pure electron beam without contamination of other type of particles. The low contamination is due to multiple coulomb scattering of electrons along the beam line.
- the chosen particles density, i.e. low density (1 e-/cm^2) or high density (100 e-/cm^2).
- the lowest possible intensity.

Several bricks with different configurations and beam energies were exposed. In this note we decide to perform the analysis and show the results of a brick made of 20 emulsion layers alternating with 1 mm thick-lead plates and exposed at a high density of 6 GeV electrons. A detailed description of the experimental set up is available in the reference [1].

The figure 1 shows the beam angle direction. The mean value for θ_{xz} is 10 mrad and for θ_{yz} is -23 mrad. This slight angle deviation allows a better study of the signal.

3 Analysis tools for MC simulations and data reconstruction

3.1 Data reconstruction

Experimental data come from Neuchatel laboratory which had scanned 6 GeV electron with dry objective. The microtracks were reconstructed on-line by SySal. Only microtracks with at least 6 grains and $|\tan(\theta_{xz})|$ and $|\tan(\theta_{yz})| < 1$ rad are kept. Then for basetrack reconstruction and plate-to-plate alignement we use FEDRA [2], a framework which allows an off-line study. The efficiency of track reconstruction is about 90 %.

In this study we keep basetracks with the following cuts :

- $|\tan(\theta_{xz})| < 400$ mrad and $|\tan(\theta_{yz})| < 400$ mrad, where z axis direction is defined as the perpendicular direction on the transverse plan of the brick.

- $\chi^2 < 0.333 \times n_{grain}$ - 4.343, in order to separate "good " basetracks (with a small χ^2 value and large number of grains) from "background" basetracks (high χ^2 value and a few grains).

The tracking is performed by the dedicated algorithm described in this note.

3.2 Monte Carlo simulations

3.2.1 Signal

The official OPERA framework [3] is used for the MC simulations of electromagnetic showers. The modular structure of the brick is defined in the OpGeom package [4] : each film has a 50 μm thick emulsion layer on both sides of a 200 μm thick plastic base, a 1 mm thick lead plate is placed between 2 films. Then the simulation MC, using OpSim package, is performed by the GEANT3 Virtual Monte Carlo. A couple of hits is produced at the begining and the end of each emulsion layer. The hits on the plastic base are smeared in x,y position and in θ_{xz} and θ_{yz} angle direction in order to have an angular resolution of 2 mrad. Microtraks are reconstructed from the pair of hits of the emulsions. The criteria for basetracks reconstruction are the following :

- $|\tan(\theta_{xz})| < 400$ mrad and $|\tan(\theta_{yz})| < 400$ mrad.
- The absolue value of the angle difference between microtracks and basetracks must be smaller than 100 mrad.

We simulate also an efficiency of 90 %. All the informations concerning a basetrack is set in a tree of ROOT.

3.2.2 Background

A 2 cm^2 area of a non-exposed zone was scanned in order to generate several small zones (1 mm^2 each), called "subzones". The subzones, from real data, are used to simulate the background. The different sources for the background come mainly from the fog, lead radioactivity, ambiant radioactivity, cosmics rays,... In this test beam, the background is much higher than in OPERA. The density of basetracks calculated from the non-exposed area after all cuts is about 65 basetraks/mm^2/film.

4 Algorithm of reconstruction : principle

4.1 Principle

When a high energy electron passes through a thick and dense absorber as lead, it initiates an electromagnetic shower ([5] and [6]) by bremsstrahlung and pair production. The two latter phenomena generate secondary particles with lower energy than the energy of the primary electron.

The algorithm of shower reconstruction, developped in the ROOT framework, is divided in two main steps :

criteria	$\chi^2_{btk/btk}$	Δr (μm)	Δθ (rad)	σ_x (μm)	σ_y (μm)
1.	≤ 5	≤ 30	≤ 0.050	≤ 30	≤ 30
2.	≤ 10	≤ 48	≤ 0.086	≤ 40	≤ 40
3.	≤ 15	≤ 90	≤ 0.150	≤ 40	≤ 40

Table 1: Table of criteria for the connection of consecutive basetracks

1. the first one is the reconstruction of the track of the primary electron which generates the electromagnetic cascade. This track is called "the primary track". The algorithm starts from the first film and proceeds to the tracking until the fifth film.

2. the second step consists in linking basetracks produced by the secondary charged particles as electrons and positrons contained in the same cylinder as for the primary track. The algorithm build the branches by starting basetrack connections in the second film until the twentieth film.

During the propagation, from the basetrack of the film (i), we predict the position of the basetrack(s) to connect in the film(i+1). Then the algorithm searches basetrack canditate(s) around the predicted position (σ_x and σ_y) with σ_x = (x_{pred} - $x_{candidate}(i+1)$) and σ_y = (y_{pred} - $y_{candidate}(i+1)$) . The algorithm keeps the basetracks if they satisfy the cuts defined on the following quantities, illustrated in the figure 2 :

1. a quality variable : $\chi^2_{btk/btk} = \frac{1}{2}$ ($\frac{\Delta\theta_x^2(i)}{\sigma_i^2}$ + $\frac{\Delta\theta_y^2(i)}{\sigma_i^2}$), with $\sigma_i = 0.014$ rad, defined from the MC simulations.

2. the position displacement $\Delta \text{r} = \sqrt{\Delta x^2(i) + \Delta y^2(i)}$

3. the angular difference $\Delta\theta = \sqrt{\Delta\theta_x^2(i) + \Delta\theta_y^2(i)}$

The table 1 presents the values of the cuts on $\chi^2_{btk/btk}$, Δr and Δθ. There are explicitly three criteria : criteria 1 and 2 concern the primary track whereas criterion 3 is applied for the branch reconstruction.

- criterion 1 : connection between 1 basetrack in film (i) and 1 basetrack in film (i+1).

- criterion 2 : this criterion is applied in order to take into account possible holes in the primary track. If the prediction or the condition of connection are not satisfied in two consecutive films (i) and (i+1), the algortihm constructs a virtual basetrack in the film (i+1) with same angles, χ^2 value, number of grains of the basetrack (i). But the position coordinates (x,y,z) of the "virtual basetrak" are calculated from the prediction, that is to say the direct projection of the coordinates of the basetrack (i) in the film (i+1). Then from this virtual basetrack, in (i+1), we search and try to connect "real" basetrack in film (i+2). This criterion connects actually basetrack in film (i) and basetracks in film (i+2). The virtual basetracks are not kept for the analysis as there are not "real" basetracks. They are considered only as a tool for the reconstruction.

- criterion 3 : connection between 1 basetrack in film (i) and 1 basetrack in film (i+1) for the reconstruction of the shower branches.

4.2 Reconstruction of the primary track

The main idea is to build the track created by the primary electron. For this reason, the algorithm tries to connect consecutive basetracks in the first five films. This part is divided in three steps. The main goal is to decrease significantly the number of combinations and so the time computation.

4.2.1 Basetrack filtering

The algorithm considers a basetrack in the first film as the starting point for the primary track. It opens a cylinder which direction axis is the same as the first basetrack. The radius (in μm) can be selected by the user. Then the algorithm tries to connect the first basetrack with all the possible basetracks in the film (2) which satisfy criterion 1. Next, it propagates itself until the film (5) and connects basetracks in film (i) with basetracks in film (i+1) with criterion 1. In this step, there isn't tracking yet : the algorithm makes only couples. Indeed a basetrack in film (i) can be assembled with several basetracks of a film (i+1). Finally, the main purpose is :

- to proceed to a "pattern recognition".
- to exclude a maximum of "fake basetracks" in order to decrease the number of combinations.

At the end of the step 1, the basetrack candidates are stored in a list of ROOT.

4.2.2 Reconstruction of the candidate tracks

The step 2 of the first main step consists in reconstructing all the possible tracks composed of 4 or 5 basetracks. So the tracking starts in this step.

1. The algorithm starts with the first basetrack. Then it tries to connect the first basetrack with a basetrack in film (2), stored in the previous list of root.
   - if there is a connection with one "real" basetrack, a track with 2 real basetracks is built.
   - if there isn't basetrack, the algorithm creates a virtual basetrack in the film (2). Indeed it allows a hole because scan efficiency is different from 100 %.
2. Once it creates tracks with 2 basetracks, the algorithm propagates itself in the third film and tries to connect basetracks of this film with a basetrack in film (2) of a track composed of 2 basetracks by applying criteria 1 or 2 (it depends on the nature of the last basetrack : virtual or not).
3. Then the algorithm continues propagating itself until the film (5) and tries to connect basetrack in film (i+1) with the last basetrack of a track in film (i). If it doesn't find a candidate in film (i+1), it makes a virtual basetrack in (i+1) and tries to search another basetrack in film (i+2) by applying criterion 2. Thus all the possibities are scanned.
4. Finally only tracks with 4 or 5 basetraks are stored in a list of ROOT and considered as potential primary tracks. Tracks with 4 basetraks contains one hole.

4.2.3 Track cut χ^2_{tk} and selection of the primary track

We define a new criteria χ^2_{tk} to select the primary track among all the candidate tracks. It is defined as $\chi^2_{tk} = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} \left(\frac{\Delta\theta_x^2(i)}{\sigma_i^2} + \frac{\Delta\theta_y^2(i)}{\sigma_i^2}. \right)$, N being the number of basetracks in a track.

The track with the lowest χ^2_{tk} value is considered as the primary track and the figure 3 shows a comparison between data and MC simulations for the χ^2_{tk} distributions.

4.2.4 Summary

To sum up, the algorithm starts with the first basetrack in film (1) and scans all the possible combinations to connect consecutives basetracks contained in a cylinder until the fifth film which satisfy criteria 1 or 2. Then it holds an ultimate track with the lowest χ^2_{tk} value and composed with 4 ou 5 basetracks : the primary track.

This algorithm is applied for each basetrack in the first emulsion and all the reconstructed primary tracks are stored in a tree of ROOT.

4.3 Reconstruction of the branches of the shower

The second main step of the algorithm consists in reconstructing the branch tracks produced by the secondary charged particles of the electromagnetic shower.

1. The algorithm opens a cylinder which direction axis is the same as the direction of the first basetrack of the primary track.

2. Inside the cylinder, it tries to connect basetracks in film (i) with basetracks in the next film (i+1) by applying criterion 3. The procedure is repeated from the film (2) until the last film.

Thus, the algorithm is able to reconstruct tracks produced by high and low energy particles (a few tens of MeV) and therefore the branches of the cascades. The criterion 3 allows to integrate a maximum of signal and minimize the connections with background ("fake basetracks"). Indeed, the values of the cuts are adapted to the physic of the particles in an electromagnetic shower (scattering,...).

Finally basetracks forming the primary track and the branches are stored in a tree of ROOT.

5 Performances of the algorithm

5.1 Data treatment and data analysis

Performances of the algorithm can be seen by comparing results from data and results from Monte Carlo simulations. As mentionned before, data come from analysis of 6 GeV high density electron exposure. But there are 2 important sources of errors :

- the first one is an important background due to a high exposure to the cosmic rays and a long time storage before and after the exposure at DESY.

- the second one is the overlapping of the showers due to the high density of the exposure (about 100 e^-/cm^2). But in the analysis we choose isolated electromagnetic showers, i.e electromagnetic showers distant enough each other and we limit our study in 400 μm radius cylinder. This algorithm cannot separate two or more showers.

E (GeV)	1	2	3	4	5
efficiency	$\frac{1461}{2000} = 73.1\%$	$\frac{1606}{2000} = 80.3\%$	$\frac{1734}{2000} = 86.7\%$	$\frac{1745}{2000} = 87.3\%$	$\frac{1706}{2000} = 85.3\%$
E (GeV)	6	7	8	9	10
efficiency	$\frac{1781}{2000} = 89.1\%$	$\frac{1753}{2000} = 87.7\%$	$\frac{1780}{2000} = 89.0\%$	$\frac{1791}{2000} = 89.6\%$	$\frac{1797}{2000} = 89.9\%$

Table 2: Efficiency of the reconstruction algorithm

E (GeV)	1	2	3	4	5	6	7	8	9	10
efficiency	0.335	0.346	0.355	0.361	0.368	0.371	0.388	0.377	0.380	0.384

Table 3: ratio of the number of basetracks associated by the algorithm over the total number of basetracks. The evaluation is done by MC simulations with 2000 electrons in a full brick.

5.2 Efficiency of the algorithm

We must evaluate the efficiency of the algrithm to see the validity of the cut values on the connection variables. The table 2 gives the efficiency of the algorithm. The efficiency is defined as the ratio between the number of events which survive on the total number of simulated events. These efficiencies can be increased by tacking only 3 or 4 basetracks for the primary track. But we choose 5 basetracks in order not to reconstruct fake primary tracks.

Then, we show in the table 3 the ratio of the number of basetracks associated by the algorithm in a full brick over the total number of basetraks created by the signal and kept after the cuts. The values presented in the table 1 are a compromise to keep a maximum of basetracks from the signal and include a minimum of fake basetracks.

5.3 Criteria of connection

All the plots show the mean distributions both for the data and MC simulations. As a convention, we choose to draw : in red, the distributions for Mont Carlo simulations with background and in blue, the data distributions without substraction of the background.

The figure 4 shows the positional displacement Δr distributions. We observe a small deviation between data and MC distribution, a few μm (see lower plots). The simulations with GEANT3 can't take into account all the defaults of the plate-to-plate alignement and the reproduction of the emulsion layers.

The angular displacement $\Delta\theta$ and the $\chi^2_{btk/btk}$ distributions are reported respectively in the figures 5 and 6. The comparision between data/MC reveals a good agreement. The angular displacement is not affected by the degradation of the alignement between data and MC simulations.

5.4 Views of reconstructed showers

The figure 7 presents 3 examples of reconstructed showers produced by 6 GeV electrons in an emulsion cloud chamber made of 20 films. One point on the figure represents a basetrack. The dimensions of the 3 axis are in μm. The beam comes from the left. On these examples, we can see the long primary track done by the primary electron. The shower development starts after the primary electron crosses a few number of plates. The next stage is to extract the mean longitunal and lateral profiles.

5.5 Profiles of the electromagnetic cascade

Other aspects of the algorithm performances can be addressed by comparing data/MC reproductions of the mean longitudinal and transversal profiles. The mean data profiles are calculated with 172 events. As a consequence, the error bars include important statistic fluctuations (about 8%) and systematic fluctuations due the background.

For this plot, we decide to choose the following convention :

- in red : MC simulations without background.
- in blue : data with background substraction. The substraction and the treatment of the background are explained in details in the reference [7].
- in green : the mean background.

The figure 8 shows the mean longitudinal profile of an electromagnetic cascade. The figure 9 presents the mean transversal profile.

In the reference [7], the mean longitudinal and the mean transversal profiles are described and explained with more details.

6 Conclusion and perspectives

A specific algorithm has been developped to reconstruct tracks of an electromagnetic cascade in an emulsion cloud chamber of OPERA. It is divided in two main steps : the reconstruction of the primary track and the reconstruction of the branches of the shower. A comparison beetween data and MC simulations shows a correct agreement except for the positionnal displacement Δr. This study can be completed and improved for other values of the electron energy and with a test beam with much lower background and low density exposure (like OPERA conditions), wich implies lower systematic fluctuations. This algorithm will be used for the reconstruction of the energy of the primary electron [7] and another complementary algorithm, a “scan back algorithm” is under development.

References

[1] D.Autiero, Y.Caffari, L.S. Esposito, A. Marotta, J. Marteau and P. Migliozzi, Characterization of T24 the electron beam line available at DESY, April 7 2004, OPERA internal note.

[2] http://nts-lab01.na.infn.it/fedra/

[3] http://operaweb.web.cern.ch/OperaWeb/internal/exchanger/home/software/ documentation/OpRelease/ (internal section of the opera home page)

[4] L. Chaussard, OpGeom, software model of the opera detector, July 1 2005, OPERA internal note.

[5] Particle data group, Particle Physics Booklet, July 2004.

[6] B. Slowinski, Phys. Part. Nucl. 25 (2), March-April 1994

[7] Y.Caffari, Neural Network for electromagnetic shower energy reconstruction, April 2006, OPERA internal note.

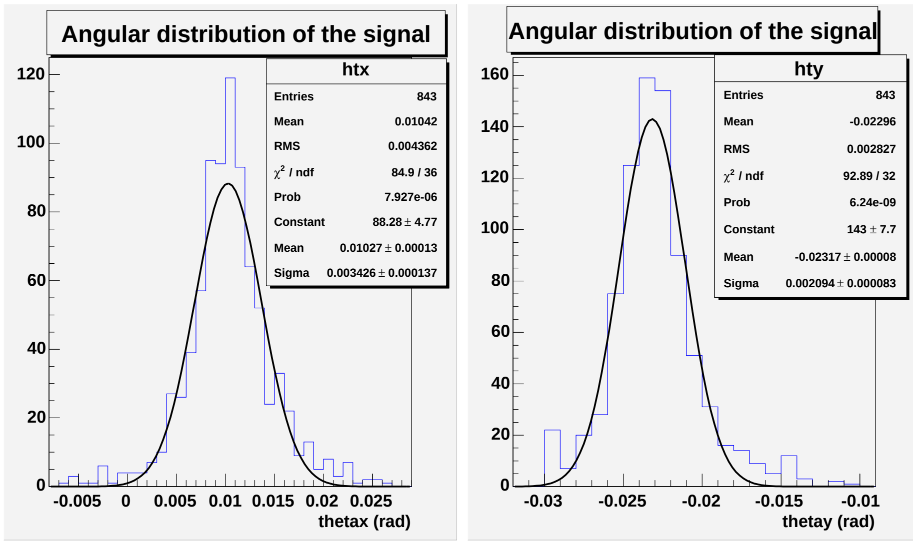

Figure 1: Angular distributions of the signal

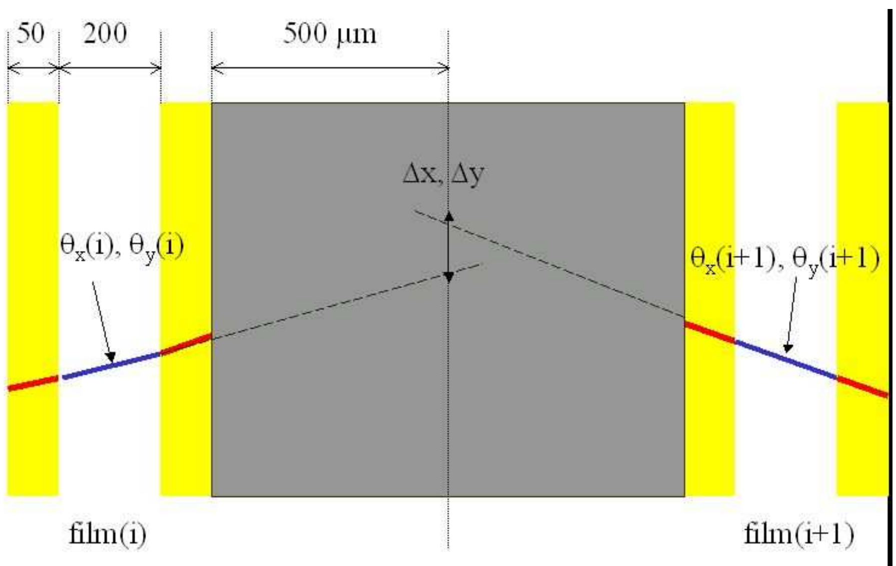

Figure 2: Schematic view of the connection of two consecutive basetracks

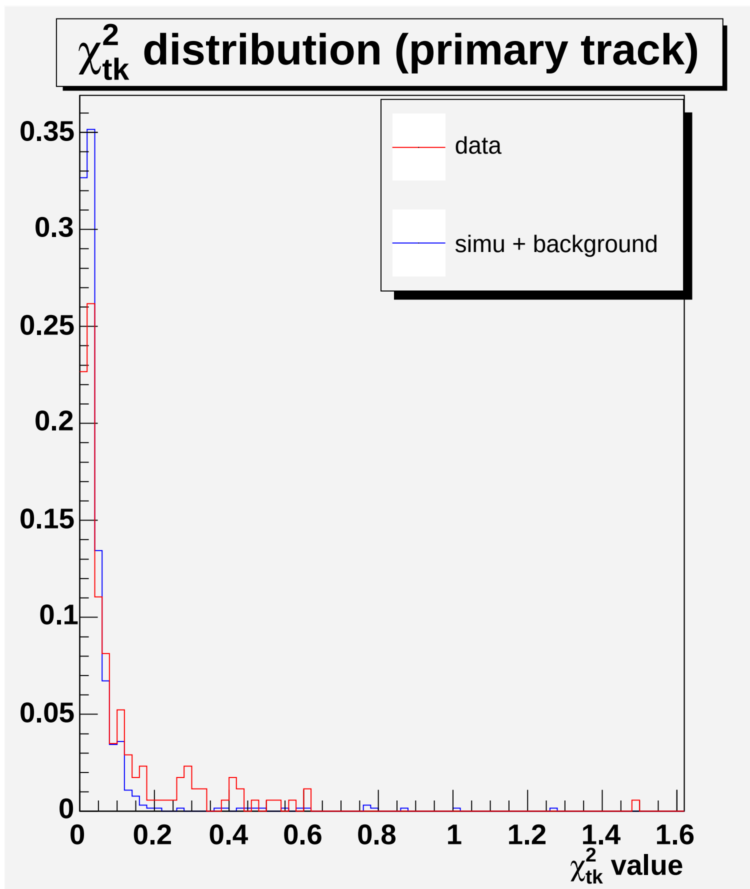

Figure 3: χ^2_{tk} distributions (primary tracks)

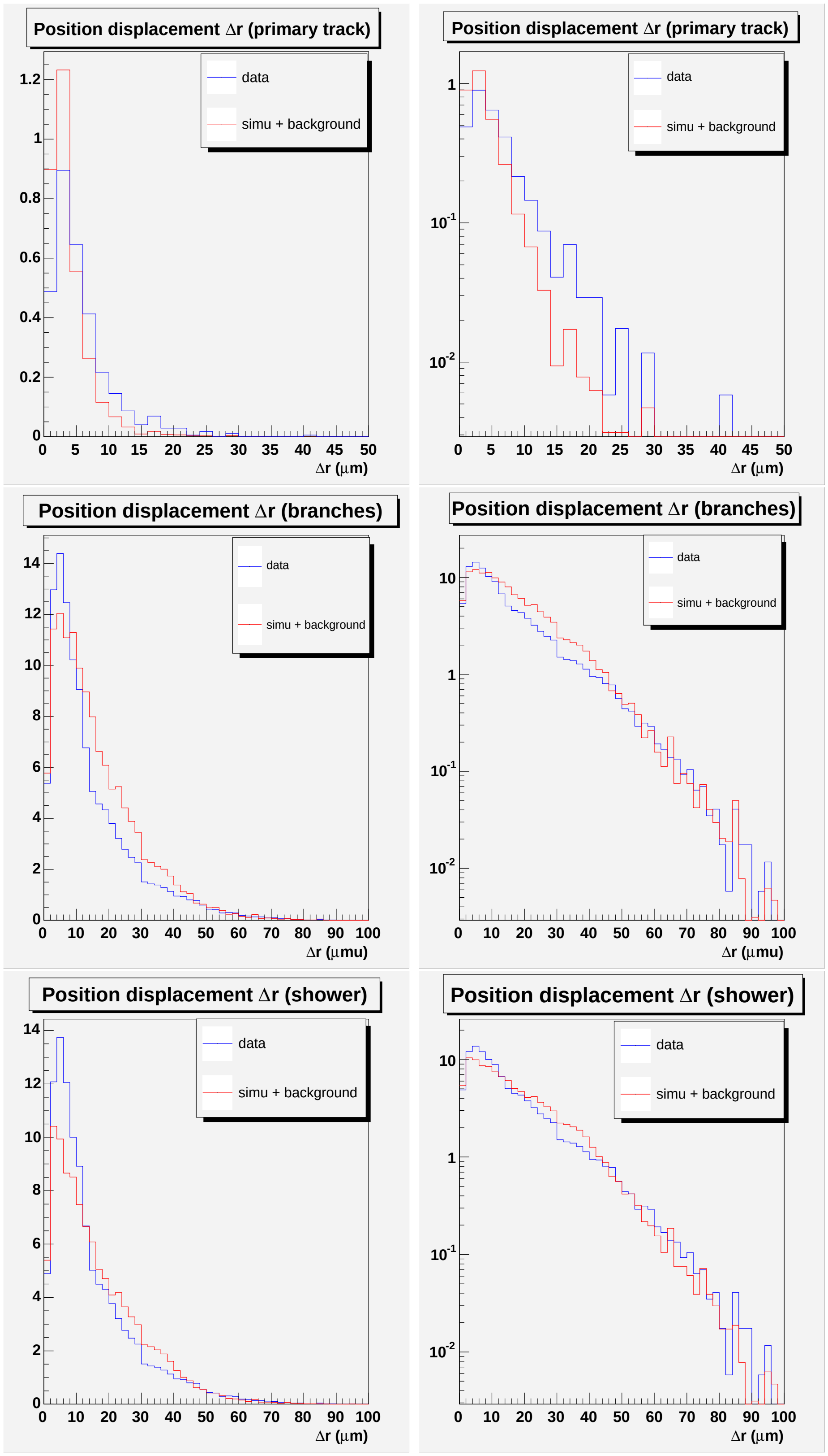

Figure 4: Δr distributions. Left plots : linear scale. Right plots : logarithmic scale. The upper plots show distributions for the primary track. The middle plots present results concerning the branches. The lower plots concerns the shower (primary track + branches).

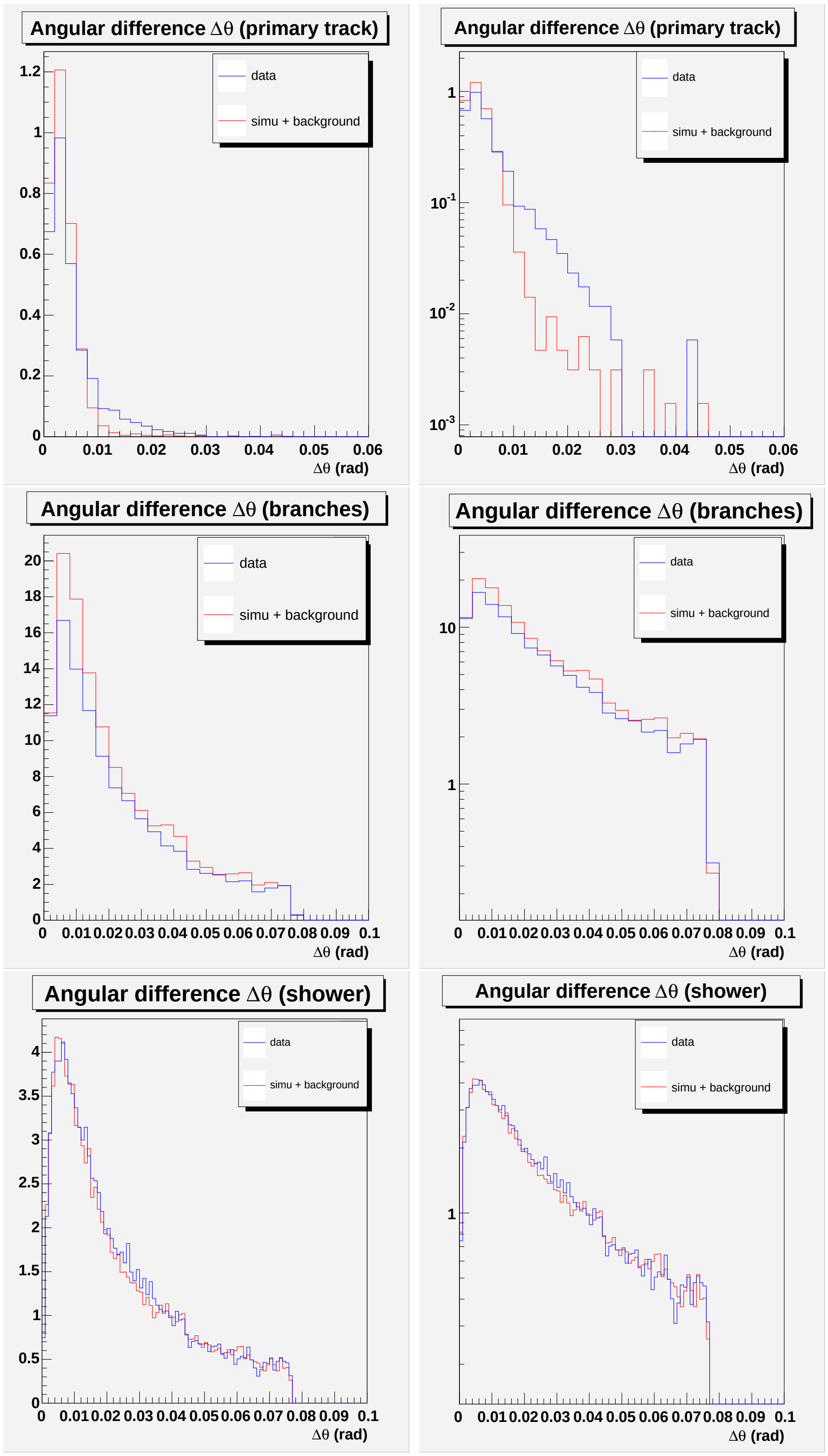

Figure 5: $\Delta\theta$ distributions. Left plot : linear scale. Right plot : logarithmic scale. The upper plots show distributions for the primary track. The middle plots present results concerning the branches. The lower plots concerns the shower (primary track + branches).

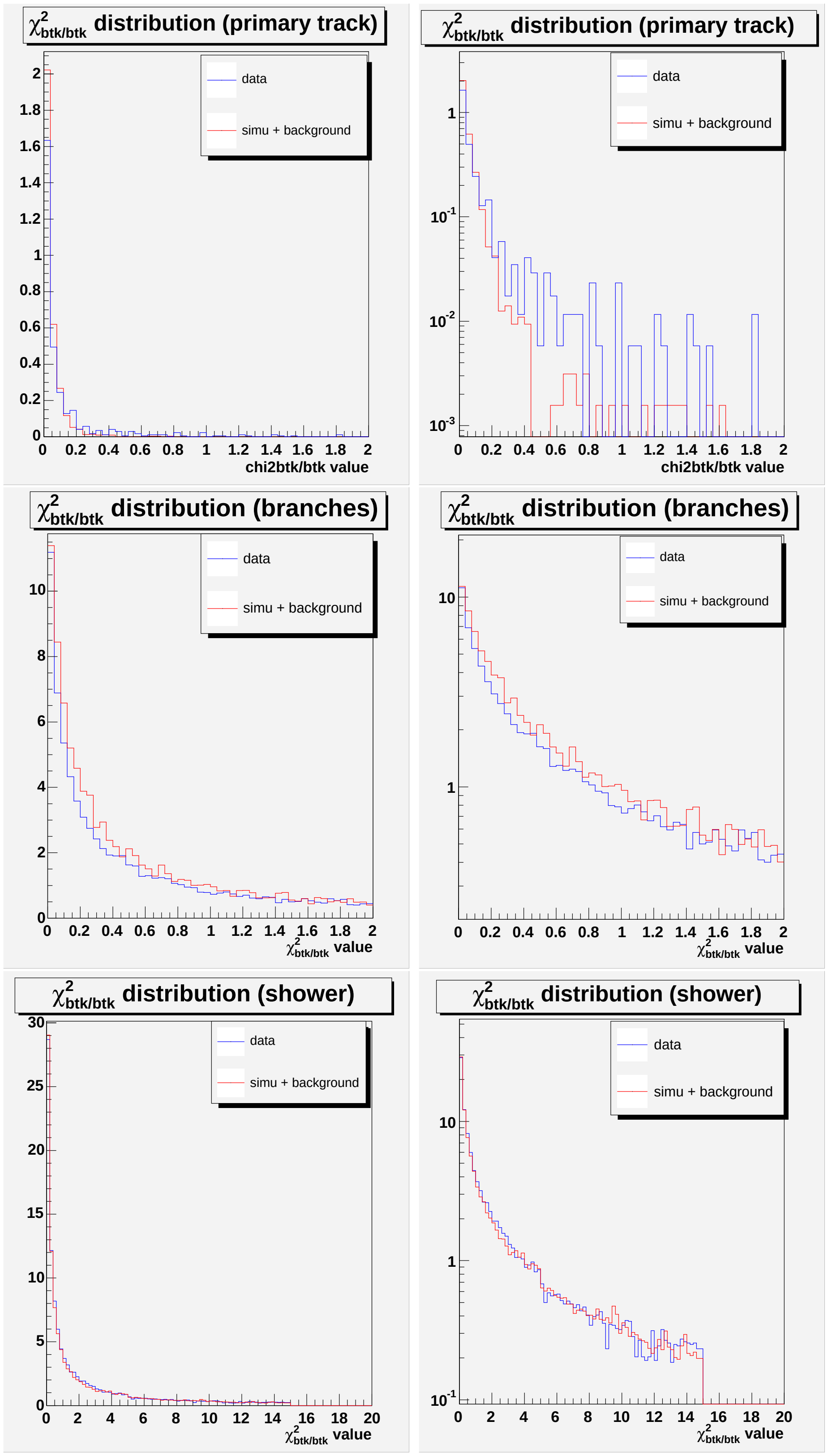

Figure 6: $\chi^2_{btk/btk}$ distributions. Left plot : linear scale. Right plot : logarithmic scale. The upper plots show distributions for the primary track. The middle plots present results concerning the branches. The lower concerns the shower (primary track + branches).

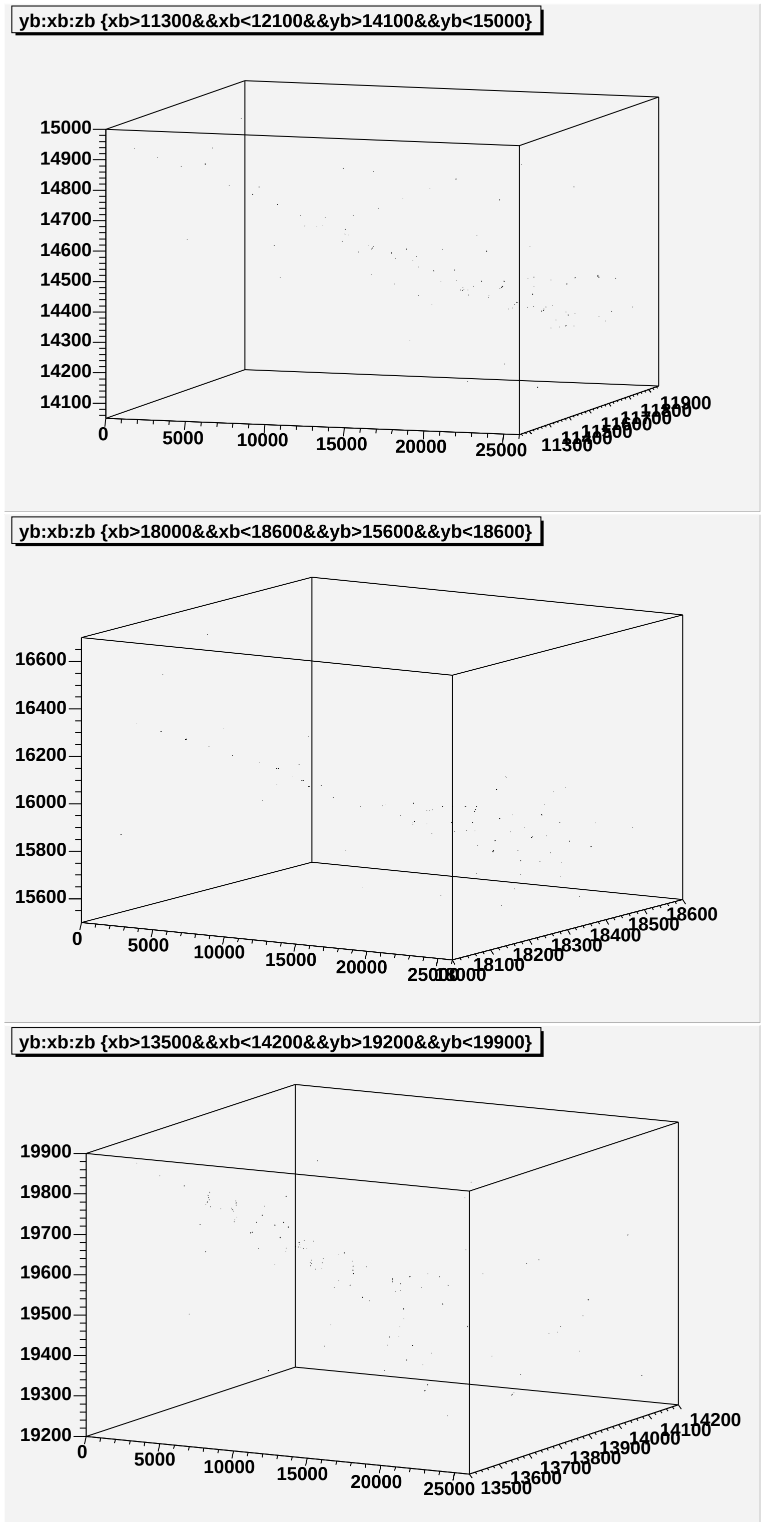

Figure 7: 3D views of reconstructed electromagnetic showers created by 6 GeV electrons. The beam comes from the left. The z-axis is defined by the direction of the beam and the plan define by x and y axis is the same as the plan define by a film. The dimensions are defined in μm. A point represents a basetrack.

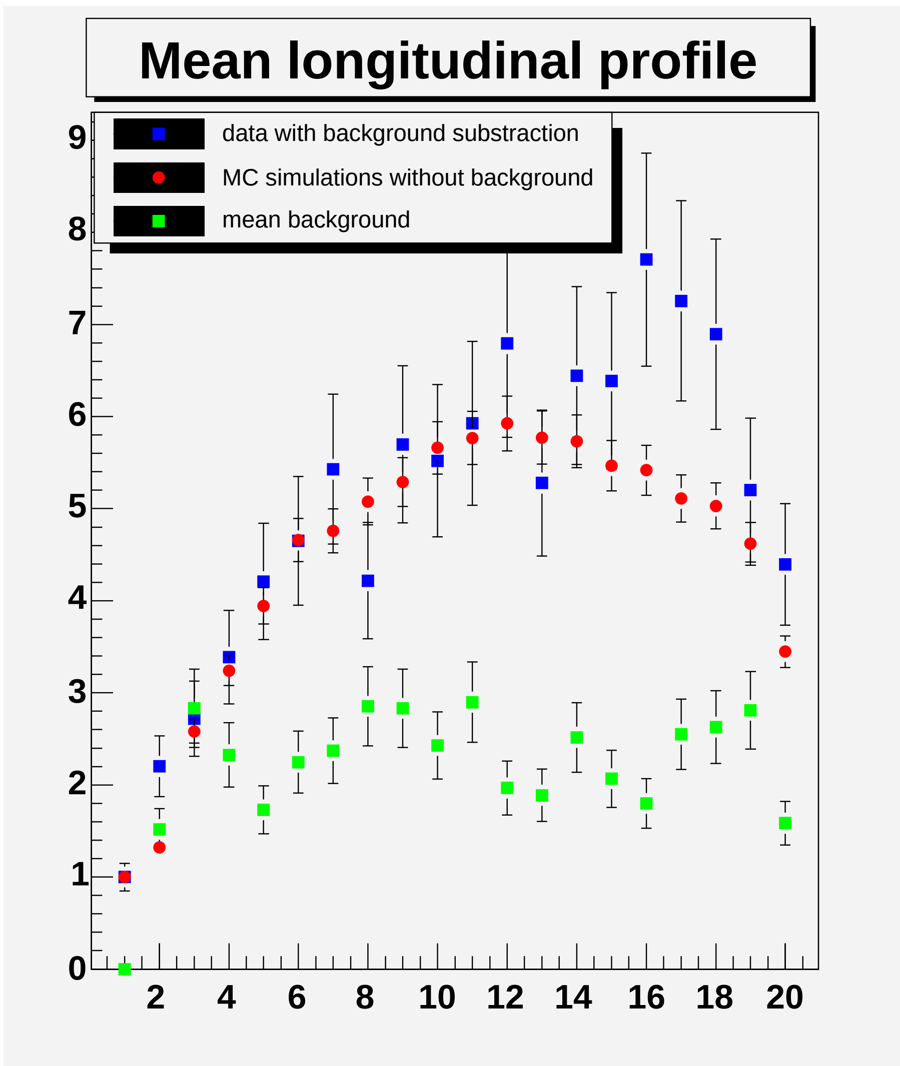

Figure 8: Mean longitudinal profile of the electromagnetic cascade. The x-axis represents the emulsion number and y-axis the mean number of basetracks per film.

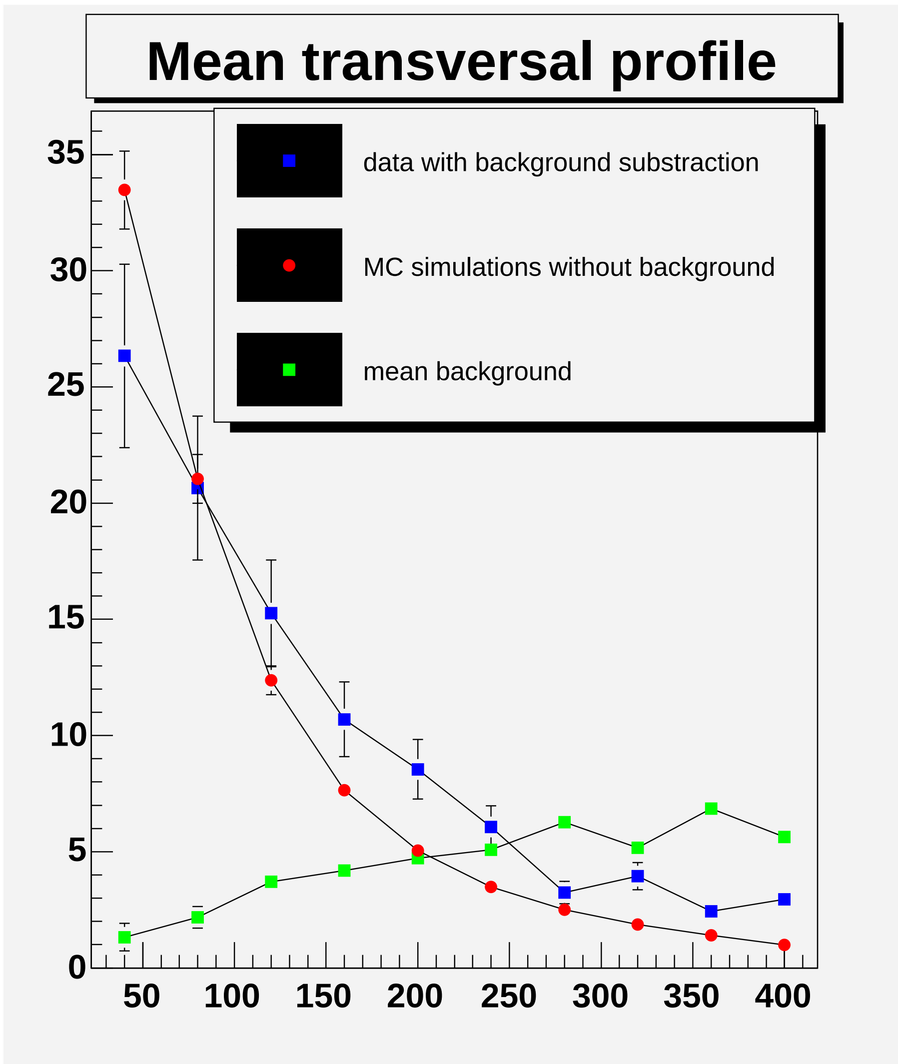

Figure 9: Mean transversal profile of the electromagnetic shower. The x-axis represents the radius r (μm). The y-axis represents the mean number of basetracks per bin. Each bin represents the number of basetracks contained between two cylinders which the difference of radius is equal to 40 μm.