DESY Summer Student Programme 2011

Performance of bremsstrahlung refits for converted photons

Luigi Longo
Supervisors: Kerstin Tackmann, Marcos Jimenez Belenguer
A thanks to Thomas Koffas for the help given.

Goal: comparison and study of the performance of the following track fitters

•Global χ^2 track fitter (default): the global fit uses a particular χ^2 form that considers also the energy losses; at this point, given an initial estimate of the track parameters, it tries to minimize the χ^2, for every hit added, updating the parameters until there's no appreciable difference between the new and the old χ^2 .

•Gaussian Sum Filter (GSF): it is a non-linear generalization of Kalman filter which works only if we have silicon hits; it approximates the Bethe-Heitler distribution by a weighted mixture of several gaussians in order to take into account the asymmetry and the low-energy tail of the distribution. For every component a different KF is executed and the results are merged together, using different weights (they're different from the previous ones), to obtain the track parameter vector and its covariance matrix. In order to avoid exponential growth of components, it's performed a component reduction at each step.

•Dynamic Noise Adjustment (DNA): At each silicon layer, a single-parameter fit is performed to estimate the increase in curvature due to possible bremsstrahlung at the current detector layer. If no bremsstrahlung is found, the KF is called. Otherwise, before calling tha KF, the result of the fit (the estimated fraction of energy reteined by the electron,Z) is used to redefine the Bethe-Heitler as function of a gaussian distribution and of the deviation of the estimated Z from the median Z_0 .

!!! All the brem refits are performed as part of the standard egamma reconstruction

Sample:

- DP samples
  mc10b_7TeV_DP17, _DP35, _DP70, _DP140, _DP280;

DP
- [20,45] GeV for _DP17
- [45,85] Gev for _DP35
- [85,200] GeV for _DP70
- [200,400] GeV for _DP140
- [400,1000] GeV for _DP280

- JF samples
  mc10b_7TeV_JF17, _JF35, _JF70, _JF140, _JF240;

JF
- [20,45] GeV for _JF17
- [45,85] Gev for _JF35
- [85,200] GeV for _JF70
- [200,400] GeV for _JF140
- [400,1000] GeV for _JF240

- Data samples
  data11_7TeV_178044, _184130;

[20,1000] GeV for the data

!!!I've considered a particular pt range for every DP samples and JF samples in order to not have overlap between the samples

Selection Cuts:
- select double-track conversions;
- good eta range: 0.<|η|<1.37 || 1.52<|η|<2.37;
- request for photons that are reconstructed by the standard photon reconstruction (author==4) or also as electrons (author==16);
- photon and OQ cleaning;
- tight selection;
- isolation cut (EtCone40 < 5000);
- Require a successful refit with both brem refitters (GSF and DNA);

DP samples:
- truthmatch cut: photons from hard process;

JF samples: from the JF I've selected 3 different sample:
- JF without truthmatch cut (I'll continue to call it only JF);
- JF with truthmatch (JFwith);
- JF with antitruthmatch (background);

!!! In every MC sample all the events are weighted in order to reproduce the same pt data shape: for every bin the weights have been estimated as the ratio of the data events over MC events.

Step 1:

-using the variables

- ph_cl_E=energy reconstructed with the calorimeter,
- ph_conv_p=momentum reconstructed by the ID using the default fitter,
- ph_conv_GSF=momentum reconstructed by the ID using the GSF fitter,
- ph_conv_DNA=momentum reconstructed by the ID using the DNA fitter,

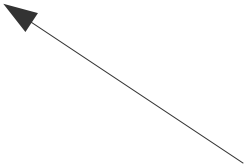
it's possible to perform the fit of the quantities ph_cl_E/ph_conv_p, ph_cl_E/ph_conv_GSF_p, ph_cl_E/ph_conv_DNA_p for different eta and pt regions using the Crystal Ball Function;

Pt regions

DP/data	JF/JFwith
[25,35] GeV	[25,35] GeV
[35,45] GeV	[35,45] GeV
[45,55] GeV	[45,55] GeV
[55,70] GeV	[55,400] GeV
[70,85] GeV	
[85,100] GeV	
[100,125] GeV	
[125,150] GeV	
[150,400] GeV	

Eta regions

DP/data	JF/JFwith
[0,0.6]	[0,0.6]
[0.6,1.37]	[0.6,1.37]
[1.52,1.81]	[1.52,1.81]
[1.81,2.37]	[1.81,2.37]

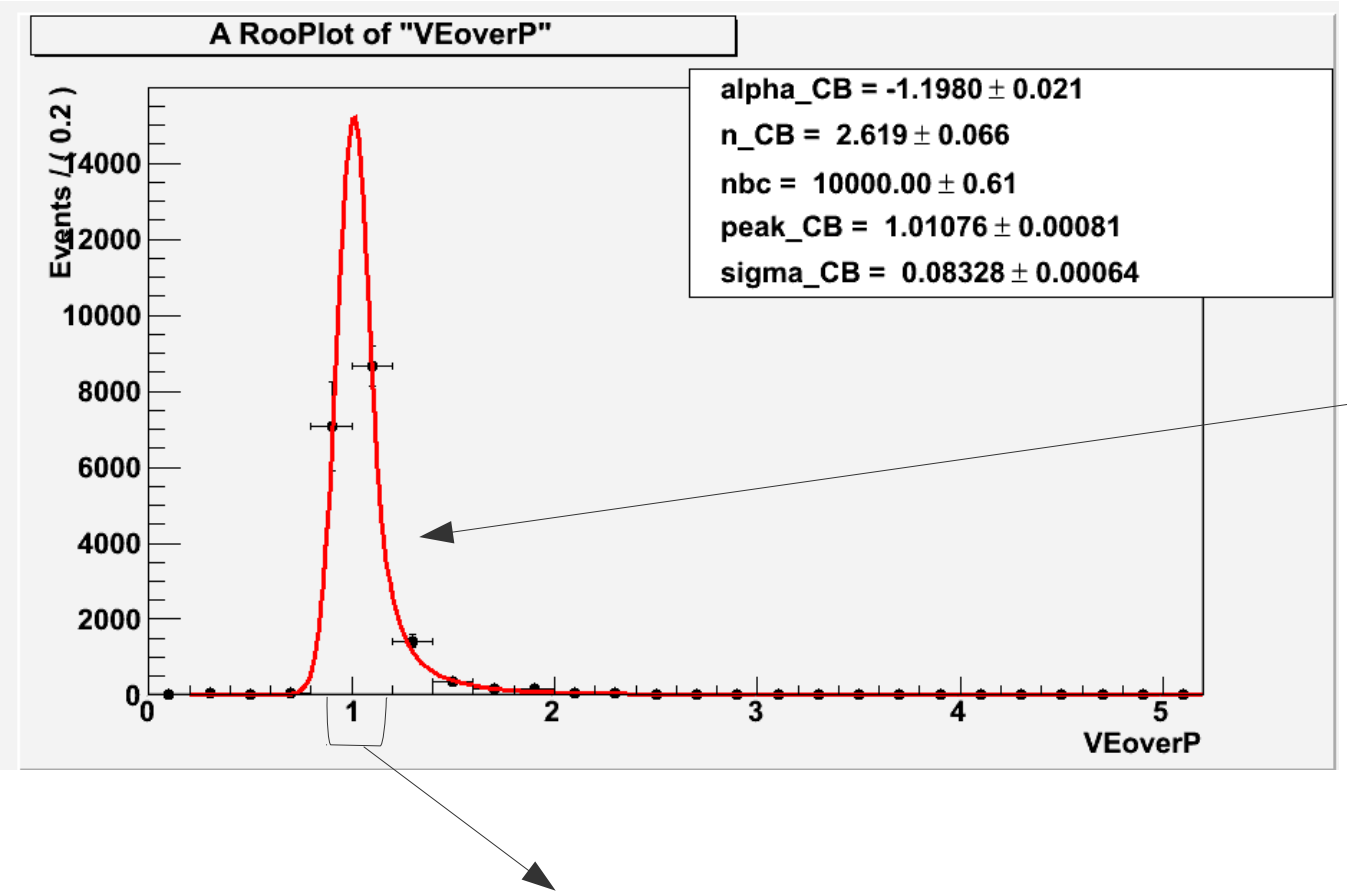


Since we have not so much statistic for the sample JF/JFwith, we have less pt regions.

Step 2:

- measure of the #events in the tail over the # all events for the different eta and pt regions;

DP sample: pt>25 GeV && pt<35 GeV && 0<|η|<0.6



Starting point of the tail is given by peak_CB-alpha_CB.*sigma_CB

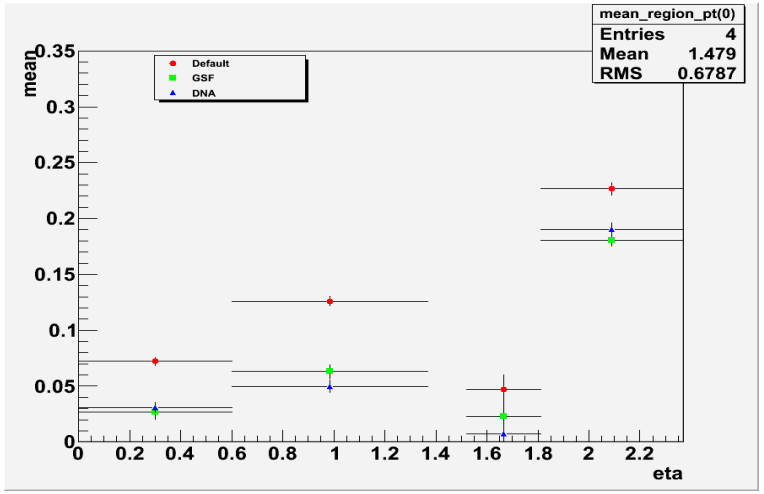
It's possible also to consider the #events in the core($\pm 2\sigma$) over #all events where the gaussian core is defined as the region between peak_CB-2*sigma_CB and peak_CB+2*sigma_CB

!!! We have used it only to check our results

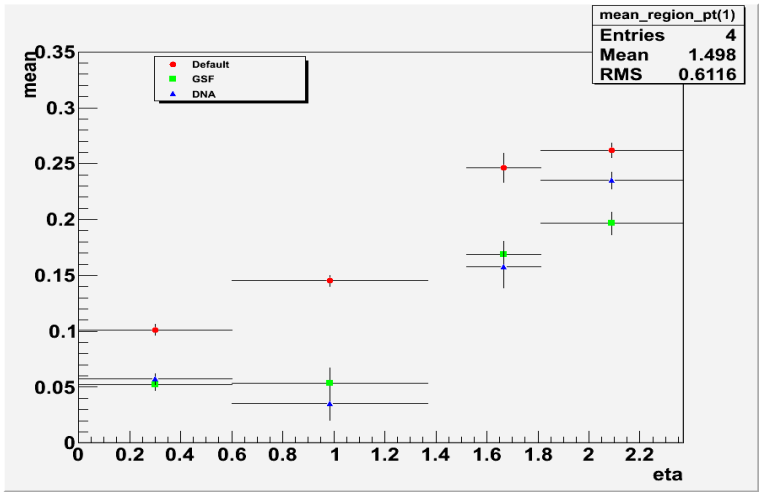
Resolution:

- using only the DP sample, for every pt and eta bin it has been estimated the mean of the resolution $\Delta = \frac{E_{truth} - p}{E_{truth}}$, with its error, to understand how the different track fitters work.

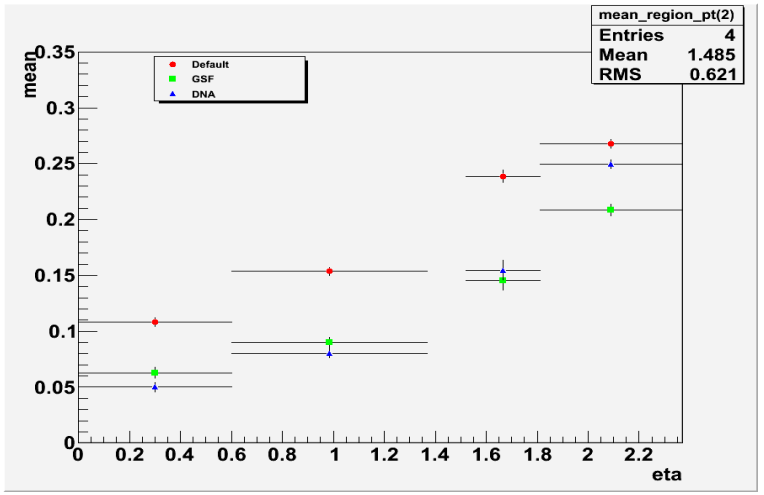
pt>25 GeV && pt < 35 GeV



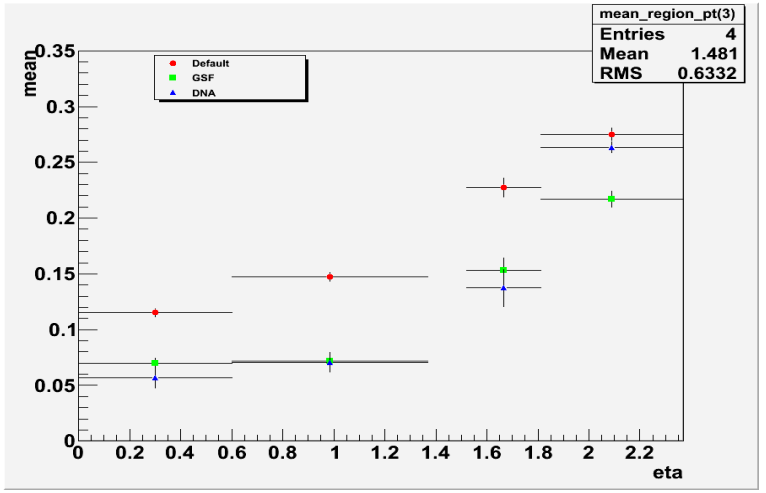
pt>35 GeV && pt < 45 GeV



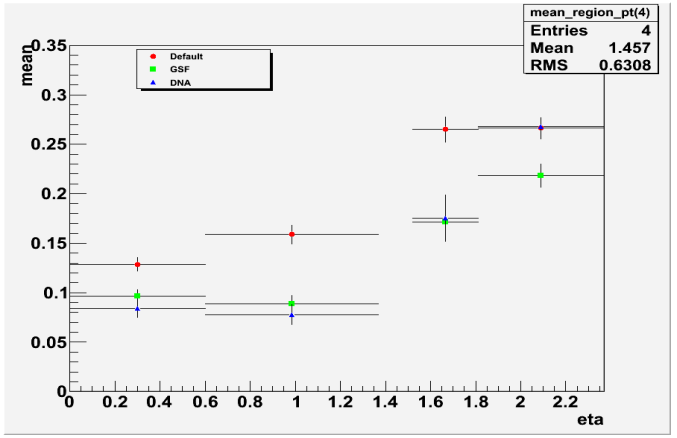
pt>45 GeV && pt < 55 GeV



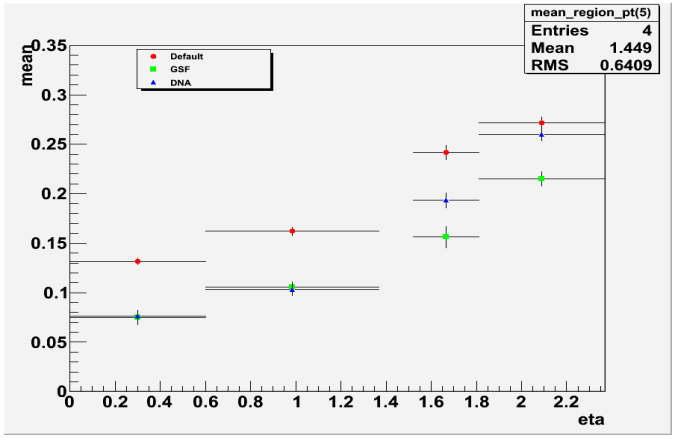
pt>55 GeV && pt < 70 GeV



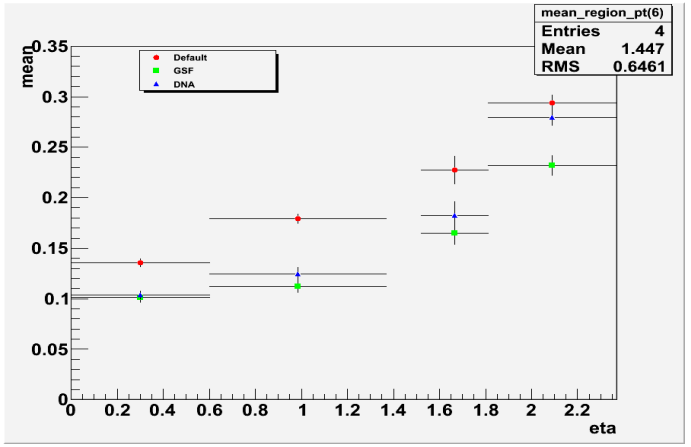
pt>70 GeV && pt < 85 GeV



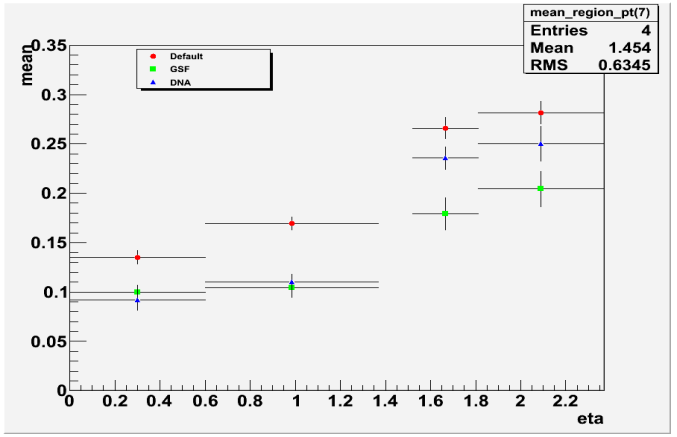
pt>85 GeV && pt < 100 GeV



pt>100 GeV && pt < 125 GeV



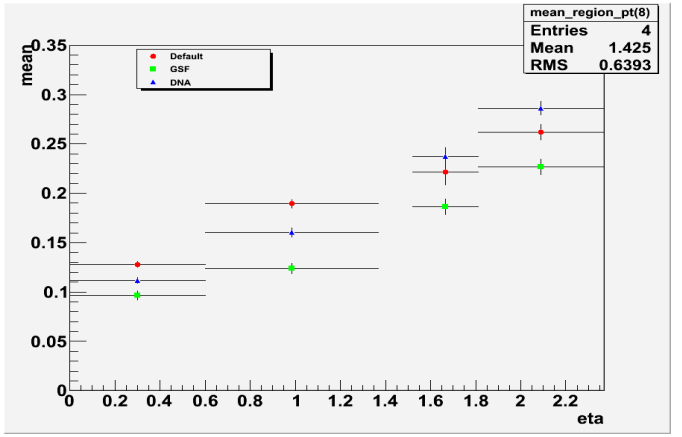
pt>125 GeV && pt < 150 GeV



Both GSF and DNA have a better resolution than the default algorithm

The GSF algorithm becomes better than the DNA rising pt and/or eta

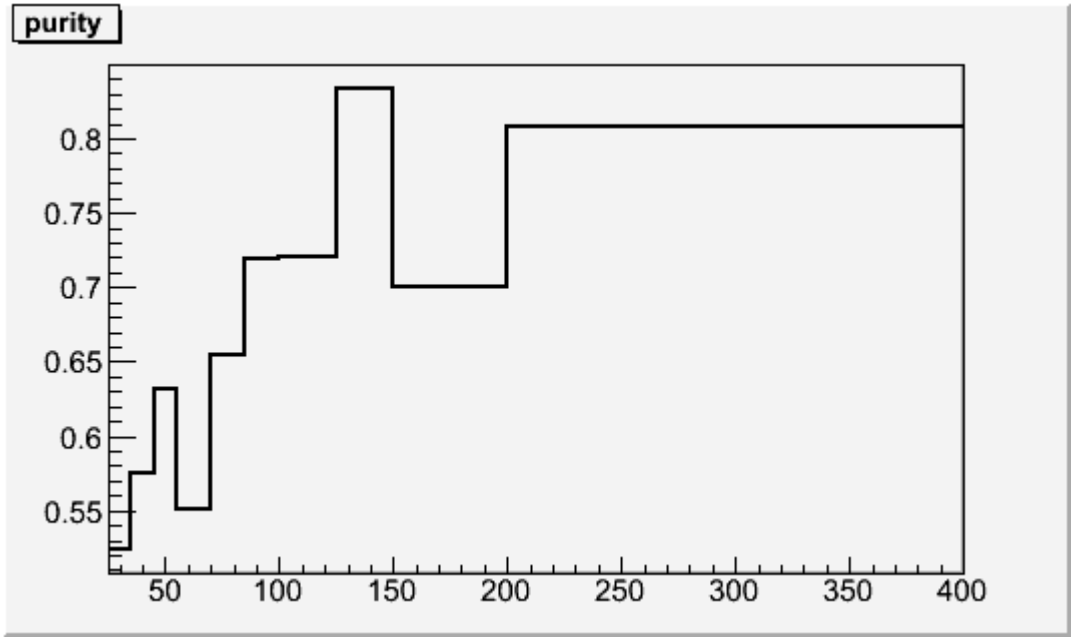
pt>150 GeV && pt < 400 GeV



!!! it's important to remember that we are working with a MC sample, so in principle we cannot exclude that the behavior is a bit different for the data sample

Background:

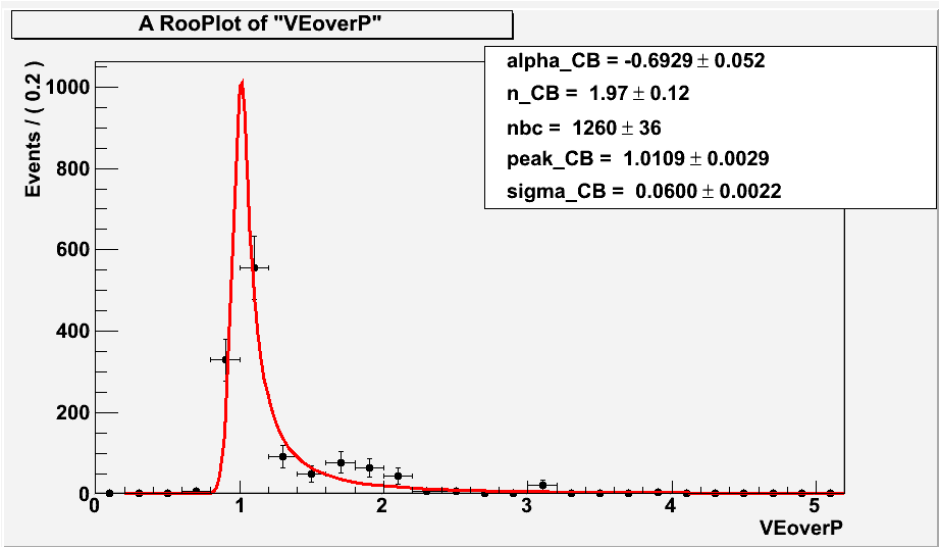
- it's useful using the JF sample to get an idea of the effect of background on our work; the purity of this sample, in fact, is far from the 100% as you can see in the following histogram



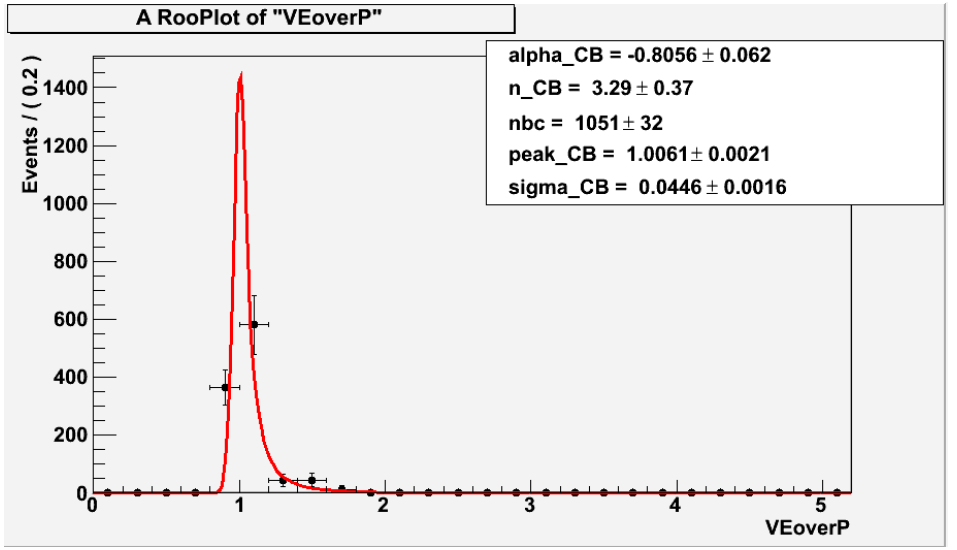
- !!! The purity has been seen to be higher in data than JF in previous studies;
- !!! In this case it has been considered also the pt bin [20,25] GeV;
- The antitruthmatch has been used to get the background;
- The default algorithm has been used for the reconstruction;

pt>25 GeV && pt< 35 GeV && 0.<|η|<0.6

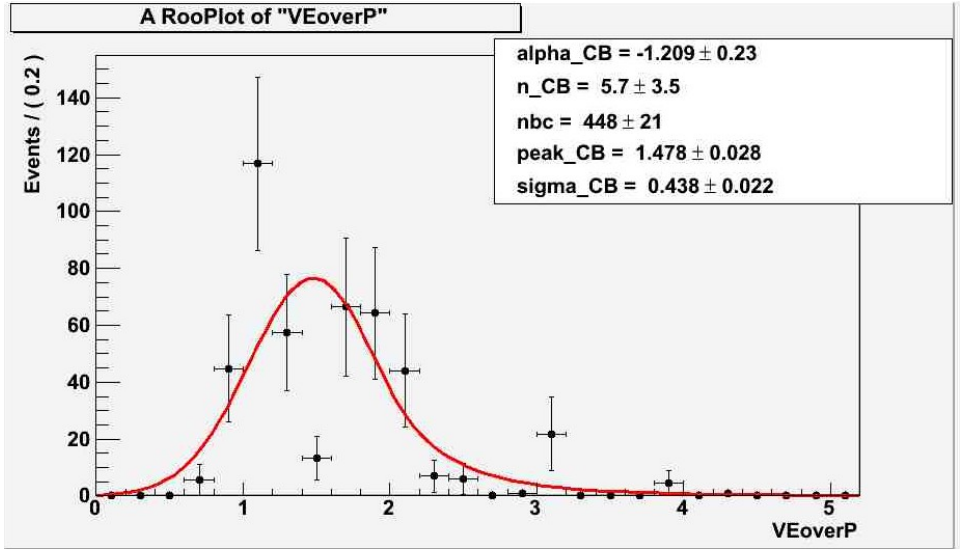
JF sample



JF sample with truthmatch



background

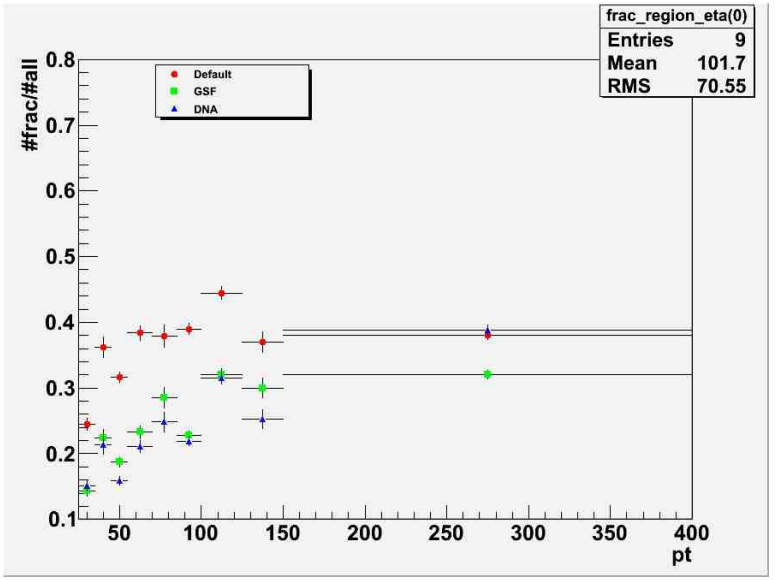


Background consequence:
- shift of the peak;
- increase of the sigma;
- increase of the events in the tail;

!!!It's true for every pt and eta region

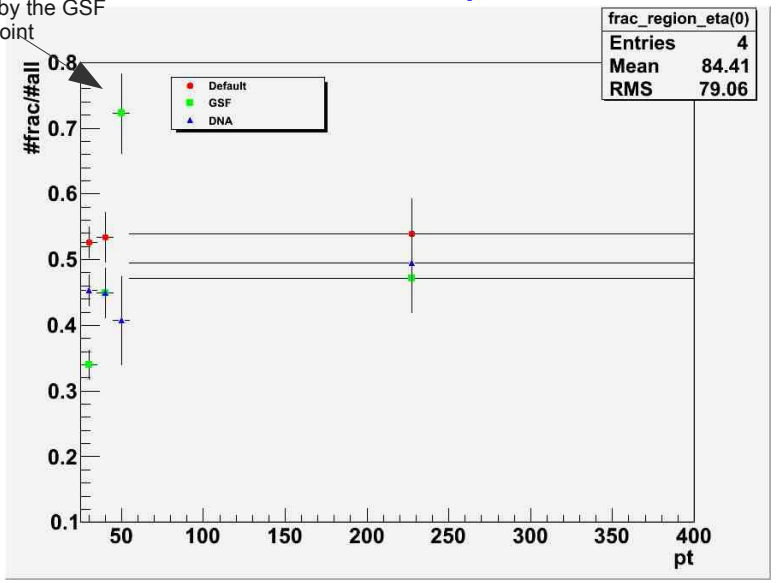
Fraction of events in the tail VS pt regions (0< |η|<0.6)

DP sample

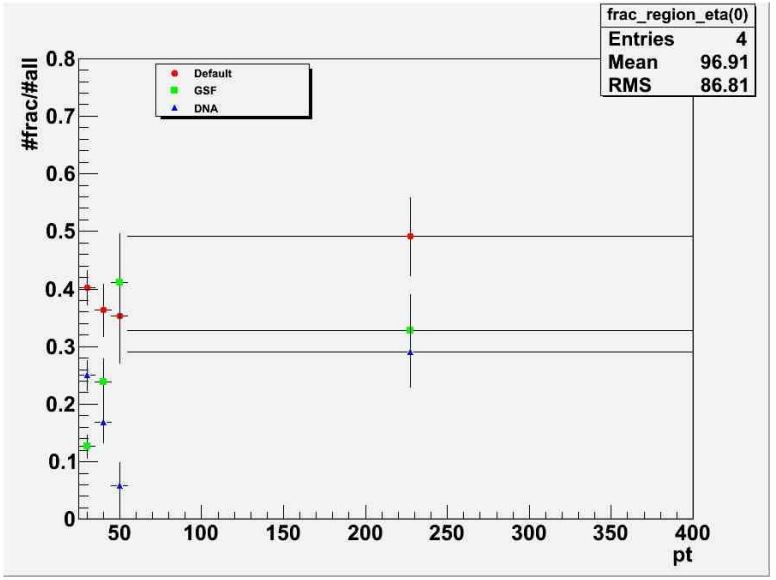


JF sample

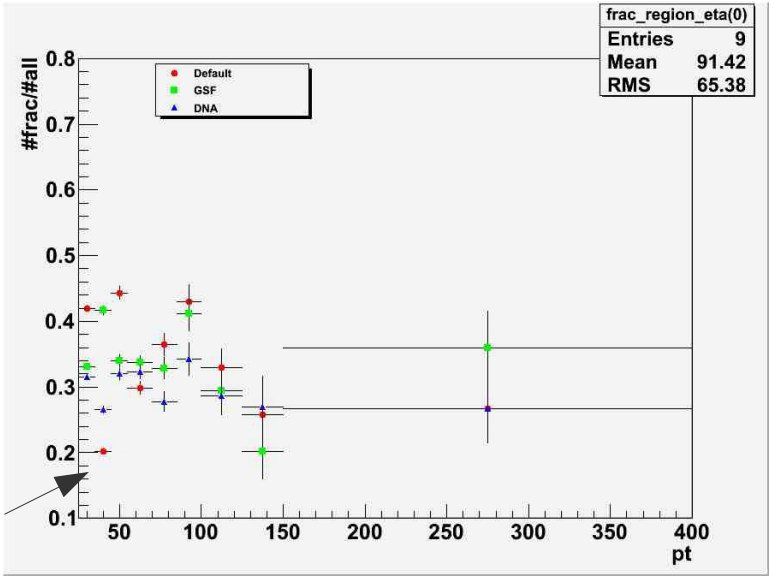
The default point is covered by the GSF point



JF sample with truthmatch



data sample

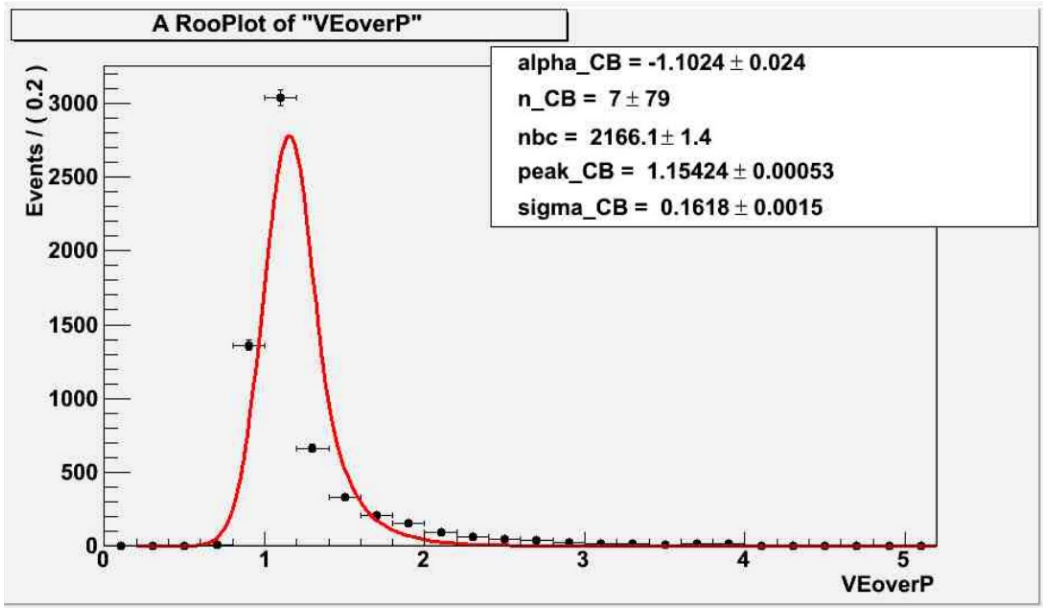


See slide 12

the same behavior shown in the resolution plots

CB fit for the quantity E/p (default) in different eta regions

pt>35 GeV && pt< 45 GeV && 0< |η|<0.6 (data)



the fit is not well done

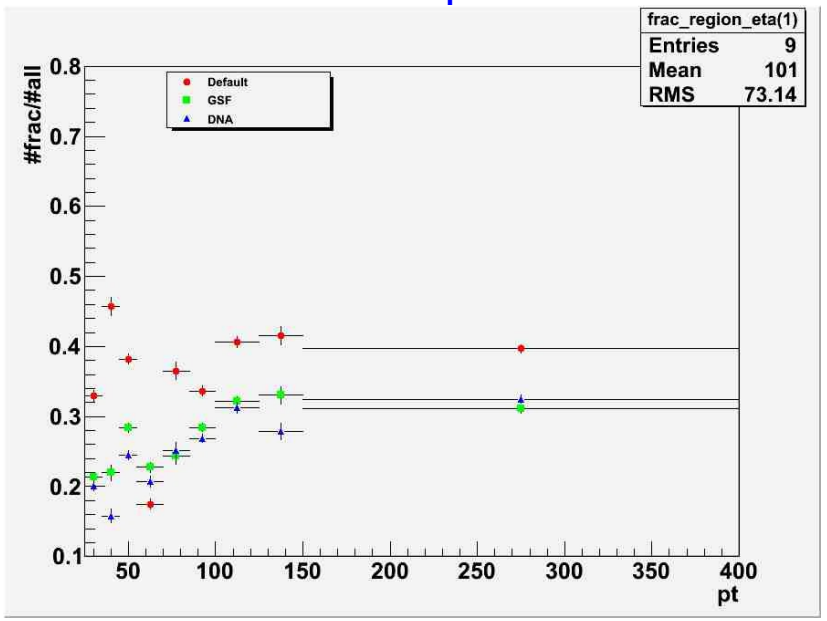
⇩

we find less events in the tail and more in the core

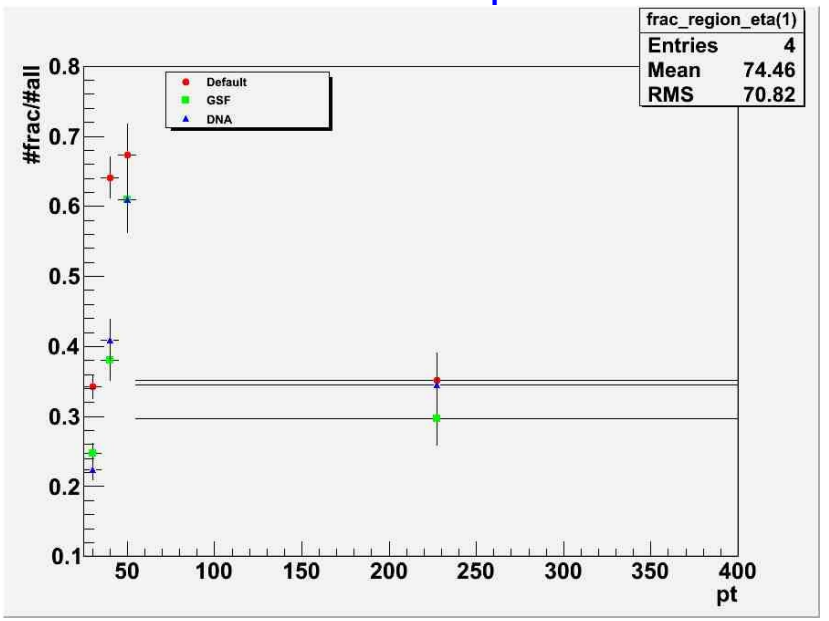
!!! For all those points that don't follow the trend, as in this case, it's possible to find that the fit is not well done.

Fraction of events in the tail VS pt regions (0.6< |η|<1.37)

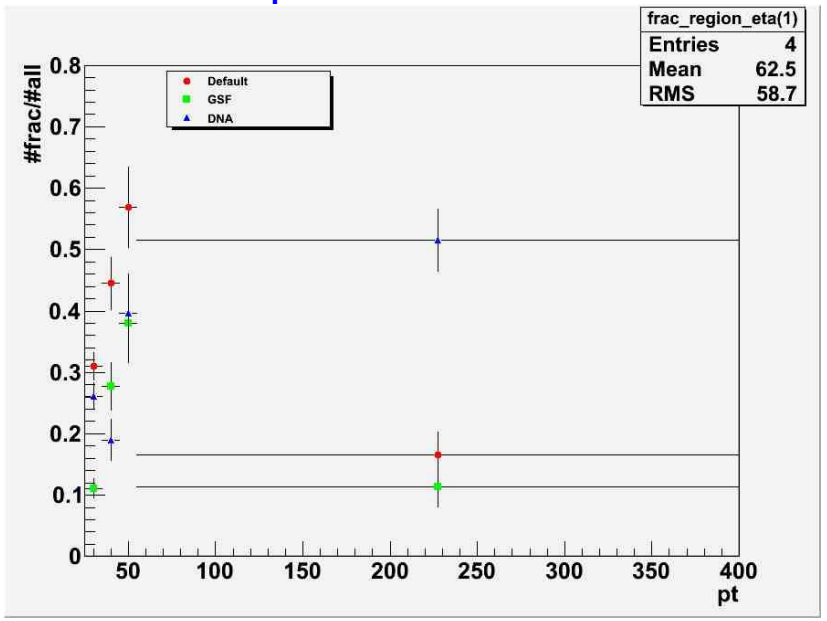
DP sample



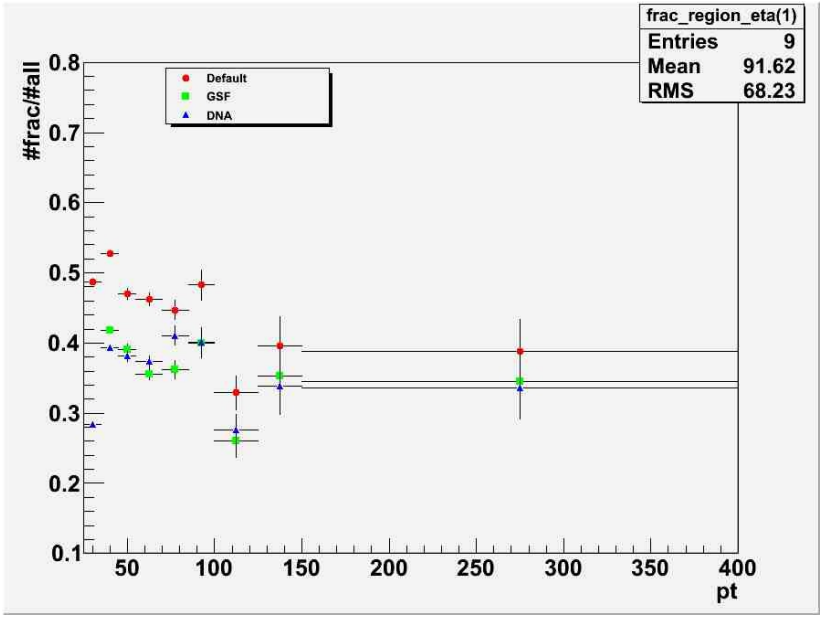
JF sample



JF sample with truthmatch

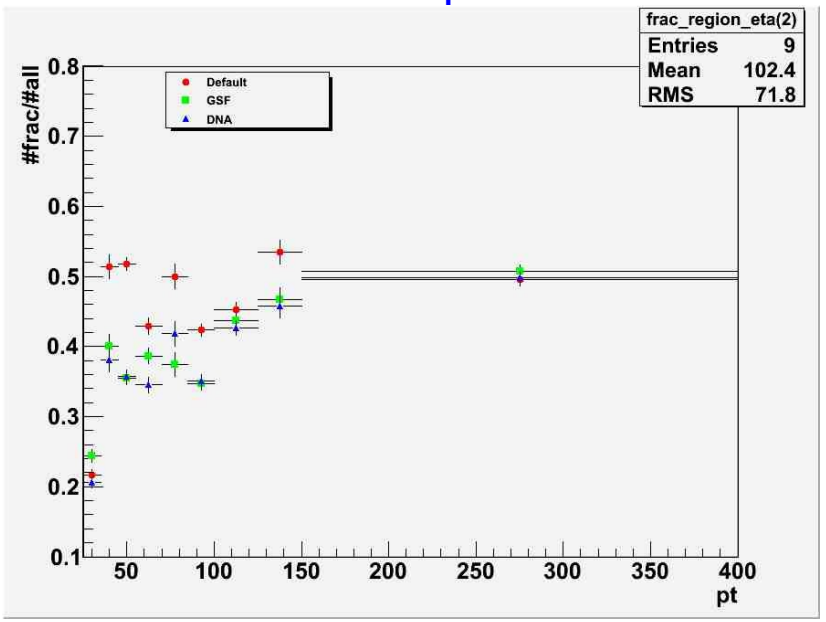


data sample

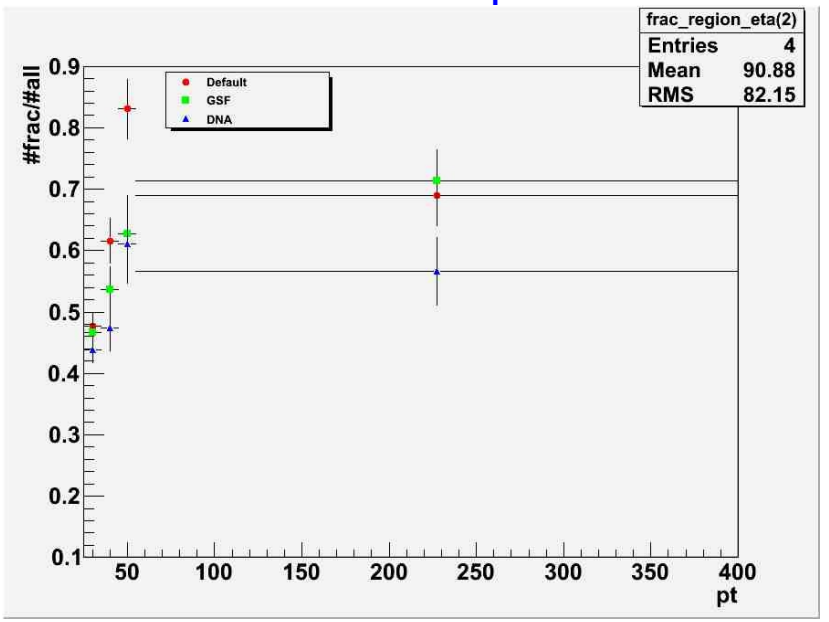


Fraction of events in the tail VS pt regions (1.52< |η|<1.81)

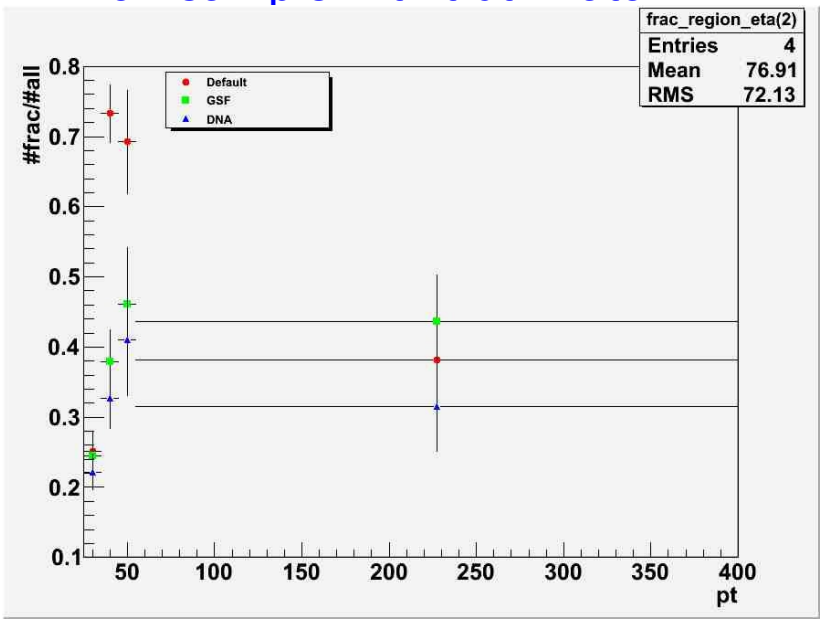
DP sample



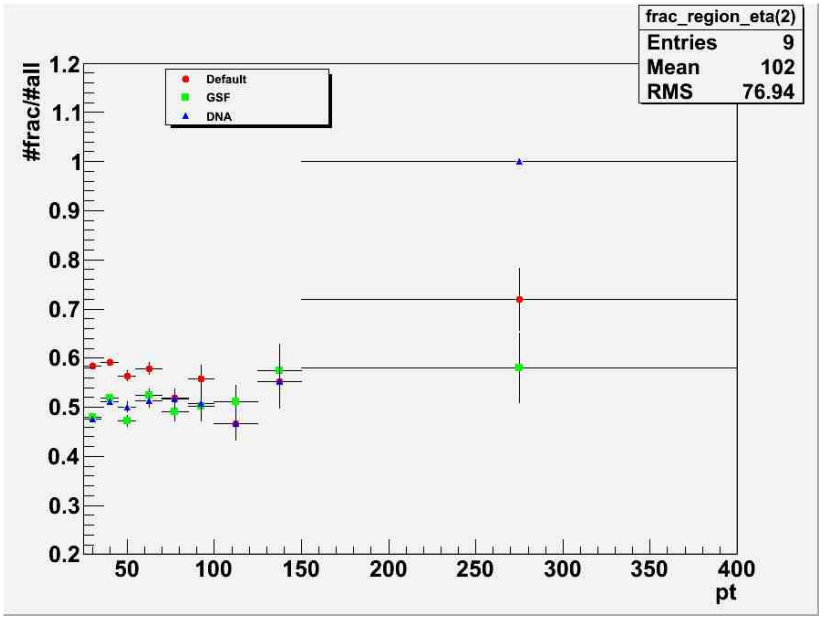
JF sample



JF sample with truthmatch

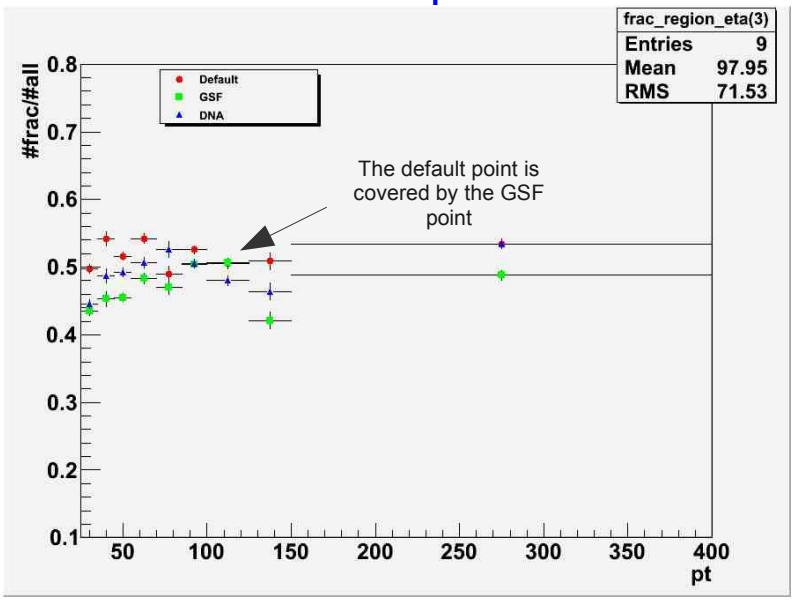


data sample

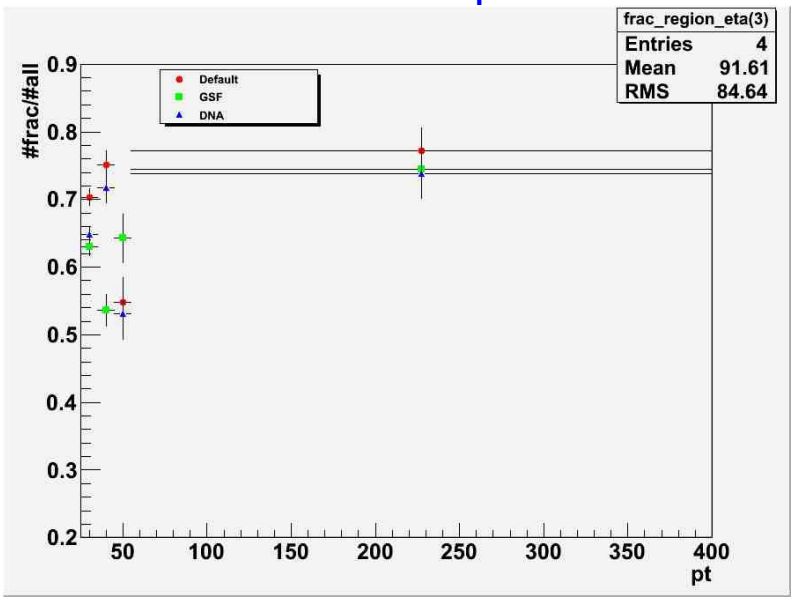


Fraction of events in the tail VS pt regions (1.81< |η|<2.37)

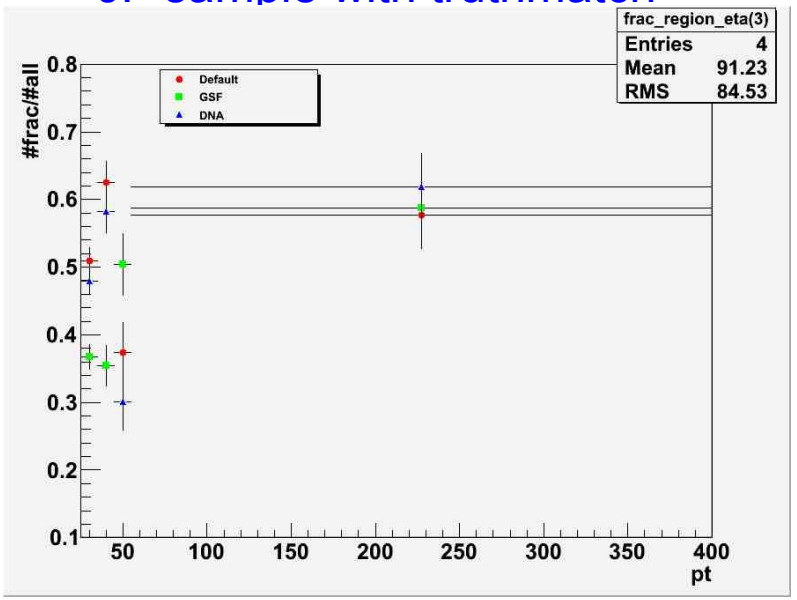
DP sample



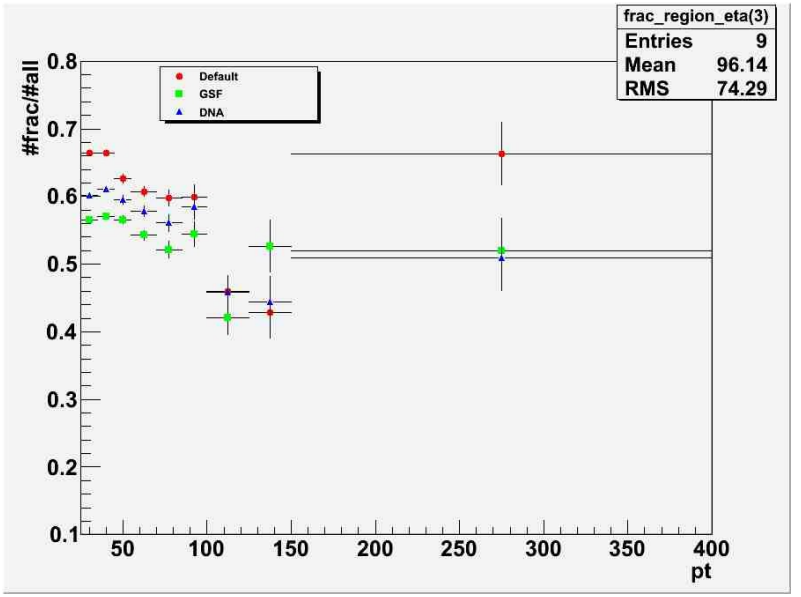
JF sample



JF sample with truthmatch

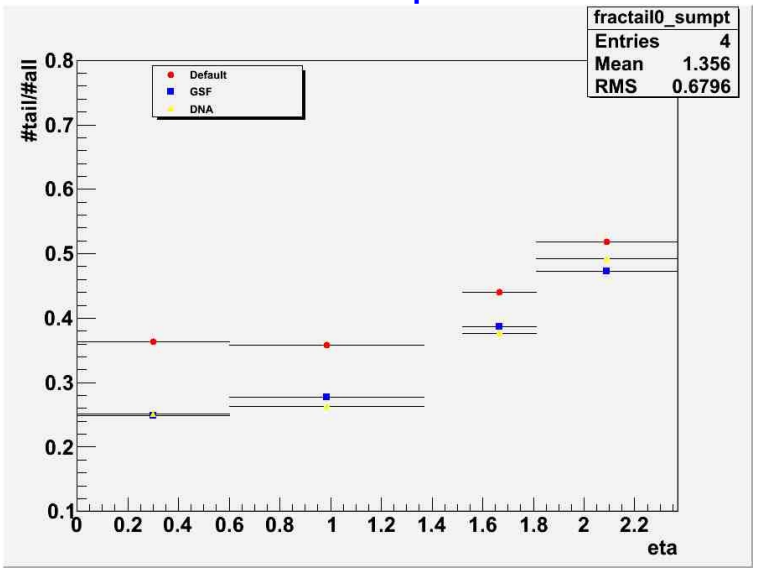


data sample

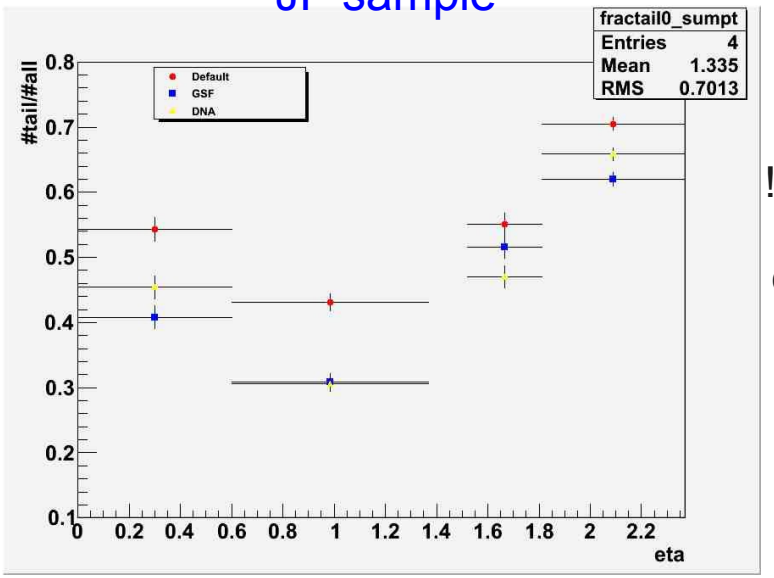


Fraction of events in the tail VS eta regions

DP sample

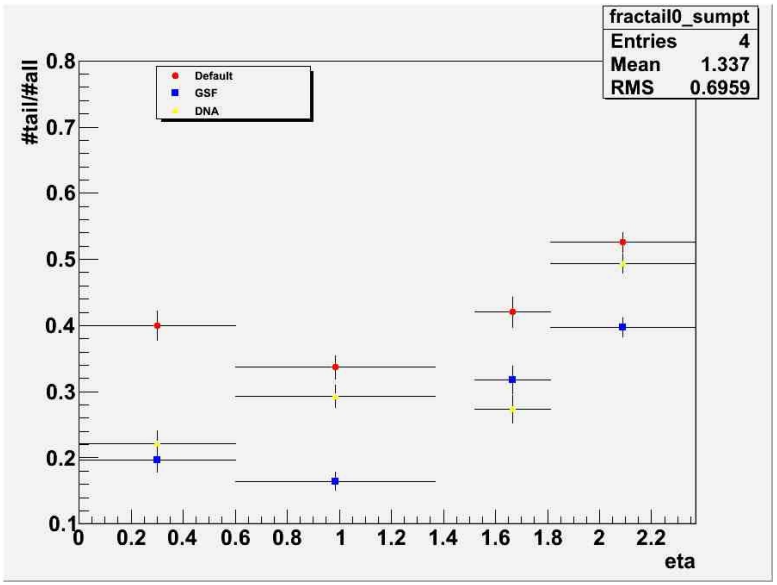


JF sample

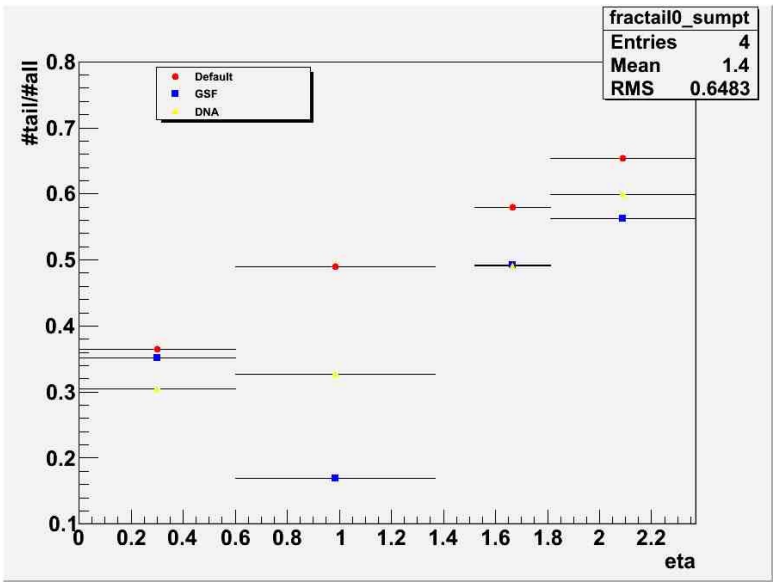


!!! The JF sample is shifted, as we expected, because of the presence of the background

JF sample with truthmatch



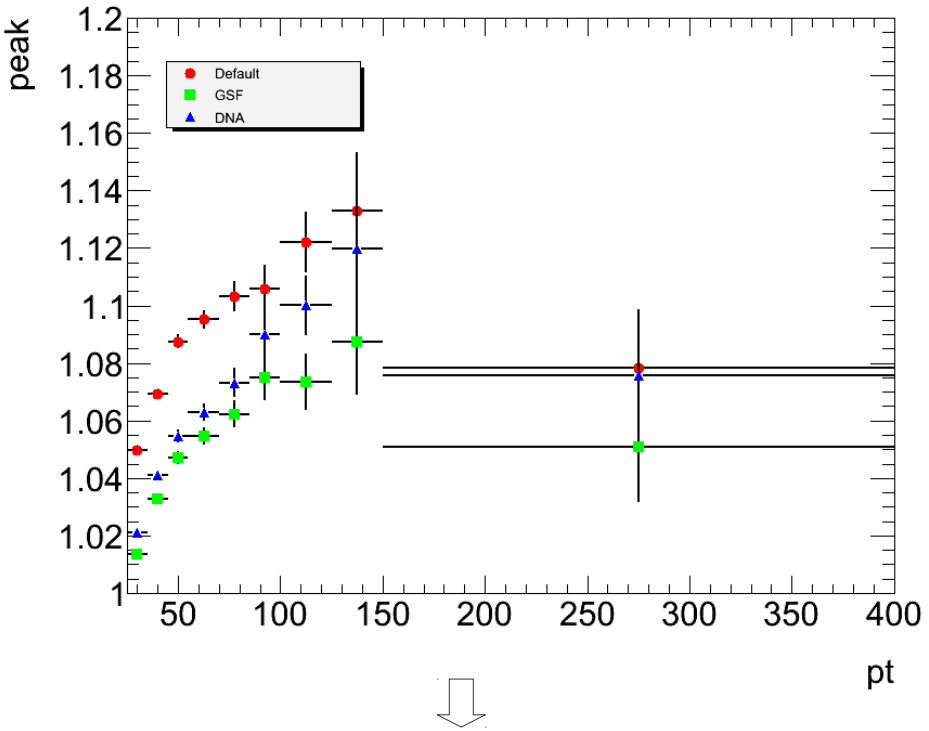
data sample

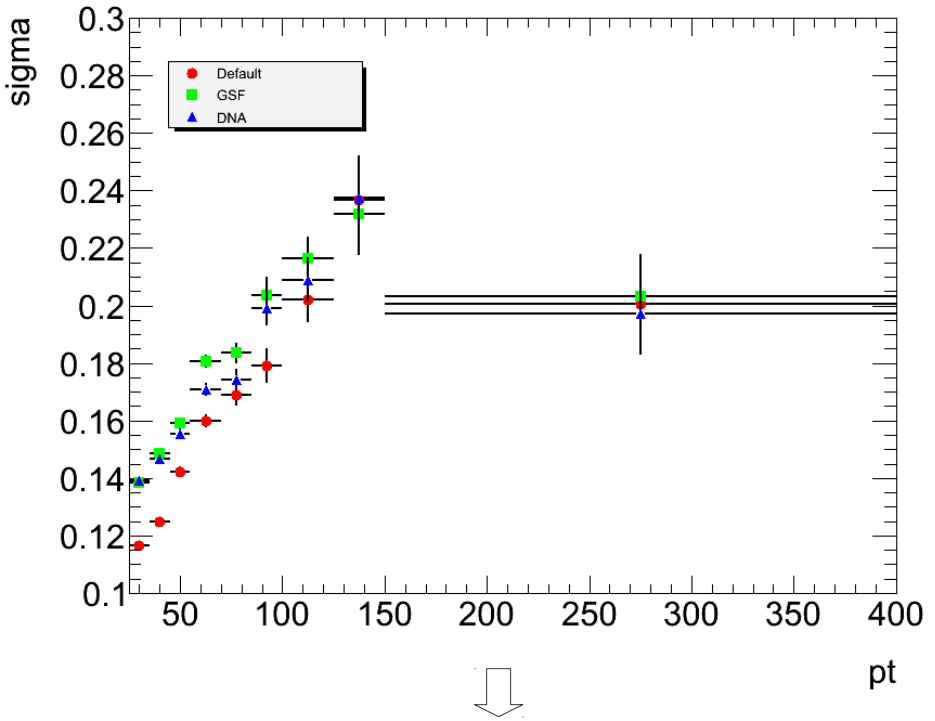


DNA and GSF seem to work better than the default algorithm

Notes: even if it's evident that the DNA and GSF are better than the default algorithm, it's hard to demonstrate that one of them is better than the other. The GSF algorithm, according to the resolution plots, seems to be better than DNA for high pt and high eta (the contrary for low pt and eta).

Crystall Ball peak and sigma for the different pt region (data sample)





- the GSF algorithm seems to be better than the DNA;

- The trend is due to worst resolution when we rise in pt;
- The GSF Gaussian core seems wider than the DNA one;

!!!The same results also for the other sample

Summary:

- we studied the performance of the bremsstrahlung refits for converted photons using E/P;
- we find a better performance for GSF and DNA compared to the default ATLAS track fit;
- we find that the GSF fitter is better than the DNA algorithm for high pt and high eta and the contrary for low pt and eta but, if we look at the peak of the fit for different pt regions, it seems that the GSF algorithm is better than DNA;

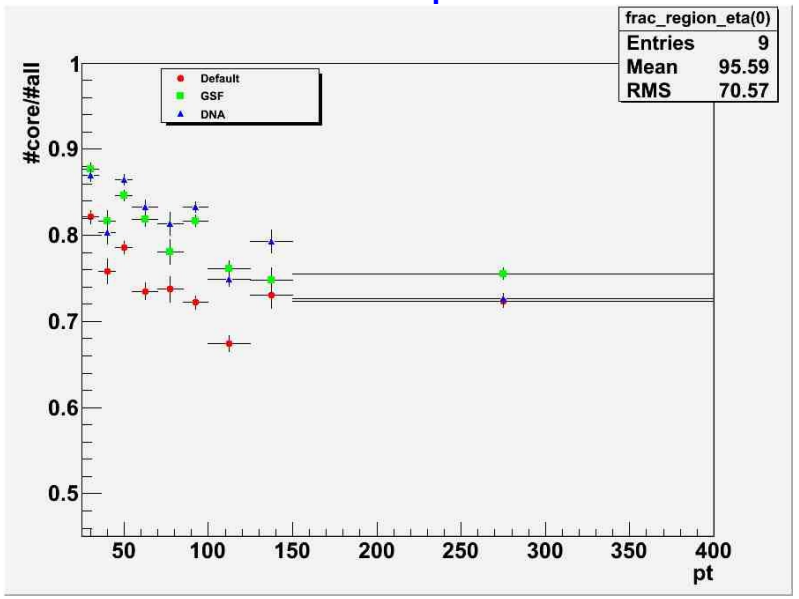
Future ideas:

- improve H->gamgam mass resolution combining calo and tracker measurement of the energy and using the brem refit conversion for the mass reconstruction;
- improve conversion vertex position could improve the mass resolution;
- improve E/p distributions for purity estimation for converted photons;

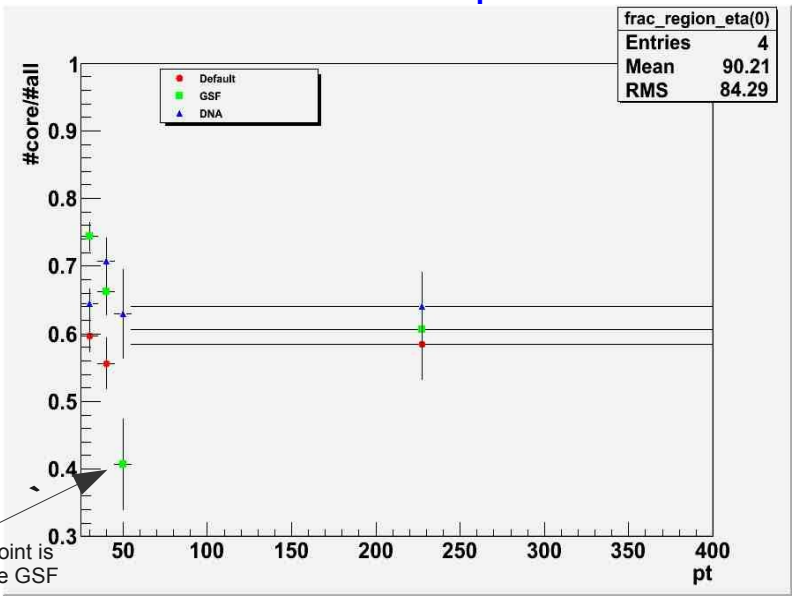
Backup

Fraction of events in the gaussian core VS pt regions (0< |η|<0.6)

DP sample

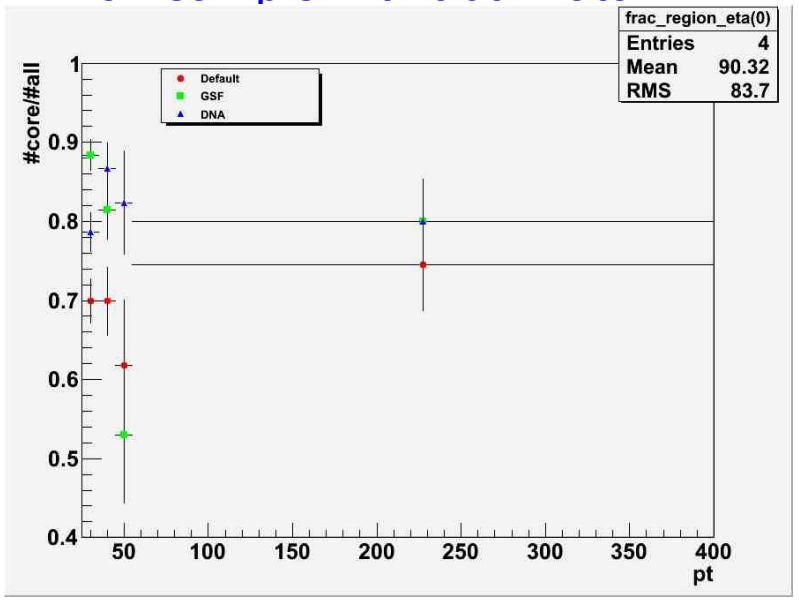


JF sample

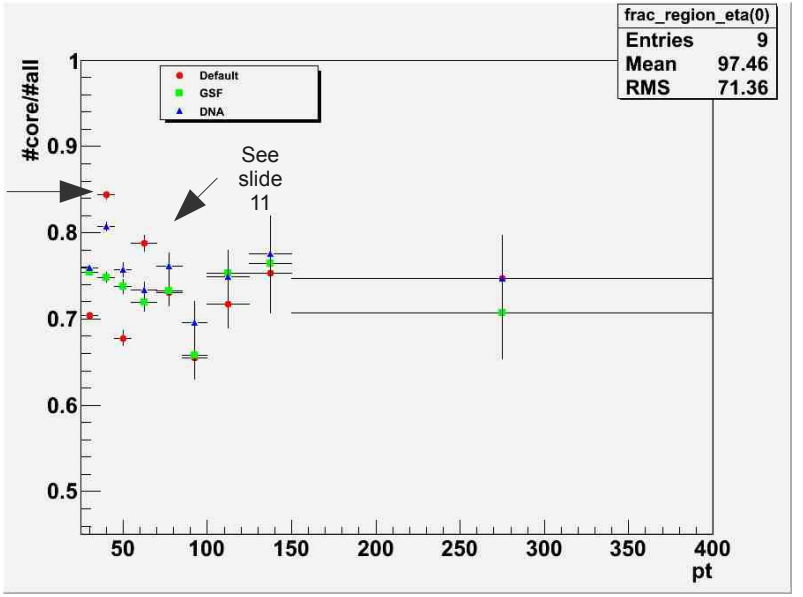


The default point is covered by the GSF point

JF sample with truthmatch



data sample

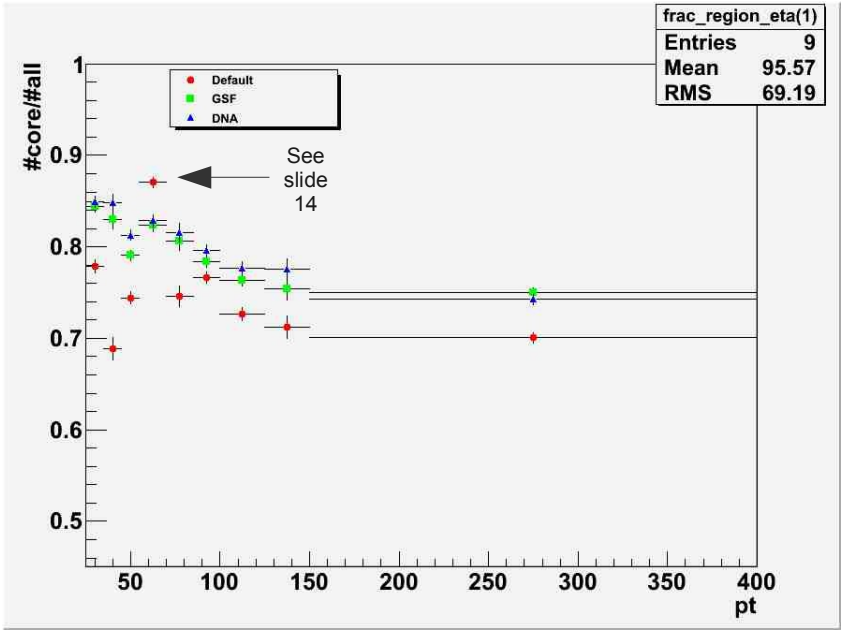


See slide 10

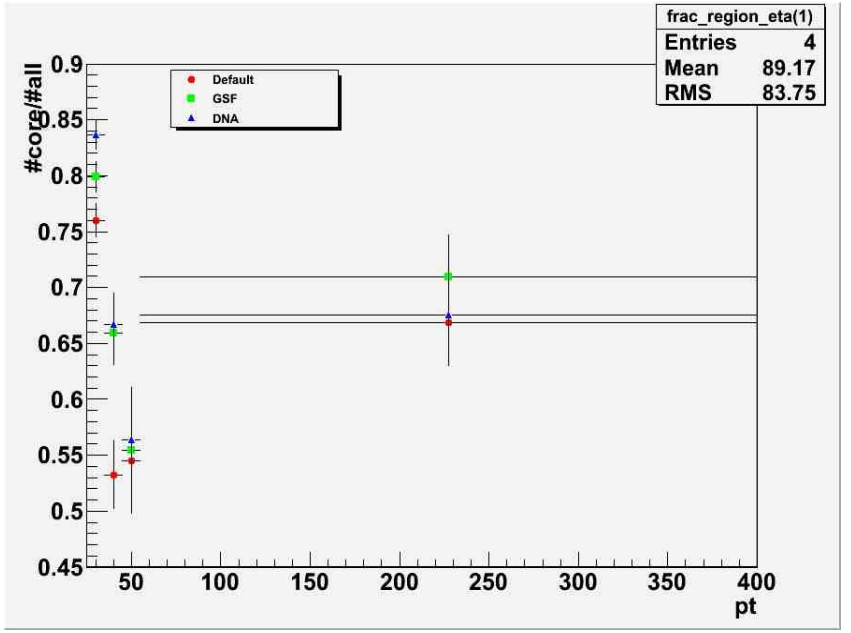
See slide 11

Fraction of events in the gaussian core VS pt regions (0.6< |η|<1.37)

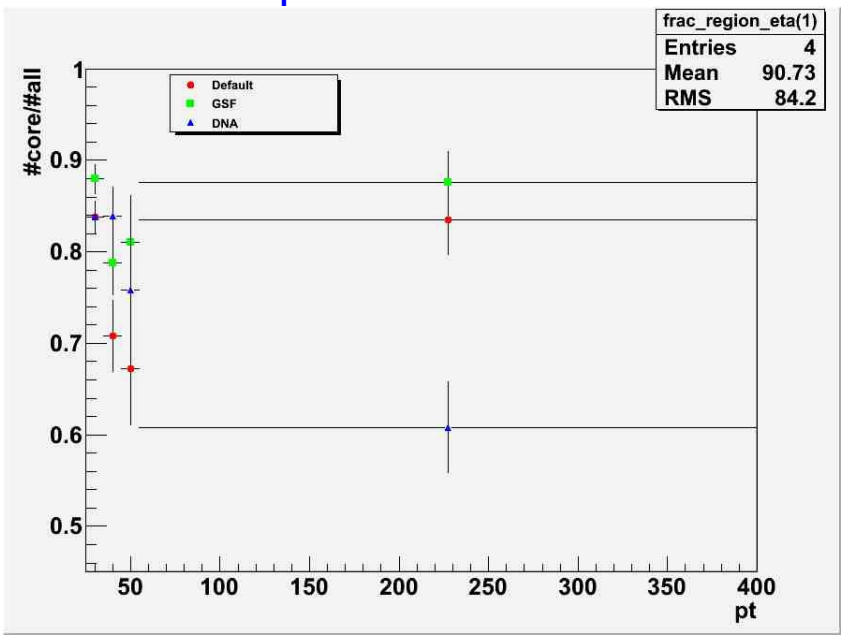
DP sample



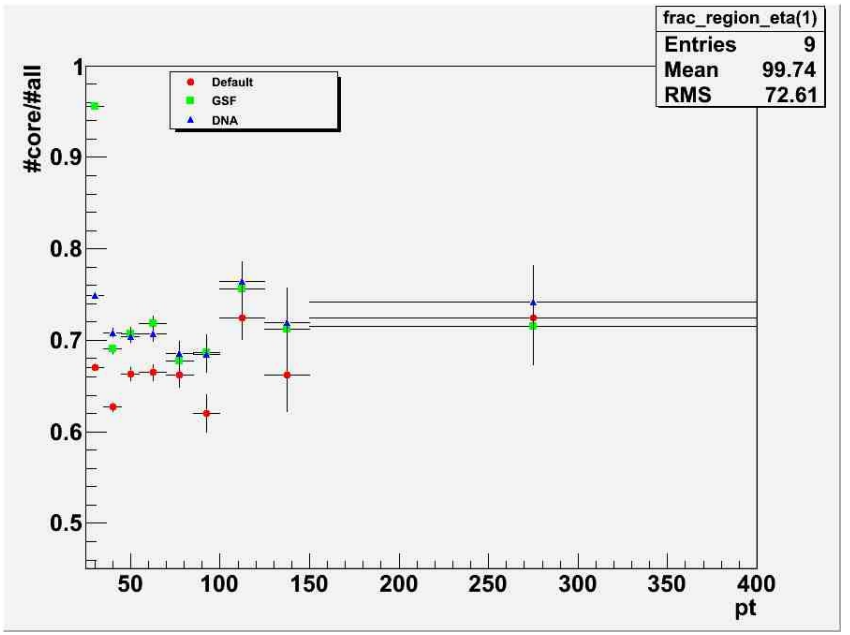
JF sample



JF sample with truthmatch

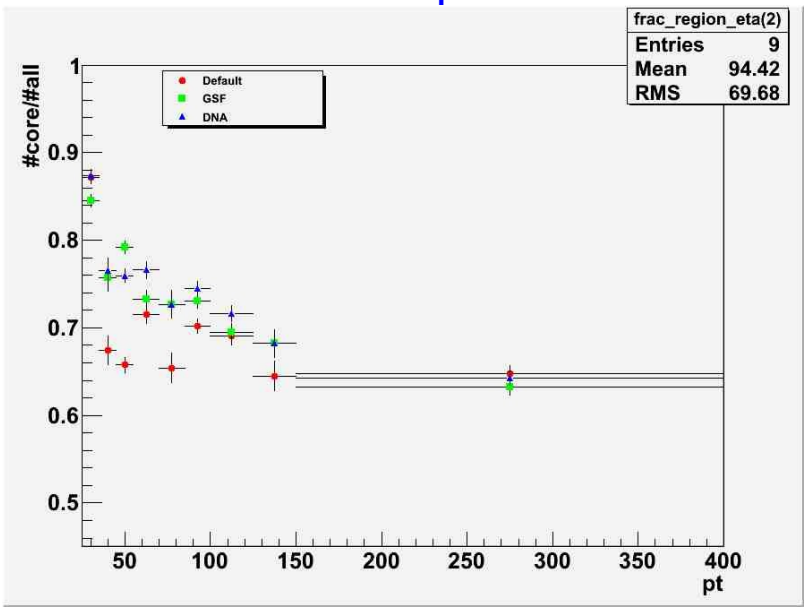


data sample

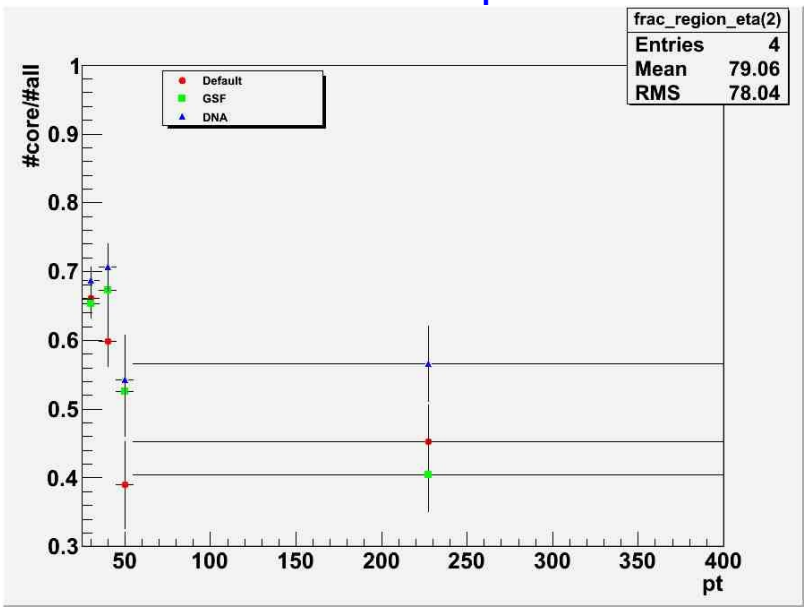


Fraction of events in the gaussian core VS pt regions (1.52< |η|<1.81)

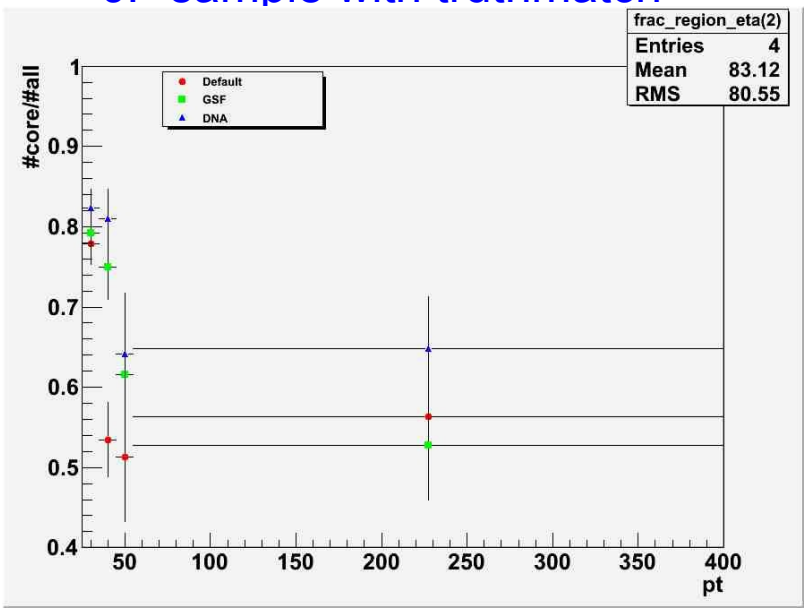
DP sample



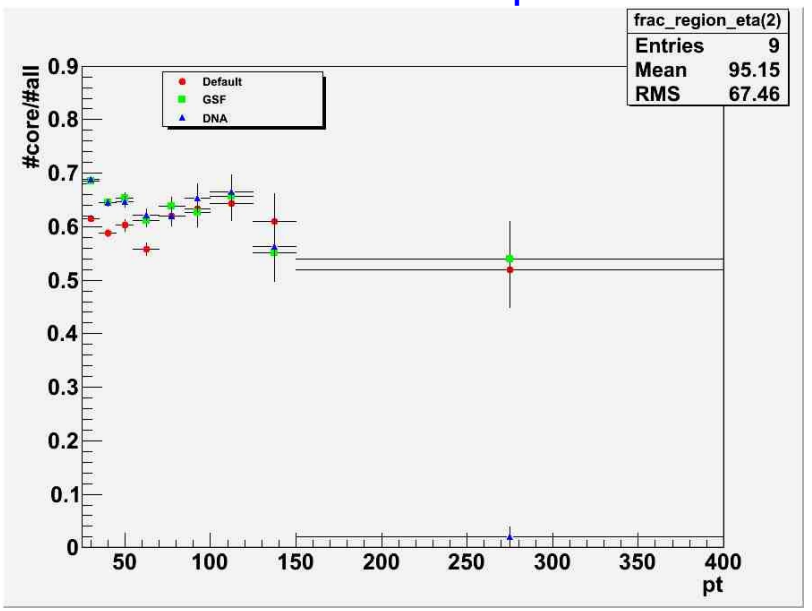
JF sample



JF sample with truthmatch

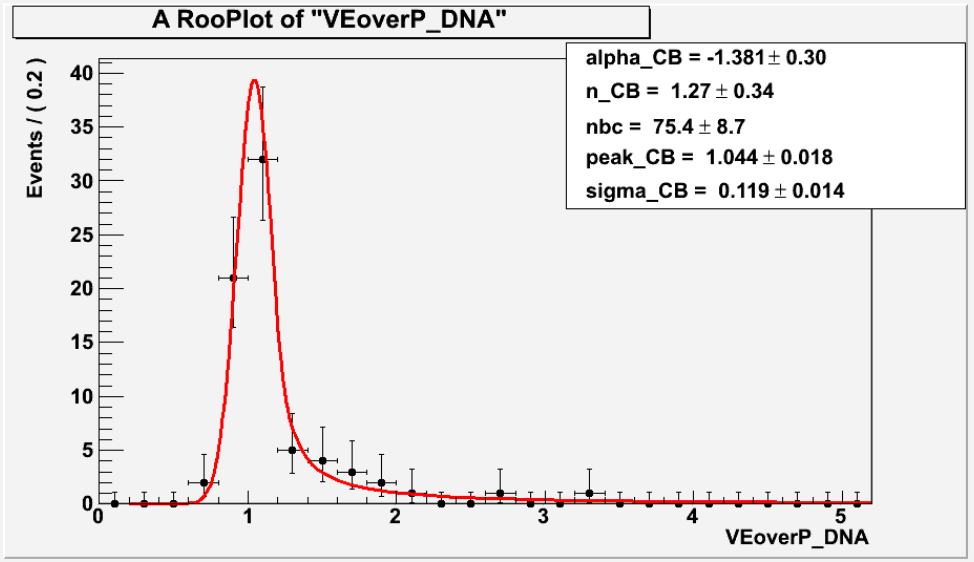


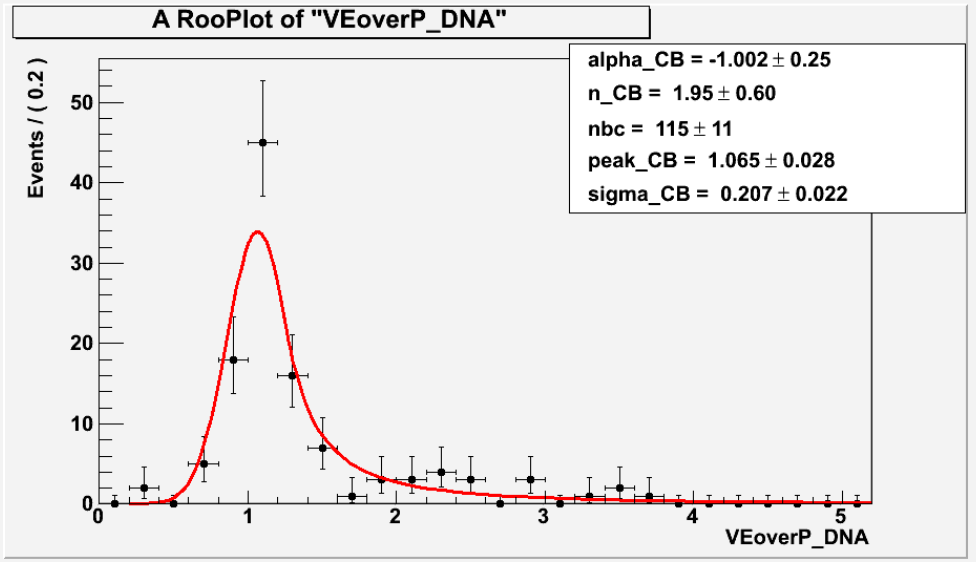
data sample

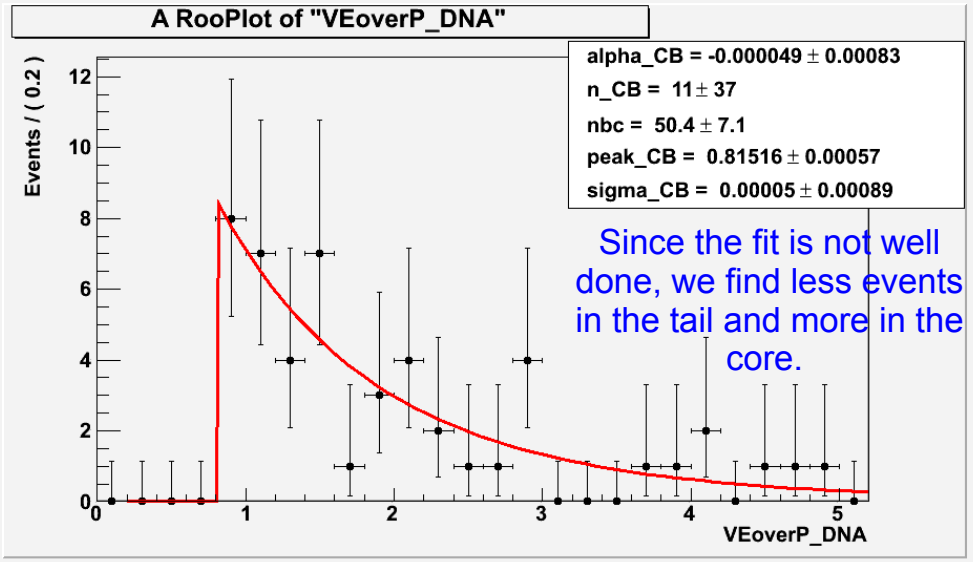


CB fit for the quantity E/p (GSF) in different eta regions

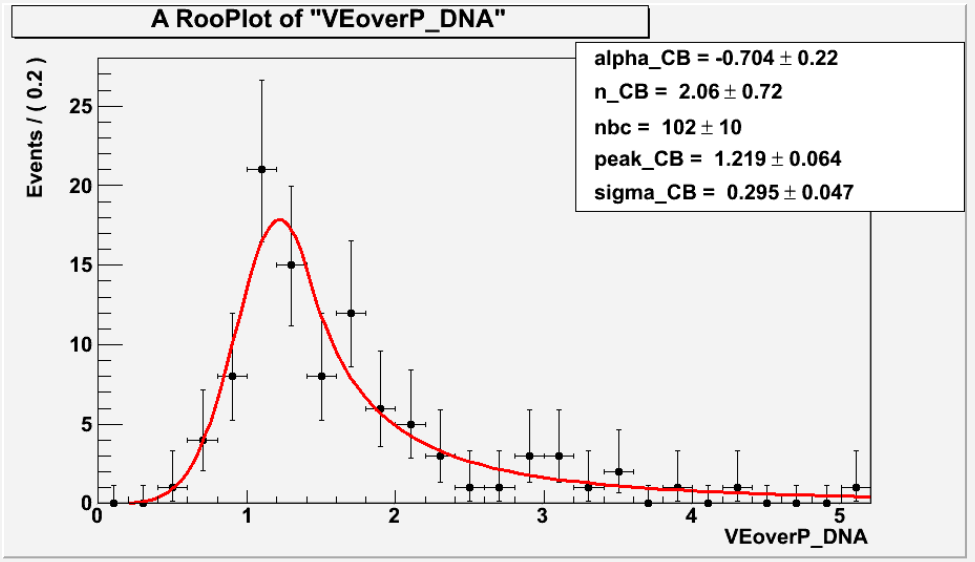
pt>150 GeV && pt< 400 GeV





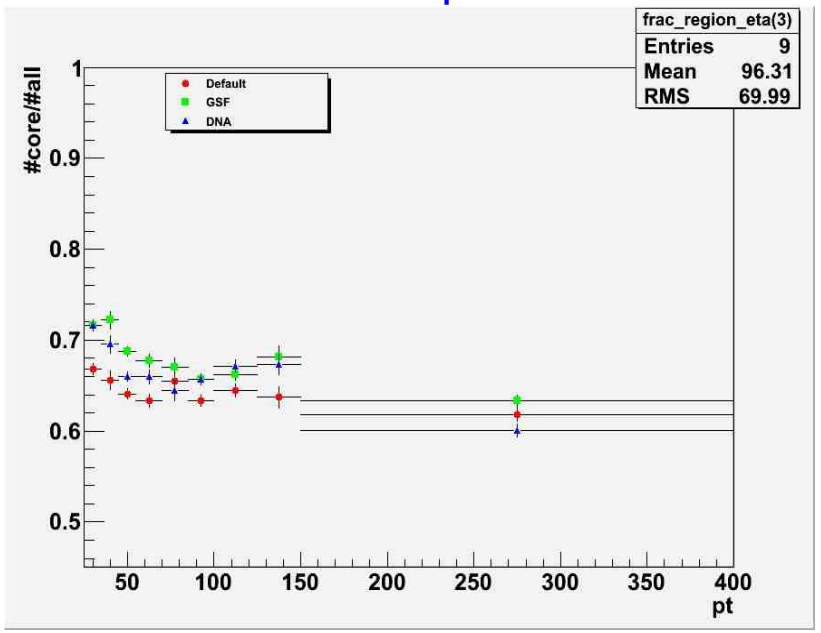


Since the fit is not well done, we find less events in the tail and more in the core.

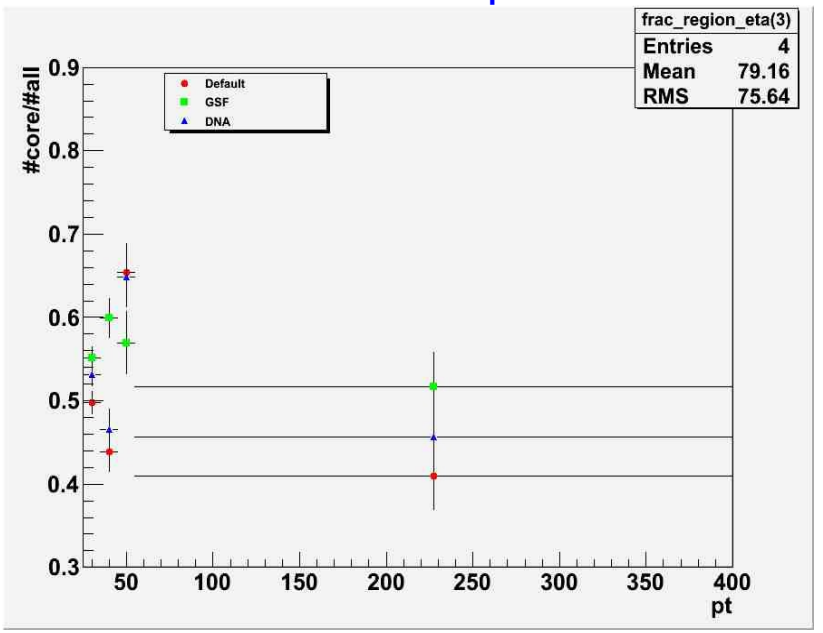


Fraction of events in the gaussian core VS pt regions (1.81< |η|<2.37)

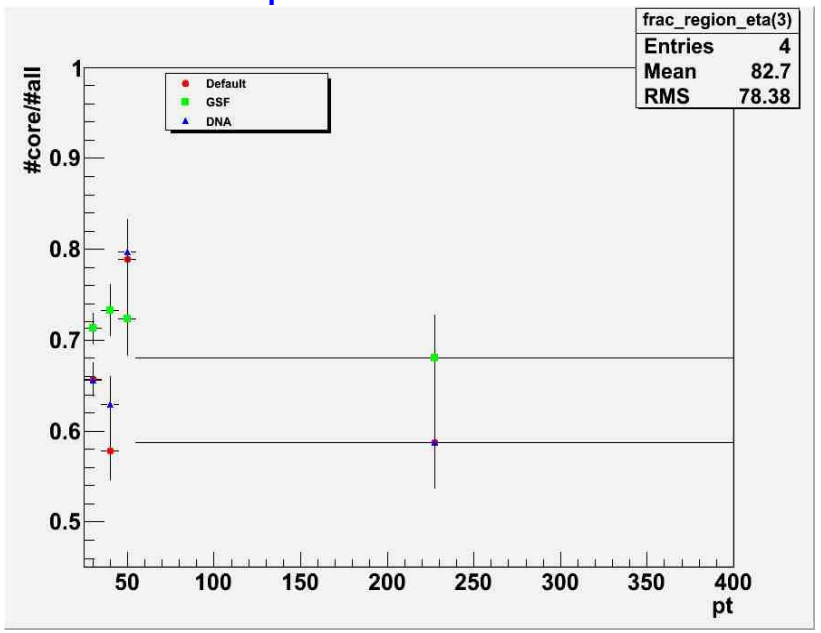
DP sample



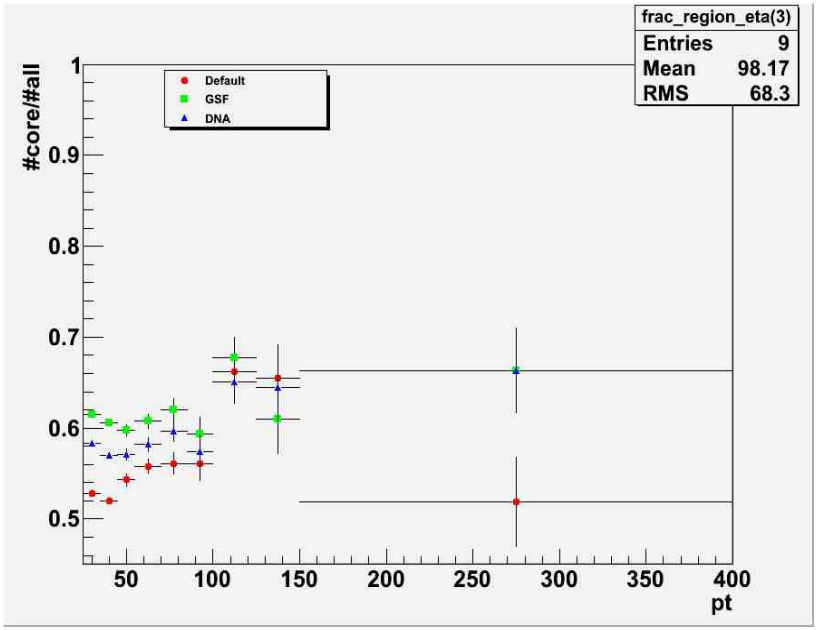
JF sample



JF sample with truthmatch

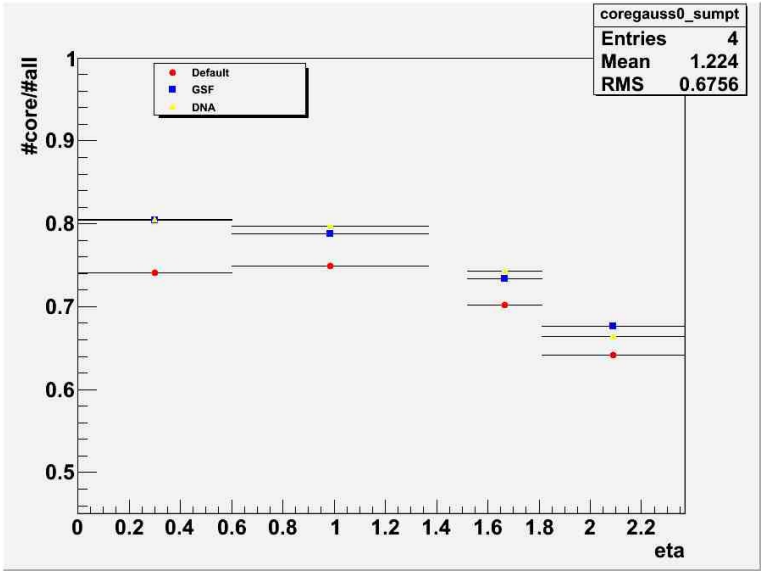


data sample

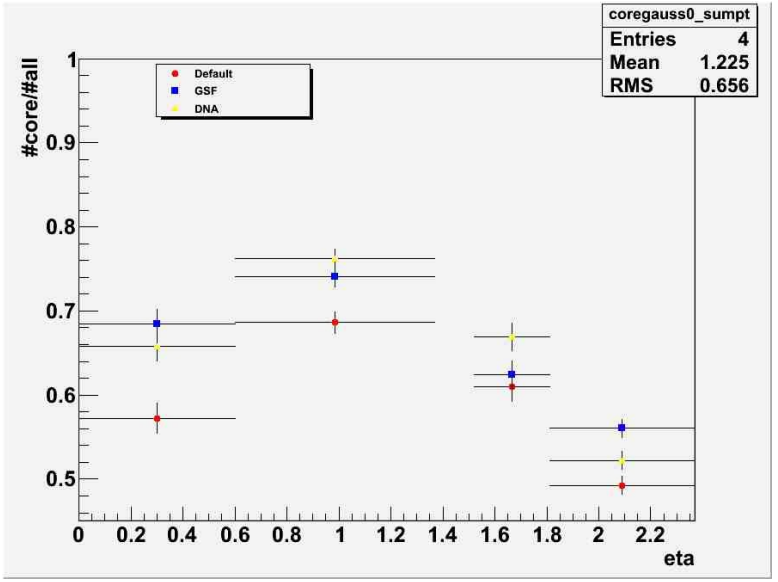


Fraction of events in the gaussian core ($\pm 2\sigma$) VS eta regions

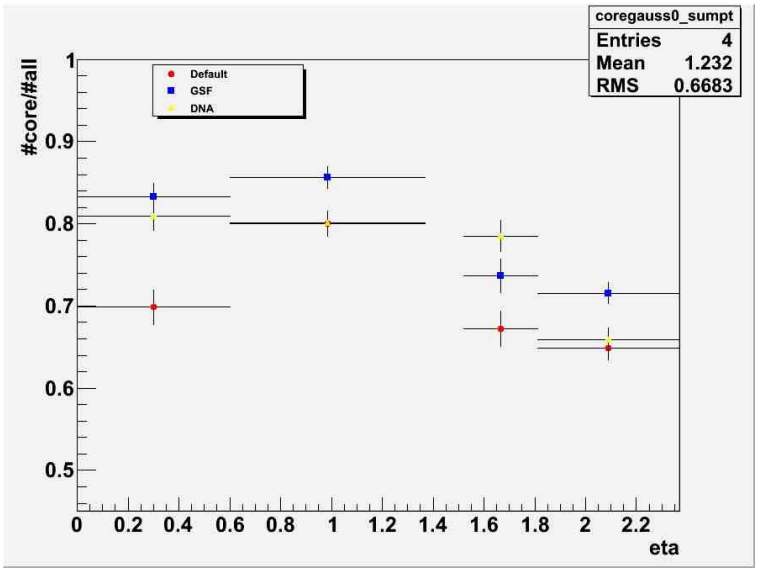
DP sample



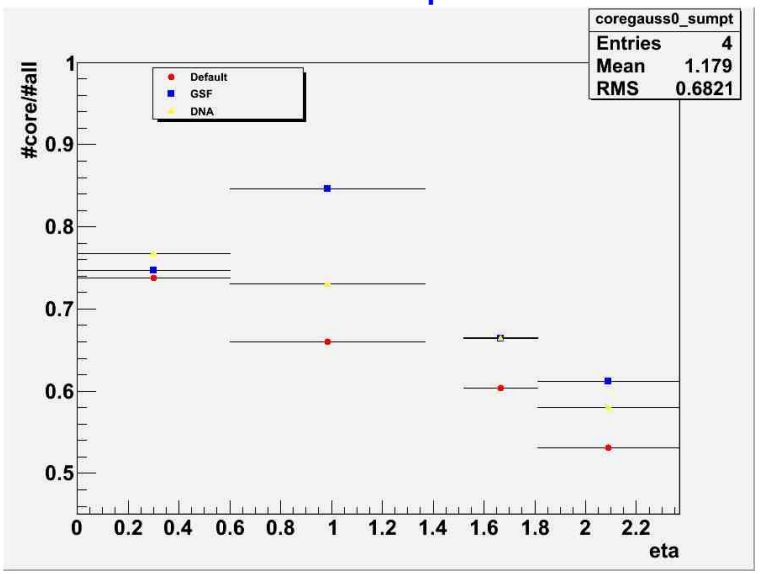
JF sample



JF sample with truthmatch

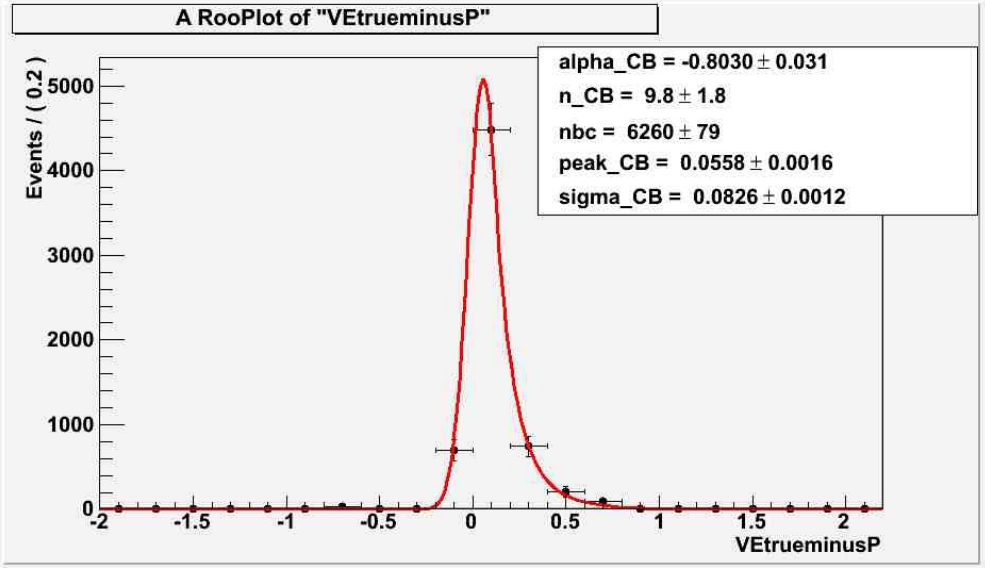


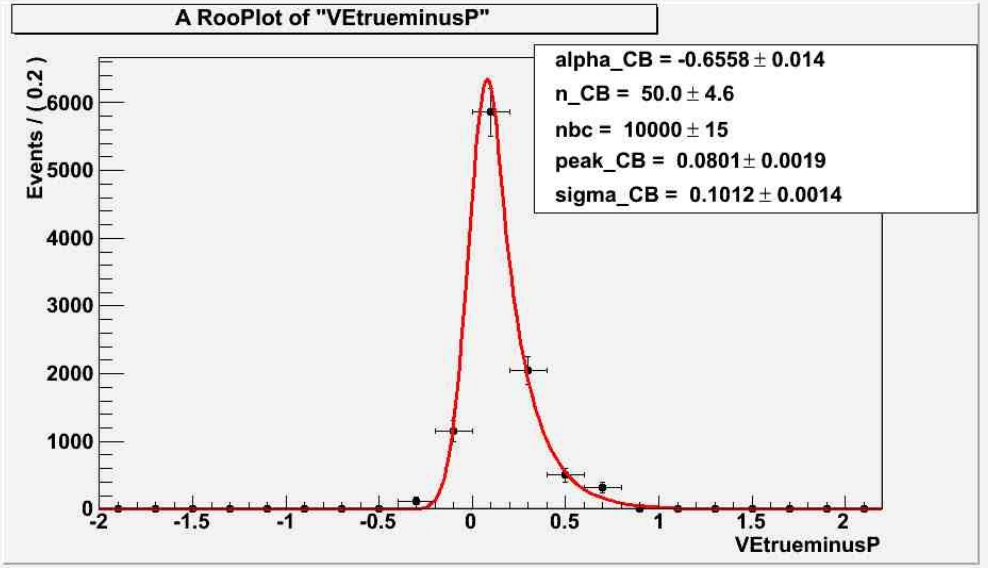
data sample

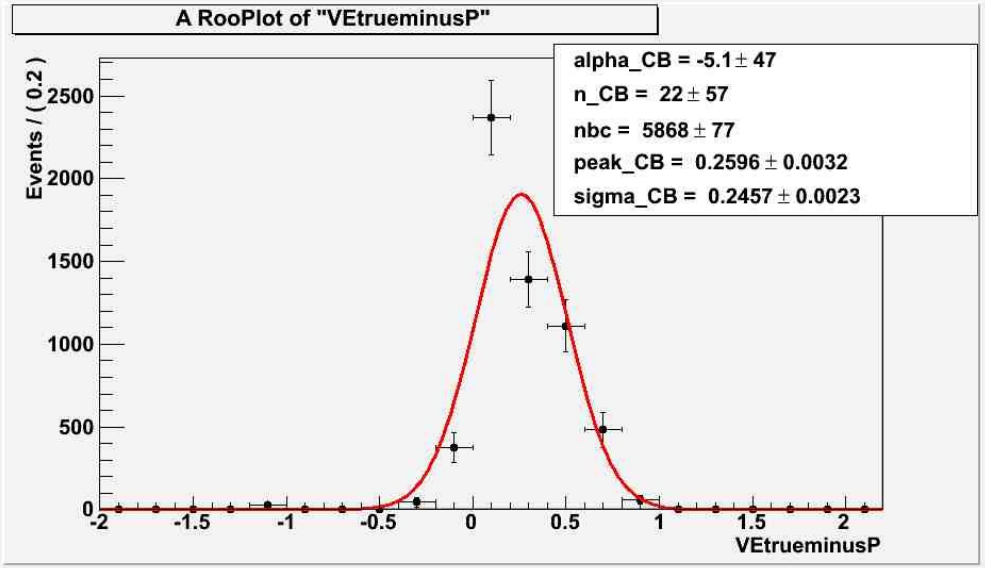


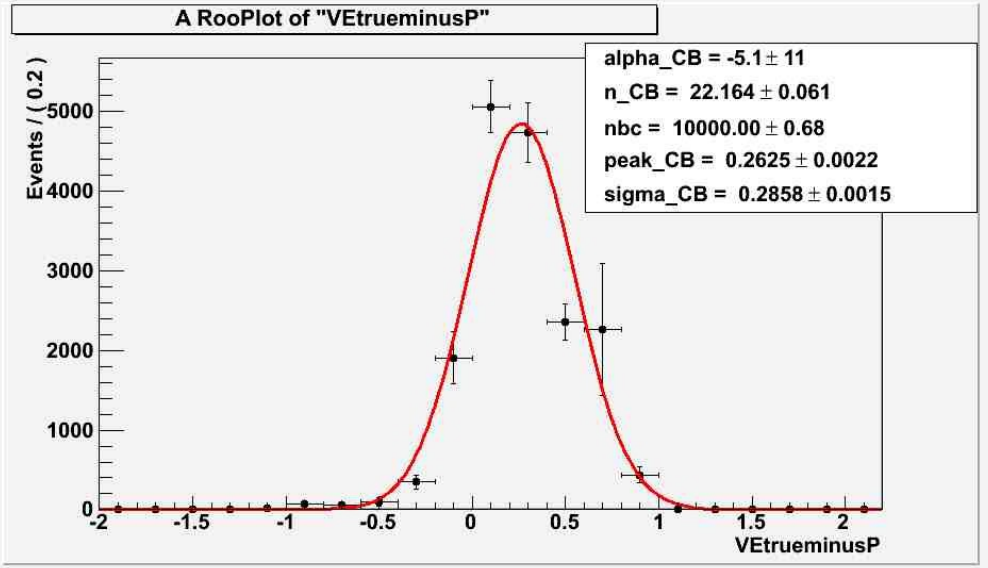
Default algorithm resolution for the DP sample

pt>35 GeV && pt< 45 GeV



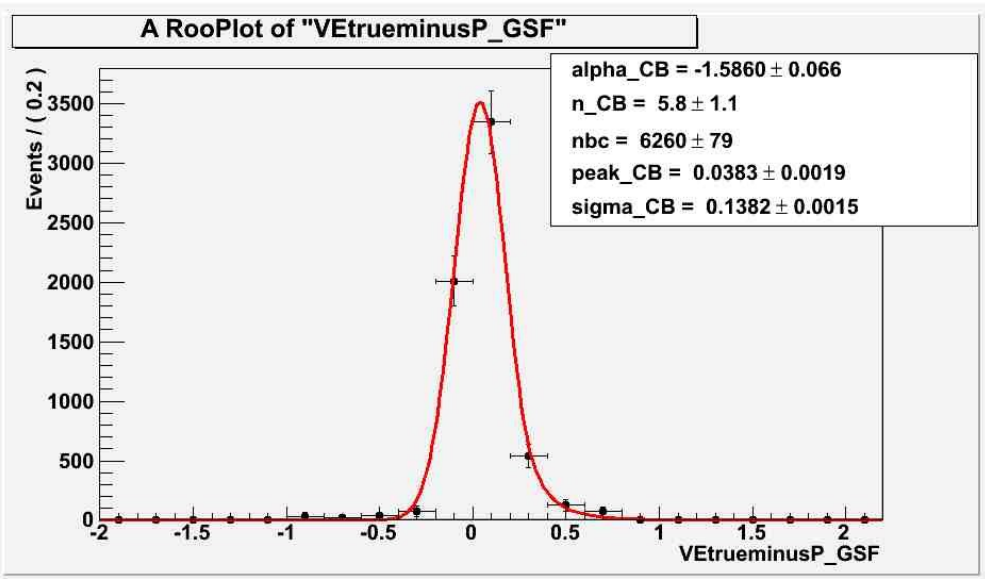


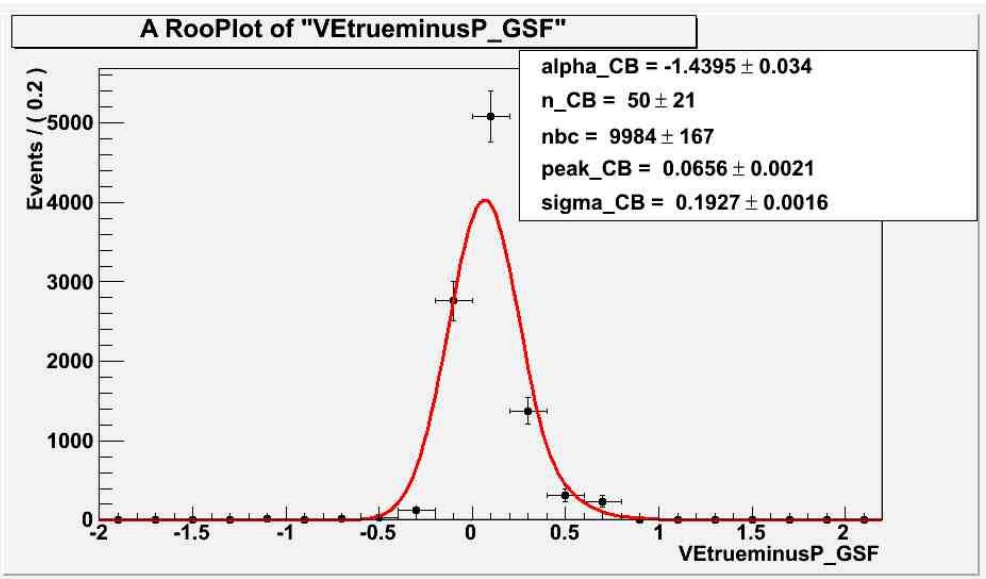


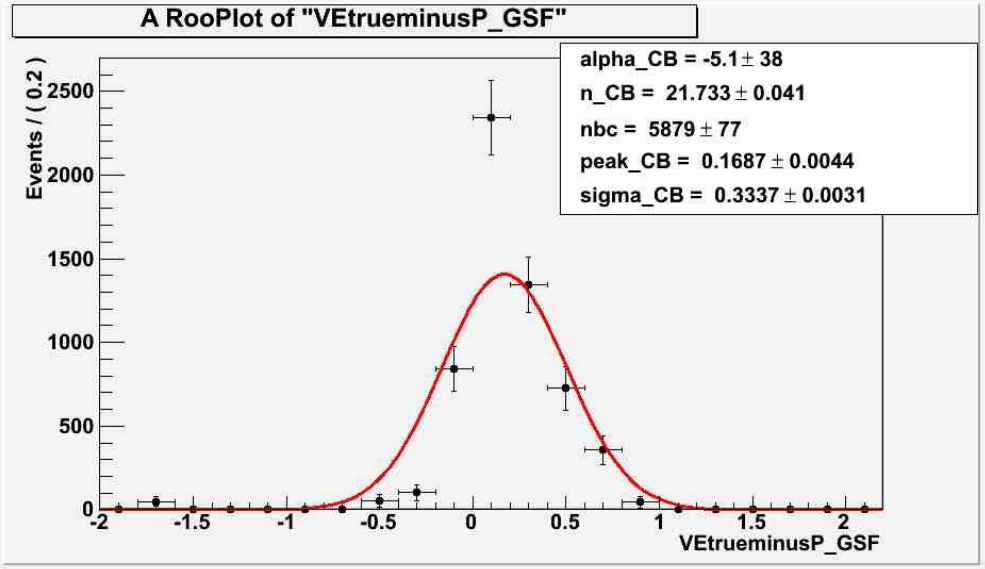


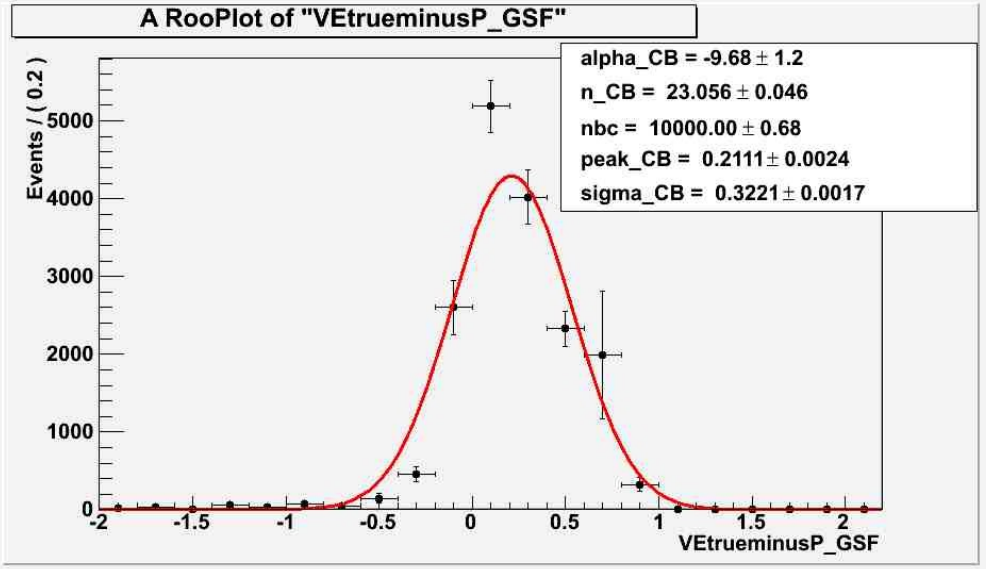
GSF algorithm resolution for the DP sample

pt>35 GeV && pt< 45 GeV



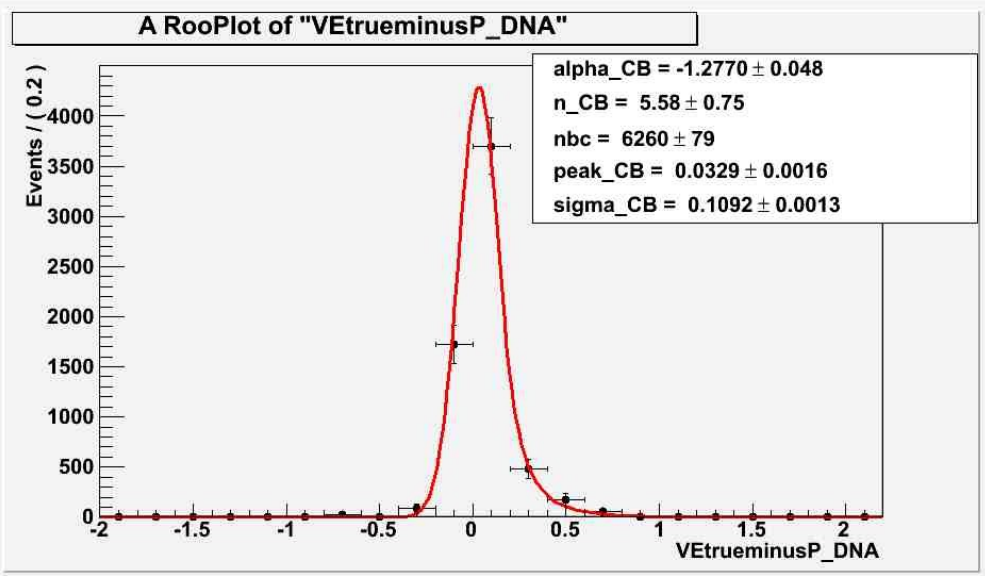


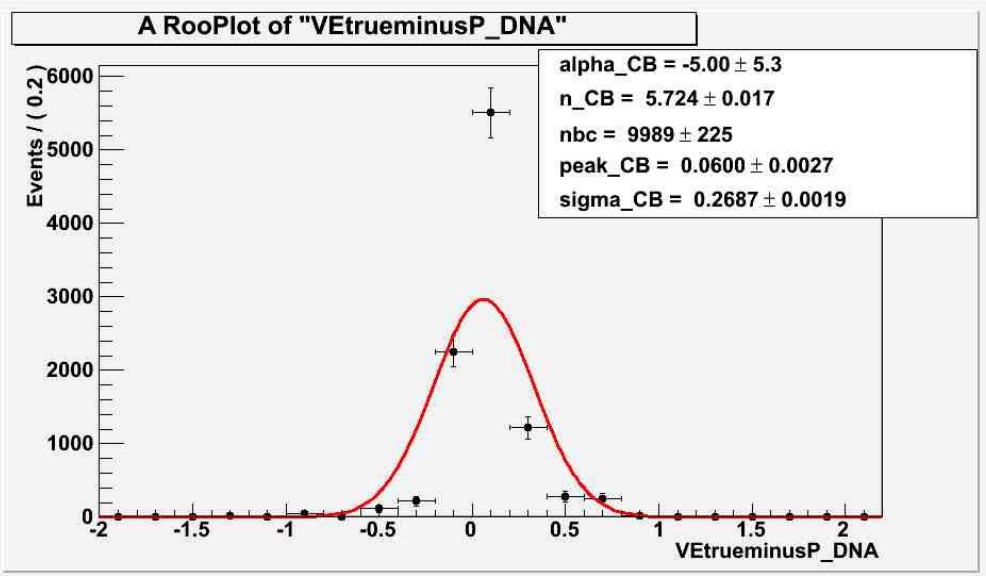


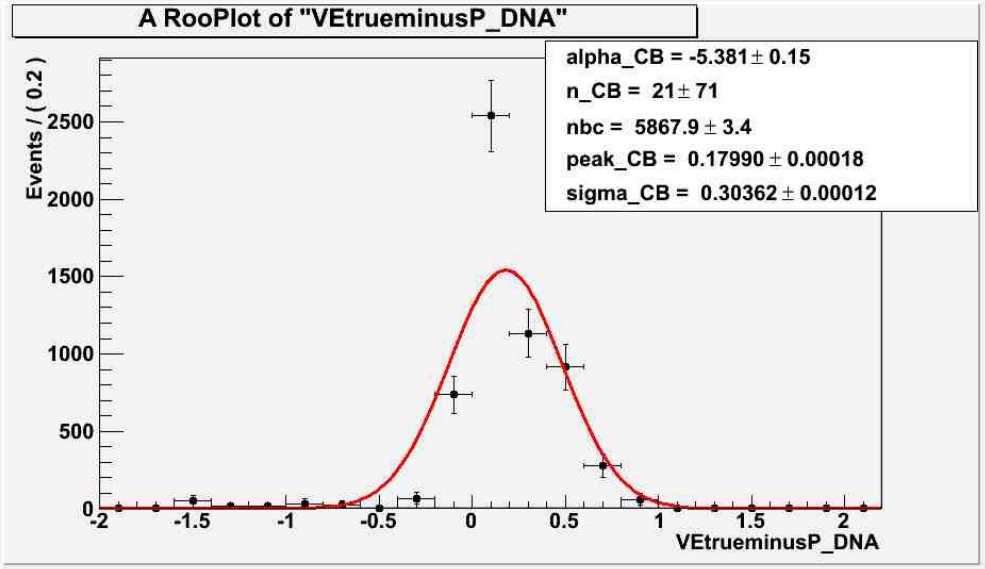


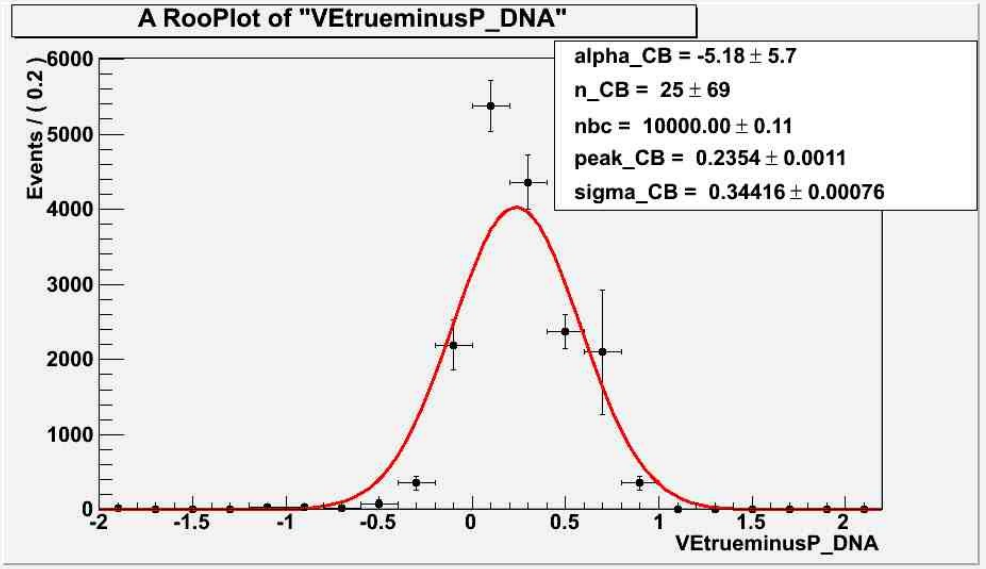
DNA algorithm resolution for the DP sample

pt>35 GeV && pt< 45 GeV



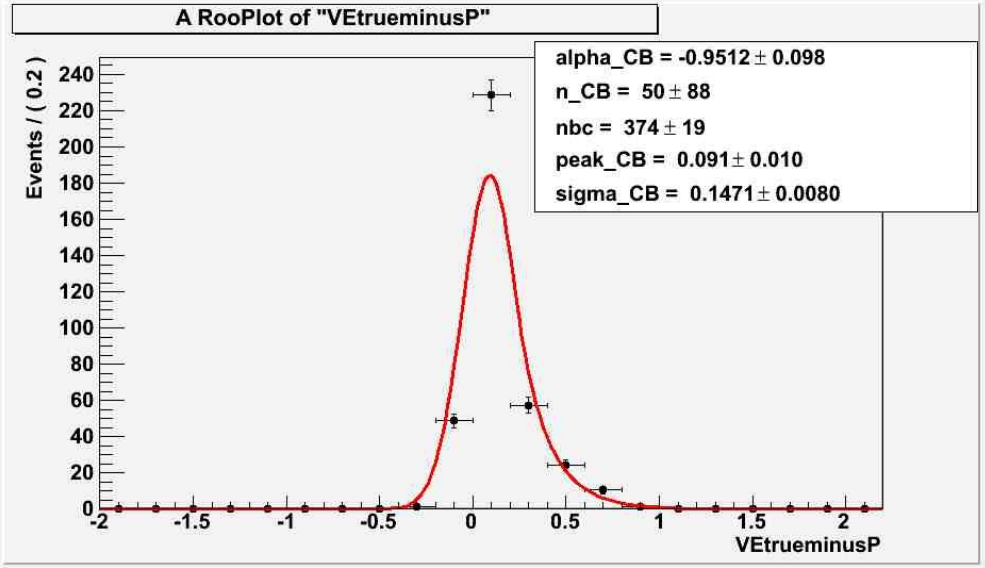


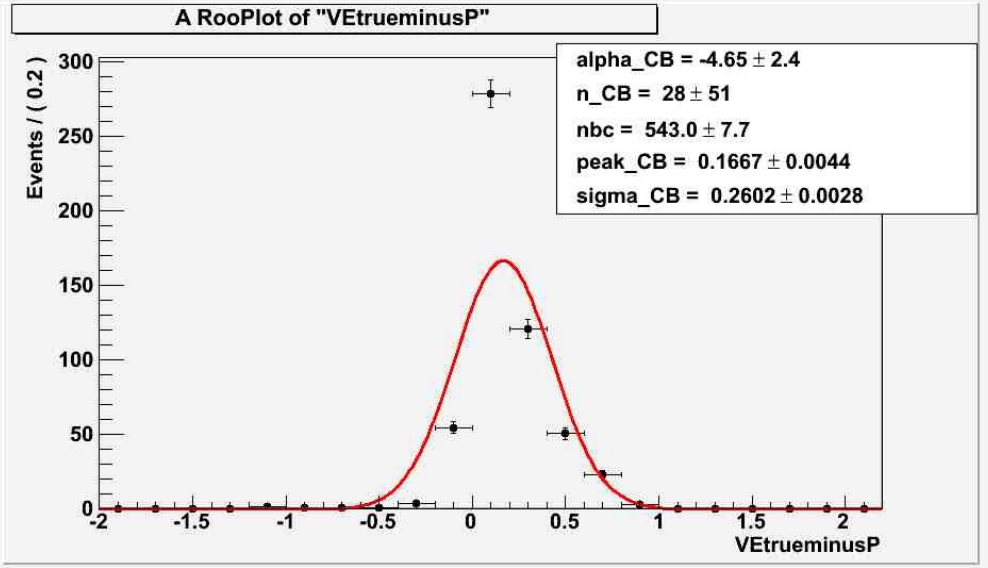


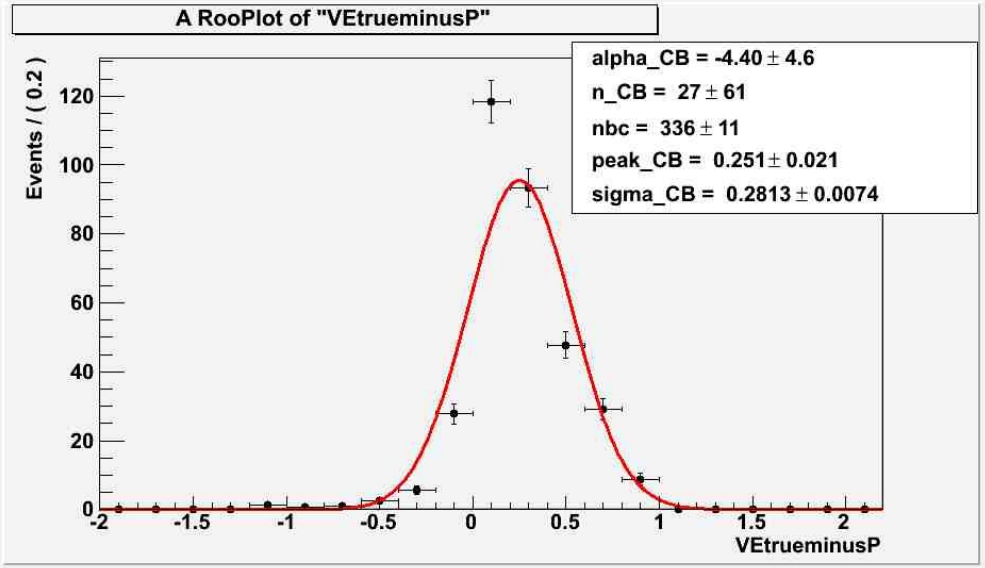


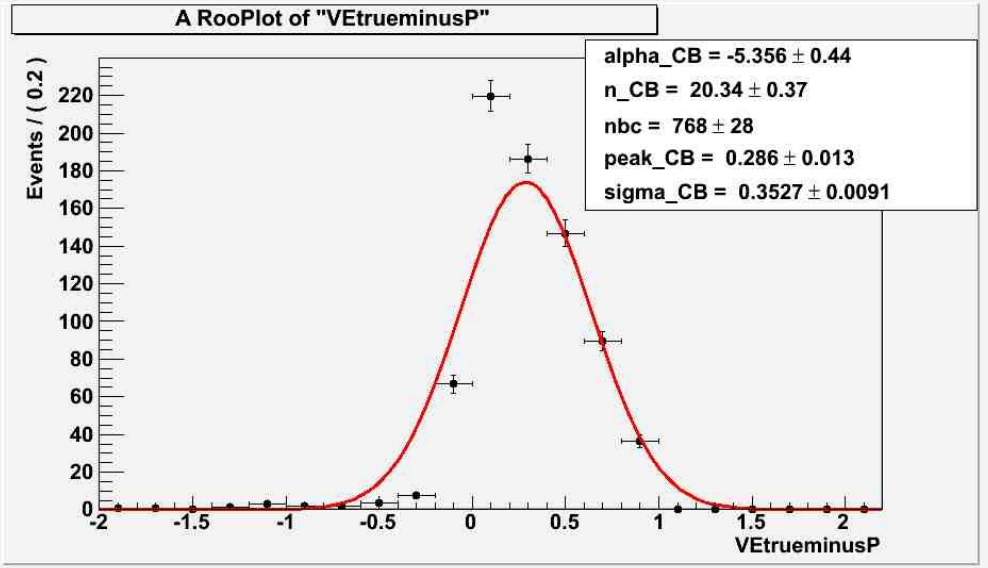
Default algorithm resolution for the DP sample

pt>85 GeV && pt< 100 GeV



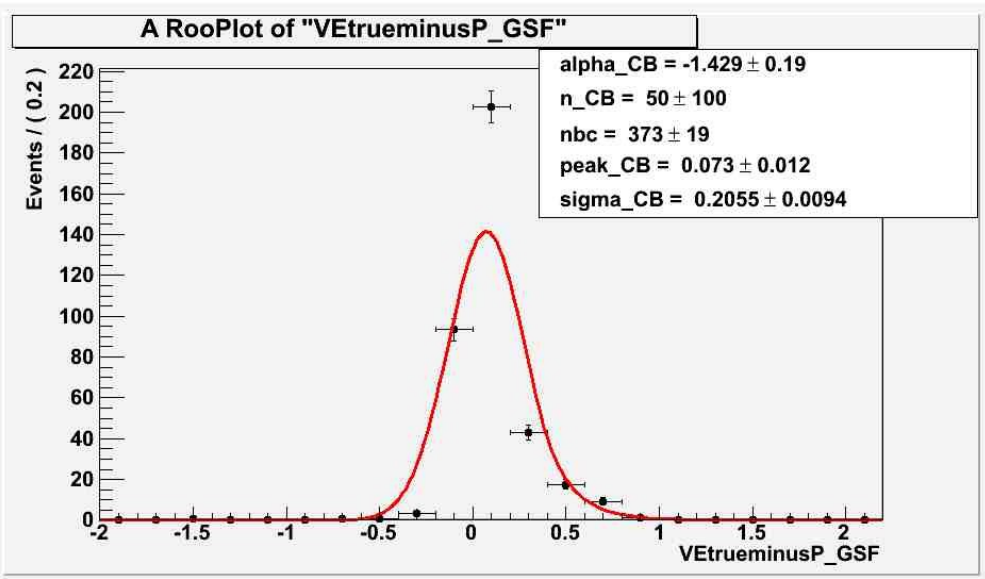


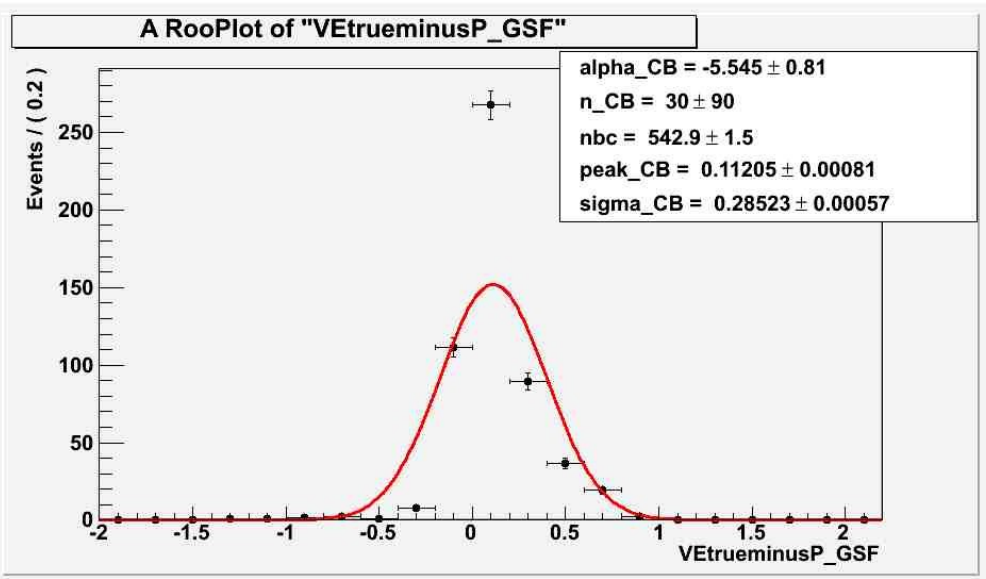


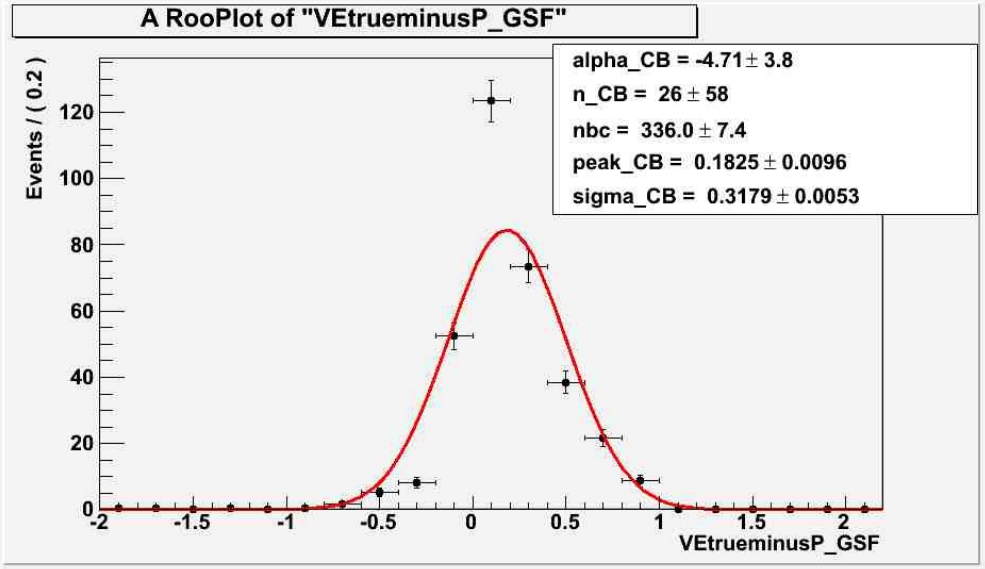


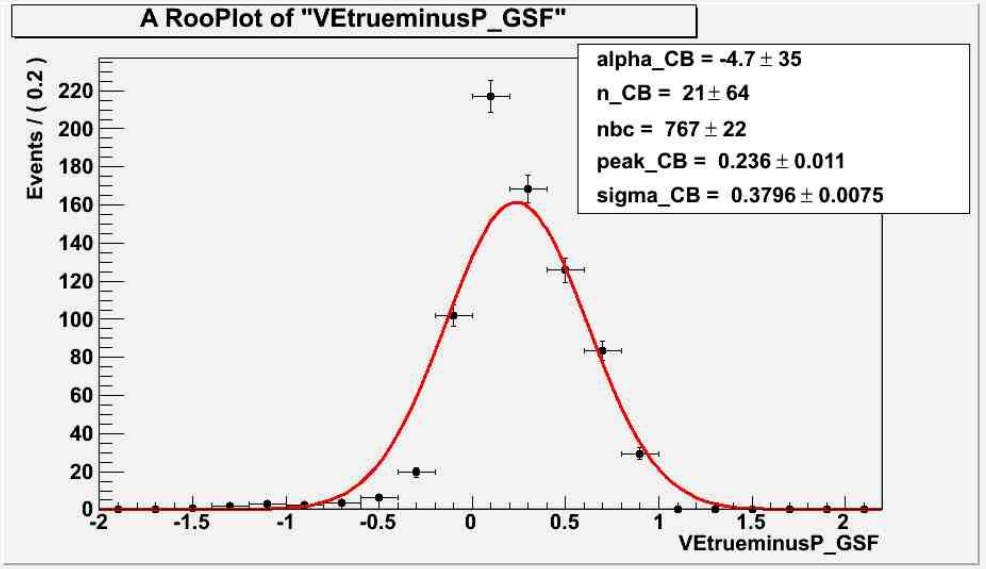
GSF algorithm resolution for the DP sample

pt>85 GeV && pt< 100 GeV









DNA algorithm resolution for the DP sample

pt>85 GeV && pt< 100 GeV