Chapter 6

Inclusive semileptonic event selection

At LEP, the B_s oscillations analyses with the highest sensitivity are obtained with inclusive semileptonic event samples. The charge of the lepton issued from the b-hadron decay is used to determine the flavour of the reconstructed b-hadron at decay time (b or $\bar{b}$). The advantage of inclusive event selections over partial decay reconstructions like $D_s\,\ell$ pairs comes from statistics. In the first case, over 50 thousand events can be selected at LEP, to be compared with only few hundreds in the second case.

In this Chapter, a new inclusive semileptonic event selection performed on the ALEPH data sample collected around the Z resonance is presented. To perform a B_s oscillation analysis some characteristics of the events need to be determined. These characteristics are mainly, *i*) the proper time of the b-hadron decay; *ii*) the probability of the lepton to originate through a direct $b \to \ell$ decay; *iii*) the probability of the b-hadron to be a B_s; *iv*) the unmixed/mixed label of the decay; and *v*) the probability of this label to be incorrect, called the "mistag probability". To achieve the best possible sensitivity for oscillations, a detailed characterization is needed. Ideally all above properties are evaluated on an event-by-event basis. All the steps of the selection and event characterization are described in this Chapter.

With respect to the previous ALEPH analysis on the subject [10], significant improvements were obtained on both the event selection and the characterization of the sample. The event selection is tailored to identify events with a good b-hadron decay vertex reconstruction. To this end, a new vertex algorithm was designed and the sample was divided in classes based on the event topological properties. Some aspects of the event characterization require the combination of several discriminant variables. This combination is in all cases based on neural networks (Ref. [80]) to obtain an optimal performance and a correct treatment of the correlations among the discriminant variables.

The data sample and the simulated event samples used in the analysis development are briefly described in Section 6.1, and the event pre-selection in Section 6.2. The new procedure to reconstruct the decay position of the charmed particle, and the b-hadron, as well as the criteria followed to reject poorly reconstructed events, are described in Section 6.3. The final

event selection is obtained after the b-hadron enrichment (Section 6.4) and the $b \to \ell$ enrichment (Section 6.5) procedures are applied. The overall selection performance is summarized in Section 6.6.

The first step of the event characterization is the proper time measurement explained in Section 6.7. The presence of a kaon in the decay products of the b-hadron, or among the fragmentation particles, can be used both to enrich the sample in B_s decays, and for the estimate of the $b/\bar{b}$ content of the candidate at production time. The procedure followed to identify charged and neutral kaons is described in Section 6.8. A technique to statistically disentangle B_s decays from other b-hadron decays was developed as explained in Section 6.9. Finally, the estimate of the b-quark content of the b-hadron at production time is described in Section 6.10.

6.1 Data and simulation samples

The data sample analysed corresponds to about four million hadronic Z decays collected by the ALEPH detector from 1991 to 1995. In 1998, these data were reprocessed with a refined version of the reconstruction program. The main improvements concern the track reconstruction and the particle identification [67]. The analysis presented here benefits from all these improvements.

During its development phase, the analysis relied exclusively on samples of simulated events. The generation of these events is done through the following process, *i*) all particles originated from the Z decay are generated according to the theoretical knowledge of the simulated process; *ii*) their interaction with the ALEPH detector is simulated using the GEANT [81] package; and *iii*) the data obtained is reconstructed with the same procedure as the *real* data. The events obtained have the same format as the *real* data, with the addition of the *true* information about the properties of all the particles in each event. This knowledge was used to optimize the selection procedure, estimate the selection efficiency, and extract all parametrizations needed for the complete event characterization.

Hadronic events were generated with JETSET [82]. Different type of events were simulated: 8.5 million unbiased $Z \to q\overline{q}$ decays, 5 million $Z \to b\overline{b}$ decays, 2.3 million $Z \to c\overline{c}$ decays, almost one million $Z \to bB_s$ ($B_s \to \ell$) decays, and about half a million $Z \to bB_s$ ($B_s \to c \to \ell$). All physical parameters used as inputs for the event generation (b-hadron fractions, decay branching ratios, fragmentation function, etc.) were adjusted to the knowledge to date, as summarized in Ref. [13].

6.2 Data sample pre-selection

The first selection criterion applied is the "hadronic selection", explained in Section 5.1, to obtain a quasi-pure sample of $Z \to q\overline{q}$ decays. The events are separated in two hemispheres with respect to a plane perpendicular to the thrust axis. A cut on the polar angle of the thrust axis is applied to reject events with most of the tracks outside the VDET acceptance: events must satisfy $|\cos\theta_{\rm thrust}| \leq 0.85$.

The events are then clustered in jets reconstructed from all energy-flow objects using the JADE algorithm (Section 5.2.1), with a jet-resolution parameter ($y_{\rm cut}$) of 0.0044 (this parameter was optimized for heavy flavour analyses with leptons in ALEPH [83]). In the following, this is the jet definition used to characterize events by their number of jets.

Events are required to have at least one lepton candidate (Section 5.3) reconstructed with a VDET hit. The transverse momentum of the lepton is defined with respect to the axis of its JADE jet (with the definition above) once the lepton itself is removed from the jet. If more than one lepton candidate is found, only that with highest transverse momentum is considered in what follows. A hemisphere is labeled as "same" or "opposite" according to whether it contains the lepton candidate or not.

The lepton requirement enriches the event sample in $Z \to b\overline{b}$ decays, from about 20% to more than 55%.

6.3 Vertex algorithm and selection

The typical topology of a semileptonic B_s decay is shown in Fig. 6.1 together with the decay of the other b-hadron in the event. The B_s decay position is the origin of the lepton track, of

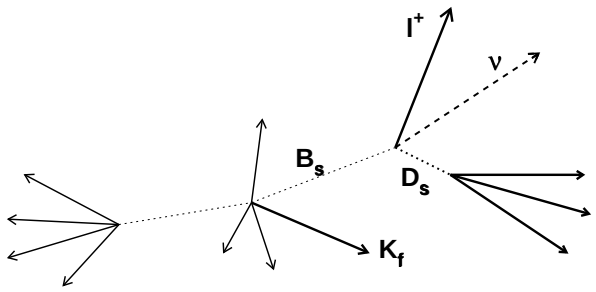

Figure 6.1: Semileptonic B_s decay schematic, tracks from the b-quark fragmentation and the decay of the other b-hadron in the opposite hemisphere of the event, are also shown.

the undetected neutrino, and of a charmed particle. The latter, with a typical decay length of 1 mm, decays at a tertiary vertex. The vertex algorithm is optimized on such topologies and it also provides criteria to select such events.

The vertex algorithm developed for this analysis starts by the inclusive reconstruction of the charmed particle. This particle is extrapolated back to the lepton and a common vertex is looked for. The B_s oscillation study requires a good estimate of the b-hadron decay length and of its uncertainty. The procedure developed for this analysis is explained in the following Sections.

6.3.1 Search for a seed

In a semileptonic B_s decay, the reconstructed decay products are of two types: *i*) the lepton, which is the only reconstructed charged particle issued directly from the meson decay; and *ii*) the particles coming from the decay of the charmed hadron. If the lepton track is ignored, the only remaining tracks with sizeable impact parameter in the B_s hemisphere are the D_s^- (charmed particle) decay products. The "search for a seed", *i.e.*, a point in space, used as a first approximation of the tertiary vertex, is based on this observation.

As explained in Section 5.6, two inclusive secondary vertex finders, VNFIT and QVSRCH, are used. The lepton candidate track is ignored and a first estimate of the tertiary vertex position is searched for with VNFIT. If the search is successful, the vertex position and uncertainties are kept as the seed for the charmed particle. When VNFIT fails, QVSRCH is used instead.

Both algorithms provide, for each track, a variable which estimates its compatibility with originating from the tertiary vertex (Sections 5.6.1 and 5.6.2). However, other track properties can be used to disentangle fragmentation particles from b-hadron decay products, as explained in the next Section.

6.3.2 The c track selection

The first step of the c-track selection consists in separating charged particles from the b-quark fragmentation from those from the b-hadron decay with the use of some of their kinematical properties. All the relevant variables are combined using a neural network. The list of discriminant variables used is given, followed by an explanation of their relevance.

- The compatibility of the track with the inclusive decay vertex (the track probability in the case of a VNFIT vertex and the impact parameter significance in the case of QVSRCH).
- The impact parameter significance of the track with respect to the primary vertex, signed using the b-jet axis.
- The particle momentum: particles from the b-hadron decay are expected to be, on average, of higher momentum than those originated in the b-quark fragmentation.
- The angle of the particle with respect to the estimated b-hadron decay direction. Three estimates of this direction are considered:
  - the thrust axis,
  - the jet axis, and
  - the direction defined by the primary and the inclusive tertiary vertices.

  The angle of the decay particles with respect to the true b-hadron decay direction is expected to be smaller than that of fragmentation particles. The quality of the direction estimates presented above depends on both the number of jets in the event and the decay length of the b-hadron. The b-hadron jet used here is a cone-based jet built around the lepton candidate. The algorithm is chosen for its performance on the direction

estimate. When the tertiary vertex is far apart from the primary vertex, the direction defined by the two vertices is the best estimate of the b-hadron flight direction. When the distance between the two vertices is small, a better estimate can be provided by the thrust axis or the jet axis. The thrust axis is found to be more accurate in two jet events, while for events with more than two jets the b-jet is a better estimator (because of hard gluon emission).

When building a combined discriminant variable with a neural network, it can happen that the best performance is obtained when the combination includes variables that have discriminating power and variables which are not discriminant by themselves, but are correlated with the relative discriminant power of the others. These auxiliary variables, called *control variables* in the following, allow an optimal combination to be obtained in the whole phase space.

In the present case, the control variables included in the neural network are, the number of jets in the event, the χ^2 of the tertiary vertex fit, the measured decay length, and the measured jet momentum.

The distributions of the most relevant variables for all tracks in hemispheres with an identified lepton are shown in Figs. 6.2 to 6.4. In each plot, the dark area corresponds to tracks originated in secondary or tertiary vertices in $Z \to b\overline{b}$ simulated events. Charged particles from $Z \to u\overline{u}$, $d\overline{d}$, $s\overline{s}$, and $c\overline{c}$ decays are displayed as "udsc". The agreement between the data and the simulation is reasonable for most of the variables. In the case of the angular variables a deviation is seen. More tracks are found at large angle in the data than in the simulation. This disagreement is probably due to the lack of accuracy of the quark fragmentation simulation. However, as the track separation combined variable is only used to select charged particles to form the charmed object, the observed disagreement is not considered to be a problem. The peak at $P_{VNFIT} \sim 0.5$ in Fig. 6.4b is mostly due to charged particles with low momentum ($p < 500\,\mathrm{MeV}/c$), and poorly reconstructed tracks.

The neural network training was carried out on simulated $B_s \to \ell$ events, all fragmentation and decay charged particles with momentum above $200\,\mathrm{MeV}/c$, with the exclusion of the lepton candidate, were used for the training. Two different trainings were performed, one for events with a VNFIT vertex and another for the remaining events.

The distribution of the combined track separation variable for events with a VNFIT vertex is shown in Fig. 6.5. The performance obtained with the other training are not as good. Those events for which the VNFIT algorithm does not provide a vertex are topologically more complicated than the average, thus with a more difficult track separation.

Two different c track selections are defined depending whether the VNFIT algorithm was successful in the event or not. In both cases, all tracks except the lepton candidate are ordered according to their compatibility with the c-hadron decay vertex, P_{track}. This compatibility is defined as the output of the track separation neural networks described above.

In the case of events with a VNFIT vertex, the track most compatible with the c-hadron decay vertex is taken. Tracks with a momentum $p > 0.5\,\mathrm{GeV}/c$, a VNFIT track probability $P_{VNFIT} > 0.7$, and $P_{track} > 0$ in decreasing value of P_{track} are added one by one till a mass of $1.8\,\mathrm{GeV}/c^2$ for the charmed candidate is reached.

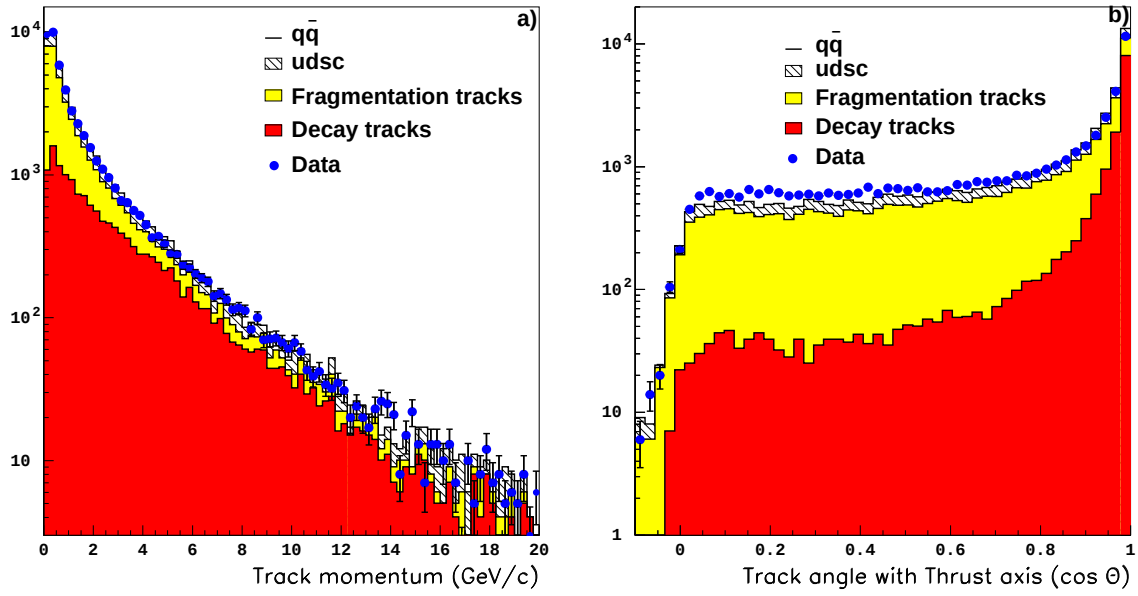

Figure 6.2: Distribution of the discriminant variables for the track separation: a) track momentum, and b) track angle with the thrust axis $\cos(\theta_{\mathrm{trhust}})$.

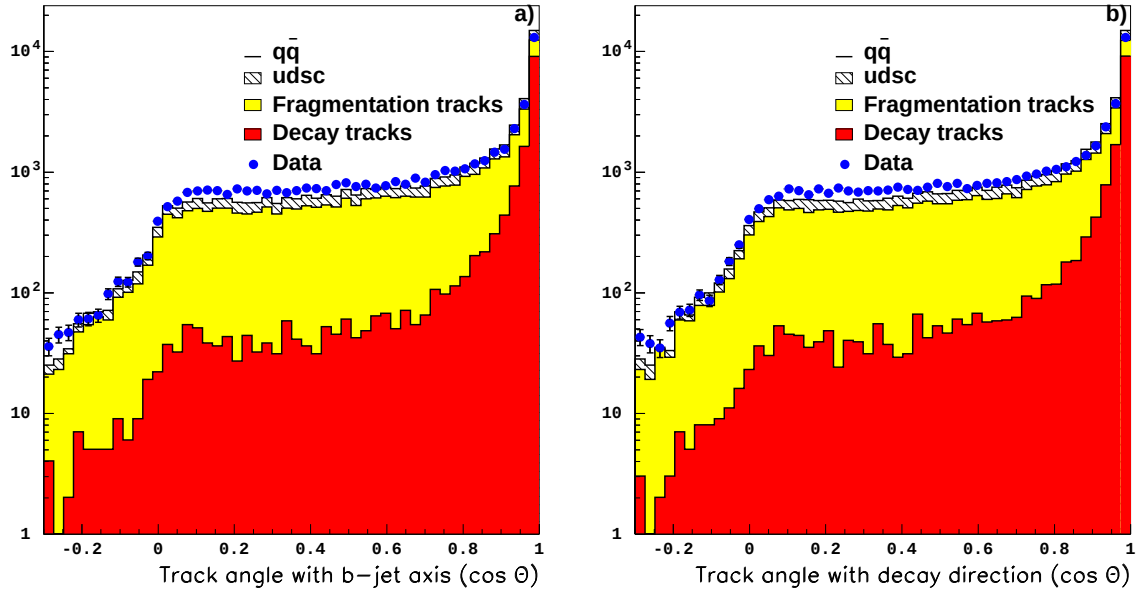

Figure 6.3: Distribution of the discriminant variables for the track separation. Track angle with a) the jet direction $\cos(\theta_{\mathrm{jet}})$, and b) the decay direction $\cos(\theta_{\mathrm{decay}})$.

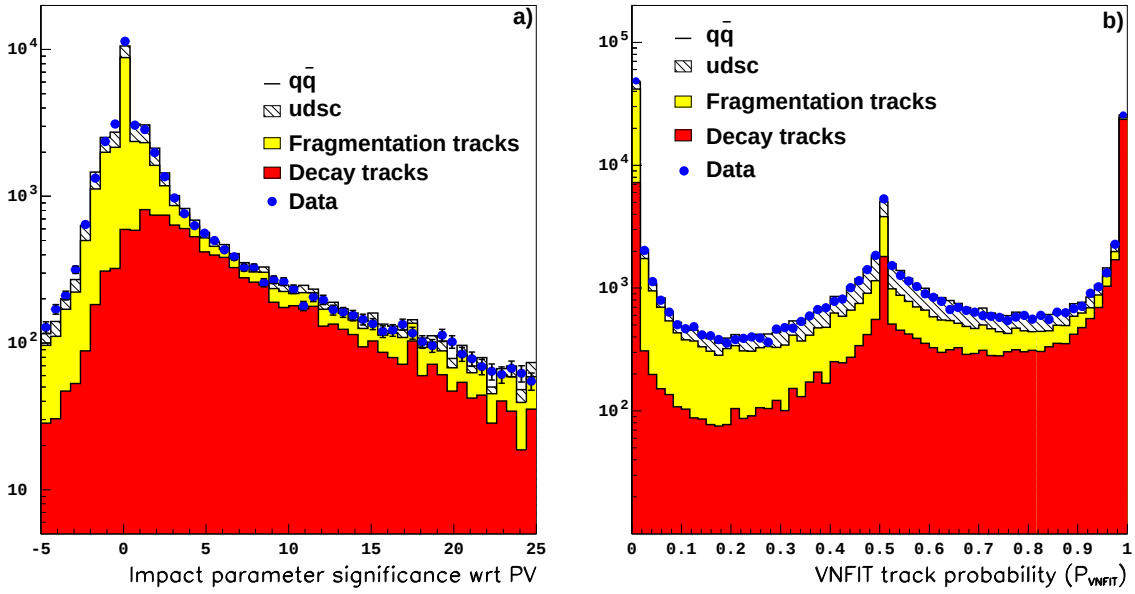

Figure 6.4: Distribution of the discriminant variables for the track separation. a) Impact parameter significance with respect to the primary vertex, and b) VNFIT track probability (compatibility with the decay vertex).

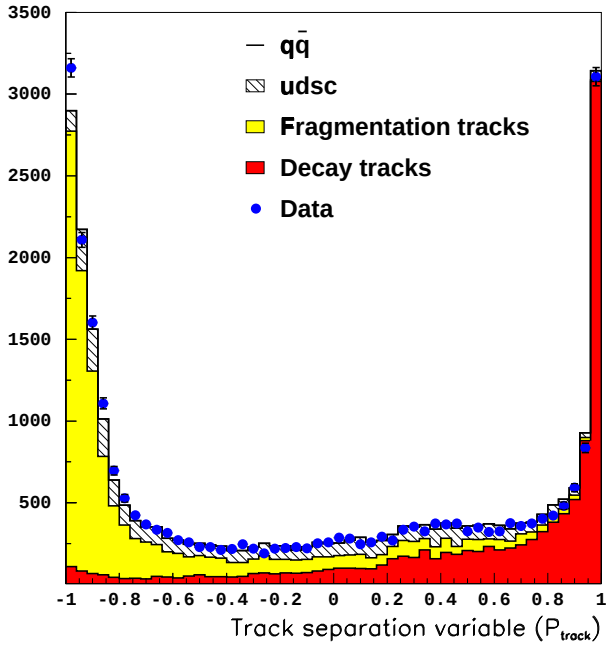

Figure 6.5: Track separation neural network output in events with a vertex reconstructed by VNFIT.

When a VNFIT vertex is not available, all tracks with an impact parameter significance with respect to the QVSRCH vertex smaller than 1.4, a momentum higher than $1.5\,\mathrm{GeV}/c$, and a c-hadron decay vertex neural network probability higher than $\mathrm{P_{track}} > 0$ are taken to form the charmed particle.

Events where no track is selected to form a charmed particle are rejected. In 28% of the events only a single track is selected to form the charmed object. If more than one charged particle is selected, the set of tracks are fit together to obtain the charm vertex.

The purity of the c-track selection is estimated to be 93% on $\mathrm{B_s} \to \ell$ simulated events, with an efficiency of 63%.

6.3.3 Photon recovery

In most of the cases, the charmed hadron decays in charged and neutral particles. In the previous Section, a procedure to select the charged particles was presented. However, if the neutral energy component is ignored, the reconstructed charm direction obtained from the charged particle fit has a larger uncertainty than that estimated from the track fit itself. This effect is translated on the b-hadron decay vertex position obtained from the fit of this charged charmed particle and the lepton. In this Section, a strategy to recover as much as possible of the information from the neutral particles is explained. An improvement on the charm decay direction is only possible when more than one charged particle is found. Indeed, neutral particles cannot help improving the charmed vertex position determination. Only the momentum direction is therefore modified by the inclusion of neutral particles.

The most energetic charged particle included in the charmed object described above is taken as a seed for a cone-based jet. This algorithm is chosen because it provides the best estimate of the decay direction. An angular aperture of 13° between two consecutive particles to be merged is taken. It is verified that all the charged particles included in the charmed object are also included in the cone-based charmed jet. Photons are looked for inside this jet. If more than one photon with energy higher than 1 GeV is found, the most energetic photon in a 16° cone around the jet direction, making an invariant mass of less than $1.8\,\mathrm{GeV}/c^2$ with the charged charmed object is taken.

By default, photons have their momentum calculated as if they were coming from the primary vertex. Here, the momentum is recomputed at the charm vertex to obtain an improvement on the decay length resolution. The position of the electromagnetic shower is defined as the energy-weighted barycentre of the relevant ECAL towers positions. The direction between the charm vertex and the shower centre defines the new photon momentum direction.

The improved charmed object momentum is obtained as the sum of the charged and photon components. Finally the covariance matrix for the new charmed object and the correlation matrix between the new charmed object momentum and the charm vertex are

calculated. The covariance matrix for the charmed object momentum reads

$$\begin{aligned} \mathrm{Cov}_{ij}^{\mathrm{new-mom}} \;=\; & \sum_{kl} \frac{\partial p_i}{\partial p_k}\, \mathrm{Cov}^{\mathrm{mom}}(p_k, p_l)\, \frac{\partial p_j}{\partial p_l} \;+ \\ & \sum_{kl} \frac{\partial p_i}{\partial p_k}\, \mathrm{Cov}^{\mathrm{mom-vertex}}(p_k, v_l)\, \frac{\partial p_j}{\partial v_l} \;+ \\ & \sum_{kl} \frac{\partial p_i}{\partial v_k}\, \mathrm{Cov}^{\mathrm{vertex}}(v_k, v_l)\, \frac{\partial p_j}{\partial v_l} \;+ \\ & \sum_{k} \frac{\partial p_i}{\partial p_k^{\mathrm{phot}}}\, \sigma_k^{\mathrm{phot}}\, \sigma_k^{\mathrm{phot}}\, \frac{\partial p_j}{\partial p_k^{\mathrm{phot}}} \;, \end{aligned} \tag{6.1}$$

where $\mathrm{Cov}^{\mathrm{mom}}$ is the covariance matrix for the charmed object momentum obtained from the fit of the selected c tracks, $\mathrm{Cov}^{\mathrm{mom-vertex}}$ is the correlation matrix between the charged charmed object momentum and the charm vertex position, $\mathrm{Cov}^{\mathrm{vertex}}$ is the covariance matrix of the charm vertex, and σ_k^{phot} is the photon momentum uncertainty in the k-th direction. The correlation matrix between the new charmed object and the charm vertex position reads

$$\mathrm{Cov}_{ij}^{\mathrm{new-vertex-mom}} \;=\; \mathrm{Cov}^{\mathrm{mom-vertex}}(p_i, v_j) \;+\; \sum_{kl} \frac{\partial p_i}{\partial v_k}\, \mathrm{Cov}^{\mathrm{vertex}}(v_k, v_l)\, \frac{\partial v_j}{\partial v_l} \;. \tag{6.2}$$

The addition of a photon to the charged charmed object is found to improve significantly the final b-hadron decay position only in some particular event topologies. In Section 6.3.7 vertex classes are defined and the class of events where a photon is added is identified.

6.3.4 The "B track"

The b-hadron vertex in an inclusive semileptonic selection is generally obtained [10] as the common vertex between the lepton and the charmed object. Here instead, information on the b-hadron flight direction is used to constrain and improve the final estimate of the secondary vertex position. The "B track" is defined as a mathematical object which aims at having the direction and momentum of the b-hadron and is used as a constraint in the global b vertex fit (see Section 6.3.5 below).

The quadri-momentum vector of the "B track" is obtained either from the estimate of the b-hadron flight direction and momentum in the case of multijet events, or from the thrust direction and the b-hadron momentum in the case of two jet events. The "B track" is imposed to originate from the reconstructed primary interaction point.

The estimate of the b-hadron flight direction is obtained from that of a cone-based b-jet formed around the lepton candidate. An angular aperture of 23°, higher than in the case of the charmed jet, is used. The energy of the jet defines the "B track" energy. In the case of two jet events, the direction estimator is chosen to be the thrust axis, which turns out to be more accurate than the jet axis.

To be used as a pseudo-track in the b-vertex fit, the "B track" tracks needs to be complemented with an estimate of its covariance matrix. The covariance matrix is built from the estimated uncertainty on the momentum p of the b-jet, and the uncertainty on the polar and azimuthal angles θ_{B}, φ_{B}, of the direction estimator. The momentum resolution is parametrized as a function of the momentum itself. Three different parametrizations are

taken, one for events with two jets (using the thrust direction and the b-jet momentum), another for events with more than two jets but a single jet in the candidate hemisphere, and a third parametrization for the remaining events (these last two using the cone-based b-jet direction). The polar and azimuthal angles resolutions are also parametrized as a function of the momentum for the same three categories of events. All the parametrizations are obtained from simulated events and optimized for b-jets in semileptonic b-decays. The estimate of the momentum and angles are considered uncorrelated. With all this information the covariance matrix in an (p_x, p_y, p_z, E) basis reads

$$\mathrm{Cov}_{ij} = \sum_{\alpha} \frac{\partial p_i}{\partial p_\alpha} \sigma_{p_\alpha} \sigma_{p_\alpha} \frac{\partial p_j}{\partial p_\alpha} \ , \tag{6.3}$$

where $p_i = p_x, p_y, p_z, E$, and $p_\alpha = p, \theta_{\mathrm{B}}, \varphi_{\mathrm{B}}$. The symbol σ_{p_α} stands for the parametrized resolutions described above.

6.3.5 Global b-vertex fit

The b-hadron decay vertex is obtained from a common fit of the lepton, the charmed object, and the "B track".

With respect to the previous analysis [10], three basic new features are introduced in the b-vertex finding procedure to improve its performance, *i*) the VNFIT inclusive vertex, used here to increase the c track selection efficiency; *ii*) the addition of photons in the charmed particle reconstruction; and *iii*) the use of the "B track" in the final secondary vertex fit.

Most of the improvement in the final estimate of the b-hadron decay position is due to the inclusion of the "B track" in the vertex fit. It introduces information on the decay direction which constrains significantly the vertex fit.

6.3.6 The decay length

The b-hadron decay length is obtained as the three-dimensional distance between the b vertex (Section 6.3.5) and the primary vertex (Section 5.4) positions. In the previous analysis [10], the decay length was estimated as the distance between the primary and secondary vertices projected on the b-jet direction. The projection was found to improve the resolution on the decay length. Here, as the information on the b-hadron flight direction is already included in the secondary vertex fit, a projection does not improve the resolution.

6.3.7 Vertex selection and class definition

A critical study of the previous analysis pointed out that the event selection and the event-by-event estimate of the decay length uncertainty were key issues to achieve an improved performance at high oscillation frequency. In particular, it was clear that defining resolution classes based on the topological properties of the events would be highly beneficial in that sense.

Some of the topological characteristics of the events which serve as basis for the event selection and for the definition of resolution classes are as follows.

- The presence of a VNFIT vertex. Events for which the VNFIT algorithm does not provide a secondary vertex have a less accurate inclusive vertex position estimate to be used as a seed.

- The χ^2 of the reconstructed b-hadron vertex and the χ^2 of the reconstructed charm vertex (if any).

- The angle between the reconstructed charmed particle and the lepton. On one hand, the bigger is the angle between the charmed particle and the lepton, the bigger is the probability of including fragmentation particles in the charm track selection; on the other hand, if the charmed particle and the lepton are close to be parallel, the reconstructed b-hadron vertex has a large uncertainty. Therefore this angle is limited from below and from above.

- The angle between the reconstructed charmed particle and the b-jet. Similar arguments as for the angle with the lepton justifies a cut on the angle of the charmed particle with the b-jet.

- The mass of the charmed particle. In signal events, the charmed particle is expected to be a D_s with a mass $m_{D_s} = 1968.6 \pm 0.6\,\mathrm{MeV}/c^2$ [13]. Therefore the reconstructed charmed particle should not be much heavier than m_{D_s}.

A detailed description of the classes is presented in Table 6.1. These classes are exclusive: if an event fulfils the requirements of more than one class, it is assigned to the first in the list. For classes 3 and 4, the b-hadron decay length is improved when photons are added to the charmed particle as explained in Section 6.3.3. The decay length resolution achieved for two of the vertex classes (the one with the best resolution and the one with the worst resolution) is shown in Fig. 6.6.

The selection criteria described above are tailored to select semileptonic B_s decays, and therefore result in a rejection of light-quark (udsc) hadron decays.

6.3.8 The bias and pull correction

The decay length evaluated as in the previous Sections is found to be slightly biased. This bias, measured in simulated events, shows a dependence on the reconstructed decay length itself. The typical dependence observed is shown in Fig. 6.7 for $B_s \to \ell$ simulated events in the first resolution class. The decay length bias is in all cases smaller than the estimated uncertainty on the decay length.

A correction for this bias is applied to all events in every vertex class. The effect of the bias correction on the decay length resolution in the overall selected sample is shown in Fig. 6.8. The resolution is slightly improved (more events are found in the core) and the bias is reduced on average from $120\,\mu$m to $28\,\mu$m.

An estimate of the decay length uncertainty is obtained, event by event, from the covariance matrices of the primary and secondary vertices, and it is dominated by the uncertainty on the secondary vertex position. The uncertainty on the secondary vertex is estimated from the uncertainty on the helix parameters of the tracks involved, but does not include any contribution from misassigned and missing particles in the charm candidate. As a consequence,

Class number	Vertex properties	Topological properties
Class 1	No VNFIT vertex $\chi^2_{\rm b} < 7.$	$N_{\rm c} = 1$ $5.7^\circ < \widehat{\rm c\,\ell} < 30^\circ$ $\widehat{\rm c\,jet} < 20^\circ$ $0 \leq m_{\rm b\ reco} \leq 8\,{\rm GeV}/c^2$
Class 2	No VNFIT vertex $\chi^2_{\rm b} < 7.$	$N_{\rm c} > 1$ $m_{\rm charm} < 3\,{\rm GeV}/c^2$ $5.7^\circ < \widehat{\rm c\,\ell} < 30^\circ$ $\widehat{\rm c\,jet} < 20^\circ$ $0 \leq m_{\rm b\ reco} \leq 8\,{\rm GeV}/c^2$
Class 3	VNFIT vertex $\chi^2_{\rm b} < 7.$	$N_{\rm c} = 2$ $5.7^\circ < \widehat{\rm c\,\ell} < 37^\circ$ $\widehat{\rm c\,jet} < 20^\circ$
Class 4	VNFIT vertex $\chi^2_{\rm b} < 7.$ $\chi^2_{\rm c} < 5.$	$N_{\rm c} > 2$ $m_{\rm charm} < 3\,{\rm GeV}/c^2$ $5.7^\circ < \widehat{\rm c\,\ell} < 37^\circ$ $\widehat{\rm c\,jet} < 20^\circ$
Class 5	VNFIT vertex $\chi^2_{\rm b} < 25.$ $\chi^2_{\rm c} < 10.$	$N_{\rm c} \geq 2$ $m_{\rm charm} < 3\,{\rm GeV}/c^2$ $\widehat{\rm c\,\ell} < 37^\circ$ $\widehat{\rm c\,jet} < 20^\circ$
Class 6	VNFIT vertex $\chi^2_{\rm b} < 7.$ $\chi^2_{\rm c} < 5.$	$N_{\rm c} = 2$ $5.7^\circ < \widehat{\rm c\,\ell} < 37^\circ$
Class 7	VNFIT vertex $\chi^2_{\rm b} < 7.$ $\chi^2_{\rm c} < 5.$	$N_{\rm c} > 2$ $5.7^\circ < \widehat{\rm c\,\ell} < 37^\circ$
Class 8	No VNFIT vertex $\chi^2_{\rm b} < 7.$ $\chi^2_{\rm c} < 5.$	$N_{\rm c} \geq 2$ $m_{\rm charm} < 3\,{\rm GeV}/c^2$ $5.7^\circ < \widehat{\rm c\,\ell} < 37^\circ$
Class 9	VNFIT vertex $\chi^2_{\rm b} < 20.$ $\chi^2_{\rm c} < 5.$	$N_{\rm c} \geq 2$ $5.^\circ < \widehat{\rm c\,\ell} < 15^\circ$
Class 10	VNFIT vertex $\chi^2_{\rm b} < 2.$ $\chi^2_{\rm c} < 1.$	$N_{\rm c} \geq 1$ $11.^\circ < \widehat{\rm c\,\ell} < 41.^\circ$

Table 6.1: Definition of the vertex classes. The number of tracks in the charged charmed object is indicated as $N_{\rm c}$; the mass of the charmed particle as $m_{\rm charm}$; the χ^2 of the b-hadron vertex as $\chi^2_{\rm b}$; the χ^2 of the charm vertex as $\chi^2_{\rm c}$; the angle between the charmed particle and the lepton as $\widehat{\rm c\,\ell}$; the angle between the charmed particle and the b-jet as $\widehat{\rm c\,jet}$; and the reconstructed mass of the b-hadron as $m_{\rm b\ reco}$.

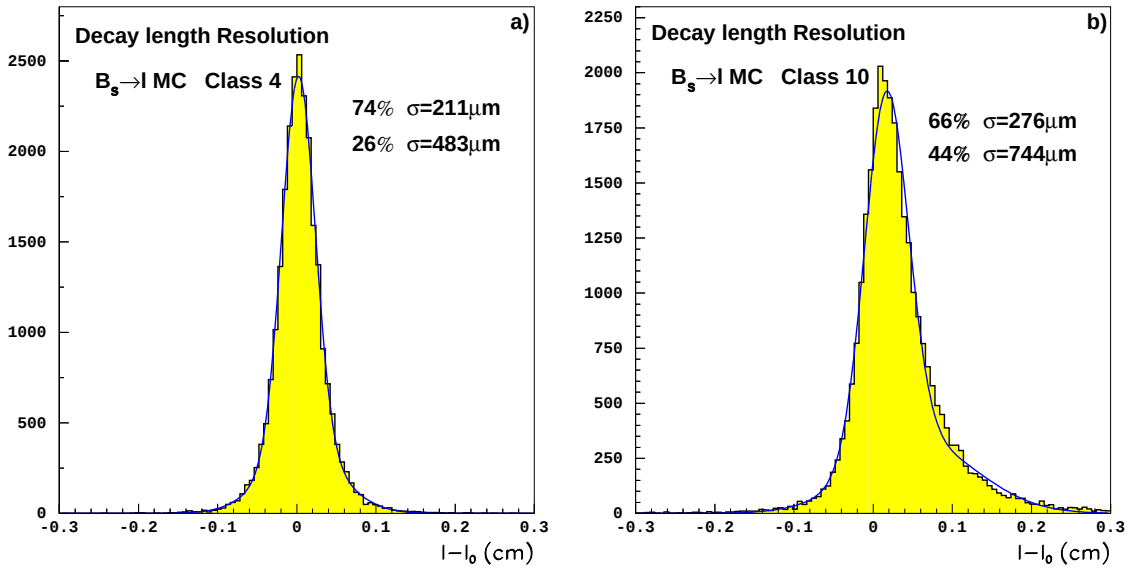

Figure 6.6: Decay length resolution on $B_s \to \ell$ simulated events in two vertex classes, a) the class with the best resolution, and b) the class with the worst resolution. The distributions are fit with the sum of two Gaussians, their fractions and widths are indicated on the plot.

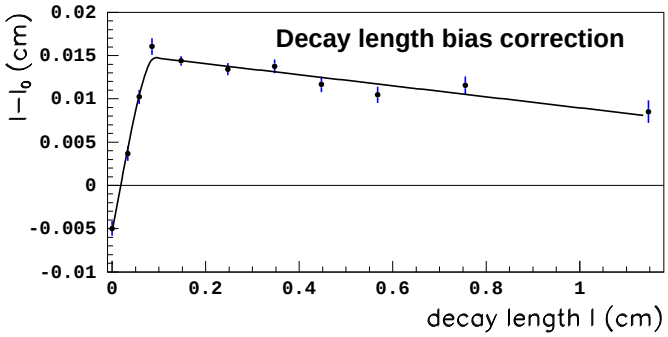

Figure 6.7: Decay length bias as a function of the reconstructed decay length.

the decay length uncertainty obtained from the track fit, is in general, underestimated and a correction based on simulated events needs to be applied. This correction is called *pull correction* in the following, because it can be illustrated with the distribution of $(l - l_0)/\sigma_l$, where l is the reconstructed decay length, l_0 is the true decay length, and σ_l is the estimated uncertainty on the decay length. If the decay length uncertainty were properly estimated for all events, this distribution would be a normal Gaussian.

In the previous analysis a global correction for the selected sample was computed and applied. Here, instead, the problem is addressed class by class to obtain a more accurate description. For the vertex classes in which the charmed particle is made of a single charged particle, the pull correction is parametrized as a function of the reconstructed decay length.

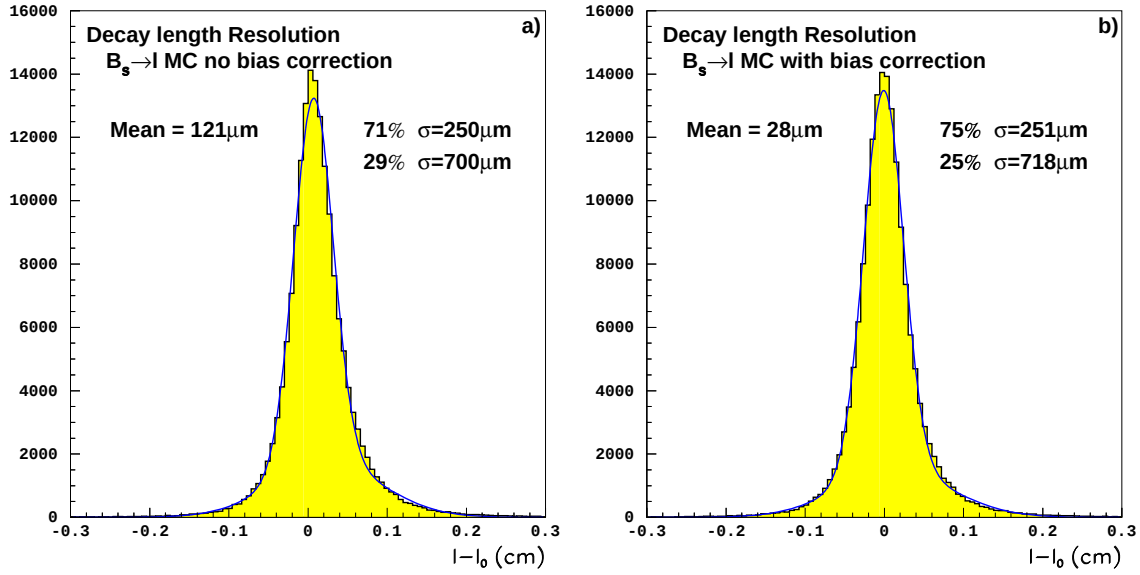

Figure 6.8: Decay length distribution for all selected events, a) before the bias correction, b) after the bias correction is applied.

For the other classes, a double parametrization is performed. The decay length uncertainty is first corrected as a function of the reconstructed charmed mass and then as a function of the reconstructed decay length. In general, as intuitively expected, the correction increases with the decay length and decreases with the reconstructed mass of the charmed particle. The pull distribution for the selected sample is shown in Fig. 6.9, both before and after the parametrized correction is applied.

The impact of the parametrizations described in this Section on the analysis sensitivity is explained in Section 8.4.2.

6.4 The b-hadron enrichment

The requirement of a lepton candidate, and the vertex selection described above, act both as rejection criteria for Z decays in light-quarks. Therefore, the sample obtained at this stage of the selection procedure is enriched in $Z \to b\overline{b}$ decays. Before any further selection is applied, the sample composition estimated from $Z \to q\overline{q}$ simulated events is shown in Table 6.2.

$(82.6 \pm 0.3)\%$ $b\overline{b}$	$(11.00 \pm 0.07)\%$ $c\overline{c}$	$(6.40 \pm 0.05)\%$ $u\overline{u}$, $d\overline{d}$, $s\overline{s}$

Table 6.2: Sample composition after the lepton and vertex selections are applied.

However, a B_s oscillation analysis requires a $b\overline{b}$ purity higher than the $\sim 83\%$ obtained. Therefore, a specific b-tagging variable is designed to increase the $b\overline{b}$ content in the final

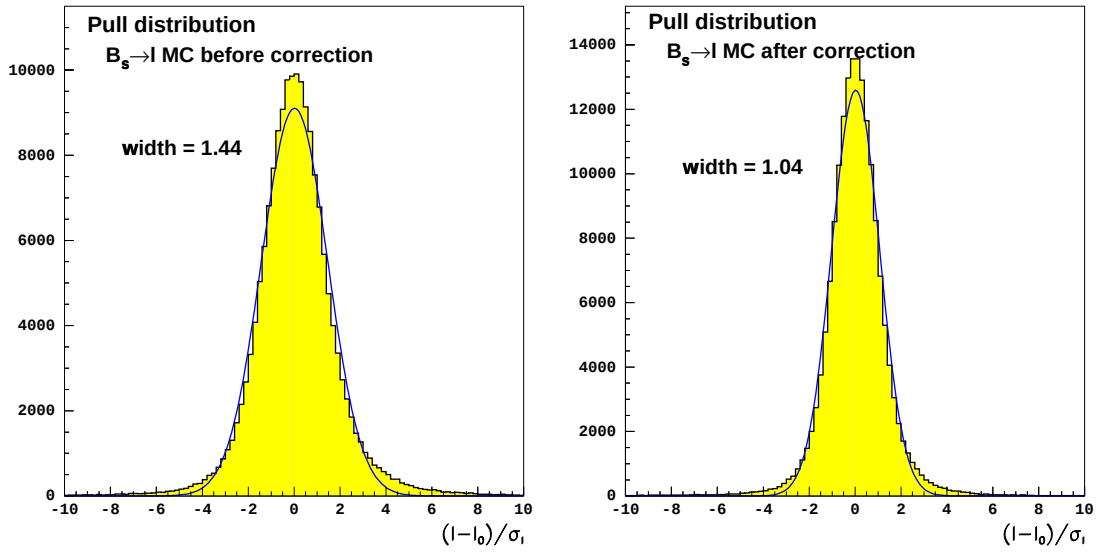

Figure 6.9: Distribution of the decay length pulls, a) before the correction, and b) after the correction is applied.

selected sample.

Several variables which distinguish between $Z \to b\overline{b}$ and other Z decays are combined with a neural network. The hard fragmentation of b quarks into b-hadrons and the high mass of b-hadrons result respectively, in a large momentum and large transverse momentum of the primary lepton compared to that of leptons produced in the decay of light-quark hadrons. The variables described in Section 5.7 for b-tagging are sensitive to both the lifetime and the mass of the hadrons produced in $Z \to q\overline{q}$. All these properties are used to define the set of variables for the b-tagging developed here. The variables are as follows. *i*) The lepton momentum; *ii*) the lepton transverse momentum; *iii*) the same side QIPBTAG probability (Section 5.7.1); *iv*) the opposite side QIPBTAG probability; *v*) the same side QBMTAG estimator (Section 5.7.3); *vi*) the opposite side QBMTAG estimator; *vii*) the same side QVSRCH estimator (Section 5.7.2); and *viii*) the opposite side QVSRCH estimator. The distributions of all the variables which are to be combined in the b-tagging neural network are shown, for events with a candidate lepton and a reconstructed vertex, in Figs. 6.10 to 6.13. Simulated events with a Z decay into light-quarks are labeled as "uds". The flavour discrimination power is shown as well as the agreement between the data and the simulated $Z \to q\overline{q}$ events.

The neural network training is performed with only events which pass both the lepton and the vertex selections. The performance of the combined b-tagging variable is shown in Fig. 6.14.

The performance of the b-tagging variable developed here can be compared with that of a more standard approach (*i.e.*, the sole QIPBTAG event probability) which is normally used to select unbiased $b\overline{b}$ samples (not semileptonic decays specifically). In Fig. 6.15, the selection efficiency on $b\overline{b}$ events versus the $b\overline{b}$ purity curves are displayed for the b-tagging presented here, N_{btag}, and for the QIPBTAG variable defined in the whole event, both applied to the

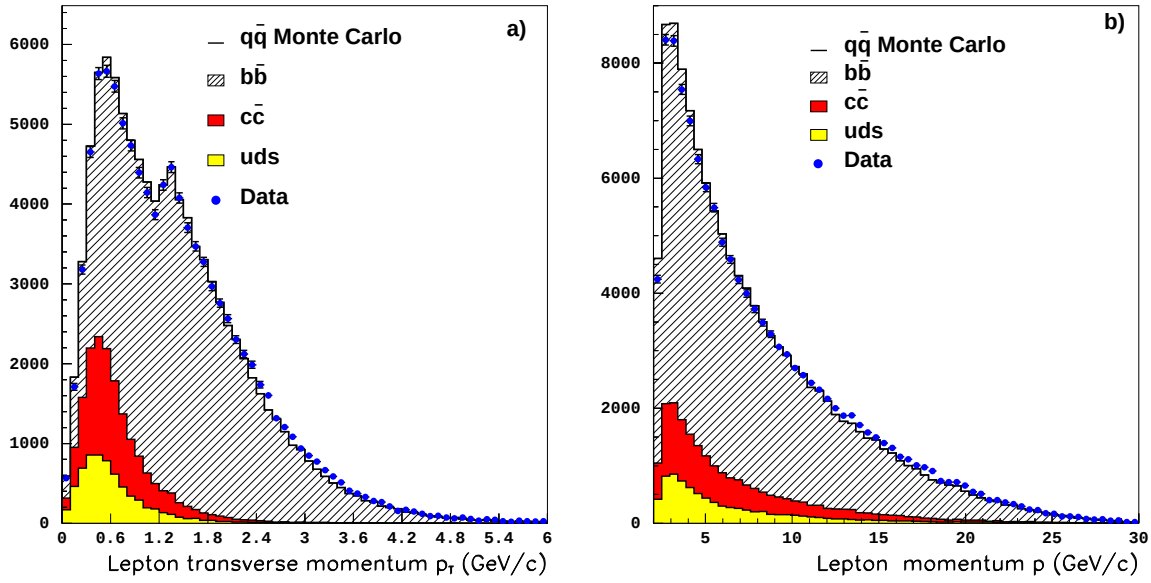

Figure 6.10: Lepton variables used for b-tagging. a) Lepton transverse momentum, b) lepton momentum

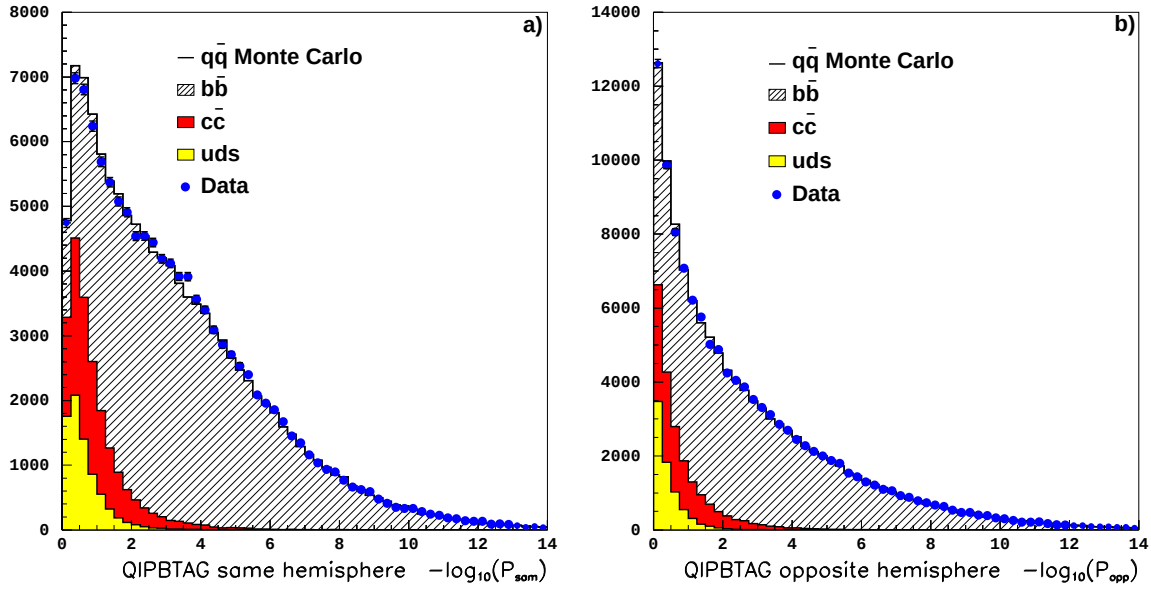

Figure 6.11: QIPBTAG variables used for b-tagging. a) Same side hemisphere probability, b) opposite side hemisphere probability, $-\log_{10}$ of the probability is plotted.

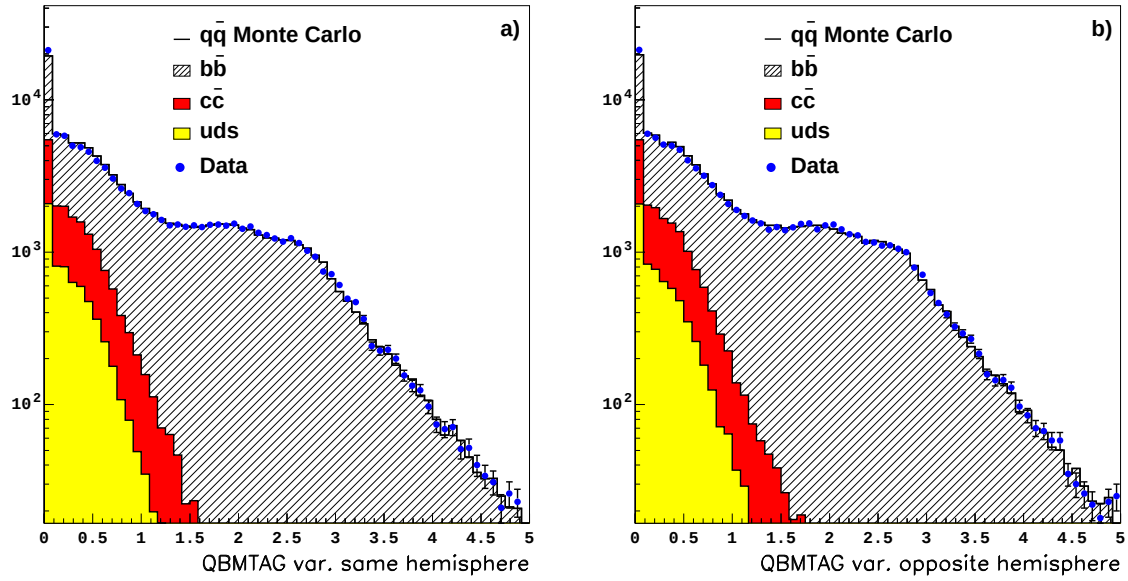

Figure 6.12: QBMTAG variables used for b-tagging. a) Same side hemisphere variable, b) opposite side hemisphere variable.

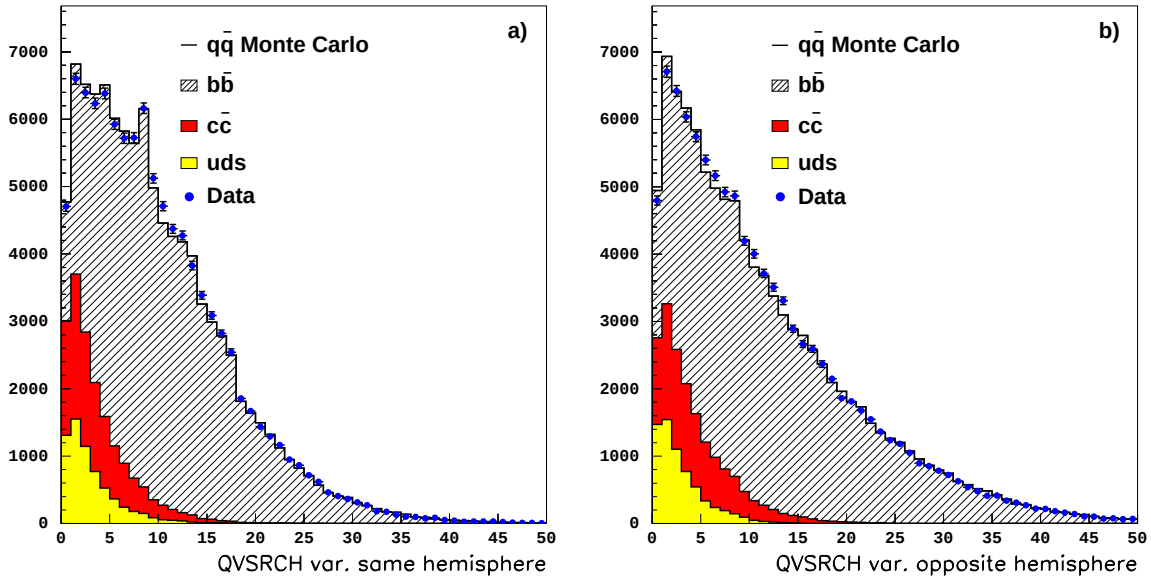

Figure 6.13: QVSRCH variables used for b-tagging. a) Same side hemisphere variable, b) opposite side hemisphere variable.

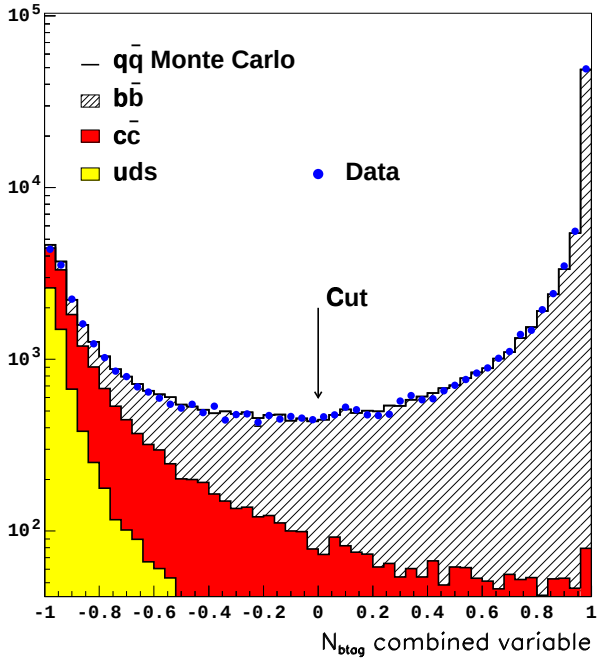

Figure 6.14: Combined b-tagging variable. The distribution of all the sample components is shown, as well as the agreement with the data.

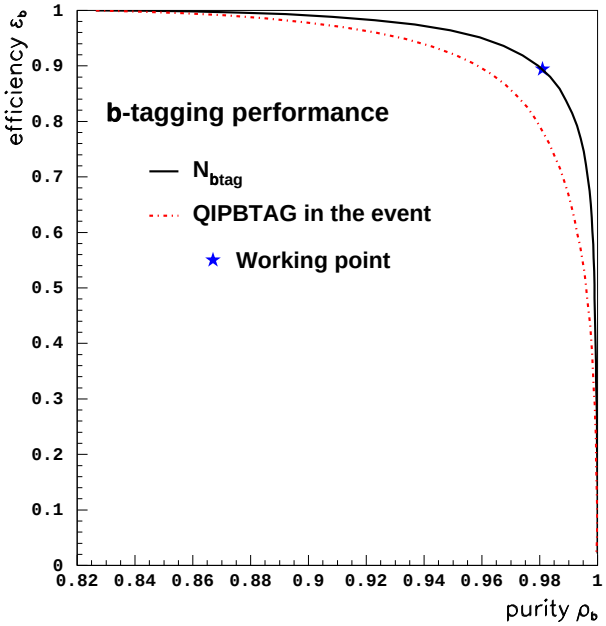

Figure 6.15: Comparison of b-tagging performance, the purity *versus* efficiency curve is plotted for the event QIPBTAG variable, and for that developed here. The working point chosen for the analysis is also indicated.

lepton plus vertex selected sample.

The improvement obtained upon the event QIPBTAG variable is mainly due to the fact that the b-tagging optimization, both for what concerns the selection of variables and what concerns the training, has been done on the specific events it is meant to be used on.

The probability of an event to originate from a $Z \to b\overline{b}$ decay is defined by the ratio of the sample components in Fig. 6.15 (the same probability definition is taken in the following). Events with very low probability of being a $Z \to b\overline{b}$ decay are rejected, an optimal cut is found at $N_{\rm btag} > 0$. The sample selected with this cut has the composition presented in Table 6.3. The selection efficiency for $Z \to b\overline{b}$ events, in a sample in which the lepton and vertex selections are applied, is estimated to be $\varepsilon_{\rm b} = 89.4 \pm 0.3\%$.

$(98.1 \pm 0.3)\%$ $b\overline{b}$	$(1.64 \pm 0.03)\%$ $c\overline{c}$	$(0.26 \pm 0.01)\%$ $u\overline{u}$, $d\overline{d}$, $s\overline{s}$

Table 6.3: Sample composition after the lepton, the vertex, and the b-tagging selections are applied.

The sample composition in $Z \to b\overline{b}$ or $Z \to u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$ is parametrized as a function of the combined b-tagging variable $N_{\rm btag}$ on simulated events. An event-by-event signal purity is defined from this parametrization and then used in the oscillation fit (Section 8.1.1).

6.5 The b → ℓ enrichment

The sign of the electric charge of the lepton candidate is used to identify the b ($\bar{\rm b}$) content of the b-hadron at decay time. As explained in Section 3.1.2 (and Figs. 3.3, 3.4, and 3.5), the presence of a negatively (positively) charged lepton is taken as a tag for a b-hadron ($\bar{\rm b}$-hadron). However, a lepton is not guaranteed to be produced via a $b \to \ell$ quark decay. Cascade decays such as $b \to c \to \ell$ and $b \to \overline{c} \to \ell$ are also possible. In the case of $b \to c \to \ell$, the charge correlation between the lepton and the b-quark is opposite to that of $b \to \ell$ decays. The process $b \to \overline{c} \to \ell$ has the same charge correlation as the $b \to \ell$. In this Section, a method to statistically separate $b \to \ell$ decays from the other decays is presented.

Several topological differences distinguish direct semileptonic decays from the rest. Eight discriminant variables are identified here and combined with a neural network to obtain an optimal separation between $b \to \ell$ and $b \to c \to \ell$ decays.

The lepton momentum and transverse momentum spectra are expected to be harder for direct $b \to \ell$ decays than for other decays. The energy of the neutrino determined as in Ref. [84] (Section 6.7.1), is harder for direct semileptonic decays as well.

Some properties of the b-hadron jet which contains the lepton candidate are also considered. The decays of the type $b \to \ell$ ($B \to \ell\nu D$) and $b \to c \to \ell$ ($B \to WD(D \to \ell\nu X)$), where D refers to either D^0, D^+, D^*, or D^{**}, lead to significantly different jet topologies which are exploited here. The boost of the b-hadron, however, tends to dilute some of these differences. It is therefore more appropriate to study the separation in the rest frame of the lepton and the charmed particle, referred as the (ℓD) rest frame in the following. All the charged and neutral particles in the b-hadron jet (defined with the cone-based algorithm)

are boosted back in the (ℓD) rest frame. Two hemispheres are defined with respect to the plane orthogonal to the lepton direction. In Fig. 6.16, a schematic of the effect of the boost on typical topologies of b $\to \ell$ and b $\to$ c $\to \ell$ events is shown.

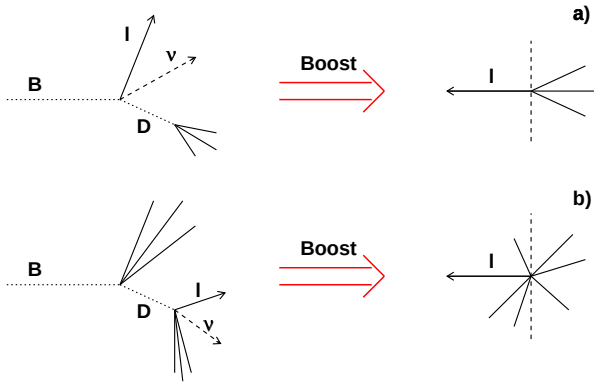

Figure 6.16: A schematic of b $\to \ell$ (a) and b $\to$ c $\to \ell$ (b) events topologies before and after applying a boost to the (ℓD) rest frame is shown. The plane used to define hemispheres in the (ℓD) rest frame is also displayed.

If no fragmentation particle were included in the b-hadron jet used to define the (ℓD) rest frame, the D and the lepton from the B $\to \ell\nu$D decay would be, by construction, produced back to back in this frame. In contrast, a lepton originating from the cascade decay B $\to$ WD(D $\to \ell\nu$X) is expected to be found close in phase space to the X system. The variable $E_{\rm CM}$ is defined as the sum of energies of the particles in the lepton hemisphere (except the lepton itself). A low value of $E_{\rm CM}$ is characteristic of a b $\to \ell$ decay.

A second shape variable, $P_{\ell \rm t}$, is designed to be sensitive to the two "directions of hadronization" present in most of the b $\to$ c $\to \ell$ decays (one direction for the c-quark decay and the other for the other b-hadron decay products), and not in b $\to \ell$ decays. This variable is defined as, $P_{\ell \rm t} = \frac{|P_+ - P_-|}{|P_+ + P_-|}$, where P_+ is the sum of the parallel momenta of the particles in the lepton hemisphere (with the (ℓD) rest frame definition) with respect to the lepton direction, and P_- is the equivalent quantity for particles in the hemisphere opposite to the lepton.

The total charged energy of the b-jet, and the lepton energy, both computed in the (ℓD) rest frame, also discriminate between b $\to \ell$ and b $\to$ c $\to \ell$ decays.

Finally, the signed impact parameter significance (see Section 5.5) of the lepton with respect to the charm vertex is computed, with the sign given by the charmed particle flight direction (Section 5.5). For events with two or more charged particles in the reconstructed charmed meson, the charm vertex (Section 6.3.2) is used. The inclusive tertiary vertex is used otherwise. The pull distributions for the two cases were checked, and the impact parameter uncertainty was corrected for the observed deviations. In the first case, the correction factor was found to be 1.19 and in the second 1.75. The lepton is expected to come prior to the charm vertex for b $\to \ell$ decays, as seen in Fig. 6.17, and therefore to have a negative impact parameter significance.

The distribution of all the variables described above are shown in Figs. 6.18 to 6.21 for

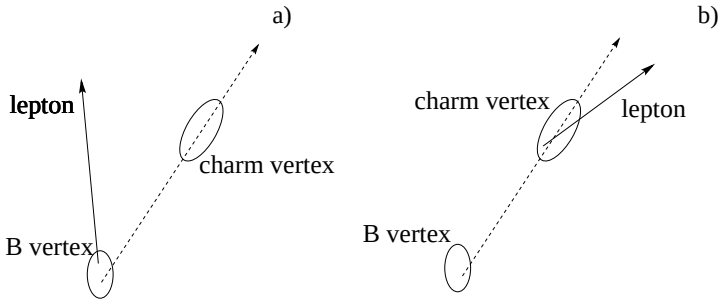

Figure 6.17: Schematic of the lepton impact parameter with respect to the charm vertex for a) a typical b → ℓ decay configuration with negative impact parameter with respect to the charm vertex, and b) a typical b → c → ℓ decay with, on average, zero impact parameter with respect to the same vertex.

events with a lepton candidate and a reconstructed secondary vertex. The simulated events are divided in four categories: *i*) events in which the lepton originates from light-quark decays or events in which a hadron is misidentified as a lepton, called "udsc and fake ℓ"; *ii*) events in which the lepton originates from a b → $\overline{c}$ → ℓ; *iii*) events in which the lepton originates from a cascade decay b → c → ℓ; and *iv*) the direct b → ℓ semileptonic decays. Agreement between data and simulation is observed for all variables.

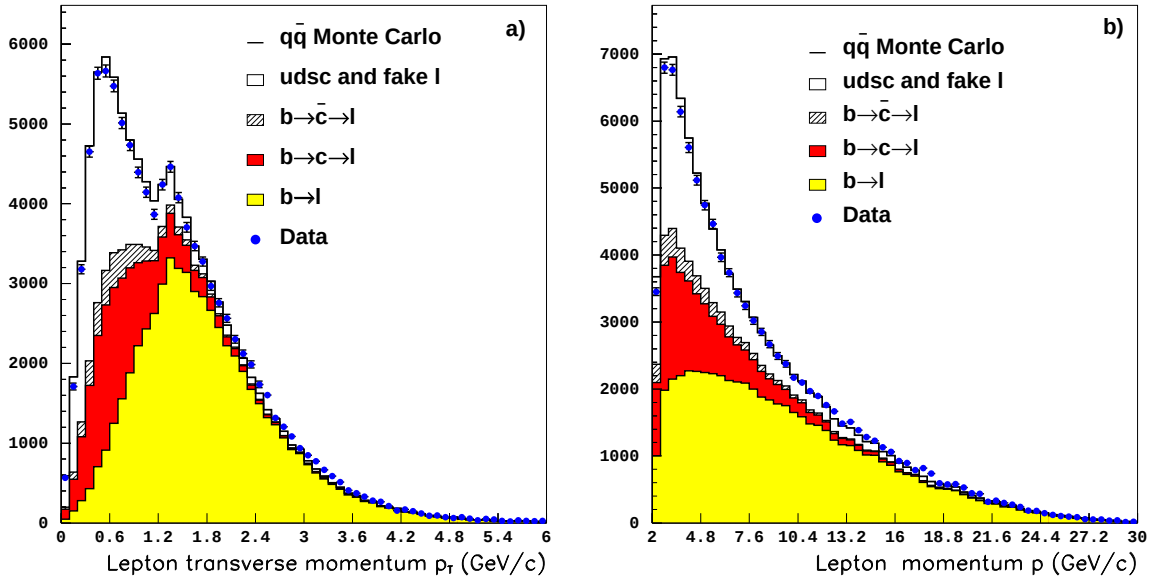

Figure 6.18: Lepton variables used for b → ℓ enrichment. a) Lepton transverse momentum, b) lepton momentum.

The b → ℓ enrichment variable combination was done with simulated B_s → ℓ events as signal and B_s → c → ℓ events as background, to obtain the best separation performance for the B_s oscillations analysis. The neural network was trained with a simulated sample in

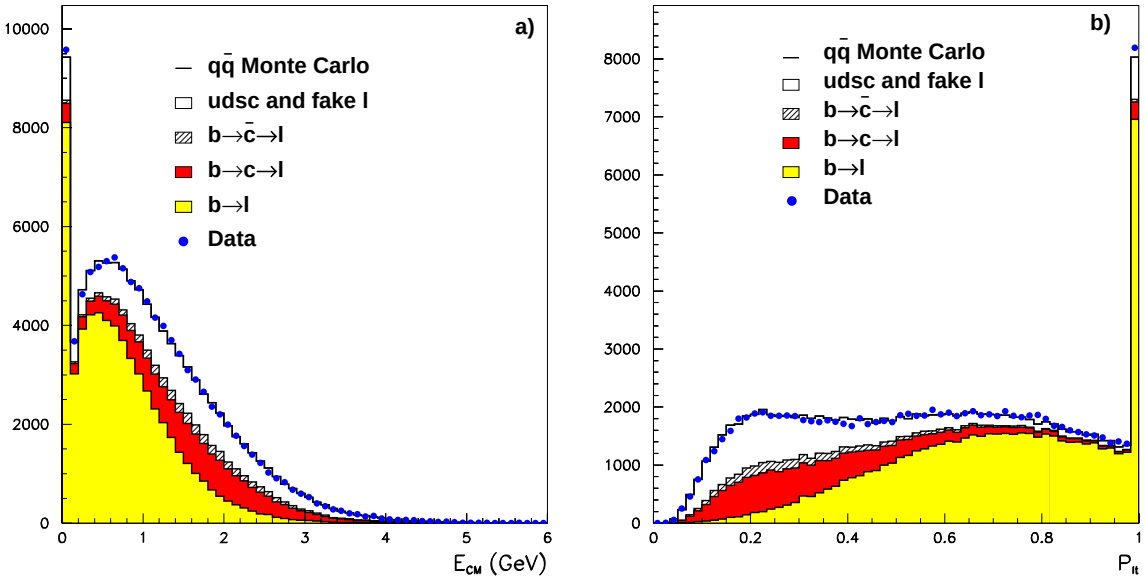

Figure 6.19: Shape variables used for b $\rightarrow \ell$ enrichment. a) Energy of the lepton hemisphere, with the lepton excluded, b) normalized sum of the track momenta parallel to the lepton.

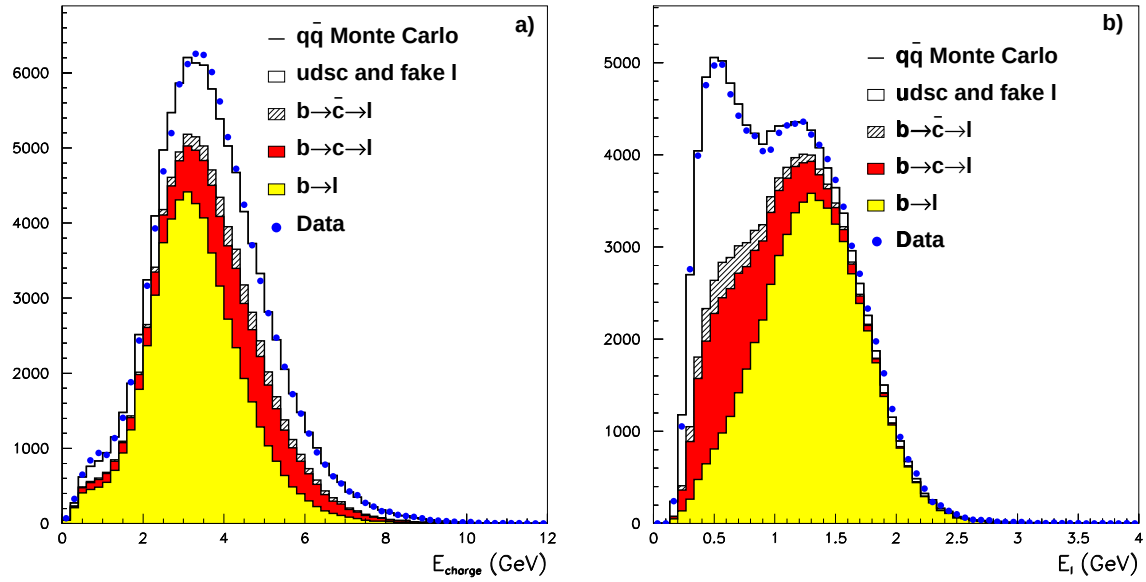

Figure 6.20: a) Charged energy of the b-jet, and b) lepton energy, both computed in the (ℓD) rest frame.

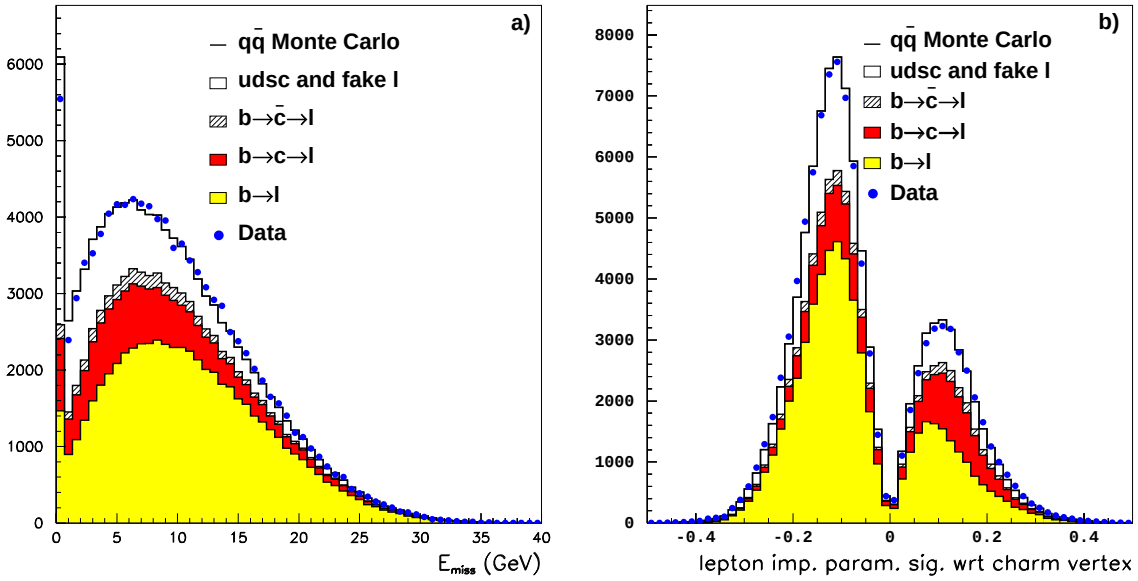

Figure 6.21: a) Neutrino energy, $E_{\rm miss}$, and b) lepton signed impact parameter significance. For the sake of visual clarity, the variable actually displayed is a function of the impact parameter significance $\mathrm{D}/\sigma_{\mathrm{D}}$ as: $\log\left[1+\left|\left(\log(1+|\mathrm{D}/\sigma_{\mathrm{D}}|)\right)\times\mathrm{sign(D)}\right|\right]\times\mathrm{sign(D)}$.

which only the lepton and the vertex selection criteria were applied.

The distribution of $N_{\rm bltag}$, the combined b → ℓ enrichment variable, is shown in Fig. 6.22, before and after the cut on $N_{\rm btag}$. In both cases a reasonable agreement between the data and the simulation is found.

A selection cut is applied on the $N_{\rm bltag}$ variable, optimized to obtain the highest statistical sensitivity for $\mathrm{B_s}$ oscillations. All events for which the probability for the lepton charge and the final state b-quark flavour to have the same correlation to that expected in b → ℓ decays is 50% or lower are rejected. The optimization is performed on simulated $\mathrm{B_s}$ decays and the cut is set at $N_{\rm bltag} > -0.5$.

The performance of the b → ℓ enrichment on $\mathrm{Z} \to \mathrm{b\bar{b}}$ simulated events are shown in Table 6.4. The sample composition and b → ℓ purity is given for the lepton plus vertex selected sample, after the b-tagging selection, and after both the b-tagging and the b → ℓ enrichment selections.

The performance of $N_{\rm bltag}$ can be compared to that of a simple cut in the lepton transverse momentum, as it was done in the previous analysis. The efficiency versus purity curve of a b → ℓ decays selection is displayed in Fig. 6.23, for a sample where the lepton and the vertex selection criteria are applied. The use of more information than the lepton transverse momentum brings a sizeable improvement on the b → ℓ enrichment performance.

The mistag probability due to b → c → ℓ decays is parametrized as a function of $N_{\rm bltag}$, and used event by event in the oscillation fit (Section 8.1.1) to increase the statistical power of the analysis. This parametrization is done independently for the four types of b-hadrons ($\mathrm{B_s}$, $\mathrm{B_d}$, $\mathrm{B^+}$, and b-baryons) because the b → c → ℓ rejection performance is not the same

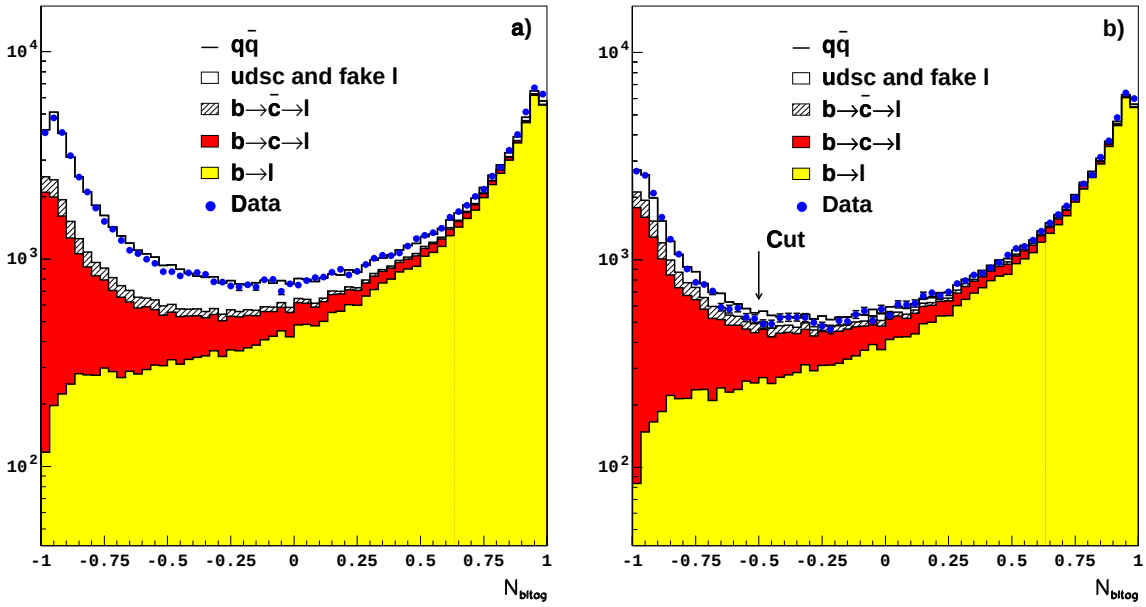

Figure 6.22: Combined b $\to \ell$ enrichment variable distribution, a) before the b-tagging cut, and b) after the b-tagging cut is applied.

	lepton+vertex	+ b-tagging	+ b $\to \ell$ enrichment
b $\to \ell$ purity on b$\overline{\text{b}}$	$(67.5 \pm 0.3)\%$	$(69.6 \pm 0.3)\%$	$(86.0 \pm 0.4)\%$
b $\to$ c $\to \ell$ fraction on b$\overline{\text{b}}$	$(22.6 \pm 0.1)\%$	$(21.1 \pm 0.1)\%$	$(9.7 \pm 0.1)\%$
b $\to \overline{\text{c}} \to \ell$ fraction on b$\overline{\text{b}}$	$(3.96 \pm 0.05)\%$	$(3.71 \pm 0.05)\%$	$(1.70 \pm 0.04)\%$
other lepton candidates	$(5.94 \pm 0.06)\%$	$(5.59 \pm 0.06)\%$	$(2.6 \pm 0.04)\%$
"correct" lepton sign fraction	$(71.5 \pm 0.2)\%$	$(73.3 \pm 0.3)\%$	$(87.7 \pm 0.3)\%$

Table 6.4: Performance of the b $\to \ell$ enrichment on Z $\to$ b$\overline{\text{b}}$. In the last row, the fraction of events in which the lepton charge has the "correct" correlation with the b-hadron flavour is indicated.

for all of them. Differences are also seen upon the vertex classes defined in Section 6.3.7, and are taken into account. For illustration, the b $\to \ell$ purity is shown in Fig. 6.24 for B_s simulated events in different vertex classes.

6.6 Final selection

The b-tagging and the b $\to \ell$ enrichment variables (N_{btag}, N_{bltag}) are treated independently in the previous two Sections, both for the cut definitions and for the parametrization of the b and b $\to \ell$ event-by-event purities. Although these variables are actually not independent, a combined treatment was found not to bring a significant improvement, and it was dropped for simplicity.

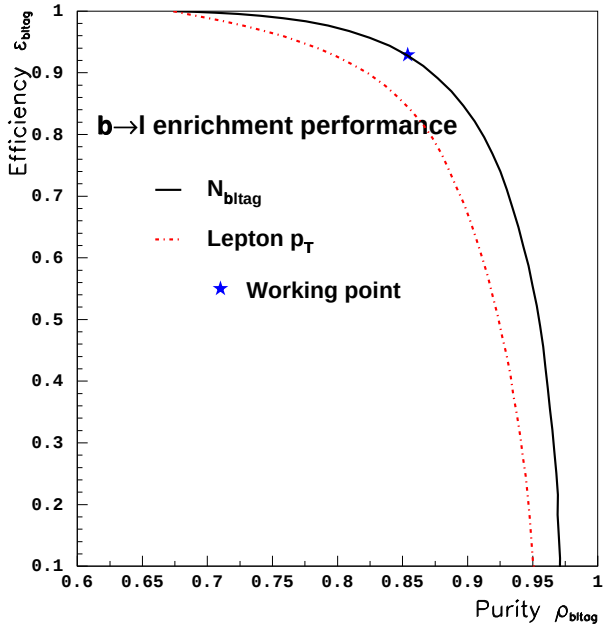

Figure 6.23: Comparison of b $\rightarrow \ell$ enrichment performance. The purity *vs* efficiency curve is displayed for the lepton p_T cut (dashed curve), and for a cut on the N_{bltag} variable (full curve).

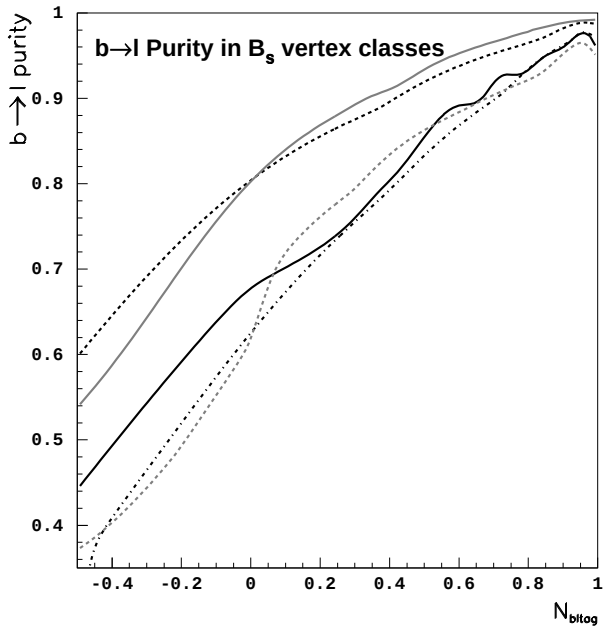

Figure 6.24: b $\rightarrow \ell$ purity as a function of N_{bltag} for $\mathrm{B_s}$ events in different vertex classes.