EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

Version l3note 2108
June 24, 1997

THE UCSD B TAG FOR 1996

UCSD L3 Group:
James G. Branson, Aaron Dominguez, Ian Fisk and Gerhard Raven.

Abstract

In this note we describe the procedure used to tag the production of b quarks in L3. This technique exploits the full 3-dimensional reconstruction of tracks from the TEC+TECZ+SMD. It improves on methods used previously in L3 and in other experiments. We also present the performance on the 1996 data sample where this tag has been used in searches for Higgs particles.

1 Introduction

Since the L3 Silicon Microvertex Detector (SMD) became operational in 1994, L3 has been improving its ability to tag b quark production by using B decay length information. A good 2-dimensional tag, which followed the methods employed in other collaborations, was developed [1, 2]. From our studies, we determined that this tag was well optimized and performed well given the information it was using. However, due to the smaller size of the L3 tracker, the low magnetic field, the scarcity of z information, and inefficiency plus noise in the SMD, the L3 tag fell far short of the performance achieved in other experiments.

The UCSD[1)] group has made a concerted effort to improve b tagging in L3 by using more information and improving on methods used in the past. Our primary motivation has been the need for a good b tag to be competitive in the Higgs search. The improvements made over previous tags in L3 include:

- Evaluation of the optimal weights of the SMD hits in the track fit.
- The implementation of a Kalman Filter track fitter to get good error estimates for tracks and to make optimal use of SMD information particularly for low momentum tracks.
- An improved 3D primary vertex fit.
- Track selection and classification in categories emphasizing the quality of 3D information.
- Use of an improved K_0 and Λ rejection from Chris Tully.
- Use of a 3D decay length by averaging between the $r\phi$ and the sz projections.
- Inclusion of an error on the B direction in calculation of the decay length error.
- Including negative decay length tracks in the calculation of an event (or jet) discriminant from track probabilities.
- Weighting tracks according to their potential for decay length measurement in the discriminant calculation.
- Enhancing the basic lifetime tag by adding information form inclusive leptons, secondary vertices, and jet variables.

After all of the above are done, each step makes only a small improvement on the tag performance. Taken together however, our ability to tag improves significantly.

The primary tagging information comes from the discriminant based on decay length and hence on lifetime information. This has been carefully checked using jet tags on Z production. We have used the basic lifetime tag to check all the other variables, allowing us to gain confidence in the optimal tag.

In all of the comparisons to MC shown below, we have used the realistic Monte Carlo which includes the measured SMD efficiency and noise rate. As will be shown, further adjustments to detector performance were needed to get a perfect match between data and Monte Carlo.

The lifetime tag starts by reconstructing the 3-D primary vertex of the event; this is described in detail in section 2. The next step is to compute, track-by-track, the decay length along the jet direction.

[1)]It is particularly ironic that a university named after all the lighter quarks should work on a b tag. However, a simple reversal of the cut on the b tag variable effectively tags the production of one of the ucsd quarks.

This is described in 5. Finally, the probability that a track is compatible with the primary vertex is derived, and the individual track probabilities are combined into a jet, hemisphere or event tag, as described in 6. Additional variables are then computed and combined with the lifetime variables. This is described in section 8.

2 Primary Vertex Reconstruction

The primary vertex reconstruction starts from the average N-event or fill vertex, which is a measure for the position of the LEP beam spot. As the vertical size of the beam spot is small compared to the attainable event vertex resolution, the primary vertex is constrained to the beam spot envelope in the plane perpendicular to the beam-direction. The beam spot position along the beam direction is not used, as it is two orders of magnitude larger than the vertex resolution based on the tracking information alone.

To reconstruct the primary vertex, the following χ^2 is minimized:

$$\chi^2_{N_{\rm trk}} = \sum_{i=1}^{N_{\rm trk}} \left(\vec{t}_i - \vec{f}(\vec{v},\vec{q}_i)\right)^T G_i^{-1} \left(\vec{t}_i - \vec{f}(\vec{v},\vec{q}_i)\right) + (\vec{v} - \vec{v}_{f\rm ill})^T V_{f\rm ill}^{-1} (\vec{v} - \vec{v}_{f\rm ill})$$

where $\vec{t}_i$ is the vector of measured parameters for the $i^{\rm th}$ track and G_i the corresponding covariance matrix; $\vec{f}(\vec{v},\vec{q}_i)$ are the predicted measurements if the track originated from the vertex $\vec{v}$ with momentum $\vec{q}_i$, $\vec{v}_{f\rm ill}$ is the fill-vertex and $V_{f\rm ill}$ is the matrix describing the size of the beam spot.

The initial fit uses all tracks which pass the following selection:

1. Track has successfully been fitted with the Kalman filter. This implies:
   (a) span > 12
   (b) number of hits > 10
   (c) $p_\perp > 0.150$ MeV
2. $|\mathrm{d}_{r\phi}| < \min(10\text{ mm}, 5(\sigma_{\mathrm{d}_{r\phi}} \oplus \sigma_{f\rm ill}))$
3. $|\mathrm{d}_{sz}| < 100$ mm

where $\mathrm{d}_{r\phi}$ and d_{sz} are the signed[2)] distances of closest approach from the track to the fill vertex in the $r\phi$ and sz projections.

For each track the probability $P_i = P(\chi^2_{N_{\rm trk}} - \chi^2_{N_{\rm trk}-1})$ is computed, and if either the probability of the vertex fit, $P(\chi^2_{N_{\rm trk}})$, is below 4% or any P_i is below 2%, the track with the lowest P_i is removed. The fit procedure is iterated until no track needs to be removed. If at the end less than three tracks remain, which happens in less than 1% of the hadronic events, the fill vertex is used as the event vertex. The vertex resolution attained on the 1996 MC is shown in Figure 1. For uds events, the RMS of the full distribution is (51,10,118) microns in the (x,y,z) direction. For b quarks these RMSs increase to (81,11,154) microns.

3 Corrections to the track fit

Two types of corrections for systematic bias are applied to the Kalman filter track fit. All corrections are determined sequentially, in the order listed, and after all preceding corrections have been applied and the primary vertex redetermined.

[2)] $\mathrm{d}_{r\phi}$ is signed according to the standard L3 convention; d_{sz} has a similar signing convention

3.1 Systematic shifts

After reconstructing the primary vertex, one expects that the mean track dca ,$d_{r\phi}$, used in the fit to be zero. However, as shown in Figure 2, this is not the case. To correct for these small biases a sample of resolution tracks is used. This sample contains tracks which have been used in the computation of the event vertex, and which have a negative apparent lifetime. The weighted mean of the $d_{r\phi}$ distributions are computed as a function of the azimuthal track angle, ϕ at which a track enters the outer TEC volume and as a function of the angle between this intersection and the closest inner TEC anode plane, $\phi_{\rm local}$. Tracks are binned in ϕ, and for each bin the tracks are corrected for the mean observed displacement. Next the tracks are binned in $\phi_{\rm local}$, and again corrected for the mean displacement. This procedure is iterated several times.

A similar procedure is applied to the d_{sz} distribution, in this case as a function of polar angle and z offset.

The weighted means before and after this correction are shown in Figure 2 for data and Figure 3. The figures shows that shifts are needed both in data and MC, but that the data shifts are much larger even though the TEC and SMD have been carefully aligned.

3.2 Resolution Corrections

The Kalman filter track fit estimates the track resolution, based on the available resolution functions of TEC and SMD, and the GEANT description of the material. As such, the track fit is capable of computing the resolution of each individual track, but is limited by the understanding of these initial assumptions. One can 'gauge' the effect of the resolution functions on the track fit by comparing the *average* resolution as observed from the $d_{r\phi}$ and d_{sz} distributions in the actual data with the fit result. This is done in a way similar to [3].

The $d_{r\phi}$ error as determined by the Kalman filter track fit, $\sigma_{\rm KF}$ versus $p_\perp^2 \sin\theta$, is shown in Figure 4.

The *average* is, as expected, well described by the following parameterization:

$$\sigma_{\rm KF}^2 = a_{\rm KF}^2 + \frac{b_{\rm KF}^2}{p_\perp^2 \sin\theta}$$

The same parameterization can be determined directly from the $d_{r\phi}$ distribution itself. One can perform an unbinned likelihood fit, with the following likelihood function:

$$\mathcal{L}(d_{r\phi}, p_\perp, \theta) = \frac{\exp\left(-\frac{1}{2}\frac{d_{r\phi}{}^2}{\sigma^2}\right)}{\sqrt{2\pi}\sigma}\ , \sigma^2 = \sigma_{\rm vtx}^2 + a_{\rm obs}^2 + \frac{b_{\rm obs}^2}{p_\perp^2 \sin\theta}$$

Both fits are performed on the sample of tracks with negative impact parameters, which were used in the primary vertex, and for both data and MC. The result of the fits is shown in table 1

The fit error for data is then corrected by a factor

$$\sigma \leftarrow \kappa\sigma_{\rm KF}\ , \kappa = \frac{a_{\rm obs}^2 + b_{\rm obs}^2/p_\perp^2 \sin\theta}{a_{\rm KF}^2 + b_{\rm KF}^2/p_\perp^2 \sin\theta}$$

this insures that the average resolution is correct, and in addition keeps the information from the track fit about the individual tracks. The correction factor applied is within 10% of unity over the entire range.

The same procedure is performed for the d_{sz} resolution.

track type	a_{KF}	b_{KF}	a_{obs}	b_{obs}
96 DATA RΦ				
TEC only	0.142	0.135	0.170	0.172
$\phi_{\mathrm{outer}}\ z_{\mathrm{outer}}$	0.030	0.122	0.033	0.133
ϕ_{inner}	0.033	0.090	0.036	0.097
$\phi_{\mathrm{inner}}\ \phi_{\mathrm{outer}}\ z_{\mathrm{outer}}$	0.033	0.076	0.034	0.076
96 Monte Carlo RΦ				
TEC only	0.153	0.133	0.186	0.190
$\phi_{\mathrm{outer}}\ z_{\mathrm{outer}}$	0.030	0.124	0.027	0.134
ϕ_{inner}	0.029	0.090	0.029	0.089
$\phi_{\mathrm{inner}}\ \phi_{\mathrm{outer}}\ z_{\mathrm{outer}}$	0.032	0.077	0.031	0.070
96 DATA SZ				
1 Z no ϕ_{inner}	0.154	0.083	0.157	0.092
$\geq$ 2 Z no ϕ_{inner}	0.137	0.112	0.128	0.124
ϕ_{inner} ; $\geq$ 2 Z	0.133	0.086	0.141	0.089
96 Monte Carlo SZ				
1 Z no ϕ_{inner}	0.150	0.083	0.152	0.073
$\geq$ 2 Z no ϕ_{inner}	0.149	0.106	0.126	0.107
ϕ_{in}; $\geq$ 2 Z	0.134	0.085	0.135	0.076

Table 1: Parameterization of the mean $d_{r\phi}$ and d_{sz} resolution from the KF fit, and the observed spread of the $d_{r\phi}$ and d_{sz} distributions.

In addition, one observes that the resolution in data and MC are not identical. Therefore the MC is adjusted to match the data. This is done by the following substitution:

$$d_{r\phi} \leftarrow d_{r\phi} + (\kappa - 1) d_{r\phi}^{\mathrm{true}}$$

where κ is given by:

$$\kappa(p_\perp, \theta) = \frac{(a_{\mathrm{obs}}^{\mathrm{data}})^2 + (b_{\mathrm{obs}}^{\mathrm{data}})^2 / p_\perp^2 \sin\theta}{(a_{\mathrm{obs}}^{\mathrm{mc}})^2 + (b_{\mathrm{obs}}^{\mathrm{mc}})^2 / p_\perp^2 \sin\theta}$$

and $d_{r\phi}^{\mathrm{true}}$ is the true $d_{r\phi}$, i.e. the impact parameter of the generated track with respect to the true primary vertex.

Once the above corrections have been performed, the tails of the $d_{r\phi}$ and d_{sz} distributions are investigated. To do this, the track selection criteria are loosened, and tracks are no longer required to be used in the primary vertex fit. The $d_{r\phi}/\sigma$, $\sigma^2 = \sigma_{\mathrm{vtx}}^2 + a_{\mathrm{obs}}^2 + b_{\mathrm{obs}}^2/p_\perp^2 \sin\theta$ distributions are fit, for both data and MC separately, with one Gaussian of unit width, a second Gaussian with a variable width and one exponential. According to the difference between data and MC, the $d_{r\phi}$ of some fraction the tracks in MC is smeared by a Gaussian. In addition, to account for the difference in the exponential tail, a very small fraction of tracks are exponentially smeared.

4 Track Selection and Classification

Not only do the $d_{r\phi}$ and d_{sz} resolution of tracks depend critically on the SMD hit pattern, as shown in the previous section, the likelihood of a very large track-finding error also depends greatly. Tracks are therefore ranked according to this hit pattern. First they are divided in 4 categories, depending only on the hit pattern in SMD and TECZ. Tracks in two out of the four categories have potentially useful in 3D information. In these two categories some additional cuts are applied to reject tracks which are misreconstructed in the sz projection. The tracks that fail these cuts could, however, still be useful in their $r\phi$ projection, and fall into two so-called demoted categories. The categories are:

1. No ϕ_{inner}
2. ϕ_{inner} or $\phi_{\mathrm{inner}}\, z_{\mathrm{outer}}\, \phi_{\mathrm{outer}}$
3. "Demoted" category (4) tracks
4. ϕ_{inner} and at least 1 Z measurement
5. "Demoted" category (6) tracks
6. ϕ_{inner} and $\phi_{\mathrm{outer}}\, z_{\mathrm{outer}}$ and (TECZ or z_{inner} or both)

Tracks which fail the following criteria are demoted:

- $|\mathrm{d}_{sz}| < 3$ mm
- $\sigma_{\mathrm{d}_{sz}} < 0.5$ mm
- $\sigma_\theta < 20$ mrad

As the primary vertex has already been determined using loose selection cuts, $\mathrm{d}_{r\phi}$ and d_{sz} are defined with respect to the event primary vertex.

In order to have optimal performance from the b tag, quality tracks must be selected. Misreconstructed tracks are much more damaging to the b tag performance than are poor precision tracks. In order to minimize the effect of misreconstructed tracks, selection cuts are applied to all six track categories. In general the selection cuts depend on track category, but typical values are listed below.

- Hit and span ratios > 64% and > 80%
- Number of inner TEC hits ≥ 1.
- $p_\perp > 300$ MeV
- $D_{\mathrm{poca}} < 650\ \mu$m for category 4 and 6,
- $|\mathrm{d}_{r\phi}| < 2$ mm
- $|\phi_{\mathrm{local}} \bmod \pi/24| > 0.005$

The hit and span ratios are defined as the number of hits divided by the maximum number of hits possible for the given θ, and the distance between the first and last hit divided by the maximum number of hits. They are chosen such that in the barrel the cuts are at 30 and 40 respectively.

The D_{poca} selection cut is defined to be the distance between the track and the jet at the point of closest approach in 3 dimensions; this only makes sense for 3D categories.

4.1 Category Migration and Track Killing

Even in the realistic MC, neither the total number of tracks, nor the relative fraction of tracks with SMD hits is identical to the real data. Starting from the resolution sample, the fraction of tracks that needs to be killed is determined. At the same time the relative fractions for given hit patterns are determined. According to the difference between data and MC, some SMD hits are artificially removed, and the $d_{r\phi}$ and d_{sz} distributions smeared accordingly using the parameterization described above.

5 The Decay Length Variable

5.1 Decay Length Calculation

To take advantage of the ability of the SMD to measure tracks in three dimensions while avoiding the detrimental effects of pattern recognition mistakes in the sz plane, the weighted average of the decay length as measured in the $r\phi$ and sz planes is used as the basic tagging variable. The calculated error of this weighted average also includes small correlations due to the primary vertex and jet direction.

Once the primary vertex is reconstructed, it is used as the reference point of all the tracks in the event (see [4] for an excellent description of the parameterization of helicoidal tracks used in the L3 software and this paper) and defines the tracks' distance of closest approach in the $r\phi$ plane, $d_{r\phi}$, and sz plane, d_{sz} . The crossing point of each track with the axis of the jet containing that track is determined in both $r\phi$ and sz neglecting the curvature of the track (which is a very good approximation in L3 near the interaction region). The distances between these crossing points and the primary vertex are then "projected" onto the jet axis and are denoted by $L_{r\phi}$ and L_{sz} . This is illustrated in Figure 6.

$$L_{r\phi} = \frac{d_{r\phi}}{\sin(\phi_j - \phi)\sin\theta_j}$$

$$L_{sz} = \frac{d_{sz}}{\sin\theta_j(\cos\theta\cos(\Delta\phi) - \sin\theta\cot\theta_j}$$

where (ϕ_j, θ_j) are the jet directions, (ϕ, θ) are the track angles and $\Delta\theta = (\theta_j - \theta)$ (likewise for $\Delta\phi$).

Defined in this way, $L_{r\phi}$ and L_{sz} represent approximately independent measurements of the true decay length of the B hadron and can be averaged to form the decay length variable, L.

$$L = \frac{L_{r\phi}/\sigma_{L_{r\phi}}{}^2 + L_{sz}/\sigma_{L_{sz}}{}^2 - \frac{(L_{r\phi}+L_{sz})c_{\phi z}}{\sigma_{L_{r\phi}}{}^2\sigma_{L_{sz}}{}^2}}{1/\sigma_{L_{r\phi}}{}^2 + 1/\sigma_{L_{sz}}{}^2 - \frac{2c_{\phi z}}{\sigma_{L_{r\phi}}{}^2\sigma_{L_{sz}}{}^2}}$$

where $\sigma_{L_{r\phi}}$,$\sigma_{L_{sz}}$ and $c_{\phi z}$ are the elements of the $(L_{r\phi}, L_{sz})$ error matrix.

As $L_{r\phi}$ and L_{sz} have the jet axis in common, the calculation of $\sigma_{L_{r\phi}}$ and $\sigma_{L_{sz}}$ includes a 15 mrad error on the jet direction. The use of this jet angle error has an impact on the tag performance. The correlations between the (x, y, z) positions of the primary vertex also add to the correlations taken into account in $c_{\phi z}$. Nevertheless, all of these correlations typically amount to only $\sim 1\%$ contribution to the error of L.

If both $r\phi$ and sz have good potential to measure the decay length, the $r\phi$ will usually dominate because the SMD (and the TEC) have better resolution in that direction. However, if for some reason the $r\phi$ information is not very useful, for example, if the track is at a very small angle to the jet direction in the $r\phi$ projection, the sz decay length measurement becomes important. Figure 7 shows the correlation between measured decay lengths in the two projections. If the probability that both the $r\phi$ and sz measurements are compatible with each other is larger than 5%, the two are averaged. Otherwise, only the $r\phi$ projection is used, minimizing the effect from mistakes in the sz pattern recognition.

5.2 Decay Length Resolution Functions

To compute the probability that a given track with decay length, L, is compatible with the primary vertex the significance, $S = L/\sigma_L$, is used (see Figure 8). The negative side of the significance distribution is used to determine the track probability function, $P(|S|)$; the probability that a track originating from the primary vertex has a significance $|S|$ or larger. The track probabilities are shown in Figure 9.

6 Combining Track Probabilities

In previous discriminants, the probabilities of all the positive decay length tracks were simply multiplied. The event (or jet) discriminant was then calculated as the probability to get a product of probabilities smaller than the one measured. This calculation can be rather simply performed analytically.

However, it is clear that this calculation is not optimal. First, it neglects entirely the number of tracks with negative decay lengths. This number is useful information for the tag. Second, all tracks are treated equally, even if they have no potential to measure lifetime. Tracks with poor decay length resolution only add noise to the discriminant. We have developed methods to use the number of negative decay length tracks and to weight tracks in the discriminant to make optimal use of all track information.

6.1 Inclusion of Negative Decay Length Tracks

In defining a track probability, the typical calculation simply integrates the positive decay length significance distribution function from the measured value to infinity[3)]. Thus, tracks with very small positive decay lengths approach probability 1, while tracks with large positive decay lengths approach 0. To extend this definition smoothly to negative decay length tracks without bias against a well measured track with zero decay length, we simply defined the probability for any negative decay length track to be equal to a track with zero decay length. That is, $P = 1$ for a negative decay length track.

We then take the same product of probabilities since factors of 1 do not matter. However, we must recalculate the probability for a uds event with the same number of measured tracks to give a smaller product. For an average uds event, with the above definition, half the tracks will have a negative decay length and hence $P = 1$. The other half of the tracks will be uniformly distributed between 1 and 0.

A similar analytical calculation gives the following result for the discriminant.

$$P_N = \frac{\Pi}{2^N} \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \begin{pmatrix} N \\ j \end{pmatrix} \frac{(-\log \Pi)^i}{i!}, \text{ where } \Pi = \prod_{j=1}^{N_+} P(S_j) \tag{1}$$

[3)]Note that the discriminant used in [1, 2] does implicitly use both the negative and positive lifetime tracks. This is achieved by changing the definition of the track probability. As a result, tracks with an impact parameter of zero get a probability of $\frac{1}{2}$ instead of 1 and tracks with negative (positive) impact parameter have probabilities $> \frac{1}{2}$ ($< \frac{1}{2}$).

and N_+ is the number of tracks with positive decay length. By construction the variable P_N is flat for events without lifetime, whereas events containing lifetime tracks show up as a peak at 0.

Studies have shown that this technique improves over the one that just uses positive decay length tracks. This makes good sense too. If there are 10 negative decay length tracks and two tracks with marginally significant positive decay lengths, the evidence for lifetime is weaker than if there are only the two positive decay length tracks. That is, there were many more tracks in the first case that could have fluctuated to give the signal tracks.

6.2 Weighted Discriminant

In addition to using the negative decay length tracks, we find it useful to assign weights to tracks based on their momentum and decay length resolution. Instead of taking the simple product of track probabilities, we compute the following product

$$\mathcal{P} = \prod_{j=1}^{N_+} P^{\alpha_j}(S_j). \tag{2}$$

The α_j are the "weights". Tracks which have full capability to measure lifetime should have $\alpha_j = 1$ while tracks which have no capability should have $\alpha_j = 0$. While the optimal α_j can be derived analytically under some assumptions, we have just used a simple analytical expression for α_j and adjusted parameters to improve tag performance.

Again the negative decay length tracks contribute a factor of 1. While we have derived a more complex analytical formula to compute the discriminant, in practice a Monte Carlo integral is faster and more accurate for large numbers of tracks like one finds in a Higgs event.

The basic integral calculated is if the form

$$C'_N = 1 - \int_{l_{N_+}}^{1} dx_{N_+} \cdots \int_{l_1}^{1} dx_1$$

where

$$l_j = \left(\frac{\mathcal{P}}{\prod_{i=j+1}^{N_+} x_i^{\alpha_i}} \right)^{\frac{1}{\alpha_j}}.$$

The inclusion of negative decay length tracks is simple in the MC calculation.

7 Performance

For the 1996 data sample the b tag was tuned on the 1.2 $\mathrm{pb}^{-1}$of Z calibration data. The result is shown in Figure 10. Compared to the 2D neural network tag, previously the best performing b tag in L3, the 3D discriminant shows a 15% relative efficiency improvement at 80% purity.

The effect of the non-optimal calibration [5, 6] of TEC due to the small size of the 96 Z sample can be seen from the difference with Figure 11, which shows the b tag, as tuned on a similar sized amount of Z data in 1995. It is not only the performance that suffers, but also the agreement between data and MC is worse in 1996.

8 Neural Net b tag

To further improve the performance of the 3D tag, the lifetime variables for each jet are combined with other distinguishing information in a feed-forward neural network to form a b-jet-tag. The neural network, based on JetNet 3.5, has three layers: one input with 14 nodes, one hidden with 28 nodes, and one output with 2 nodes. Two nodes, a b and c node, are used in the output layer to help distinguish between b and c jets. Lifetime variables, inclusive lepton information, secondary vertex information, and jet shape variables, form the input layer. The number of hidden nodes was chosen to attempt to reduce the generalized error.

We have used almost 200,000 jets broken into three samples to train, test, and verify the neural net. To increase the speed of training Langevin updating is used. One third of the events are used to train the neural network[4)]. The next third is used to test the network and choose the configuration. When the performance on this sample has reached its maximum value the training is frozen. This prevents the neural network from being over optimized to the training sample and not a general sample of jets. The third sample is used to quote the performance of the neural network; because the maximum value of the second sample is used to choose the net configuration, it is not reasonable to use it to quote the performance. The chosen neural network configuration is applied to the third sample and used gauge performance.

8.1 Inputs

Lifetime Variables

The lifetime variables are the standard jet tag and the weighted jet tag described earlier. While they are highly correlated, both are used in the neural network.

Secondary Vertex Variables

Secondary vertices are found using the decay length and decay length error of each track in the jet. The routine attempts to maximize a likelihood function by comparing the compatibility of the decay lengths of the tracks with the primary vertex and with each other. A penalty is applied to the likelihood function when forming a secondary vertex to reduce the number of mismeasured tracks forming erroneous single track vertexes. Once a group of tracks is identified as having compatible decay lengths, they are passed to a fitting routine. The vertexes are fit using the same technique as used for the 3D primary vertex (See section 2), but in addition the position is constrained to be along the jet axis. The efficiency to find a secondary vertex in a b jet is approximately 50%.

Once the vertex has been found, the information is given to the neural network. The number of vertexes and the number of tracks assigned to secondary vertexes are provided for each jet; shown in Figure 8.1. If a secondary vertex is found then invariant mass with and without missing momentum, decay length, and sigma decay length are also given to the neural network for the most significant secondary vertex. All four are shown in Figure 13 for the most significant secondary vertex. The missing vertex momentum is obtained by adding the $p_\perp$ needed to point the secondary vertex back to the primary vertex:

$$M_c^2 = M_v^2 + |\vec{q}_v|^2 - \left(\vec{q}_v \cdot \vec{j}\right)^2 \quad \text{where} \quad \vec{j} = \frac{\vec{x}_s - \vec{x}_p}{|\vec{x}_s - \vec{x}_p|}$$

[4)]Adjustments were made in the concentrations of b and c jets compared to usdg jets in an attempt to improve the networks ability to distinguish signal from background. In the end no improvement was seen over the natural admixture of b,c and udsg jets.

The error on the missing momentum is also calculated. If the magnitude of the missing momentum is less than the error on the missing momentum, no momentum is added.

Lepton Variables

Electrons and Muons are identified using the APL3 routines provided by Franz Muheim. A lepton must pass selection cuts and is then assigned to the nearest jet, a lepton cannot be its own jet. Lepton momentum and transverse momentum, with respect to the jet axis, are used. The leptonic information is particularly useful in the high purity region. Unfortunately, the number of selected leptons, due to physics and to detector inefficiencies, is small: approximately 3% of b jets contain an identified electron and 6% contain a muon. Distributions of momentum and transverse momentum for the identified leptons are shown in Figure 14.

Jetshape Variables

Three jet shape variables are used as inputs to the neural network. $\beta\gamma$ calculated using all the clusters in a jet, $\beta\gamma$ calculated using only the four most energetic clusters of the jet, and boosted sphericity [7]. Boosted Sphericity, S_{boost}, is defined as

$$S_{boost} = \frac{\sum p_{\perp}^2}{\sum p^2},$$

where the momenta are given in the rest frame of the jet.

Plots for of all three variables for data, all Monte Carlo, b jets, c jets, and udsg jets are found in Figure 15.

Auxiliary Variables

To account for the degradation of the tracking system in the forward direction, the θ of the jet is used as input. While there is no distinguishing power in the θ distribution, the θ information helps the neural net to place cuts on other variables. There is good agreement between data and Monte Carlo in the θ distribution.

8.2 Combination of Jets

The neural network output for a set of jets is combined into an event tag by computing the probability that each jet is compatible with the distribution for light quarks determined from Monte Carlo. Once the jet probabilities are determined, they are combined using the same technique as the individual track probabilities of the b tag. The neural network event tag variable is defined as the negative log-likelihood of the combined probabilities.

8.3 Performance

Once the jet tags have been combined into an event tag, a direct comparison can be made with the event discriminant described in section 6. The neural network achieves a 10% relative improvement at 80% purity over the weighted event discriminant. The performance and agreement for the neural net event tag is shown in Figure 16. This performance improvement was duplicated in 95 where the efficiency at 80% purity jumps from 70% to 77%. The performance of the neural network and standard event discriminants for 80%, 90%, and 95% purity is compared in Table 2.

	b efficiency			
b purity	3D ’95 plain	3D ’95 NN	3D ’96 plain	3D ’96 NN
0.80	0.70	0.77	0.66	0.73
0.90	0.58	0.65	0.52	0.60
0.95	0.46	0.52	0.40	0.45

Table 2: The performance for the discriminant event tag and the Neural Net event tag for both a subset of the ’95 data and the ’96 calibration sample at $\sqrt{s}$ = 91GeV. The numbers are quoted for events within the barrel region of the detector, i.e. $|\cos\theta_{\mathrm{thrust}}| < 0.72$.

9 Outlook

Both the SM Higgs search and MSSM Higgs search require good b tagging capability at LEP 200. This tag was used in the Standard Model Higgs search in 1996 [8–10] (where it is referred to as the variable B_{tag}). It is also being used for MSSM Higgs searches where b tagging is even more important. L3 has produced very competitive Higgs limits [11].

The Kalman filter and primary vertex reconstruction have been in production in L3 since version 213 and the b tag is being integrated into the standard L3 production software at this time. The b tag variables described here are available on the NP ntuples for general use.

We expect to improve the tag performance somewhat in the near future again with the goal of improving our Higgs sensitivity.

At this point, the UCSD tag has not been optimized to make a precise measurement of R_{b} although we have tested the tag by computing R_{b} from our hemisphere tag. Work in that direction would be interesting particularly in a multiple tag approach. The tag should already be of general use in other B physics measurements.

10 Acknowledgments

The UCSD tracking and tagging software is self contained and separate from the BTL3 software. However, we have benefited from continuing discussions with others working on b tagging in L3, particularly Riccardo Faccini and Maurizio Biasini. Our b tag is intimately connected to the SMD pattern recognition software of Juan Alcaraz. The tag performance is very sensitive to how well the pattern recognition performs. In particular, Juan has done an excellent job in the z-coordinate where there is very little reliable information other than in the SMD itself.

Chris Tully’s K_0 and Λ identification code has been incorporated to reduce backgrounds from these sources. It makes a small improvement in the tag[5)] whereas other routines we have tried caused our results to deteriorate.

We have also incorporated Franz Muheim’s inclusive lepton code. This gives us a simple clean interface to the inclusive lepton information.

References

[1] B. Bertucci, Measurement of the Z Branching Fraction into Bottom Quarks Using Lifetime Tags, Internal report, The L3 Collaboration, 1996, L3 Note 1968, Analysis submitted to ICHEP

5) For this, Chris has been given the high honor of the UCSD “Better Than Nothing” award.

96 (Warsaw).

[2] J. Alcaraz, B.Bertucci, M. Chamizo, S.Easo, J.Goldstein, D.Kamrad, M.Maity, J.Rodin, Measurement of the Z Branching Fraction into Bottom Quarks Using Double Tag Methods, Internal report, The L3 Collaboration, 1997, L3 Note 2114, Analysis submitted to 97 summer conferences.

[3] G.Borisov,C.Mariotti, Nucl. Inst. Meth. **A 372** (1996) 181.

[4] J. Alcaraz, Helicoidal tracks, Internal report, The L3 Collaboration, 1994, L3 Note 1666.

[5] D. Bourilkov, TEC Calibration at LEP2 using High Energy Lepton Pairs; Part I - The Methods, Internal report, The L3 Collaboration, 1997, L3 Note 2040.

[6] D. Bourilkov,G. Rahal-Callot, TEC Calibration at LEP2 using High Energy Lepton Pairs; Part II - 1996 Results, Internal report, The L3 Collaboration, 1997, L3 Note 2047.

[7] L3 Collab., O. Adriani *et al.*, Phys. Lett. **B 307** (1993) 237.

[8] M.Biasini, J.Branson, F.Di Lodovico, A.Dominguez, R.Faccini, A.Favara, M.Felcini, I.Fisk, L.Lugnier, J.P.Martin, M.Pieri, G.Raven, M.Rescigno, Preliminary results of Higgs Searches at $\sqrt{s}$ = 161–172 GeV, Internal report, The L3 Collaboration, 1997, L3 Note 2067.

[9] J.Branson, A.Dominguez, I.Fisk, A.Kunin, L.Lugnier, J.P.Martin, G.Raven, I.Vorobiev, J.Xu, B.Zhou, Search for the Standard Model Higgs Boson at LEP2, Internal report, The L3 Collaboration, 1997, L3 Note 2086.

[10] J.Branson, F.Di Lodovico, A.Dominguez, R.Faccini, A.Favara, I.Fisk, M.Pieri, G.Raven, Final results on the Standard Model Higgs search at $161 \leq \sqrt{s} \leq 172$GeV, Internal report, The L3 Collaboration, 1997, L3 Note 2107.

[11] The L3 Collaboration, PLB **??** (1997) ??, Submitted to Phys.Lett. B.

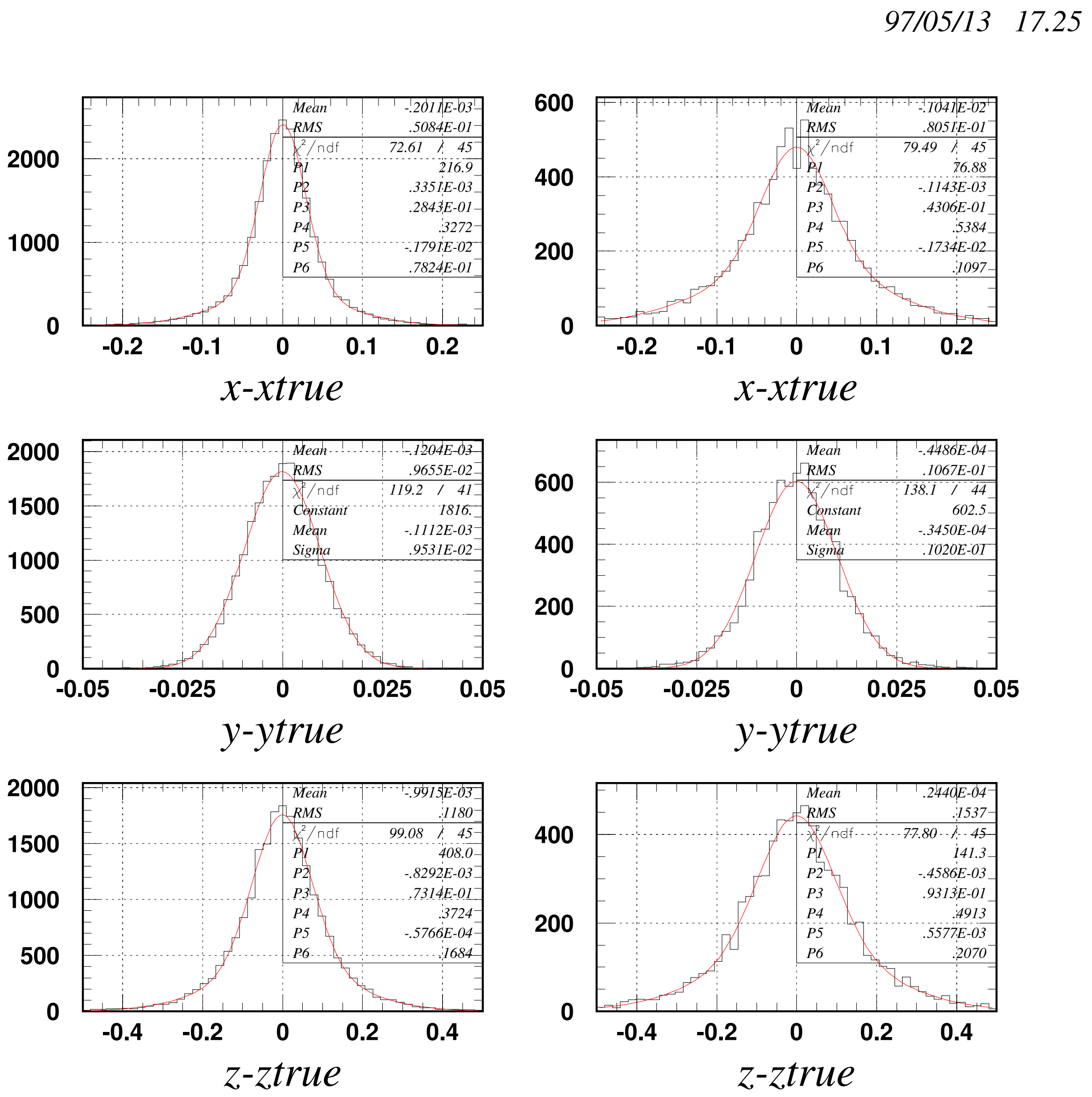


Figure 1: The vertex resolution obtained on the 1996 Monte Carlo for x,y and z. On the left are uds events, on the right b events.

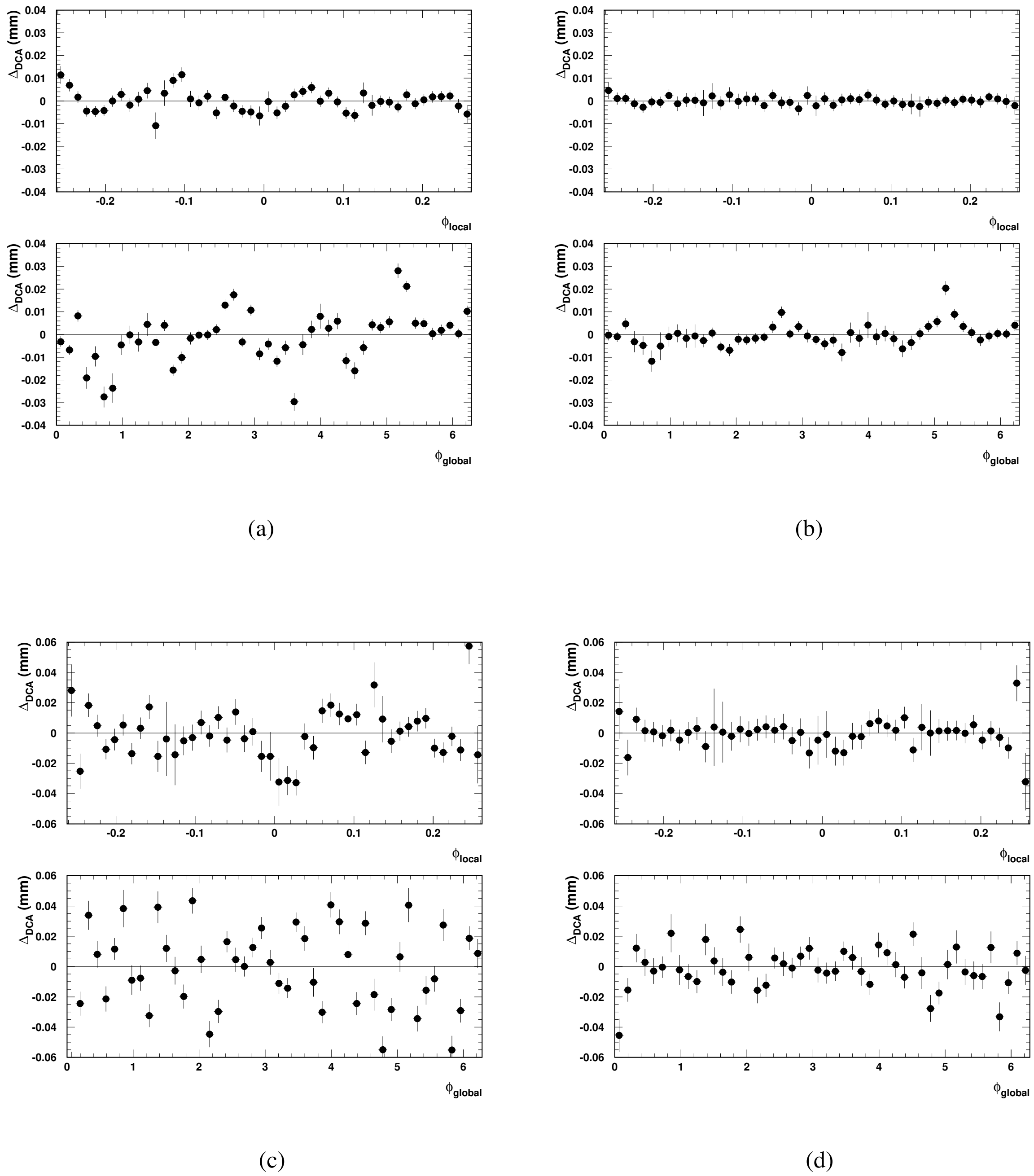
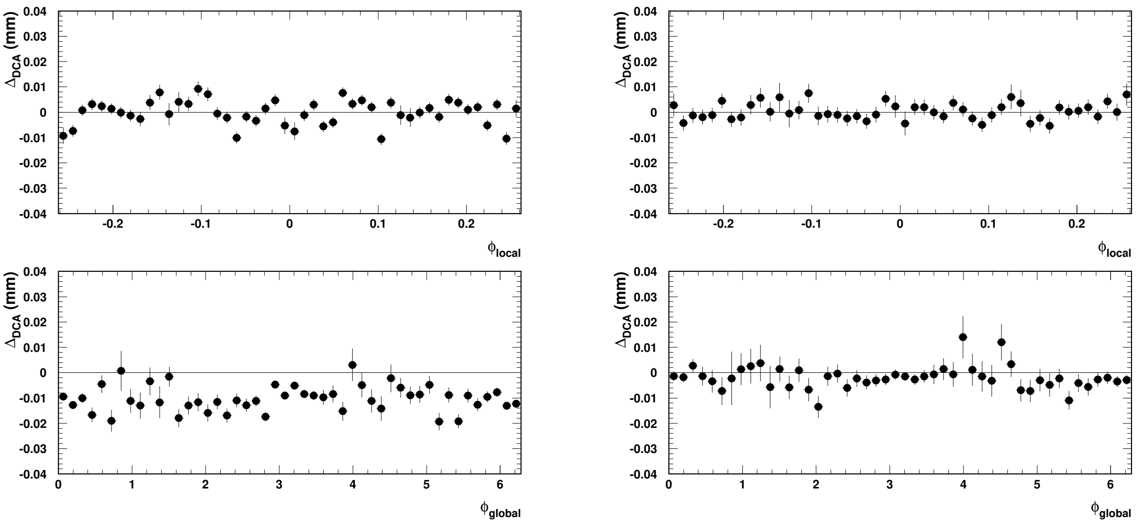


Figure 2: The mean $d_{r\phi}$ for the '96 data resolution sample before (a),(c) and after correction (b),(d), for tracks with all SMD hits (a),(b) and no SMD hits (c)(d).



(a) (b)

Figure 3: The mean $d_{r\phi}$ for the '96 MC resolution sample before (a) and after correction (b), for tracks with all SMD hits.

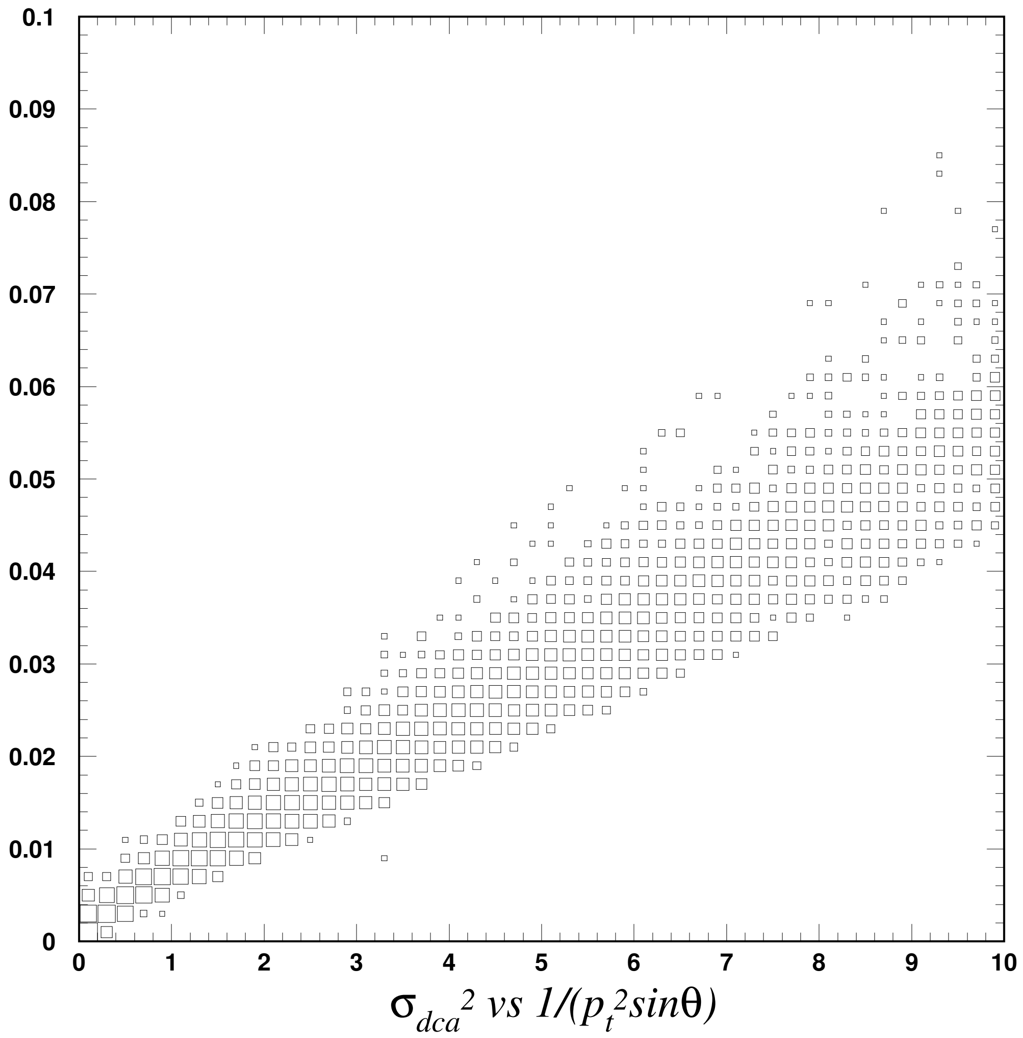


Figure 4: The square of the fit error versus $1/p_\perp^2 \sin\theta$ for tracks with all SMD hits.

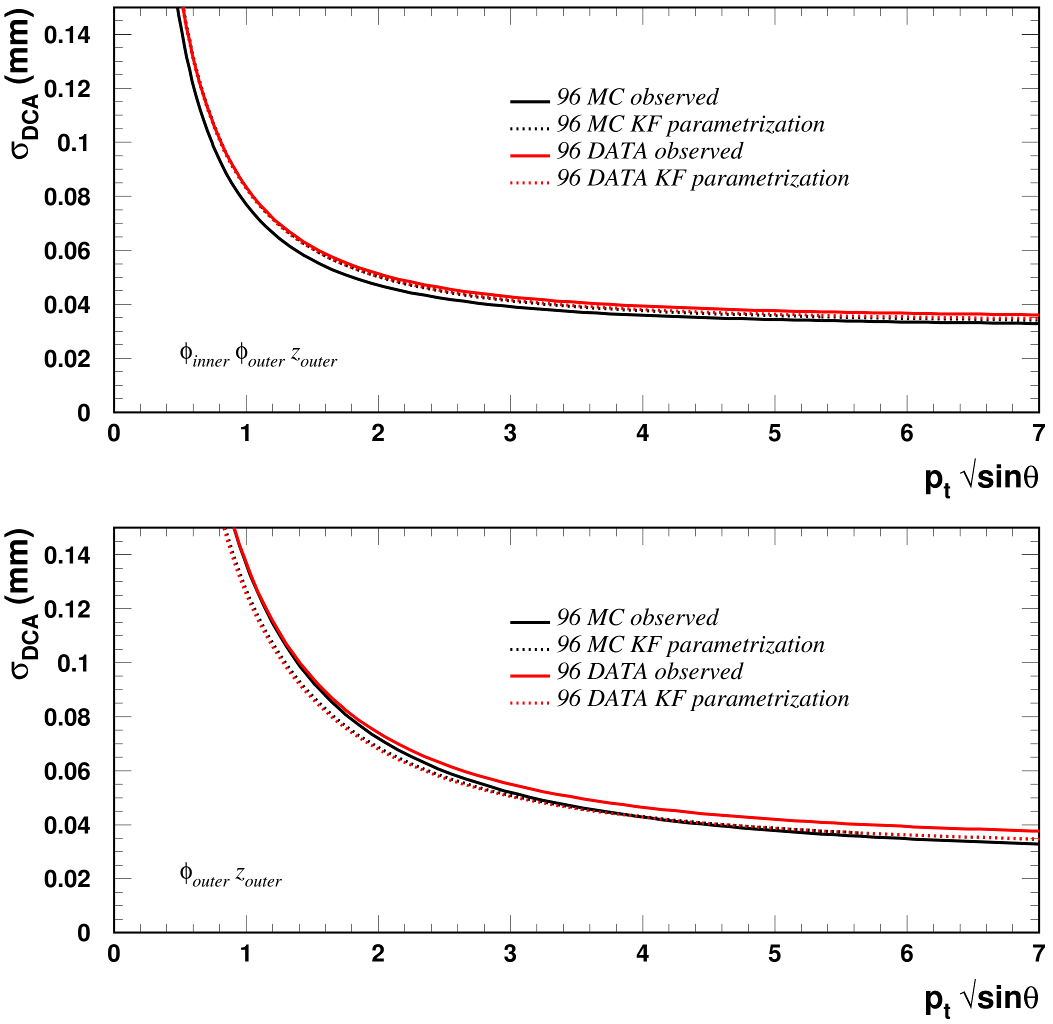
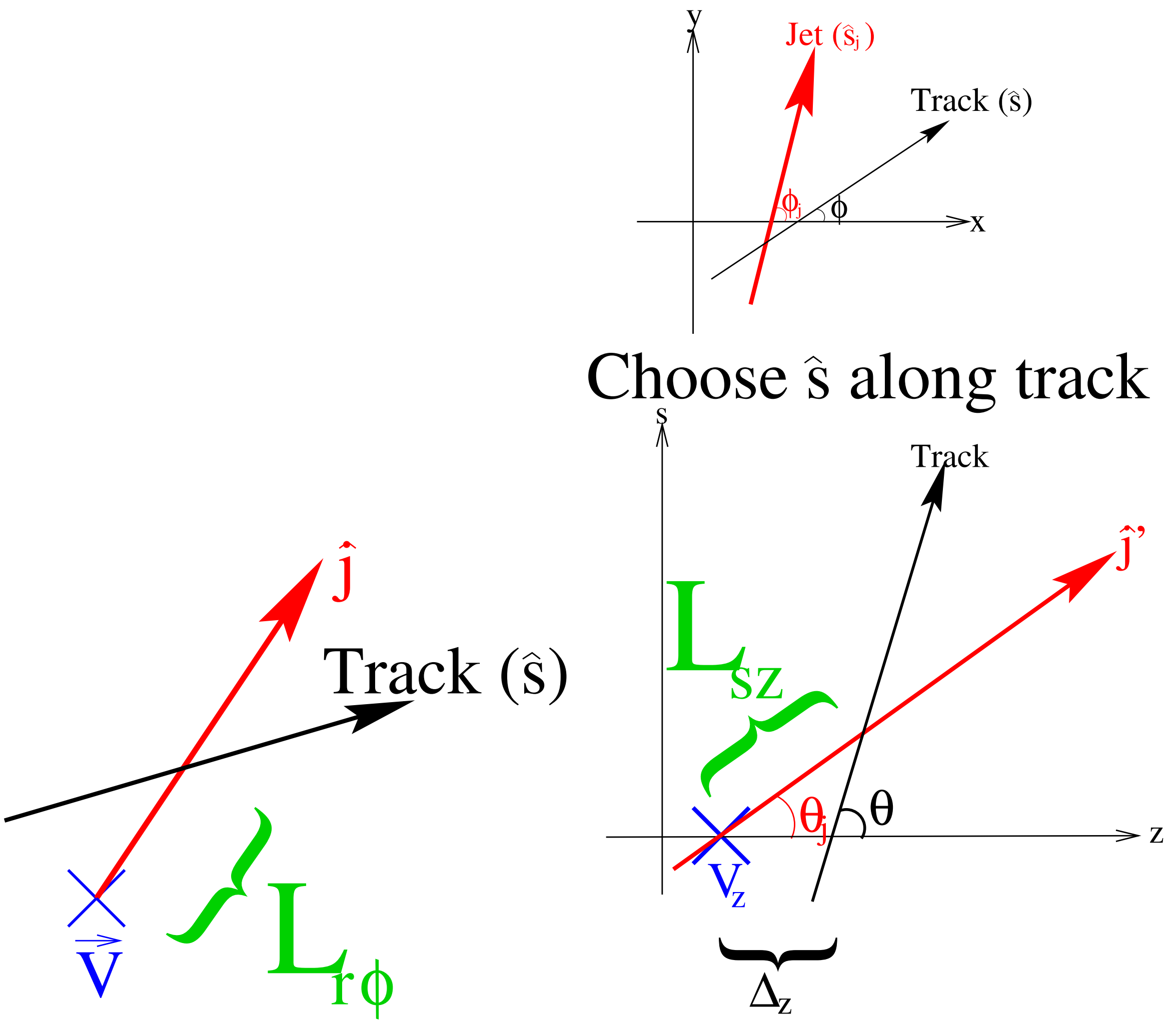


Figure 5: Parameterization of the fit error and the observed average resolution for both data and mc. The top plot is for tracks containing ϕ_{inner} ,ϕ_{outer} ,z_{outer} SMD hits, the bottom plot for ϕ_{outer} ,z_{outer} .



(a) $L_{r\phi}$ (b) L_{sz} .

Figure 6: The definition of the tagging variable in the $r\phi$-plane, $L_{r\phi}$ (a) and in the sz-plane, L_{sz} (b).

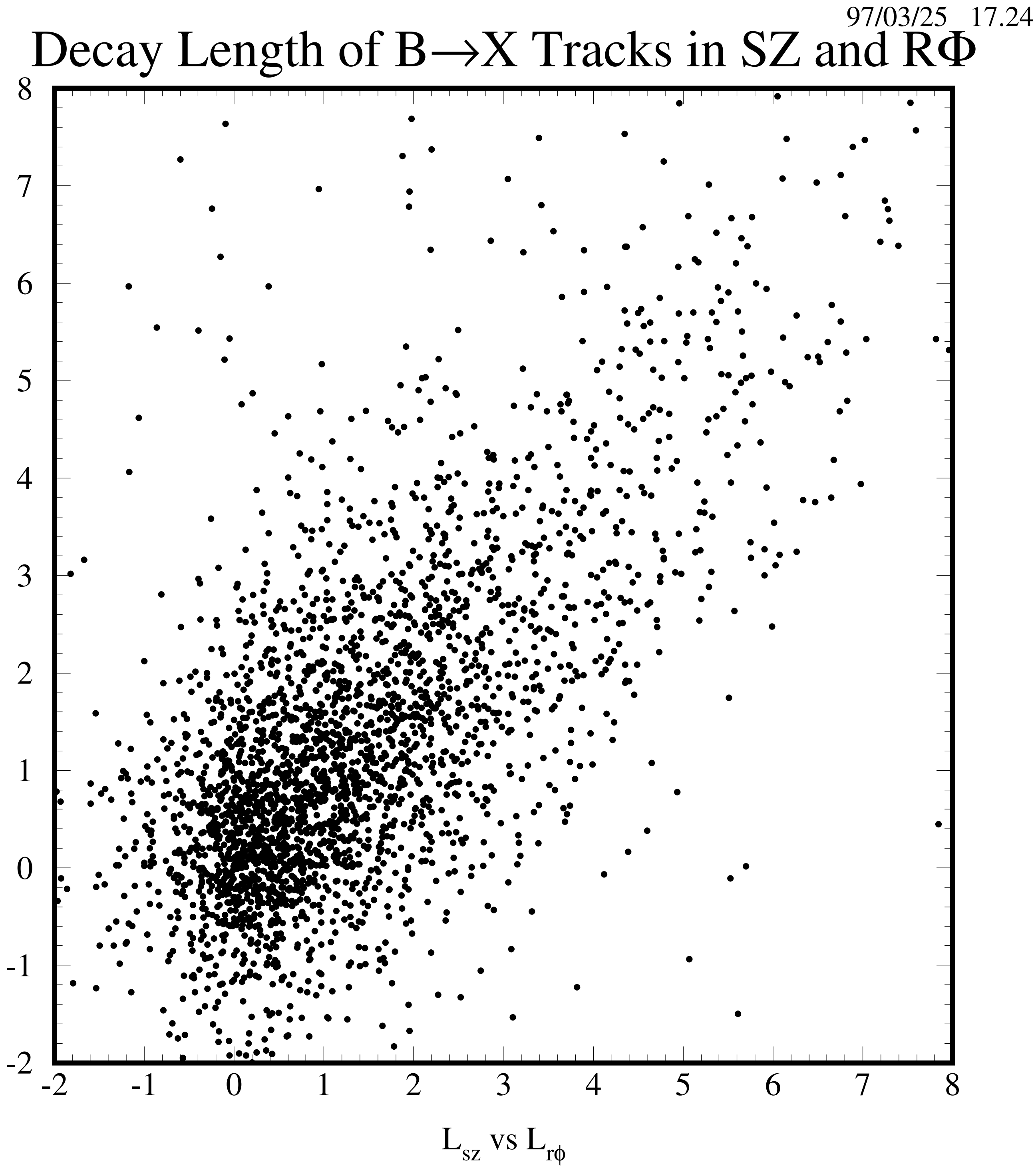


Figure 7: L_{sz} versus $L_{r\phi}$ for tracks coming from decay products of B hadrons in Monte Carlo. The axes are in mm.

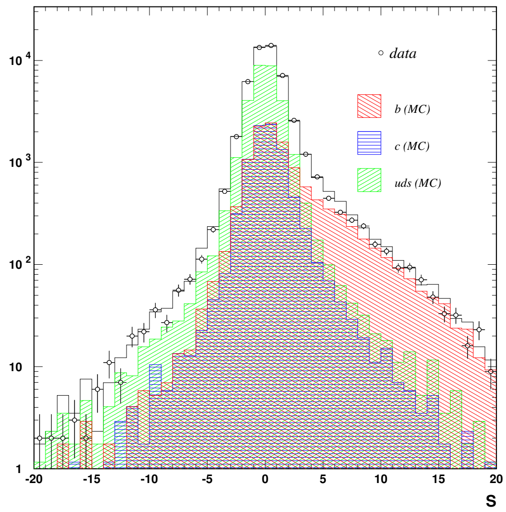


Figure 8: The decay length significance for tracks with all SMD hits in 1996

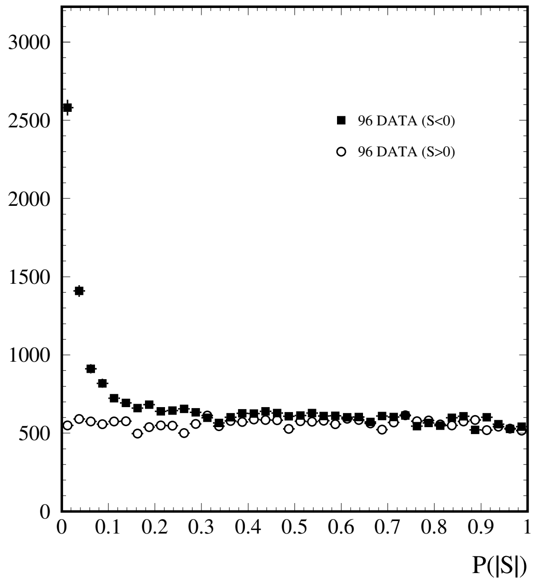


Figure 9: The CL for the decay length significance for the resolution sample with negative decay length and the sample with positive decay length for tracks with all SMD hits in 1996

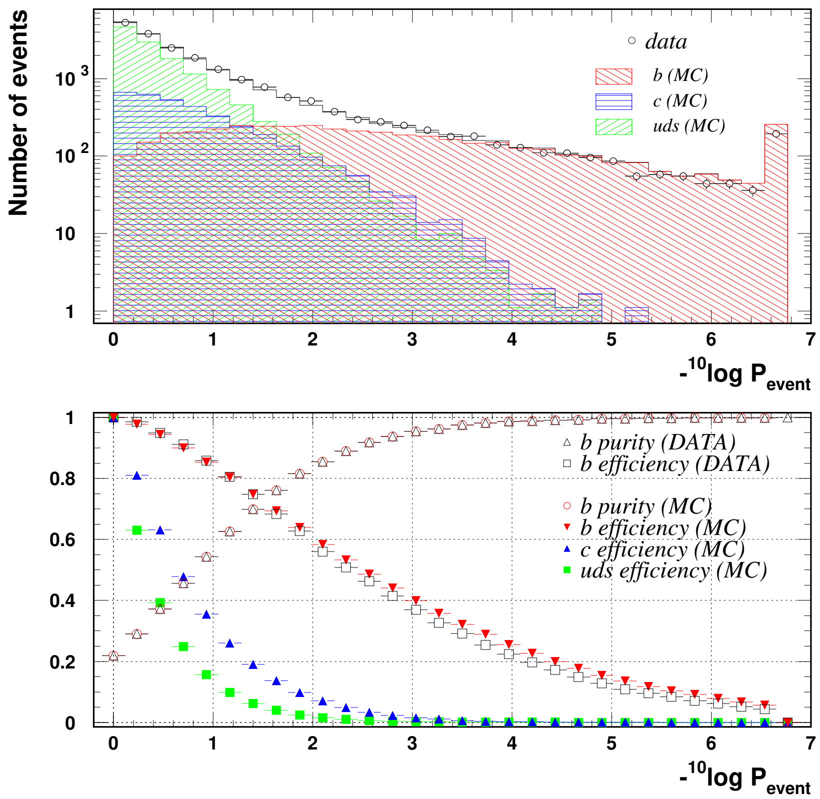


Figure 10: The event discriminant distribution for the 1996 Z calibration data. The overflows are put in the last bin. The efficiency and purity as a function of a cut on the discriminant are shown in the bottom plot.

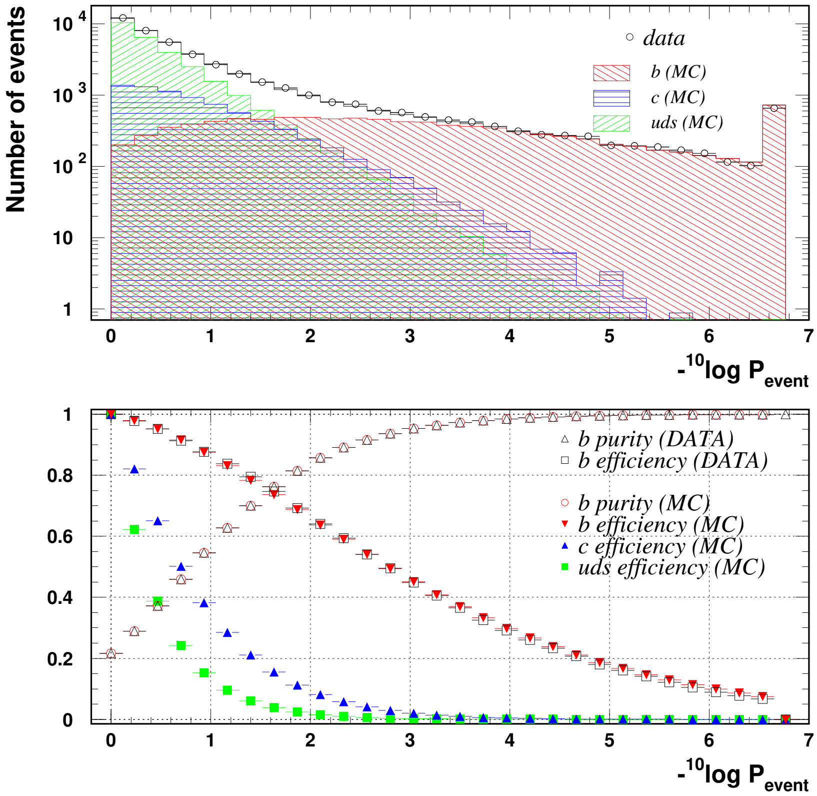
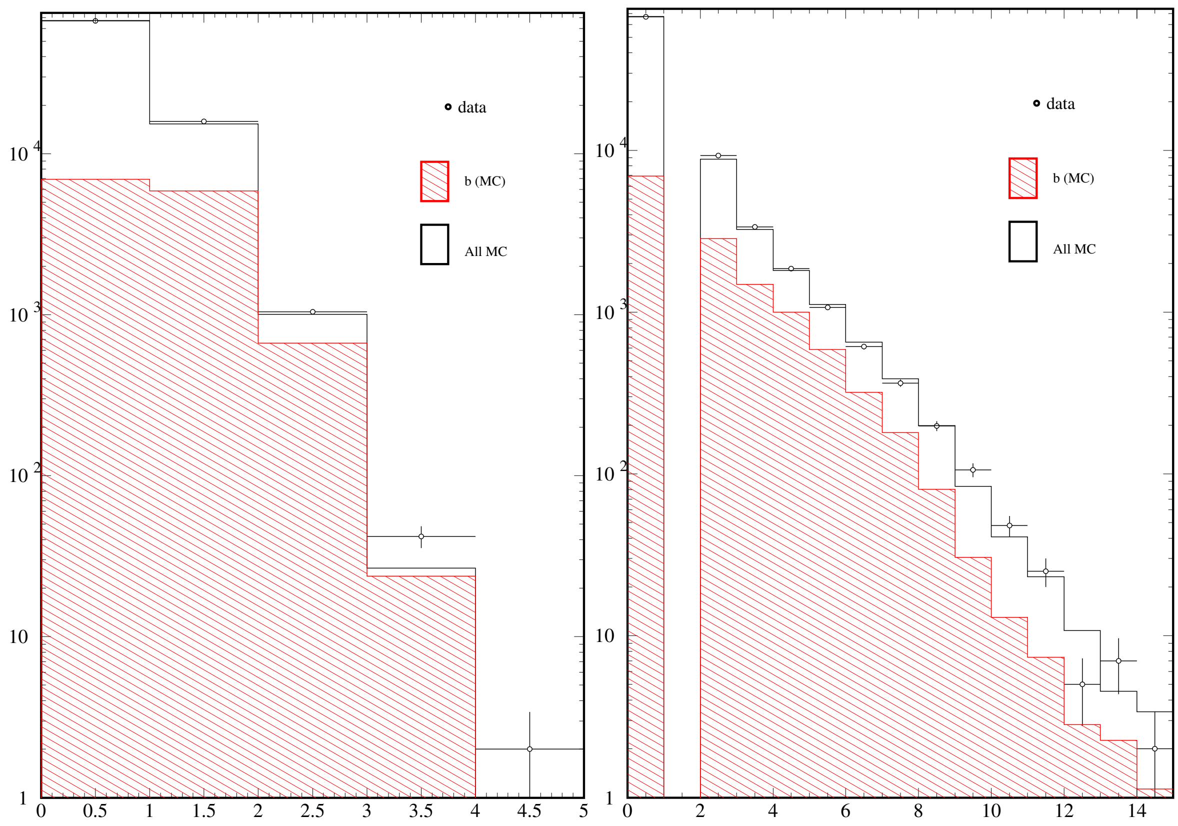


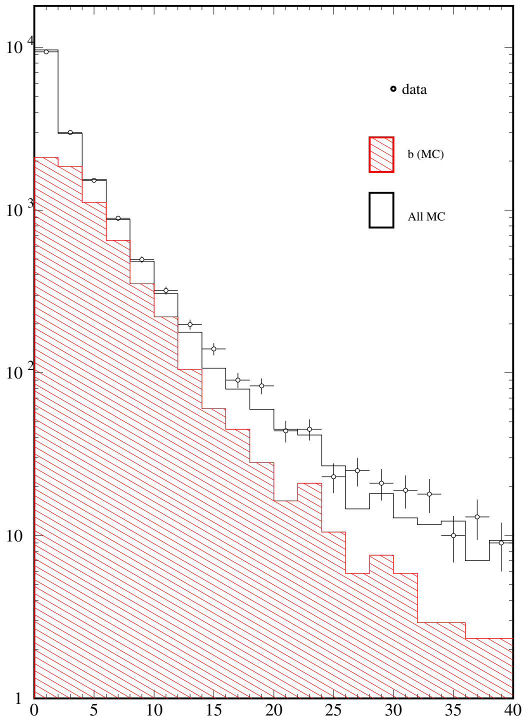
Figure 11: The event discriminant distribution for a subset of the 1995 Z data. The overflows are put in the last bin. The efficiency and purity as a function of a cut on the discriminant are shown in the bottom plot.



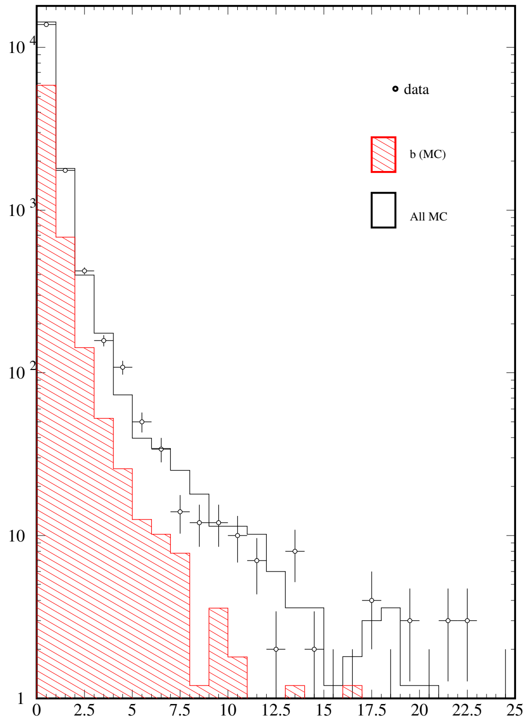
(a) The number of secondary vertexes per jet.

(b) The number of tracks assigned to secondary vertexes.

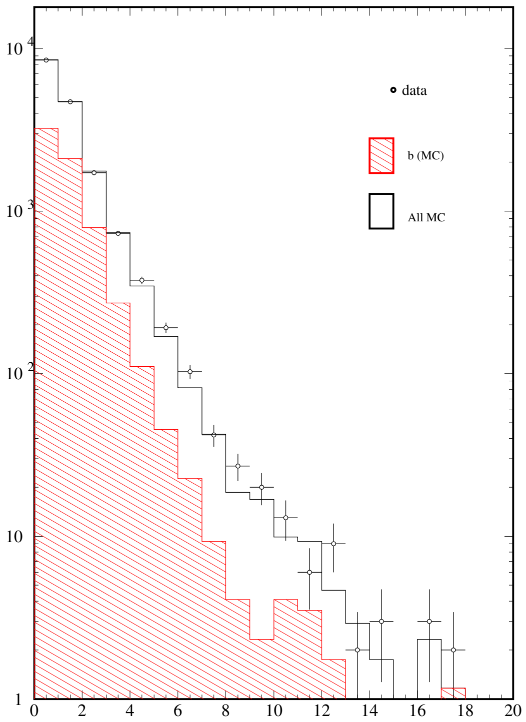
Figure 12: Reconstructed secondary vertex variables.



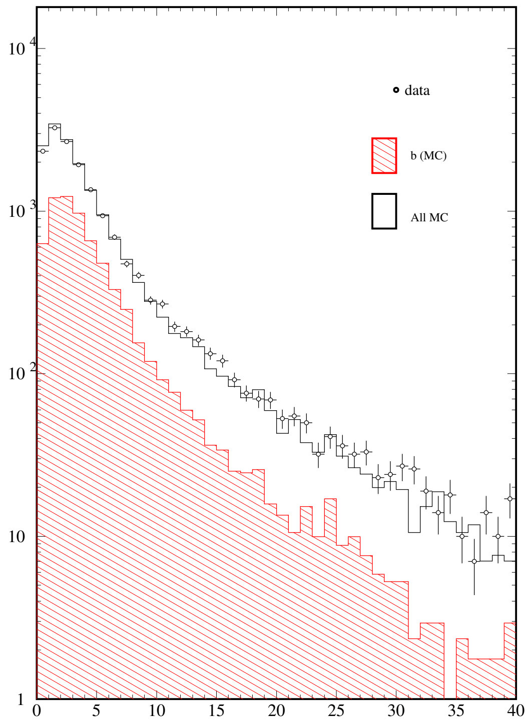
(a) Decay length of the secondary vertex.



(b) Error on the decay length of the secondary vertex.

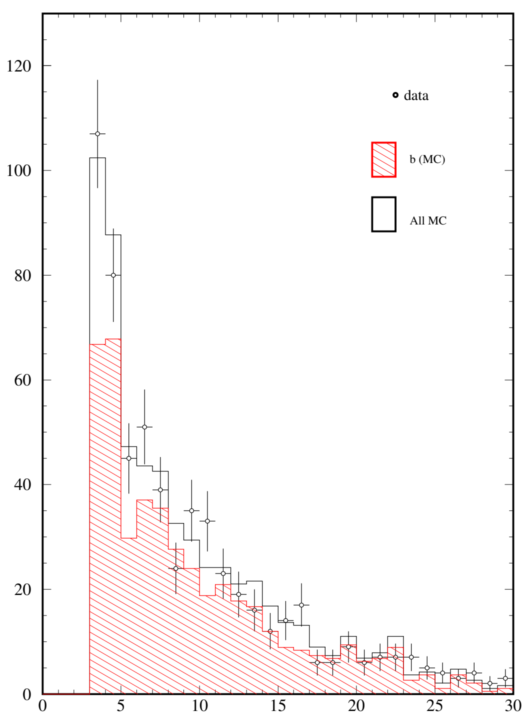


(c) Vertex mass without missing momentum.

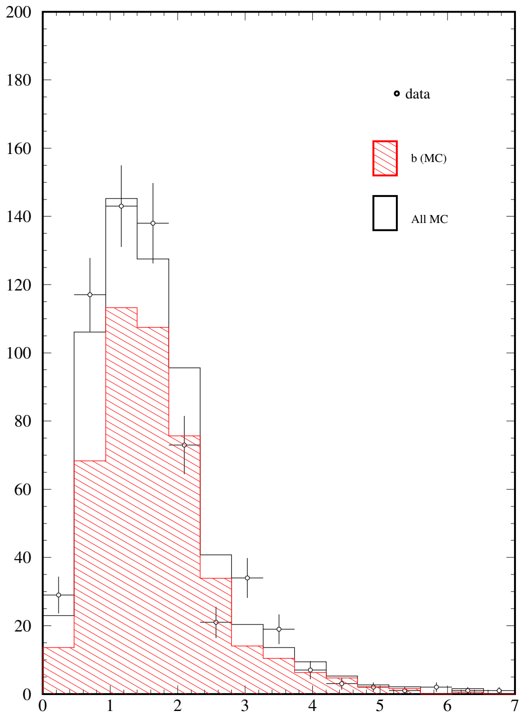


(d) Vertex mass with missing momentum.

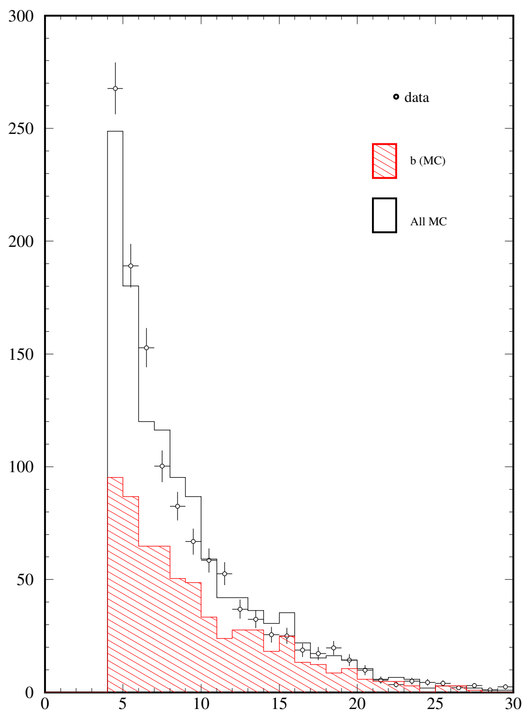
Figure 13: Secondary Vertex Information



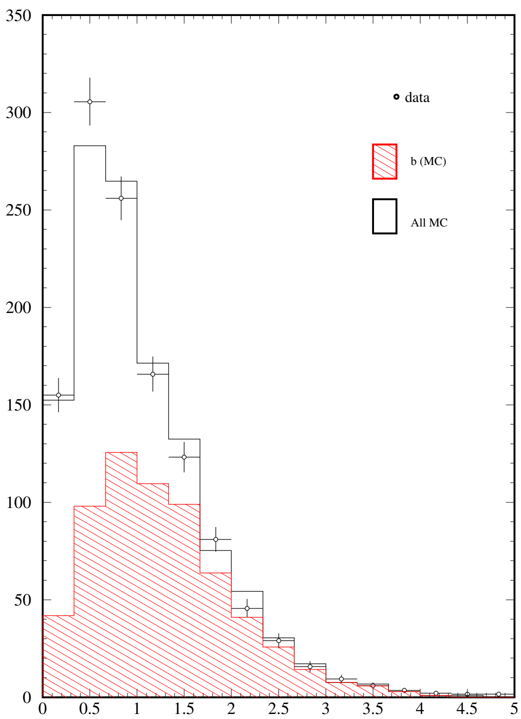
(a) Momentum of electrons.



(b) Transverse momentum of electrons with respect to jet axis.

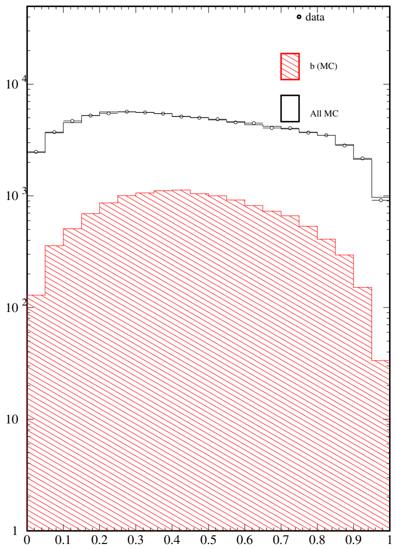


(c) Momentum of muons.

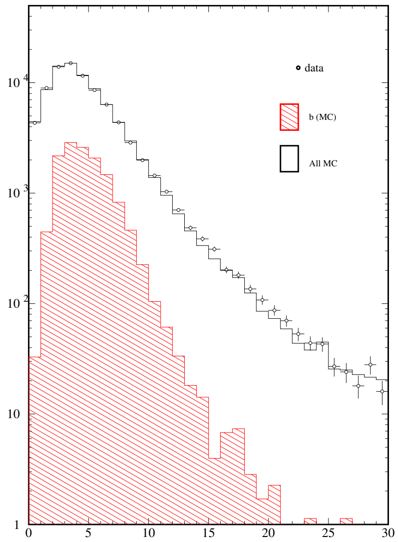


(d) Transverse momentum of muons with respect to jet axis.

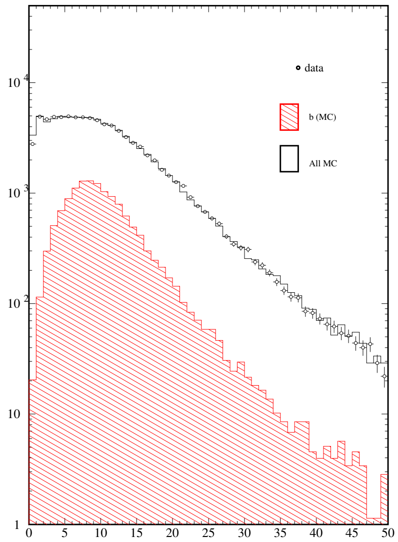
Figure 14: Inclusive Leptons Variables



(a) Boosted Sphericity.



(b) $\gamma\beta$ of jet.



(c) $\gamma\beta$ calculated from 4 most energetic clusters.

Figure 15: Jet Shape Variables

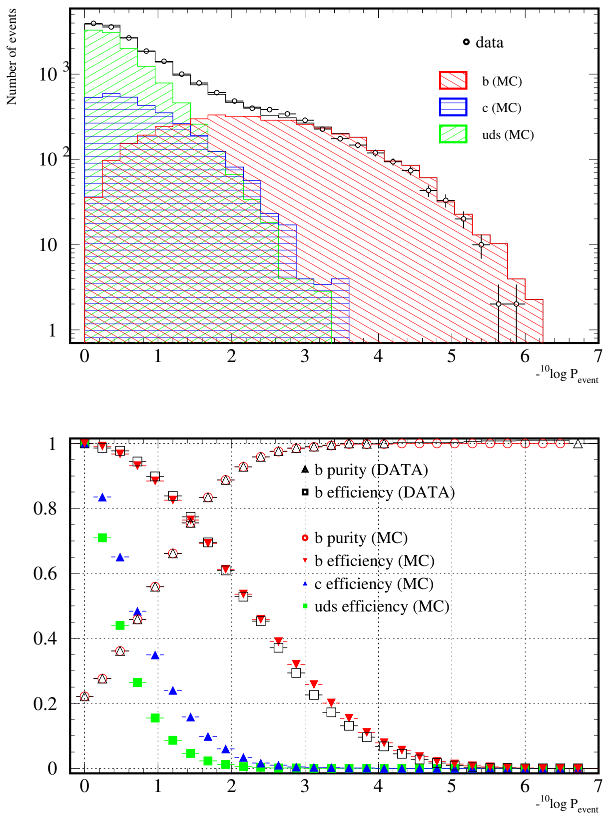


Figure 16: The Neural Net Event tag for the 96 calibration sample taken at $\sqrt{s} = 91$ GeV