CDF/DOC/TOF/PUBLIC/7696
Version 1.0
16th June 2005

Jet Probability Mistag Asimmetry in Gen5

G. Gómez, E. Palencia, T. Rodrigo, L. Scodellaro[1], R. Vilar
Instituto de Física de Cantabria

Abstract

We present a study of the light quark mistag asymmetry for the Jet Probability (JETPROB) tagger in 5.3.3. Jet probability cut values of 1% and 5% are considered. By using the JET50 data sample, we measure the ratio β between the number of positive and negative tagged light quark. We find a value $\beta_1 = 1.57 \pm 0.10$ and $\beta_2 = 1.27 \pm 0.10$ for jet probability cut at 1% and 5% respectively.

1 Introduction

The estimate of the contribution of light flavoured jets to tagged samples (*mistag*) is an important ingredient for analysis based on b-tagging algorithms. The number of negative tags provides a natural estimate of the mistags due to detector resolution effects. Unfortunately, long-lived light particle decays and secondary interactions in detector material provide an additional contribution to the mistags.

In this note, we describe a measurement of the asymmetry between the numbers of positive and negative tags in light jets for the JETPROB tagger[1, 2, 3, 4]. Jet probability cut values of 1% and 5% are considered. We use Gen5 standard ntuples of JET50 samples for data and Monte Carlo (HERWIG generator). The method we use follows the guidelines of similar studies that have been performed for the SECVTX tagger [5, 6].

2 Strategy

The strategy we use has been described in CDF note 7585. In that work, the distribution of SECVTX pseudo-$c\tau$ for c, b and no-matched jets in Monte Carlo events were used to fit data distribution and determine the flavour composition of tagged jets samples. scshapeSecVtx mass and momentum were also considered. Unfortunately, the JETPROB algorithm does not provide a secondary vertex, so there is not a natural jet variables to characterize the behaviour of light and heavy flavoured tagged jets. We decided to use the set of jet tracks passing the jet probability good quality criteria to build a jet variable sensitive to the flavour content of the jet itself. We selected the following variables:

- maximum track impact parameter d_0 in the jet,

[1]Luca Scodellaro is supported by the European Union funding under the Research Training Network contract: HPRN-CT-2002-00292, *Probe for New Physics*

- maximum significance S_{d_0} of the track impact parameter in the jet,
- mass of the system of tracks with $|d_0| > 0.01\ cm$,
- mass of the system of tracks with $S_{d_0} > 2\ cm$,
- transverse momentum $P_T^{rel.}$ with respect to the jet direction of the system of tracks with $|d_0| > 0.01\ cm$,
- transverse momentum $P_T^{rel.}$ with respect to the jet direction of the system of tracks with $S_{d_0} > 2\ cm$.

We use only tracks with positive (negative) impact parameter for positive (negative) tagged jets. Since the impact parameter resolution is overestimated in Monta Carlo, and since it could not describe correctly the ratio between the tags due to resolution effects and the ones coming from tracks really desplaced, we subtract the distribution obtained for negative tags to the one we get for the positive tags. In this way, we construct templates for c, b no-matched jets in Monte Carlo. We use these templetes to fit the data distribution and obtain the flavour composition of the excess of postive tags observed in data. In order to compute the total numbers of positive and negative tags from light, c and b-jets, we use their Monte Carlo distributions to derive the numbers of negative tags from their contribution to the positive-negative excess. Successively, we normalize the sum N^-_{MC} of the numbers of negative tags obtained for light, c and b-jets to the number N^-_{data} of negative tags observed in data. We call the ratio between N^-_{data} and N^-_{MC} the *negative scale factor* SF^-. We scale our estimates of the negative tags for light, c and b-jets by SF^-. Finally, we sum the fitted values of positive tags excess to the negative tags to obtain the numbers of positive tags from light, c and b-jets in data. We can in this way measure the ratio β of positive and negative tags from light jets.

3 Fit Results

3.1 Jet Probability Cut 1%

Figs. 1 to 12 shows the distributions of the selected variables for data and c, b and no-matched jets in Monte Carlo and the results of the fits of the positive tag excess observed in data. Table 1 summarizes the results of the fits. The fit quality is reasonable. The values of the fractions of light, c and b-jets fluctuate over the different fits, but the value of β_1 is not too sensitive to these fluctuations. We average the differents determination of β_1 and we obtain $\beta_1 = 1.57 \pm 0.10$.

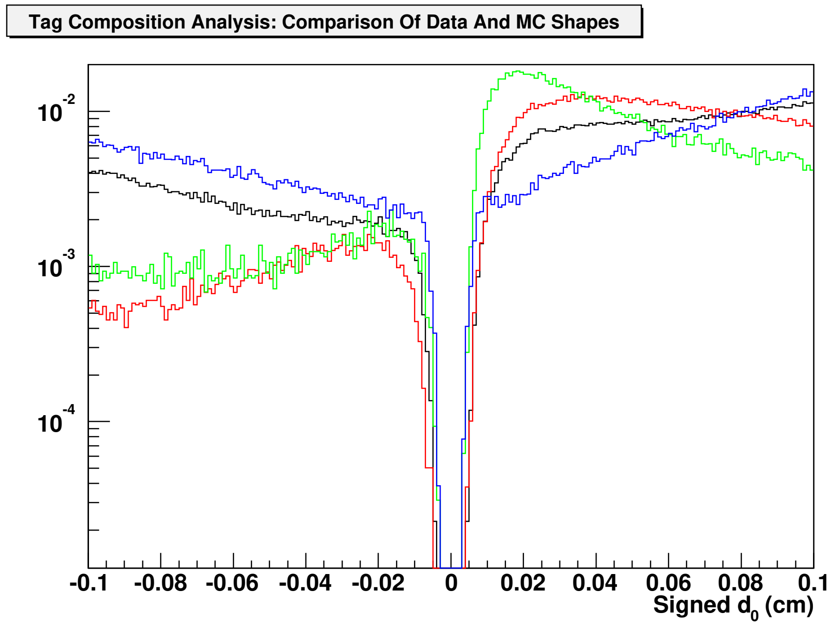

Figure 1: Tamplates for tag composition fit: maximum impact parameter d_0 of the good jet probability tracks inside tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 1%.

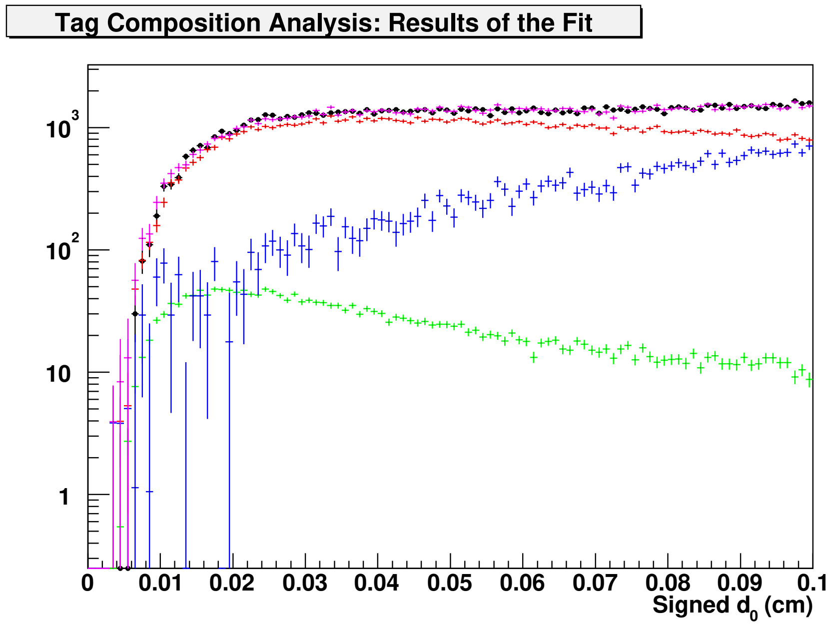

Figure 2: Result of the fit of the positive tag excess in JET50 data as a function of the maximum impact parameter d_0 of the good jet probability tracks inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 1%.

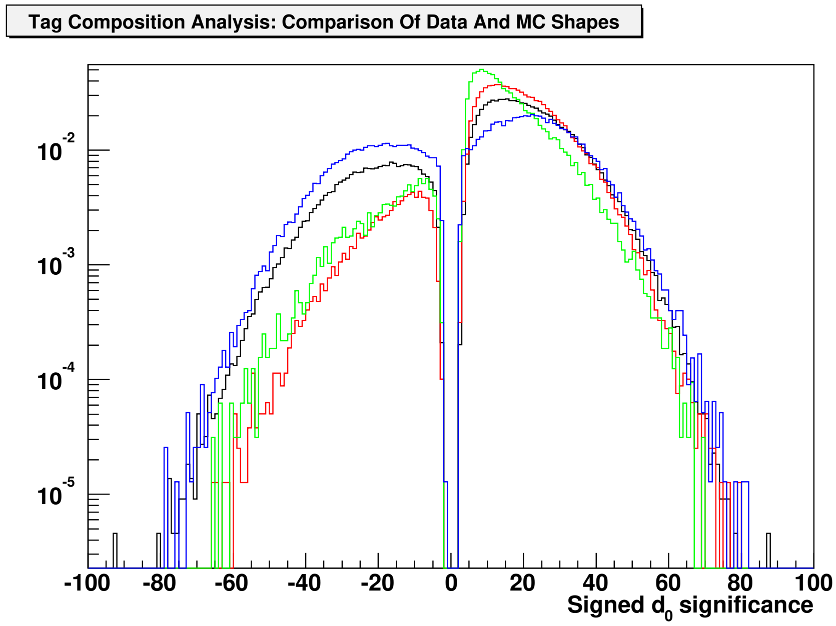

Figure 3: Tamplates for tag composition fit: maximum impact parameter significance S_{d_0} of the good jet probability tracks inside tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 1%.

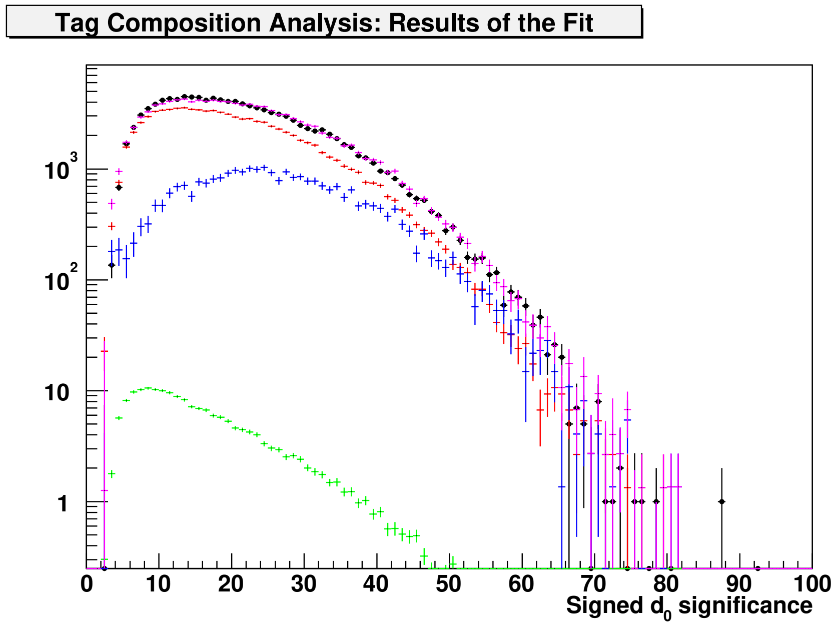

Figure 4: Result of the fit of the positive tag excess in JET50 data as a function of the maximum impact parameter significance S_{d_0} of the good jet probability tracks inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 1%.

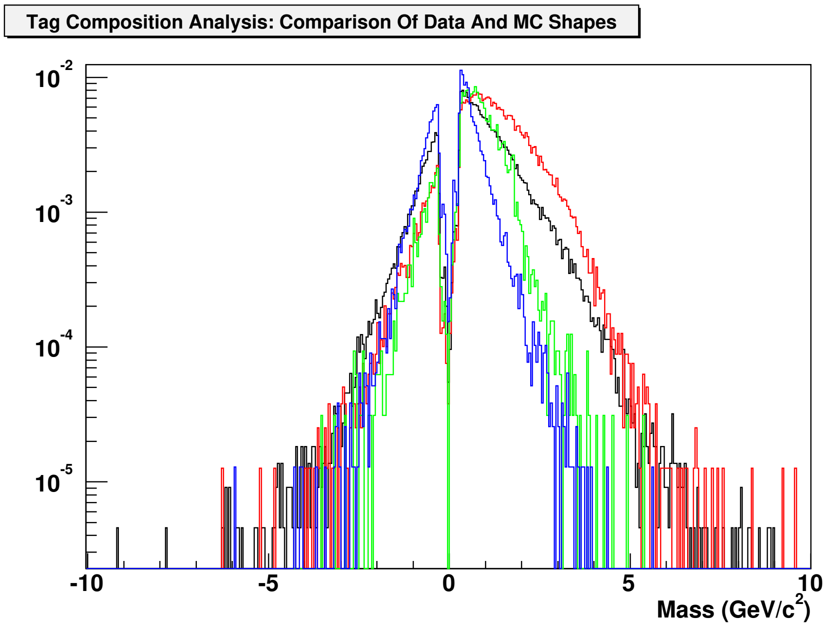

Figure 5: Tamplates for tag composition fit: mass of the good jet probability tracks with $d_0 > 0.01$ cm inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 1%.

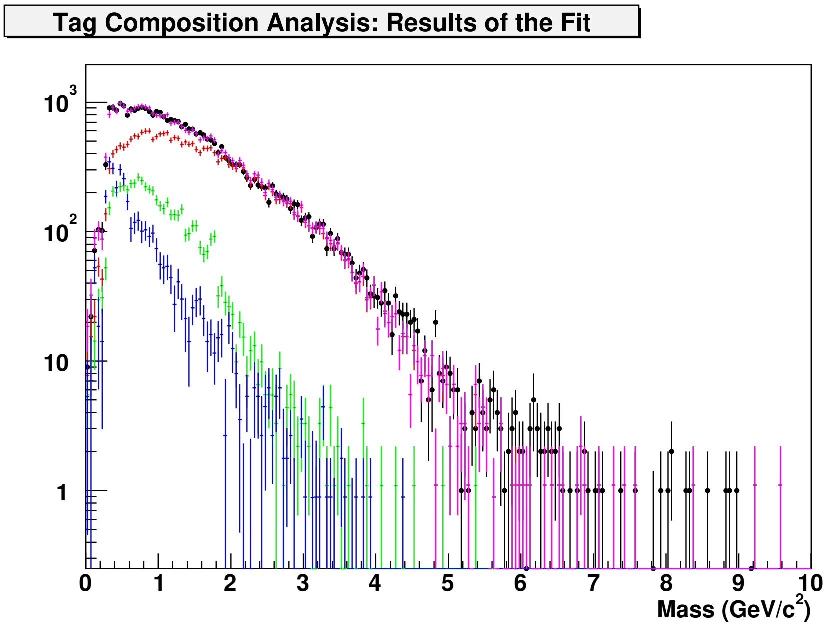

Figure 6: Result of the fit of the positive tag excess in JET50 data as a function of the mass of the good jet probability tracks with $d_0 > 0.01$ cm inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 1%.

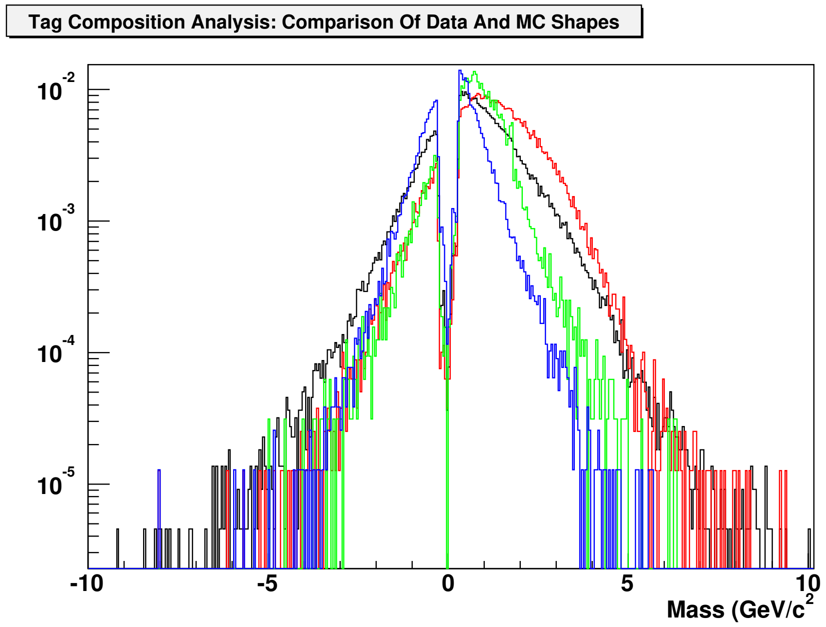

Figure 7: Tamplates for tag composition fit: mass of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 1%.

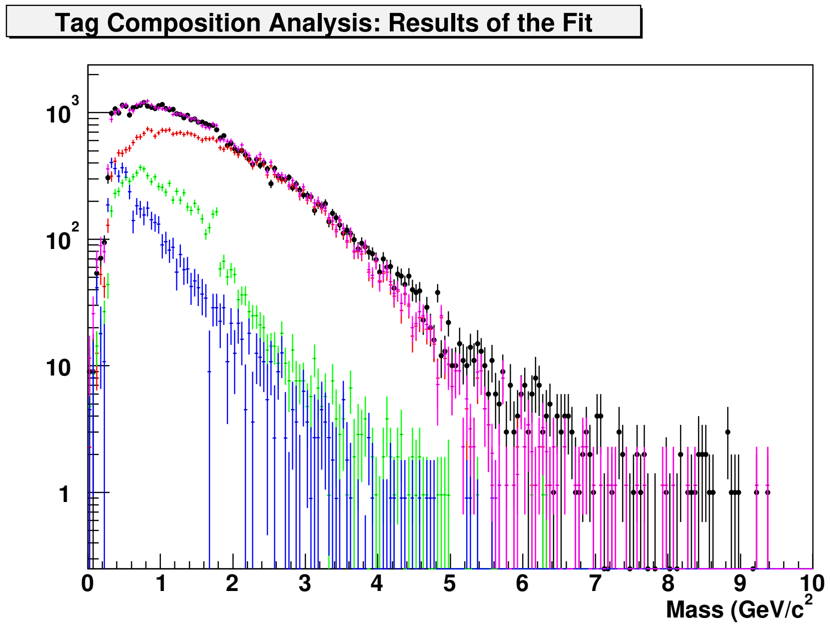

Figure 8: Result of the fit of the positive tag excess in JET50 data as a function of the mass of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 1%.

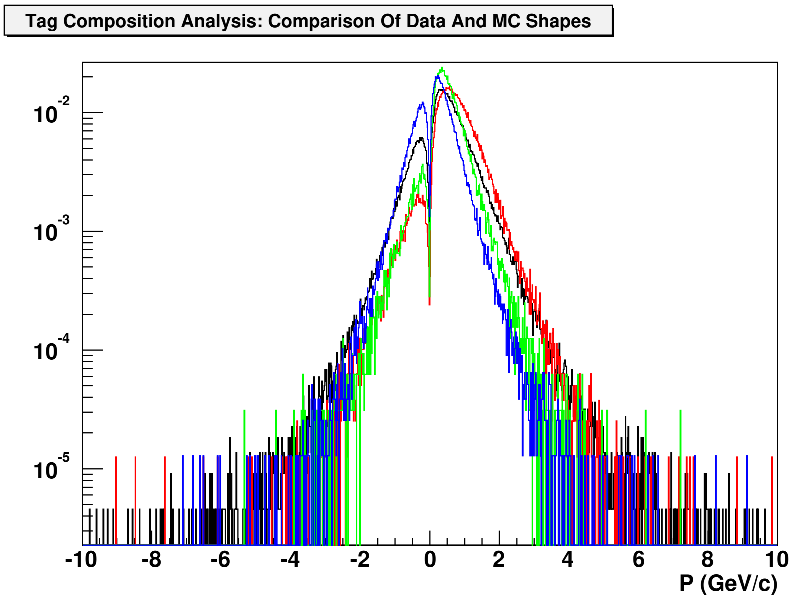

Figure 9: Tamplates for tag composition fit: relative transverse momentum of the good jet probability tracks with $d_0 > 0.01$ cm inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 1%.

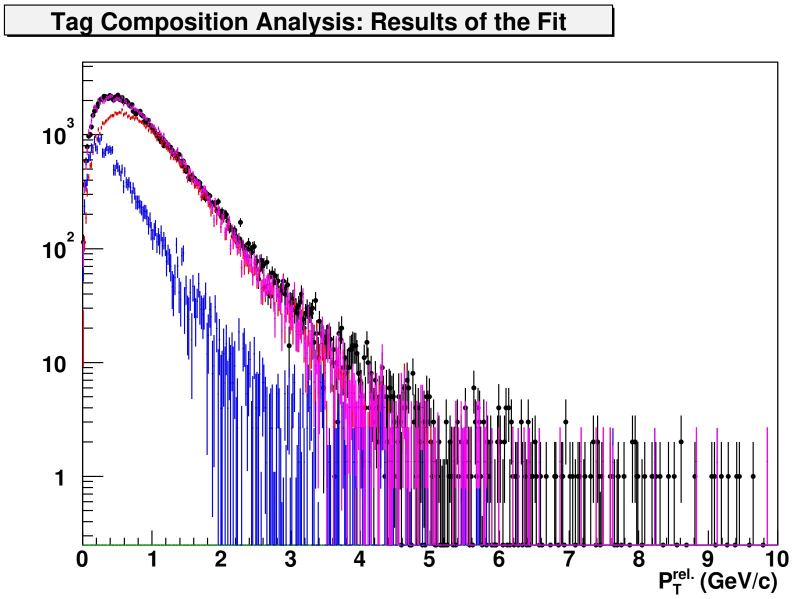

Figure 10: Result of the fit of the positive tag excess in JET50 data as a function of the relative transverse momentum of the good jet probability tracks with $d_0 > 0.01$ cm inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 1%.

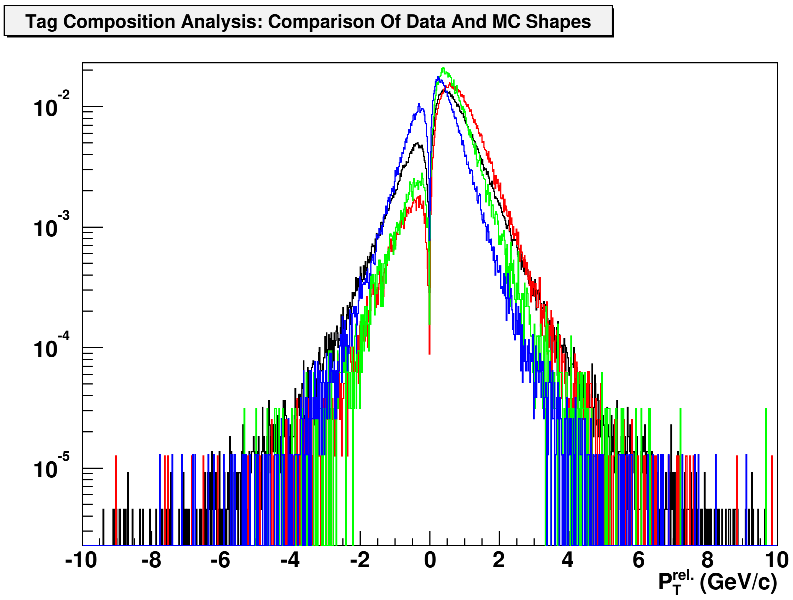

Figure 11: Tamplates for tag composition fit: relative transverse momentum of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 1%.

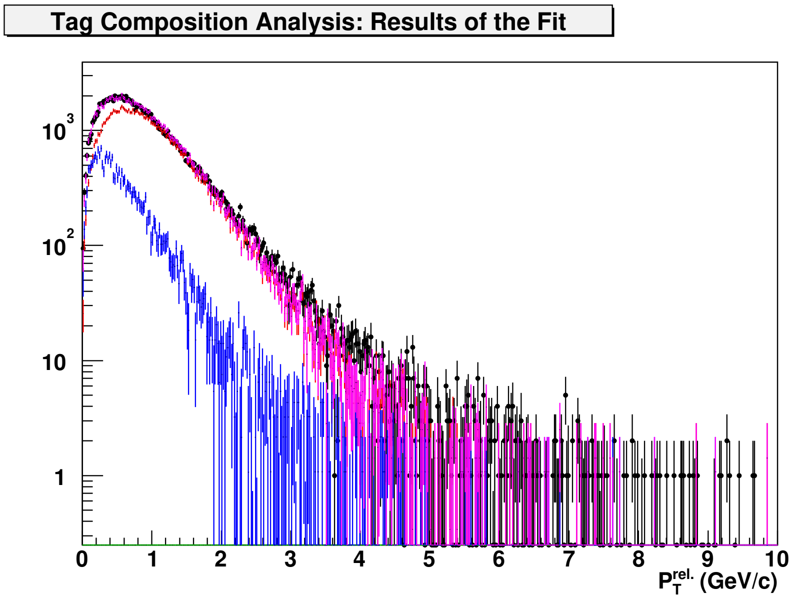

Figure 12: Result of the fit of the positive tag excess in JET50 data as a function of the relative transverse momentum of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 1%.

	Max. d_0	Max. S_{d_0}	Mass ($d_0 > 0.01\ cm$)	Mass ($S_{d_0} > 2.$)	$P_T^{rel.}$ ($d_0 > 0.01\ cm$)	$P_T^{rel.}$ ($S_{d_0} > 2.$)
χ^2/dof	1.07515	3.4165	0.980093	1.46056	1.82739	1.72597
data positive	167964	167964	42141	61891	166872	167912
data negative	51267	51267	10861	17929	50598	51237
data excess	116697	116697	31280	43962	116274	116675
b fraction	0.753737	0.758266	0.723413	0.711912	0.765048	0.808699
c fraction	0.0195661	0.00158336	0.170116	0.181386	4.69904e-07	0
light fraction	0.226697	0.240151	0.106471	0.106702	0.234952	0.191301
b excess	87967.5	88488.2	22630.6	31300.4	88963.8	94364.2
c excess	2283.62	184.771	5321.99	7975.24	0.0598812	1.03581e-07
light excess	26491.5	28028.3	3334.85	4698.04	27355.8	22351
neg. SF	1.12388	1.0792	1.47006	1.53577	1.10635	1.27491
b positive	97680.7	97870.5	25859.5	36428.8	98557.2	106184
c positive	2637.77	212.286	6441.03	9731.05	0.0689326	1.21803e-07
light positive	67691.2	69885.5	9847.89	15742.8	68360.5	61768.2
b negative	9713.2	9382.34	3228.92	5128.39	9593.36	11819.8
c negative	354.142	27.5152	1119.04	1755.81	0.00905148	1.82219e-08
light negative	41199.7	41857.1	6513.04	11044.8	41004.6	39417.2
β	1.643	1.66962	1.51203	1.42536	1.66714	1.56704

Table 1: Summary of the results of the fit for the tag composition for jet probability cut 1%.

3.2 Jet Probability Cut 5%

Figs. 13 to 24 shows the distributions of the selected variables for data and c, b and no-matched jets in Monte Carlo and the results of the fits of the positive tag excess observed in data. Table 2 summarizes the results of the fits. The fit quality is reasonable. The values of the fractions of light, c and b-jets fluctuate over the different fits, but the value of β_2 is not too sensitive to these fluctuations. We average the differents determination of β_2 and we obtain $\beta_2 = 1.27 \pm 0.10$.

4 Conclusions

We measured the mistag asimmetry for the JetProb tagger by fitting the distributions of positive tag excess for several variables. We obtain an average value of $\beta_1 = 1.57 \pm 0.10$ and $\beta_2 = 1.27 \pm 0.10$ for a jet probability cut of 1% and 5% respectively.

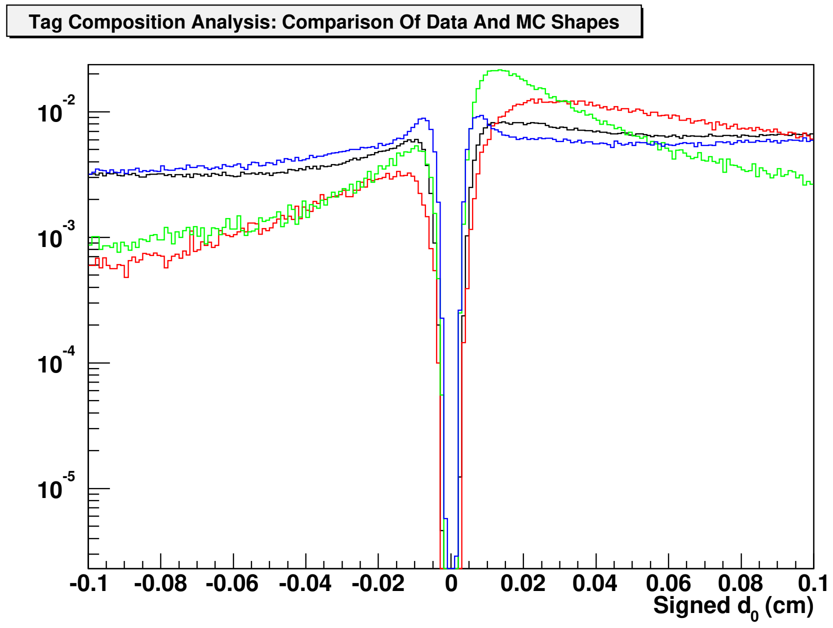

Figure 13: Tamplates for tag composition fit: maximum impact parameter d_0 of the good jet probability tracks inside tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 5%.

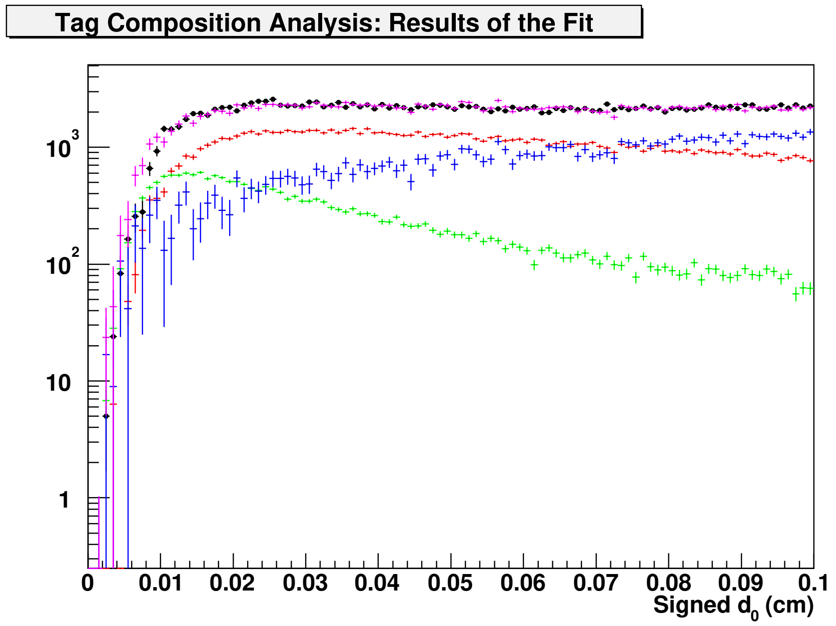

Figure 14: Result of the fit of the positive tag excess in JET50 data as a function of the maximum impact parameter d_0 of the good jet probability tracks inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 5%.

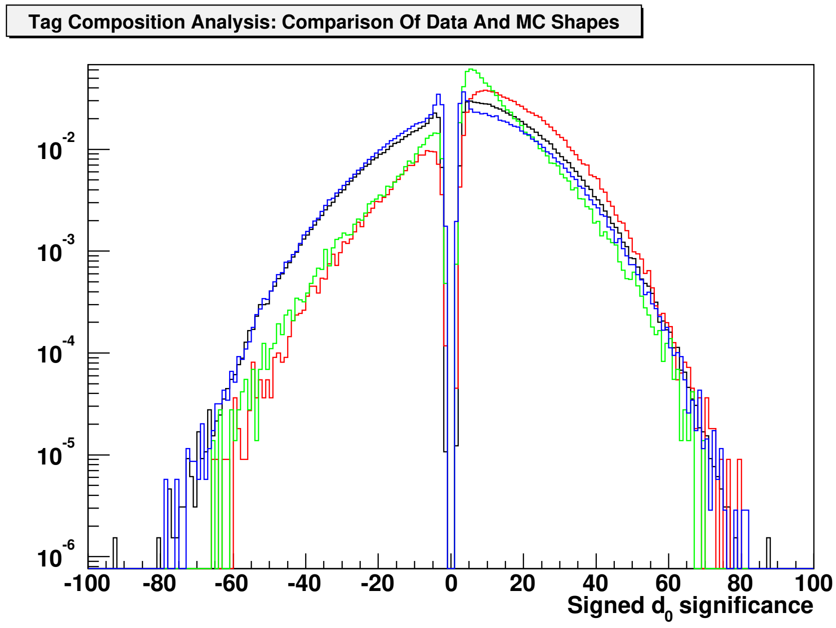

Figure 15: Tamplates for tag composition fit: maximum impact parameter significance S_{d_0} of the good jet probability tracks inside tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 5%.

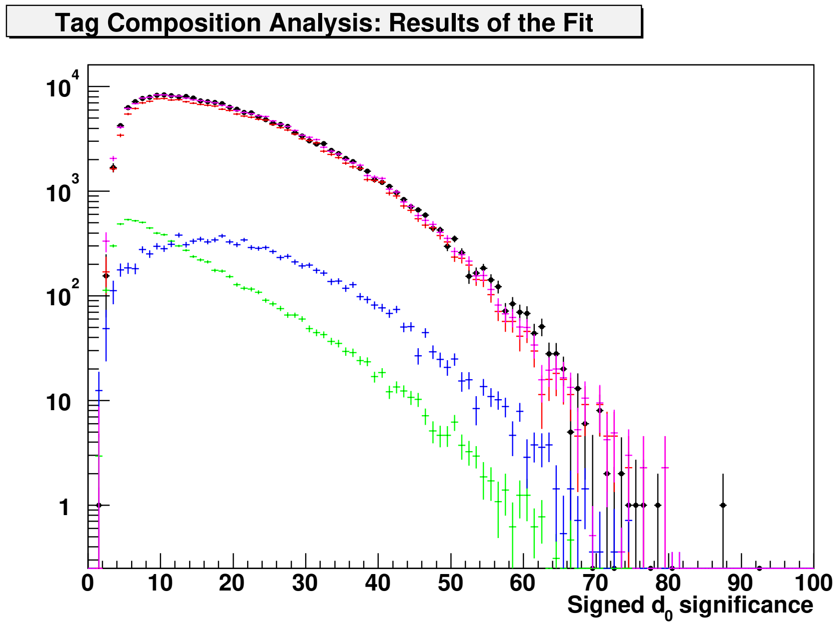

Figure 16: Result of the fit of the positive tag excess in JET50 data as a function of the maximum impact parameter significance S_{d_0} of the good jet probability tracks inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 5%.

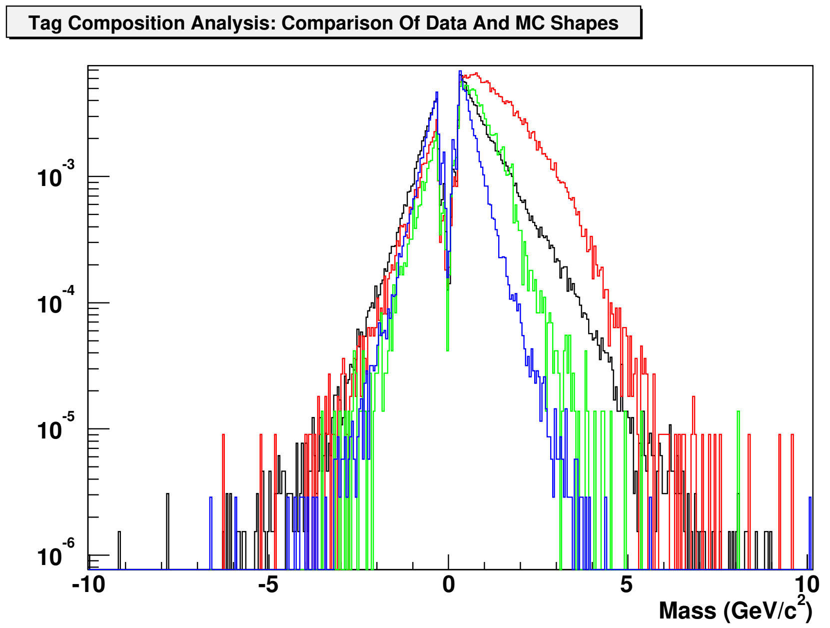

Figure 17: Tamplates for tag composition fit: mass of the good jet probability tracks with $d_0 > 0.01\ cm$ inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 5%.

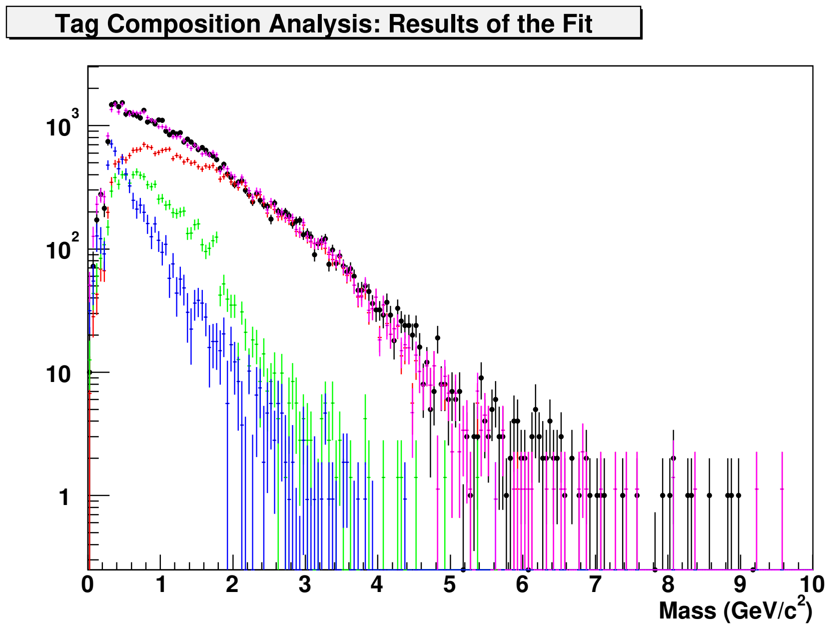

Figure 18: Result of the fit of the positive tag excess in JET50 data as a function of the mass of the good jet probability tracks with $d_0 > 0.01\ cm$ inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 5%.

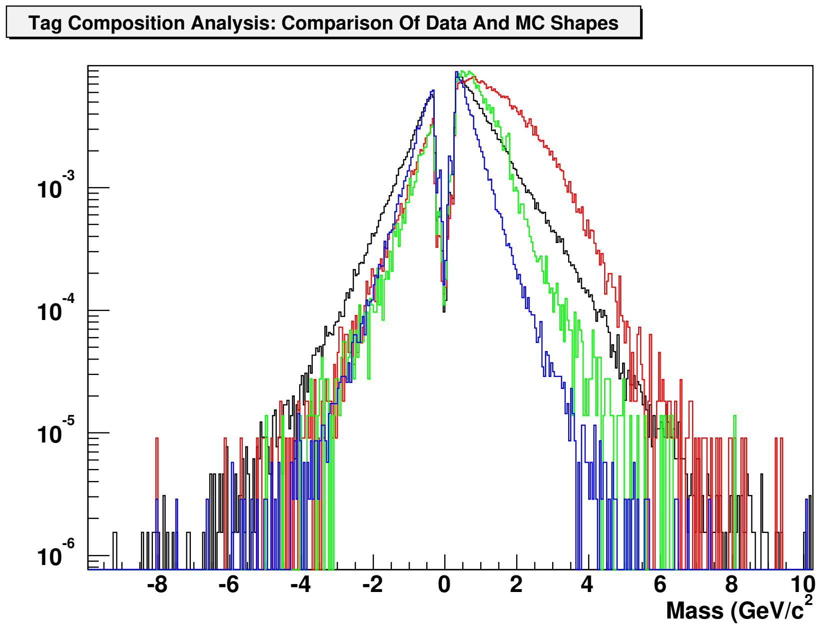

Figure 19: Tamplates for tag composition fit: mass of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 5%.

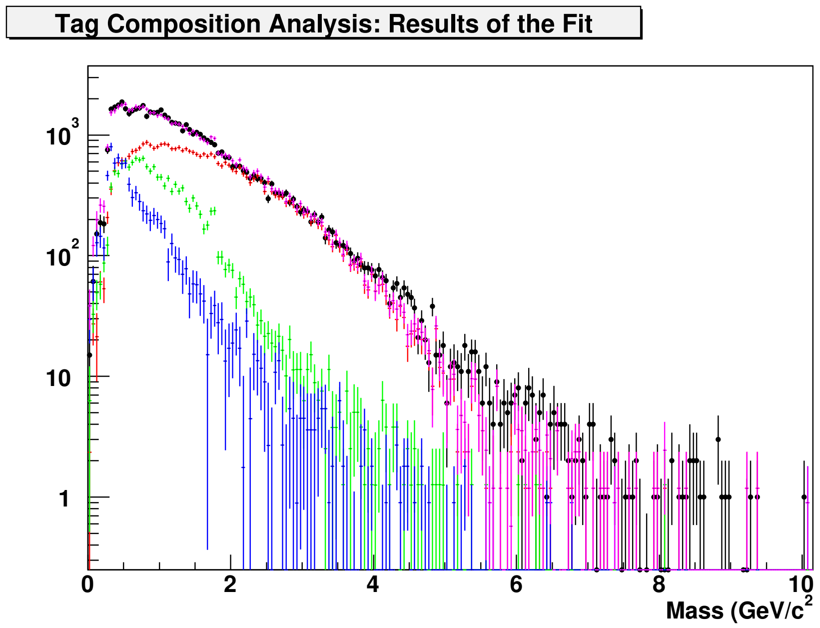

Figure 20: Result of the fit of the positive tag excess in JET50 data as a function of the mass of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 5%.

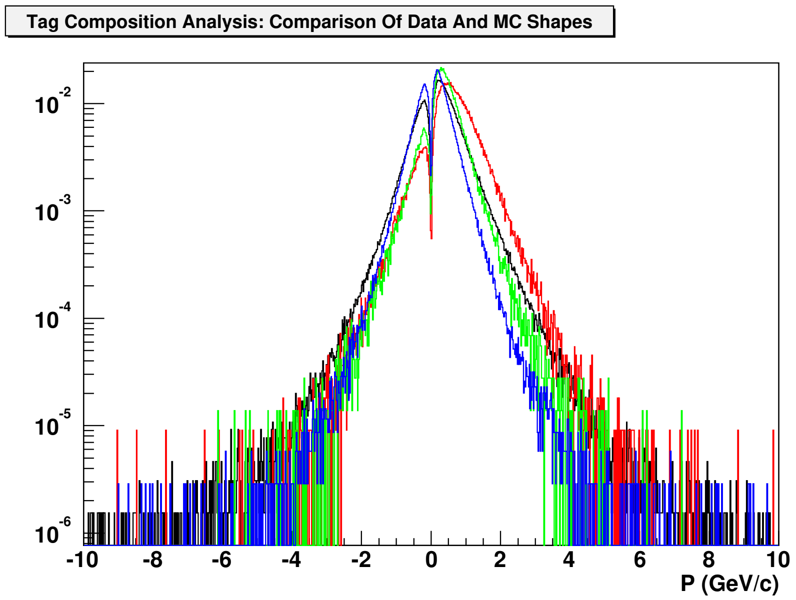

Figure 21: Tamplates for tag composition fit: relative transverse momentum of the good jet probability tracks with $d_0 > 0.01$ cm inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 5%.

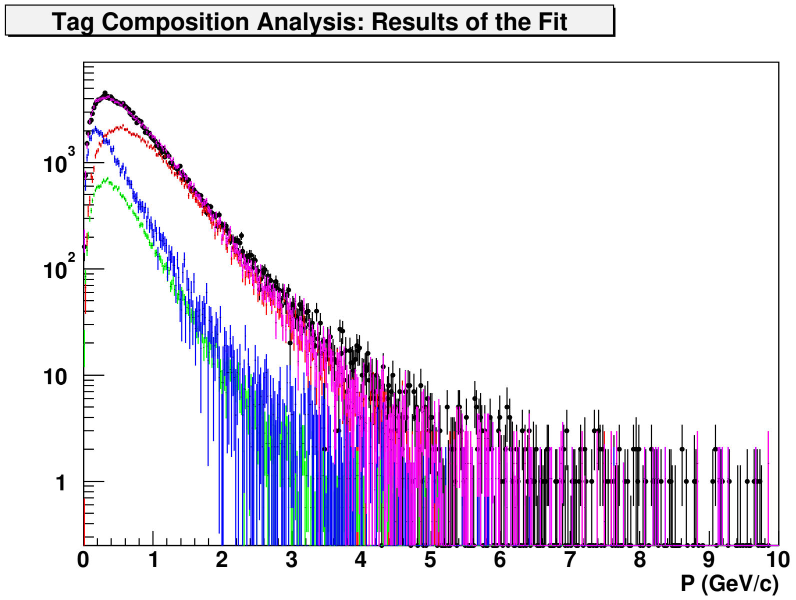

Figure 22: Result of the fit of the positive tag excess in JET50 data as a function of the relative transverse momentum of the good jet probability tracks with $d_0 > 0.01$ cm inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 5%.

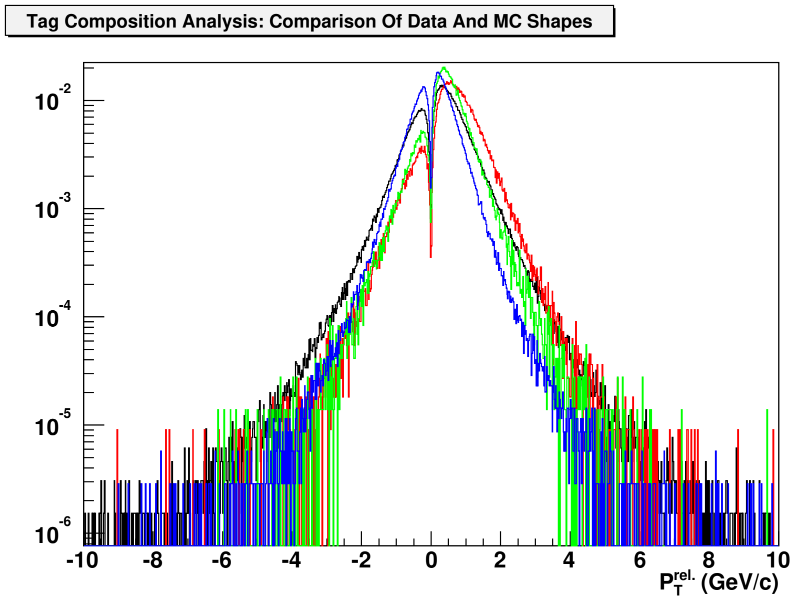

Figure 23: Tamplates for tag composition fit: relative transverse momentum of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets in JET50 data (black histogram), b-matched jets (red), c-matched jets (green) and no-matched jets (blue). Jet Probability cut is 5%.

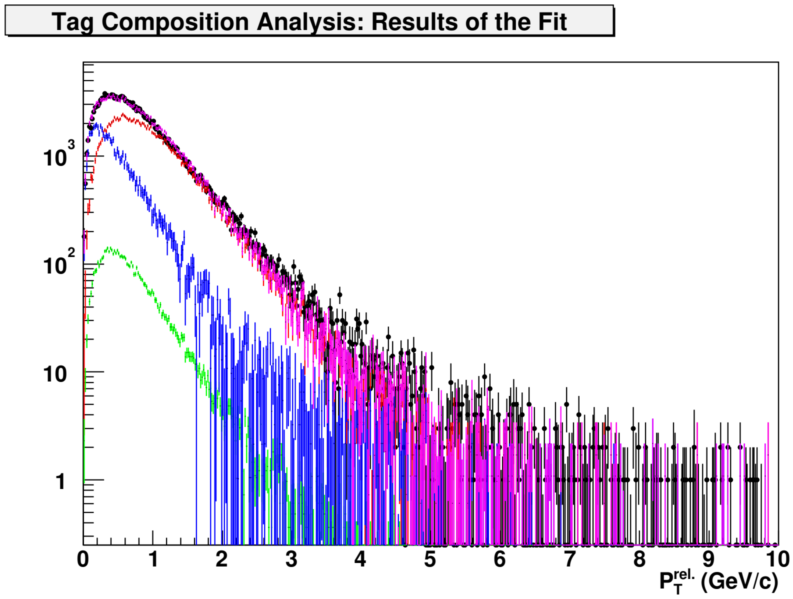

Figure 24: Result of the fit of the positive tag excess in JET50 data as a function of the relative transverse momentum of the good jet probability tracks with $S_{d_)} > 2.$ inside the tagged jets: observed data are the black histogram, while the fitted component from b, c and no-matched jets are the red, green and blue histograms respectively. The purple histogram is the sum of the three fitted components. Jet Probability cut is 5%.

	Max. d_0	Max. S_{d_0}	Mass ($d_0 > 0.01\ cm$)	Mass ($S_{d_0} > 2.$)	$P_T^{rel.}$ ($d_0 > 0.01\ cm$)	$P_T^{rel.}$ ($S_{d_0} > 2.$)
χ^2/dof	1.4788	1.3713	1.2042	1.533	1.73537	1.81718
data positive	423764	423764	69491	112818	405803	423628
data negative	226568	226568	29847	55742	211016	226448
data excess	197196	197196	39644	57076	194787	197180
b fraction	0.508019	0.915631	0.626568	0.614697	0.593409	0.68017
c fraction	0.113542	0.0362404	0.215326	0.243708	0.122934	0.0283206
light fraction	0.378439	0.0481286	0.158106	0.141595	0.283657	0.291509
b excess	100199	180594	24843.2	35090.1	115610	134142
c excess	22396.3	7148.49	8538.5	13913.4	23952.4	5585.86
light excess	74834.2	9517.16	6281.21	8104.42	55390.7	57639.2
neg. SF	0.969834	3.53326	1.57823	1.74315	1.27066	1.20861
b positive	119416	306779	30447.7	45004.2	142938	166173
c positive	28562.2	14318.3	11846	20093.3	32133.5	7497.97
light positive	276019	102731	27216.3	47752.4	230898	250144
b negative	19217	126184	5604.49	9914.11	27328	32031.1
c negative	6165.87	7169.84	3307.46	6179.9	8181.12	1912.11
light negative	201185	93214	20935.1	39648	175507	192505
β	1.37197	1.1021	1.30003	1.20441	1.3156	1.29942

Table 2: Summary of the results of the fit for the tag composition for jet probability cut 1%.

References

[1] D. Gerdes, *B identification Using Jet Probability*, CDF Note 2023.

[2] D. Acosta et al., *Introduction to Run II Jet Probability Heavy Flavor Tagger*, CDF Note 6315.

[3] A. Sukhanov et al., *Efficiency of Jet Probability Heavy Flavor Tagger Used for 2005 Winter Conferences*, CDF Note 7444.

[4] G. Gómez et al., *JetProbability Tag Rate Matrix for 5.3.3*, CDF Note 7442.

[5] H. Bachacou, SecVtx *Tag Composition and Heavy Flavor Fraction Studies in QCD Jets*, CDF Note 6739.

[6] J. G. da Costa, *Measurement of the* SecVtx*/ Mistag Asymmetry in 5.3.3*, CDF Note 7585.