LHCb-PUB-2009-032
January 13, 2010

Particle identification for charmless two-body B decays at LHCb

A. Carbone, S. Perazzini, V. Vagnoni.

Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Italy

Abstract

Due to the possibility to produce B^0, B_s^0 and Λ_b hadrons in the proton-proton collisions at the LHC, the mass peaks of many two-body B decays overlap. For this reason, the particle identification capabilities of the LHCb detector will play a crucial role in the measurements related to these decay modes. In this document we propose a method for combining different observables provided by the LHCb particle identification system into a single generalized discriminating variable, useful when performing the statistical separation of the various channels in event-by-event fits.

1 Introduction

The large beauty production cross section $\sigma_{b\bar{b}} \simeq 500\ \mu$b expected at the LHC [1] will allow the LHCb detector [2] to reconstruct some 10^5 $H_b \to h^+h'^-$ events per year, where h and h' stand for a pion, kaon or proton. Due to the production of B^0, B_s^0 and Λ_b hadrons in the 14 TeV proton-proton collisions, the mass peaks of many two-body B decays overlap (see Fig. 1) and an excellent control of the particle identification (PID) observables is required in order to isolate the various decay modes.

PID at LHCb is based on the information provided by the RICH system [3], the electromagnetic and hadronic calorimeters [4], including pre-shower and scintillator pad detectors, and the muon chambers [5]. The variable used to discriminate between different particle hypotheses is $\Delta \log \mathcal{L}_{AB}$, that for each track is defined as:

$$\Delta \log \mathcal{L}_{AB} = \log \mathcal{L}_A - \log \mathcal{L}_B, \tag{1}$$

where $\mathcal{L}_A$ and $\mathcal{L}_B$ are the likelihoods for particle hypotheses A and B, respectively.

For each track, the PID system gives an answer in terms of the four observables $\Delta \log \mathcal{L}_{e\pi}$, $\Delta \log \mathcal{L}_{\mu\pi}$, $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$. In the studies presented in this document, we only consider the two observables $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$. These observables are sufficient to discriminate particles coming from all the $H_b \to h^+h'^-$ decays, since signal decays contain neither electrons nor muons.

The $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$ distributions for true pions, kaons and protons from a sample of offline selected $H_b \to h^+h'^-$ events, including all the different decay modes, are shown in Fig. 2. The offline event selection is described in Ref. [6].

2 Combining $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$

As is apparent from Fig. 3, the $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$ variables are strongly correlated. Hence the employment of such observables in order to discriminate between the various mass hypotheses in event-by-event fits would require the use of 2-dimensional joint probability density functions (p.d.f.'s) of $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$, properly taking into account the correlation between the two. Although this is formally simple to state, it would considerably complicate the calibration of such distributions from data, since it would require a two-dimensional analysis. We propose instead to merge the information of the $\Delta \log \mathcal{L}_{K\pi}$ and $\Delta \log \mathcal{L}_{p\pi}$ observables into a single observable Δ, defined below.

The exponentiation of the $\Delta \log \mathcal{L}_{h\pi}$ gives by definition the ratio of the likelihoods $\mathcal{L}_h$ and $\mathcal{L}_\pi$, i.e.:

$$\frac{\mathcal{L}_K}{\mathcal{L}_\pi} = e^{\Delta \log \mathcal{L}_{K\pi}} \tag{2}$$

and

$$\frac{\mathcal{L}_p}{\mathcal{L}_\pi} = e^{\Delta \log \mathcal{L}_{p\pi}}. \tag{3}$$

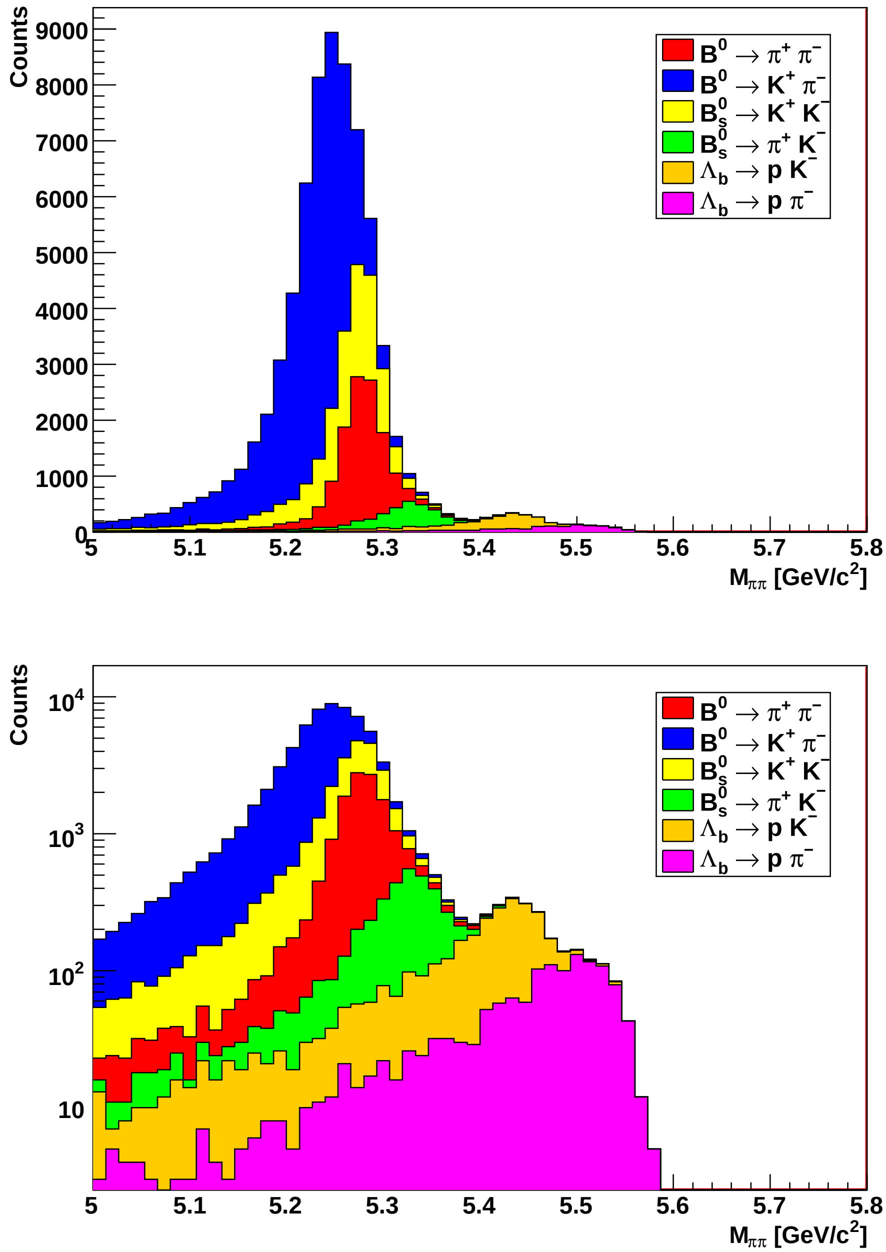

Figure 1: Invariant mass distribution, under the $\pi^+\pi^-$ hypothesis, for all the decay modes after the offline event selection: linear scale (top) and logarithmic scale (bottom). The histograms of the various channels are cumulatively added up to form the overall mass line shape. Background events are not included. See Ref. [6] for details on the event selection procedure.

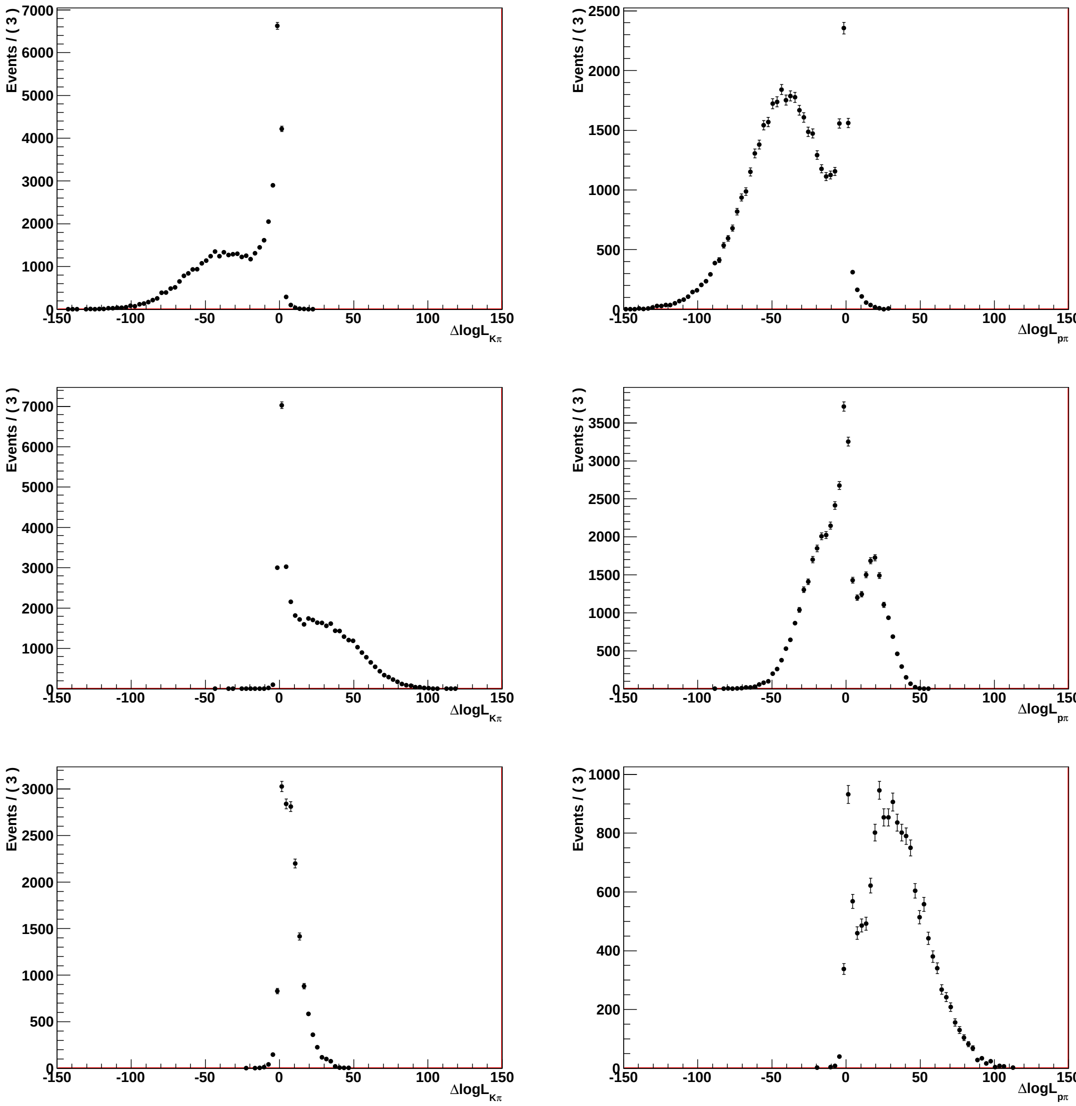

Figure 2: Distributions of $\Delta \log \mathcal{L}_{K\pi}$ (left) and $\Delta \log \mathcal{L}_{p\pi}$ (right) for pions (top), kaons (middle) and protons (bottom) from a sample of offline selected $H_b \to h^+ h'^-$ events including all the different decay modes.

One can then define the *confidence level* for a given particle to be a pion, kaon or proton, respectively, as:

$$P_\pi = \frac{\mathcal{L}_\pi}{\mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p} = \frac{1}{1 + e^{\Delta \log \mathcal{L}_{K\pi}} + e^{\Delta \log \mathcal{L}_{p\pi}}}, \tag{4}$$

$$P_K = \frac{\mathcal{L}_K}{\mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p} = \frac{e^{\Delta \log \mathcal{L}_{K\pi}}}{1 + e^{\Delta \log \mathcal{L}_{K\pi}} + e^{\Delta \log \mathcal{L}_{p\pi}}} \tag{5}$$

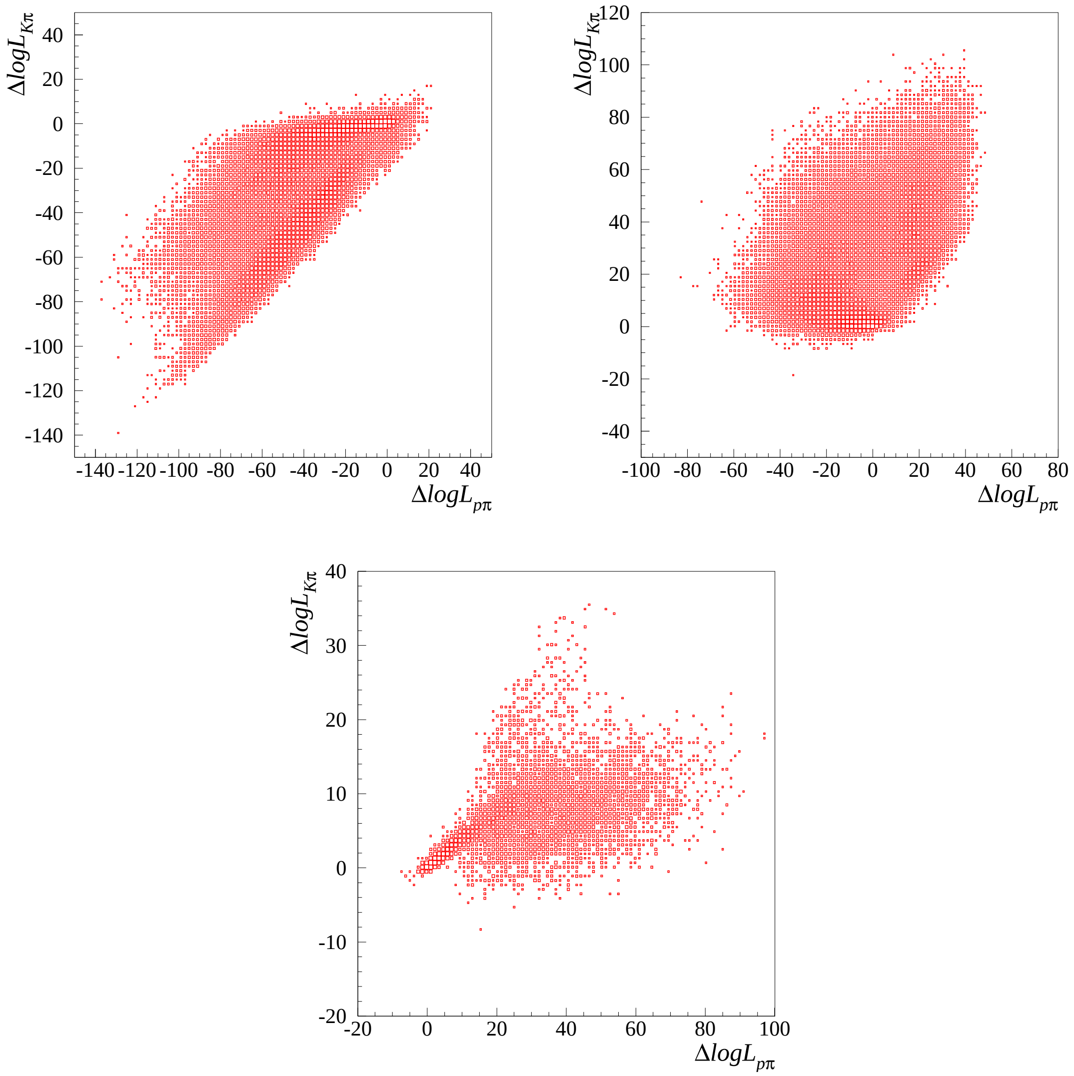

Figure 3: Distributions of $\Delta \log \mathcal{L}_{K\pi}$ versus $\Delta \log \mathcal{L}_{p\pi}$ for pions (top left), kaons (top right) and protons (bottom) from a sample of offline selected $H_b \to h^+h'^-$ events including all the different decay modes.

and

$$P_p = \frac{\mathcal{L}_p}{\mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p} = \frac{e^{\Delta \log \mathcal{L}_{p\pi}}}{1 + e^{\Delta \log \mathcal{L}_{K\pi}} + e^{\Delta \log \mathcal{L}_{p\pi}}}. \tag{6}$$

The three confidence levels are not independent, since by construction their sum is identically equal to 1. The distributions of P_π, P_K and P_p for pions, kaons and protons from offline selected $H_b \to h^+h'^-$ decays are shown in Fig. 4. As expected, the distributions are peaked at 1 when the corresponding mass hypothesis is correct, while they are peaked at 0 otherwise.

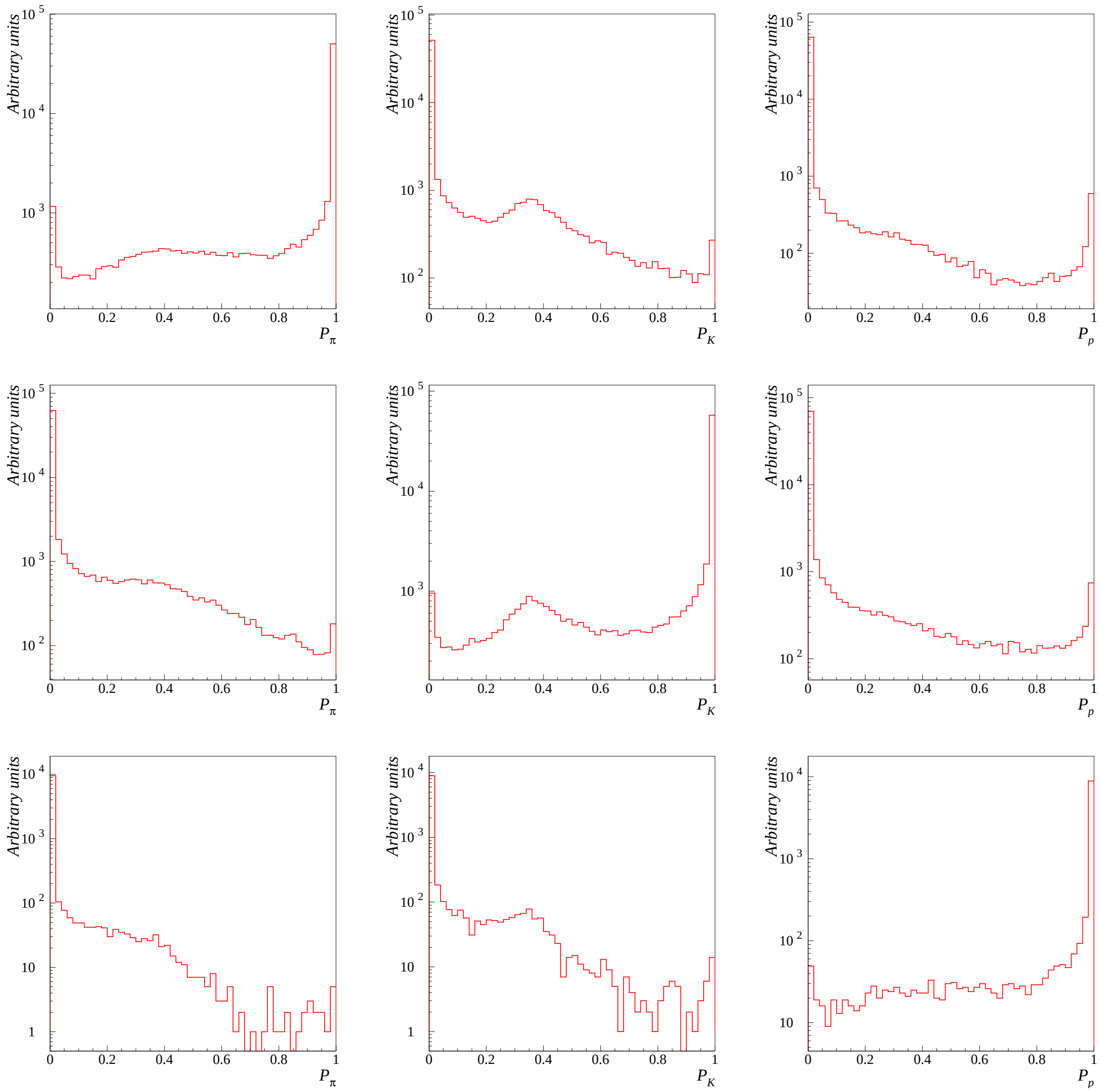

Figure 4: Distributions of P_π, P_K and P_p for pions (top), kaons (middle) and protons (bottom) from a sample of offline selected $H_b \to h^+h'^-$ events including all the different decay modes.

We can now define a generalized PID discriminating variable Δ as follows:

$$\Delta = 2 - P_\pi + P_p, \tag{7}$$

which by construction lies between 1 and 3. As it is apparent in Fig. 5, the distribution of Δ is peaked at 1 for true pions, at 2 for true kaons, and at 3 for true protons.

In order to characterize the Δ observable, Fig. 6 shows the efficiencies and misidentification probabilities achievable by cutting on Δ, as a function of the cut value C_Δ. The

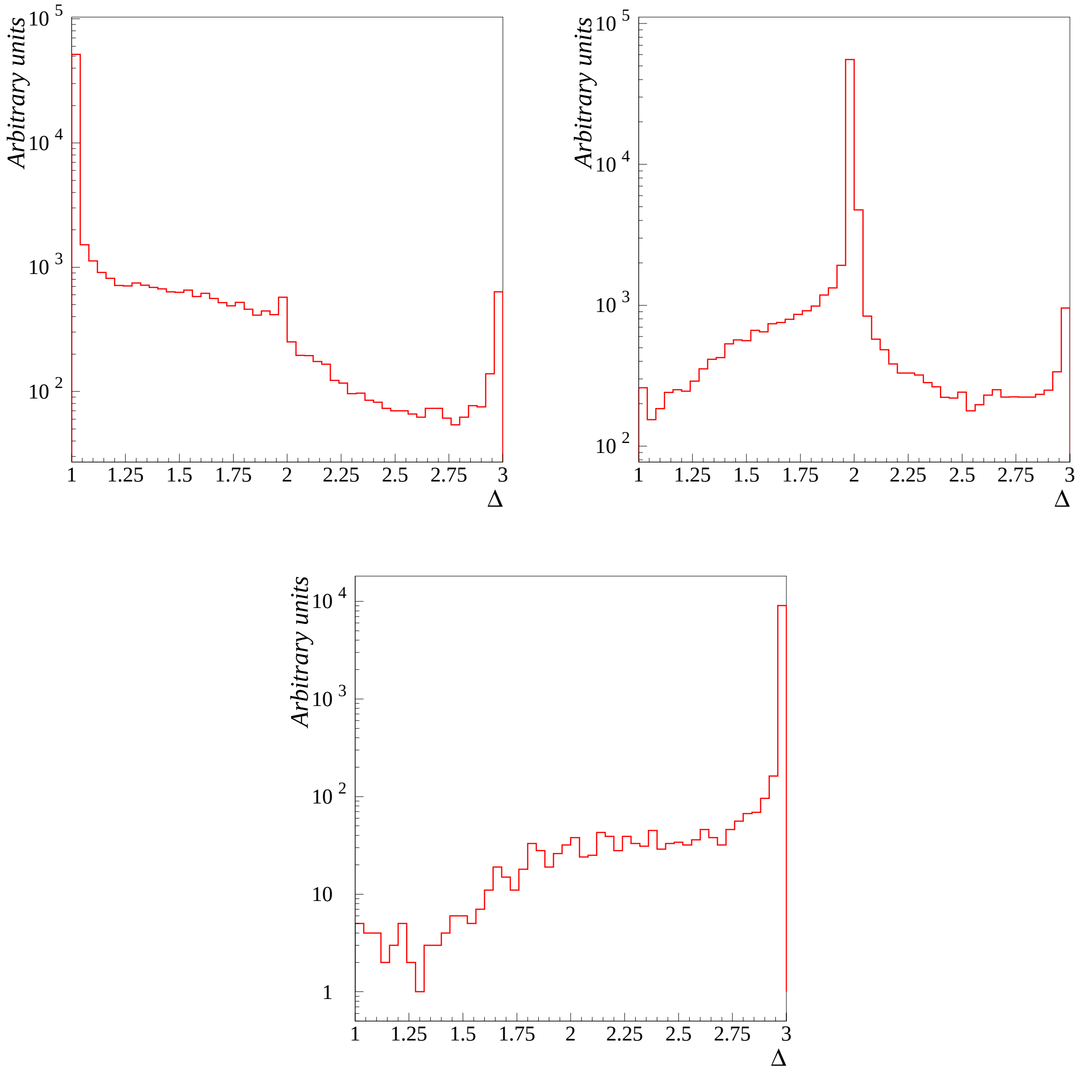

Figure 5: Distributions of Δ for pions (top left), kaons (top right) and protons (bottom) from a sample of offline selected $H_b \to h^+h'^-$ events including all the different decay modes.

cut for true pions, kaons or protons has been defined differently in order to have the same range in each of the three cases:

- when calculating the efficiency for the identification of pions, the event is accepted if $(\Delta - 1) \leq 2 \cdot C_\Delta$ AND $|\Delta - 2| > 0.05$ AND $(3 - \Delta) > 0.05$;
- when calculating the efficiency for the identification of kaons, the event is accepted if $(\Delta - 1) > 0.05$ AND $|\Delta - 2| \leq C_\Delta$ AND $(3 - \Delta) > 0.05$;

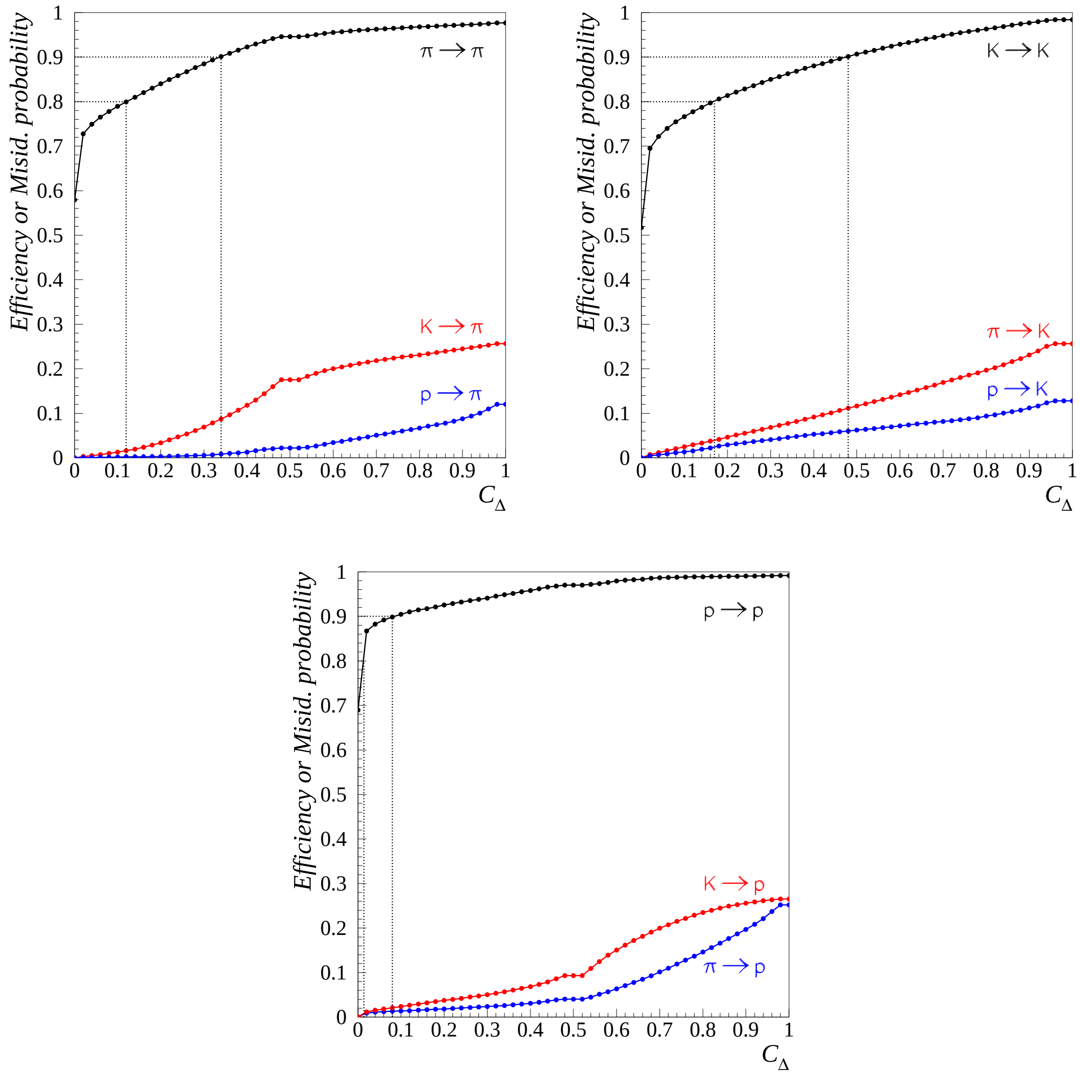

Figure 6: Efficiency and misidentification probabilities as functions of the cut value C_Δ for pions (top left), kaons (top right) and protons (bottom) from a sample of offline selected $H_b \to h^+h'^-$ events including all the different decay modes. As a reference, the dotted lines indicate the points where the efficiency reaches 80% and 90%.

- when calculating the efficiency for the identification of protons, the event is accepted if $(\Delta - 1) > 0.05$ AND $|\Delta - 2| > 0.05$ AND $(3 - \Delta) \leq 2 \cdot C_\Delta$.

With these definitions, C_Δ always lies in the range between 0 and 1, with $C_\Delta = 0$ corresponding to the minimum efficiency and $C_\Delta = 1$ to the maximum efficiency. We emphasize that the efficiencies are calculated for samples which span a significant range in momentum for the particles in question, from a few GeV/c up to more than 100 GeV/c.

3 Conclusions

We have derived a method to combine different correlated $\Delta \log \mathcal{L}$ variables coming from the PID system of the LHCb detector into a single variable, so-called Δ. Such a reduction of the dimensionality of the PID parameter space simplifies in a considerable way the employment of the PID information in maximum likelihood fits including all the two-body B decay modes, as discussed in Ref. [6].

References

[1] P. Nason *et al.*, arXiv:hep-ph/0003142.

[2] A. A. Alves *et al.* [LHCb Collaboration], JINST **3** (2008) S08005.

[3] [LHCb Collaboration], CERN-LHCC-2000-037.

[4] [LHCb Collaboration], CERN-LHCC-2000-036.

[5] [LHCb Collaboration], CERN-LHCC-2001-029, CERN-LHCC-2003-002, CERN-LHCC-2005-012.

[6] B. Adeva *et al.* [LHCb Collaboration], LHCb-PUB-2009-029 [arXiv:0912.4179].