Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Florian Reiß

born in Heidelberg

2017

Measurement of efficiency corrected yields of the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ - a first step towards a pentaquark search in the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$

This Master thesis has been carried out by Florian Reiß

at the

Physikalisches Institut

under the supervision of

Prof. Dr. Stephanie Hansmann-Menzemer

Messung der Effizienz-korrigierten Anzahl der Ereignisse des Zerfalls $\mathbf{\Lambda_b^0 \to \Lambda_c^+ D_s^-}$ - ein erster Schritt in Richtung Pentaquark-Suche im Zerfall $\mathbf{\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-}$:

Diese Arbeit stellt die Messung der Effizienz-korrigierten Anzahl von Λ_b^0 Baryonen im Zerfall $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ vor, bei dem das Λ_c^+ im Kanal $\Lambda_c^+ \to pK^-\pi^+$ und das D_s^- im Kanal $D_s^- \to K^-K^+\pi^-$ rekonstruiert wird. Benutzt werden Proton-Proton-Kollisions-Daten des LHCb Experiments, aufgenommen bei den Schwerpunktsenergien 7 TeV und 8 TeV in 2011 and 2012, entsprechend einer integrierten Luminosität von 3 fb^{-1}. Die korrigierte Anzahl ist:

$$N_{\Lambda_b^0} = (4.5009 \pm 0.1044(stat.) \pm 0.2052(syst.)) \cdot 10^6.$$

Dieser Wert wird in einer zukünftig Analyse als Referenz-Wert für die Messung des Verzweigungsverhältnisses des Zerfalls $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ dienen, der wiederum interessant im Kontext von Pentaquark-Suchen ist.
Um kombinatorischen Untergrund zu reduzieren werden multivariate Klassifizierungsalgorithmen benutzt, die auf unabhängigen Kalibrierungskanälen trainiert werden. Ihre Effizienzen werden auf Daten validiert, wobei die klare Signatur von $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ benutzt wird.

Measurement of efficiency corrected yields of the decay $\mathbf{\Lambda_b^0 \to \Lambda_c^+ D_s^-}$ - a first step towards a pentaquark search in the decay $\mathbf{\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-}$:

This thesis presents the measurement of the efficiency corrected Λ_b^0 yield measured in the channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$, where the Λ_c^+ is reconstructed in the channel $\Lambda_c^+ \to pK^-\pi^+$ and the D_s^- in the channel $D_s^- \to K^-K^+\pi^-$. The analysis is performed using proton-proton collision data from the LHCb experiment taken at center-of-mass-energies of 7 TeV and 8 TeV in 2011 and 2012, respectively, corresponding to an integrated luminosity of 3 fb^{-1}. The corrected yield is:

$$N_{\Lambda_b^0} = (4.5009 \pm 0.1044(stat.) \pm 0.2052(syst.)) \cdot 10^6.$$

This yield will serve in an upcoming analysis as reference value for the branching fraction measurement of the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$. This decay is relevant in the context of further pentaquark searches.
To reduce the combinatorial background multivariate classifiers trained on independent calibration channels are employed and their efficiencies are validated in a data-driven way using the clean signature of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

Contents

1 Introduction

When the quark model was established only hadrons consisting of three quarks (anti-quarks) or a quark-antiquark pair were known. Already in 1964 it became clear that in principle nothing forbids the existence of particles containing more quarks. Experimental experience shows that, if something is not explicitly forbidden, it is realized in nature. In 2015 this assumption seemed to be true: The LHCb collaboration published the observation of two puzzling resonances in the $J/\psi p$ system in an amplitude analysis of the decay $\Lambda_b^0 \to J/\psi K^- p$ [1]. They were found to be consistent with pentaquarks made up of four quarks and one anti-quark ($uudc\bar{c}$) [1]. This measurement, however, is not conclusive regarding the interpretation of these newly found resonances. The quantum number assignment could not be measured unambiguously. These newly found states were unexpected, sparking a wide theoretical interest. If the new states are indeed consisting of five quarks, their binding mechanism has to be understood. For this purpose additional measurements are necessary. One approach is to look for the same states in different decay channels. An example for such a channel is the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$, where the a pentaquark with the same quark content as the recently discovered pentaquark could be found in the $\Lambda_c^+ \overline{D}^0$ system. This Λ_b^0 decay itself has not yet been observed before and thus before an amplitude analysis is performed first the decay needs to be established and its branching ratio (BR) will be measured. A BR measurement of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ relative to the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ is foreseen:

$$\mathrm{BR} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}. \tag{1.1}$$

This thesis deals with the reference channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. This channel is chosen because it is well-established with a relative large branching fraction, $\mathrm{BR}(\Lambda_b^0 \to \Lambda_c^+ D_s^-) = 1.10 \pm 0.10\%$ [2]. Additionally, both reference channel and the signal channel $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ are reconstructed in final states with the same six hadrons, namely one proton, three kaons and two pions, so a majority of systematic uncertainties are expected to cancel. Identifying these hadrons is one of the main challenges of determining the relative branching fraction. Having this many particles in the final state leads to large backgrounds from random particle combinations. In addition to particle identification variables defined on a per-track basis for the final state particles, multivariate classifiers trained on independent control channels are employed. The aim of these classifiers is to provide a variable to identify particles containing charm quarks. For both cases, the efficiencies of selections using these

[1] Charge-conjugation is implied throughout this document.

variables have to be evaluated on independent control channels. If the efficiencies are valid on other channels is tested using the clean signature of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

This chapter gives a short introduction in the theoretical foundation of pentaquark searches and the LHCb experiment. In Chapter 2 the analysis strategy is detailed. Chapter 3 describes the event selection of the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. Chapter 4 is devoted to the Λ_b^0 yield extraction and Chapter 5 introduces the selection efficiencies. An important component of the analysis is the usage of multivariate classifiers trained on independent control channels to select particles containing charm quarks (open-charm particles). These classifiers are described in more detail in Chapter 6, where their their performance is validated in a data-driven way using the clean signature of the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay. The thesis is completed by an estimation of the systematic uncertainty in Chapter 7 and the conclusion in Chapter 8.

1.1 The Standard Model

The Standard Model (SM) of particle physics is a quantum field theory describing the fundamental particles that constitute matter and their interactions. The interactions are the strong, weak and electro-magnetic interaction. The fundamental spin $\frac{1}{2}$ particles (fermions) are divided into three generations of quarks and leptons, where a quark generation is divided in up- and down-type quarks and a lepton generation is divided in a neutral neutrino and a charged particle (see Fig. 1.1). The fermions are characterized by the charges they carry, which determines which interactions they take part in. For each particle there is an anti-particle carrying the opposite charges. Interactions in the SM are mediated by gauge bosons:

The photon mediates the electro-magnetic interaction, the $W^{+/-}$ and Z^0 bosons the weak interaction, gluons the strong interaction. The SM is completed by the discovery of the Higgs boson in 2012 [3] which gives mass to fundamental particles.

Neutrinos only interact by the weak interaction, charged leptons additionally by the electromagnetic interaction. The key difference of quarks from leptons is that they also carry color, the fundamental charge of the strong interaction. The field theory describing strong interactions is quantum chromodynamics (QCD). It is governed by the $SU(3)_c$ color symmetry, which has 8 generators corresponding to 8 gluons. Gluons themselves carry color charge (in fact they carry color and a different anti-color) leading to self-interaction between them, which is different from quantum electrodynamics, where the photon is not self-interacting. This can lead to confinement, although an analytic formalism is not known up to date. Confinement is the observation that there are no free color-charged systems, only color neutral bound states. Quarks carry color, therefore they have never been observed as free particles, only bound in hadrons. This is caused by the potential energy between two quarks rising linearly with distance at larger distances. It can be imagined as 'flux-tube' formed by self-interacting gluons spanning between the two quarks. The field lines are 'pulled together' by the self-interacting gluons forming a field with constant energy density giving rise to an potential energy rising linearly with dis-

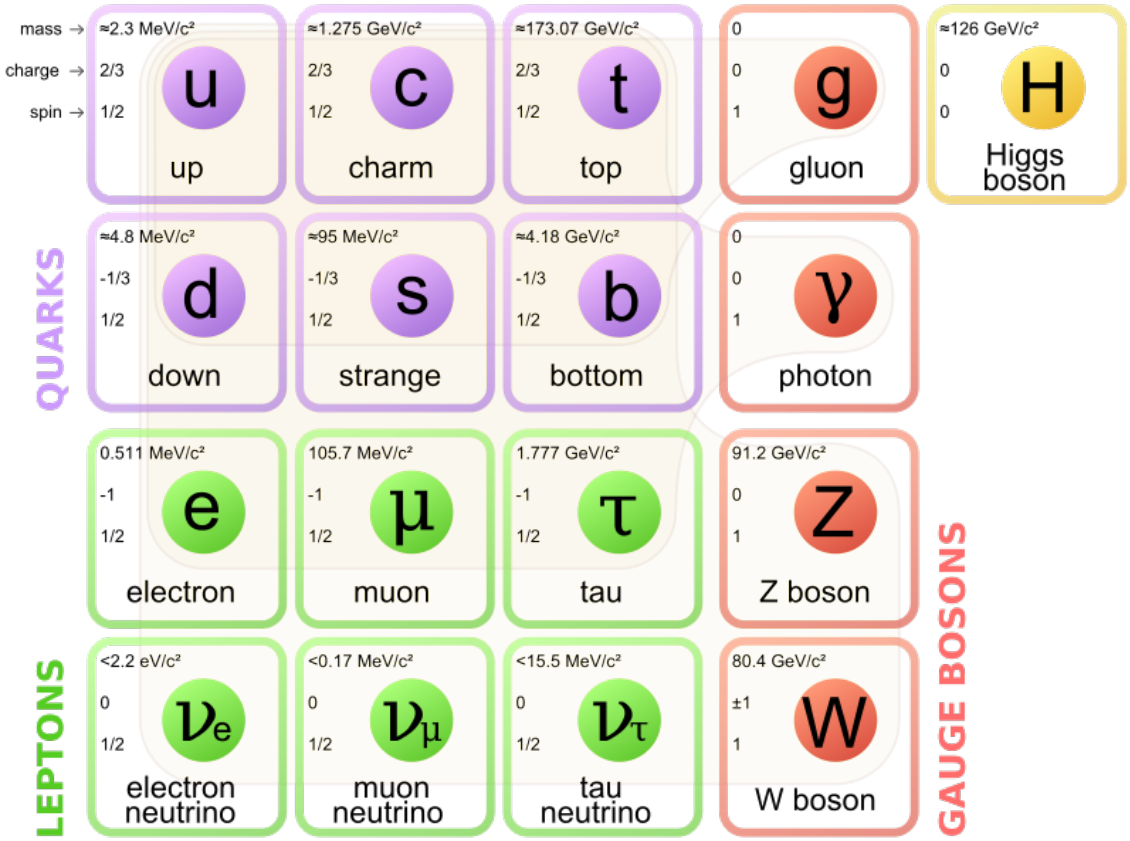


Figure 1.1: The Standard Model of particle physics. Taken from [4].

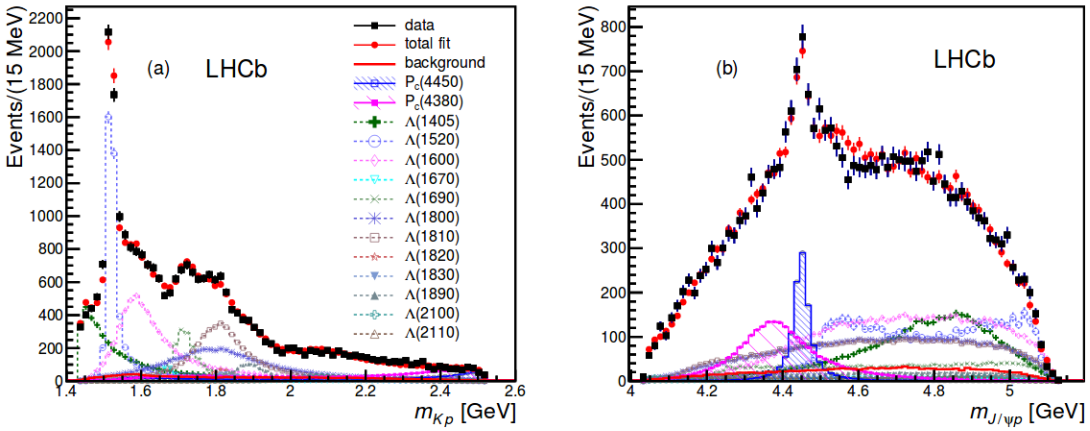


Figure 1.2: Projections of the Kp and $J/\psi p$ mass spectra of the decay $\Lambda_b^0 \to J/\psi K^- p$. The blue and purple shaded histograms show the newly found pentaquark resonances [1].

tance. At some point the energy is large enough that this 'flux-tube' breaks apart creating a quark anti-quark pair from vacuum, which then form colorless bound states again. Colorless states can be formed by a quark anti-quark pair or three quarks (or anti-quarks) called meson and baryon, respectively, which constitute the matter found in nature. Gell-Mann and Zweig already remarked in 1964 [5][6] that in principle colorless systems containing four or five quarks are allowed in QCD. In recent years candidates for such states have been observed and been given the name tetraquarks and pentaquarks [7].

1.2 Pentaquark Searches

The search for so-called 'exotic' hadronic states consisting of more than three quarks has a long experimental — and sometimes controversial — history [7] [8]. In 2015 the LHCb collaboration published the discovery of two resonances consistent with pentaquarks [1] with the quark content $uudc\bar{c}$ in the $J/\psi p$ system in the channel $\Lambda_b^0 \to J/\psi K^- p$ (see Fig. 1.3) with the masses $4380 \pm 8 \pm 29$ MeV ($P_c(4380)^+$) and $4449.8 \pm 1.7 \pm 2.5$ MeV ($P_c(4450)^+$). Their widths were found to be $205 \pm 18 \pm 86$ MeV and $39 \pm 5 \pm 19$ MeV, respectively. An amplitude analysis of this channel was done and found these resonances with a significance of more than 9 standard deviations. Their spin and parity J^P are not unambiguously determined by this measurement, but the preferred assignment is for them to have opposite parity P and one state having spin $J = \frac{3}{2}$, the other $J = \frac{5}{2}$. Fig. 1.2 shows the projections of the $K^- p$ and $J/\psi p$ mass spectra including the newly found pentaquark resonances.

Several theories describe such resonant structures. For a comprehensive summary, see [9]. If one describes them as pentaquarks, their inner structure has to be understood. In the molecule model pentaquarks consist of a baryon and a meson loosely bound by meson exchange. In the compact model all five quarks share the same

color 'bag' and are tightly bound by gluon exchange. Fig. 1.4 illustrates these two models. Because of confinement the pentaquark has to be color-neutral. This is achieved by three of the quarks having the color-charge red, green and blue and the anti-quark having the anti-color of the remaining quark.

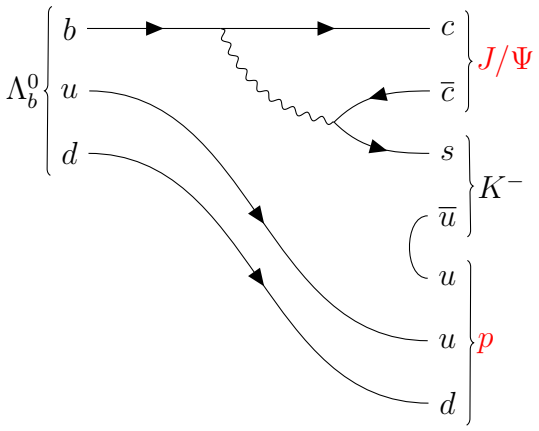


Figure 1.3: Pentaquark discovery channel. The pentaquark resonances were found in the $J/\psi p$ system.

To get a better understanding of the nature of these states, additional measurements have to be made [11]. One possibility is to search for these particles in different decays channels. The decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ is an interesting choice to look for pentaquark resonances, as it has the same quark content in its final state as the pentaquark discovery channel (see Fig. 1.5). The main difference is that the charm anti-charm pair is contained in two separate hadrons compared to the J/ψ in the discovery channel. This means that the same pentaquark resonances could possibly be observed in the $\Lambda_c^+ \overline{D}^0$ system. An analysis of this decay using data recorded with the LHCb experiment is in preparation, which aims in a first step to establish the channel and measure its branching fraction relative to the reference channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and in a second step to do a full amplitude analysis to search for pentaquark resonances in the $\Lambda_c^+ \overline{D}^0$ system.

Closely related to this channel is the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{0*} K^-$. Depending on the model the expected branching fraction is large for pentaquarks decaying to $\Lambda_c^+ \overline{D}^{0*}$. For example, if the $P_c(4380)^+$ is a molecule formed from $\overline{D}^0 \Sigma_c^+ (\frac{3}{2}^-)$, its decay width to $\Lambda_c^+ \overline{D}^{0*}$ is about 35-times the decay width of the decay to $J/\psi p$, where it was discovered [11]. Analyzing this channel in addition to the channel $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ would further help to determine the properties and advance the theoretical understanding of the pentaquarks.

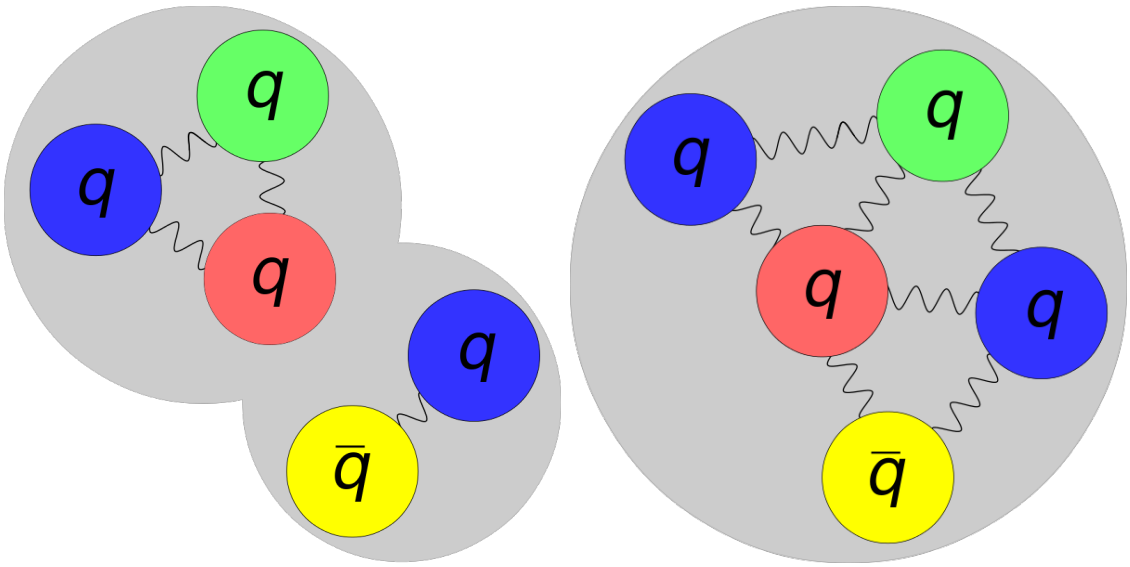


Figure 1.4: Illustration of the pentaquark as molecule (left) and compact particle (right). The pentaquark needs to be color-neutral. Anti-blue is shown as yellow [10].

1.3 The LHCb Experiment

The Large Hadron Collider (LHC) located at CERN provides a good environment to search for pentaquarks in exclusive decay chains, as it is the only accelerator reaching a large enough center-of-mass energy to provide Λ_b^0 baryons with sufficient high statistics to study their decays.

At the LHC protons and heavy ions are accelerated and brought to collision at different points. Various experiments are set-up to detect the particles produced in such collisions. The LHCb experiment [12] is one of these experiments. It is a dedicated flavor physics experiment optimized for the detection of hadrons containing charm and bottom quarks.

When bottom (charm) quarks are produced in the proton collision, the $b\bar{b}$ ($c\bar{c}$) pair is boosted in either forward or backward direction as the interacting partons have in general different proton momentum fractions. The LHCb detector is laid out to cover the cone where the hadrons from the $b\bar{b}$ ($c\bar{c}$) pair are expected to fly. Therefore it is a forward spectrometer. The coordinate system of the detector is definied by the beam axis being the z-axis, the vertical direction the y-axis and the horizontal direction the x-axis. It consists of several sub-systems described in the following (also see Fig. 1.6):

1.3.1 Charged Track Reconstruction

Closest to the interaction point is the **Vertex Locator** (Velo), a silicon-strip tracker, to reconstruct tracks of charged particles. Before the magnet further tracking is

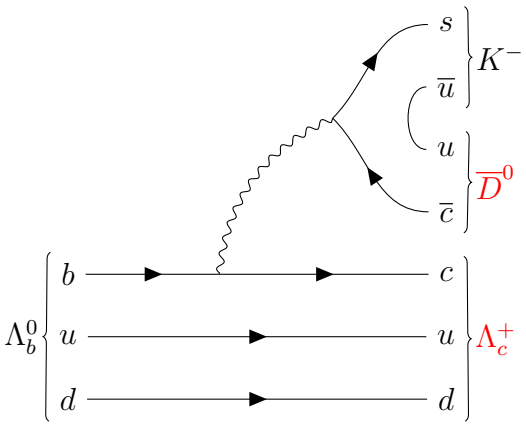


Figure 1.5: Feynman diagram for $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$. Pentaquark resonances could be found in the $\Lambda_c^+ \overline{D}^0$ system.

provided by the **Tracker Turicensis** (TT), also a silicon-strip detector. After the magnet providing a magnetic field with bending power of about 4 Tm three tracking stations, divided into **Inner Tracker** (IT) and **Outer Tracker** (OT), are located. The IT covers the high occupancy inner region closest to the beampipe and also uses silicon microstrips, while in the OT straw-tubes are utilized to cover the outer region. The TT and each tracking station consist of four layers, where the inner two layers are slightly tilted by $\pm 5°$ as indicated in Fig. 1.7 for the TT. The trajectory of charged particles is bent by the magnetic field which allows the measurement of their momenta. The tracks are built up from measurements in the different tracking systems and propagated through the magnetic field. Matching track segments before and after the magnet are then combined to so-called long tracks (see Fig. 1.8).

Hadrons containing charm or bottom quarks decay via the weak interaction and have large enough lifetimes and a boost in the experiment to travel for a measurable distance before they decay. The Velo provides a good resolution of decay vertices of 13 μm transverse to and 71 μm along the beam axis [13]. Long-lived particles produced in the collision can be identified by the displacement of their decay vertex from the primary vertex (PV), where the proton-proton interaction takes place. B hadrons in the LHCb experiment travel on average ~ 1 cm before decaying leading to a displaced vertex, which is used to select such decays. This is exploited in this thesis, as the Λ_b^0 decays at a vertex separated from the primary vertex and its daughter particles containing charm quarks also decay at separated tertiary vertices.

LHCb has many subsystems dedicated to identifying and differentiating between particle species (pions, kaon, protons, electrons and muons), as described in the following section.

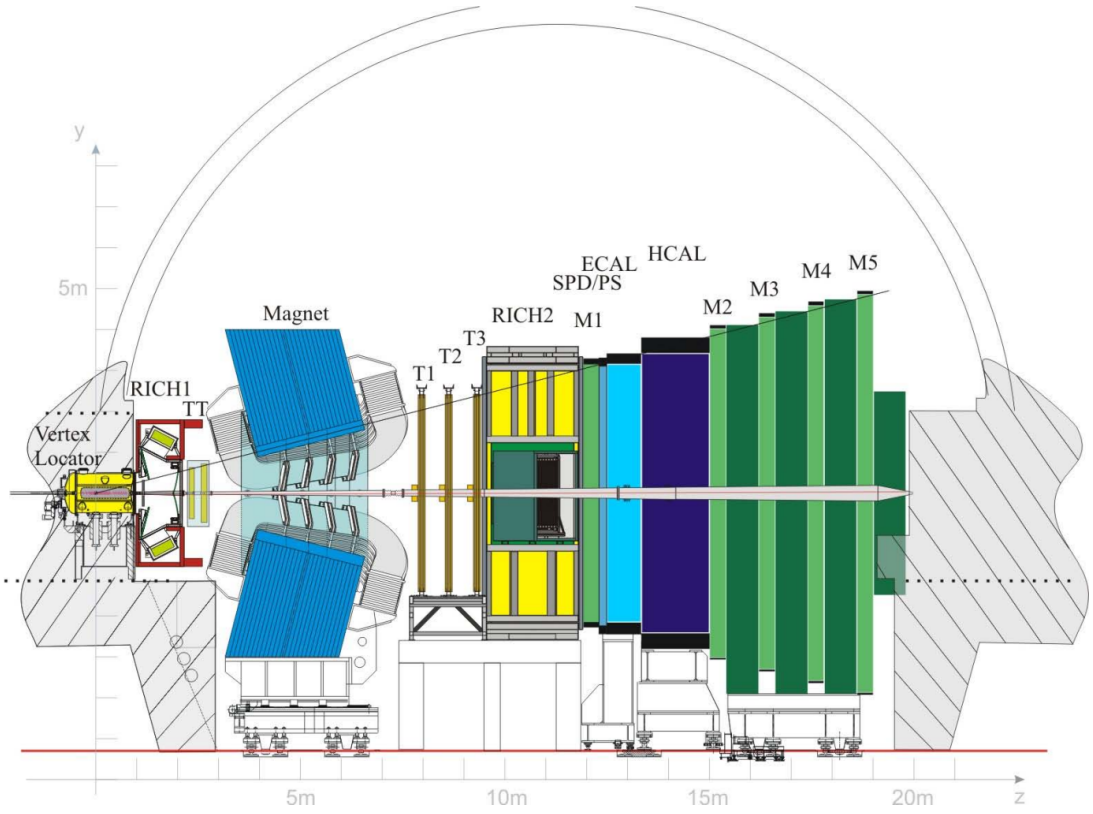


Figure 1.6: The LHCb detector. On the left is the Vertex Locator around the collision region. To the right follows the first RICH detector and the Tracker Turicensis. After the magnet the tracking stations are placed. The final part consists of the second RICH detector, electronic and hadronic calorimeters and muon stations [14].

1.3.2 Particle Identification (PID)

Particle identification (PID) is provided by two **RICH** systems (ring-imaging Cherenkov light detectors), an electronic calorimeter **ECAL**, a hadronic calorimeter **HCAL** and muon stations **M1-M5**.

The RICH systems exploit the Cherenkov effect. When a charged particle traverses a material with a velocity v greater than the speed of light in that medium, it radiates light at an angle θ depending on its velocity and the refraction index n of the medium:

$$\cos(\theta) = \frac{1}{n\beta}, \tag{1.2}$$

where $\beta = \frac{v}{c}$ is the velocity of the particle relative to the vacuum speed of light, c. Cherenkov light is only produced if the momentum p of the particle exceeds a certain threshold depending on the particle mass m:

$$m < p\sqrt{n^2 - 1}, \tag{1.3}$$

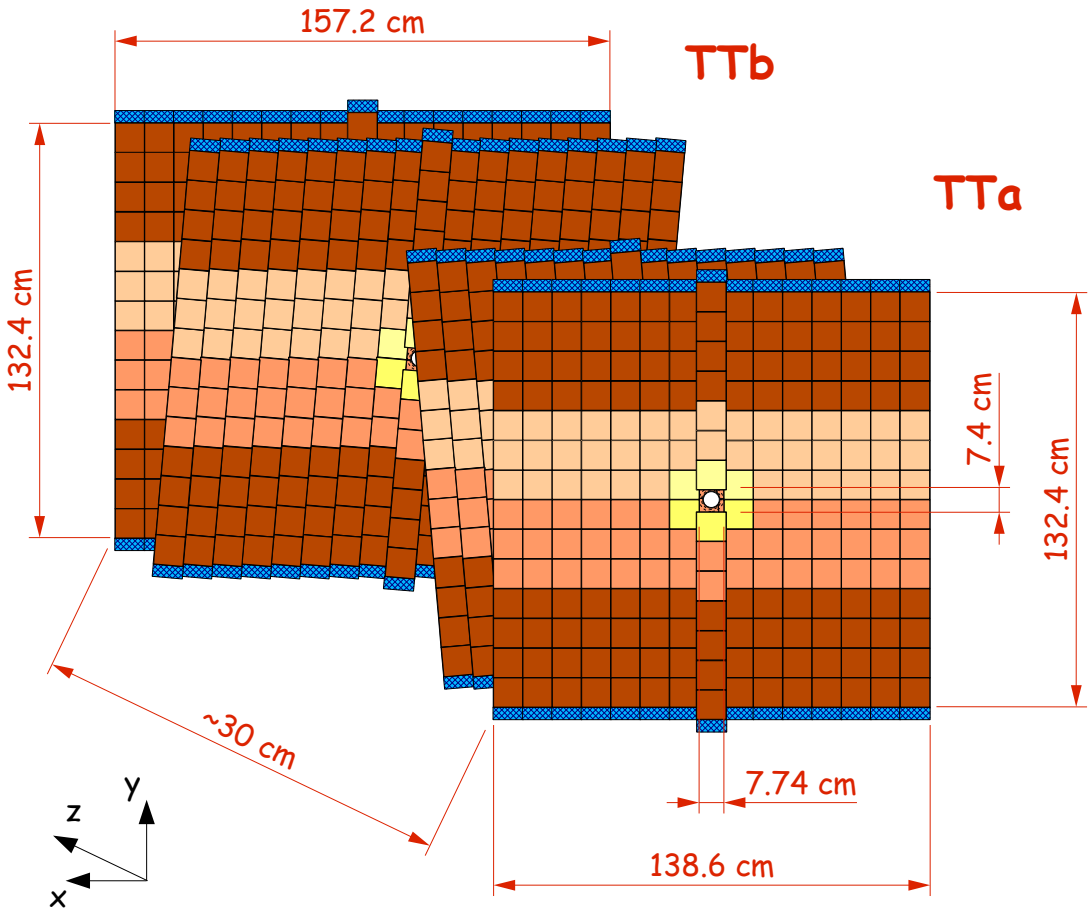


Figure 1.7: Layout of the four TT layers. The inner layers are rotated with respect to the other ones [15].

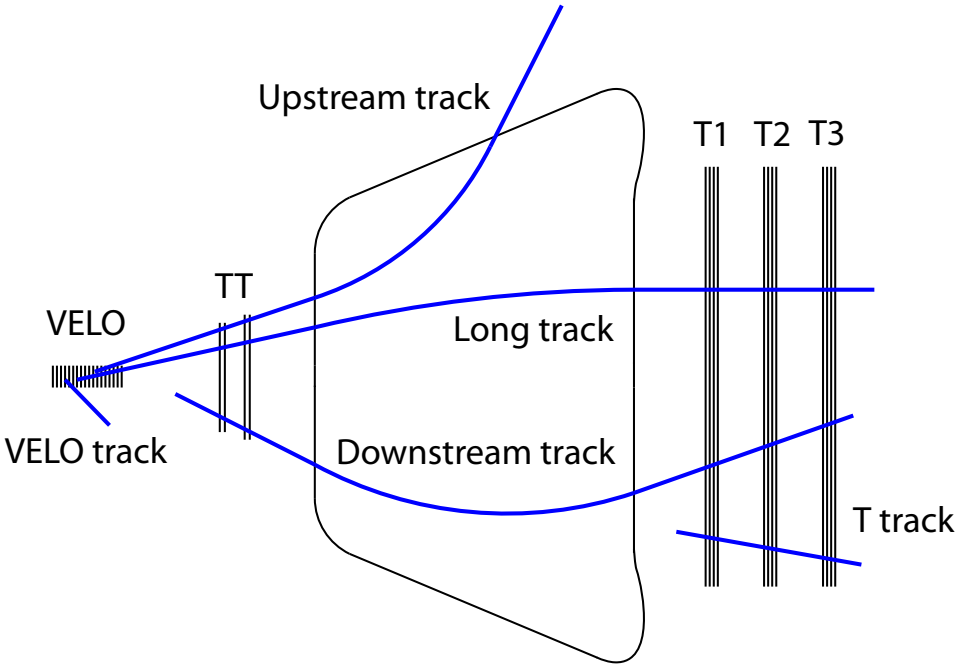


Figure 1.8: Different track types at the LHCb detector [16].

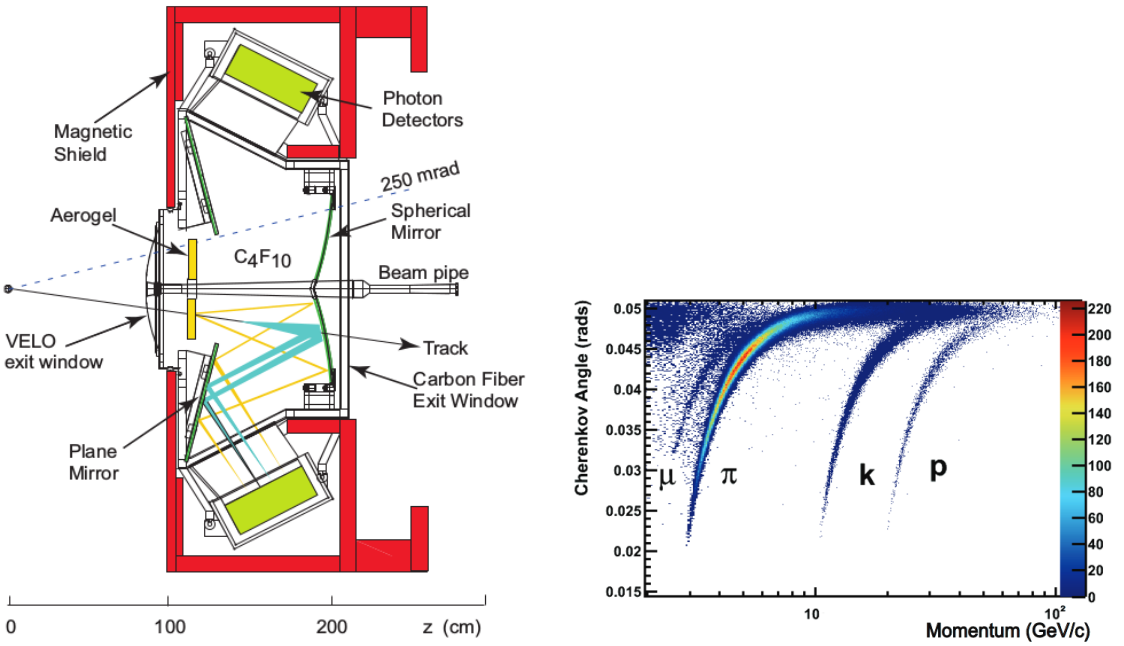


Figure 1.9: Left: Schematic view of the RICH1 system [12]. Right: Cherenkov angle in the RICH system as function of the particle momentum for muons, electrons, kaons and protons [17].

where the speed of light c is set to $c = 1$. Measuring the angle of the Cherenkov light provides a measurement of the velocity of the particle. Combined with the momentum measurement from the tracking system a mass hypothesis can be assigned to the particle. This is done in the following way: For each track hitting the RICH systems in a single event the hypothesis that the corresponding particle is a pion is assumed and an overall event likelihood $\mathcal{L}$ is calculated. This is done because pions are the most abundant charged particles produced in high energy collisions. Then for each particle the hypothesis is changed to different particles (proton, kaon,electron and muon) and the change in the likelihood calculated. The final mass assignment is then the hypothesis, which maximizes the likelihood. For each particle DLL (delta log likelihood) variables are defined, describing the change in likelihood compared to the pion hypothesis:

$$DLLX = \Delta \log \mathcal{L}(X - \pi) = \log \mathcal{L}(X) / \log \mathcal{L}(\pi), \tag{1.4}$$

where X is the electron, muon, kaon or proton hypothesis. It should be noted that this likelihood is maximized for all tracks in an event at once, which could introduce correlations between the DLL variables of different particles.

At LHCb two RICH systems are used. RICH1 is placed after the Velo, RICH2 after the tracking stations. They cover different momentum ranges: RICH1 covers the low momentum range of $\sim 1 - 60$ GeV, RICH2 the high momentum range from ~ 15 GeV beyond 100 GeV. Fig. 1.9 shows a schematic view of the RICH1 system.

It uses an aerogel and C_4F_{10} gas as radiators to produce Cherenkov light. The light is then reflected and collected by mirrors and detected by Hybrid Photon Detectors.

The performance of the RICH system depends on the momentum of the particle as shown in Fig. 1.9. The difference between the Cherenkov angle of pions, kaon and protons is small for large particle momenta resulting in an inefficient particle identification. The quality of identification also depends on the angle of incidence of the particle and on the track multiplicity in an event. If many particles hit the RICH at once the Cherenkov cones of different particles can overlap making it more difficult to identify a single particle.

The electronic calorimeter ECAL measures the energy of photons and electrons using bremsstrahlung and electron-positron pair production. The hadronic calorimeter determines the energy of hadrons, which produce hadronic showers in it. Muons with a energy typical for LHCb are minimally ionizing and traverse the whole detector without losing much energy. Therefore the muon stations are placed at the end to detect them.

Similarly to the RICH system, for other PID subsystem also likelihood variables are defined, which are all added to a combined likelihood (PIDX), where X is the electron, muon, kaon or proton hypothesis. An alternative PID variable is provided by an artificial neural network, which combines information from the different PID systems to a single variable ($\mathrm{Prob_{NN}(X)}$), where X is a pion, electron, muon, kaon or proton. The advantage of using a neural network is to exploit the correlations between different PID variables.

1.3.3 Data Flow

At LHC, protons are collided at a rate of 30 MHz. Collecting all the information from the detector would require a large bandwidth of about 3 TB/s and much disk space. In order to reduce the data output multiple mechanism are employed at LHCb. First is the hardware trigger (L0 trigger) reconstructing the hadron, electron and photon clusters in the calorimeters with the highest transverse energy E_T and the two muons in the muon system with the highest transverse momentum p_T. The event is only kept if at least on of those values surpass a certain threshold. This reduces the rate to 1 MHz. Events selected by the L0 trigger are then reconstructed by the high-level trigger (HLT), a software trigger, using the full event data reducing the rate to a manageable level of 5 kHz corresponding to a bandwidth of 0.5 GB/s, which is written to disk.

Events are then reconstructed offline by the *BRUNEL* software [18], which transforms detector information to objects used in the analysis, for example combining measurements in the tracking stations to tracks. The reconstructed data is then analyzed with the *DaVinci* software [19] forming physical objects from the tracks. To further reduce the data used in an analysis to a more manageable level a collaboration-wide preselection stream called *stripping* is applied selecting events likely to contain particular decays. The full reconstruction takes a large amount of

time and is only redone if there are major changes in the calibration and configuration.

In addition to the data produced by the detector, events from Monte-Carlo (MC) simulations are used, especially to determine various efficiencies, like the efficiency of the reconstruction algorithm or the trigger. Particles are generated by the *Pythia* [20] software, decays are simulated by *EvtGen* [21] and the decay products are propagated through a simulation of the whole detector using *Geant4* [22]. The simulated detector responses are then transformed by the *Boole* application [23] to closely resemble real data. The same software used for reconstructing real events is then used to reconstruct simulated data. With MC simulation the simulated (true) values of the properties of a particle can be compared with the values after it is reconstructed by matching a reconstructed track with its corresponding generated particle (truth-matching).

2 Analysis Strategy

This analysis uses data taken in 2011 and 2012 with the LHCb detector at a center of mass energy $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of 1 fb^{-1} and 2 fb^{-1}, respectively. The channel studied is $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ (see Fig. 2.1) where the Λ_c^+ and D_s^- are reconstructed in the decays $\Lambda_c^+ \to pK^-\pi^+$ and $D_s^- \to K^+K^-\pi^-$.

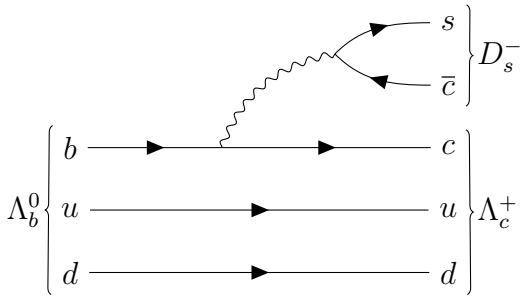


Figure 2.1: Lowest level Feynman diagram for $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

This channel will be used in the context of an upcoming pentaquark search as reference channel for the branching fraction ratio (or short branching ratio BR) measurement of the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$:

$$\mathrm{BR} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = \frac{N(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-)}{N(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} \cdot \frac{\epsilon(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}{\epsilon(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-)} \cdot \frac{\mathcal{B}(D_s^- \to K^+K^-\pi^-)}{\mathcal{B}(\overline{D}^0 \to K^+\pi^-)}, \tag{2.1}$$

where the Λ_c^+ will be reconstructed in the decay mode $\Lambda_c^+ \to pK^-\pi^+$ and the $\overline{D}^0$ in the mode $\overline{D}^0 \to K^+\pi^-$.

The yields N and selection efficiencies ϵ of both decay channels need to be known as well as the branching fractions of the decays $D_s^- \to K^+K^-\pi^-$ and $\overline{D}^0 \to K^+\pi^-$, which have been previously measured [24] [25]. The advantages of measuring a Λ_b^0 branching fraction relative to another Λ_b^0 decay is that the b-quark production and Λ_b^0 hadronization fractions, which are only known with large uncertainties, are canceled. Also, both channels are reconstructed in the same final state, so that systematic uncertainties like the tracking efficiencies are expected to cancel.

The aim of this thesis is to provide the efficiency corrected yield $\frac{N(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}{\epsilon(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}$ for the branching fraction measurement. Both signal and reference channel contain two

'open-charm' hadrons, Λ_c^+ and $\overline{D}^0$ in the signal and Λ_c^+ and D_s^- in the reference case. The name 'open-charm' refers to the fact that those hadrons have charm quantum number ± 1. Both channels have the same final state consisting of a proton, three kaons and two pions. Selecting the open-charm hadrons in final states with many hadrons is one of the main challenges for the branching fraction measurement, as the high number of final state particles lead to a large combinatorial background contribution. To reduce this background, multi-variate classifiers trained on independent calibration samples are employed using mostly PID information of the open-charm decay products. The selection efficiencies of these classifiers are measured on the calibration samples giving rise to the question whether they are applicable on other channels. A method to check this is developed exploiting the clean signature of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

The analysis proceeds as follows:

- Λ_b^0 candidates are selected in the decay $\Lambda_b^0 \to \Lambda_c^+(\to pK^-\pi^+)D_s^-(\to K^-K^+\pi^-)$ using multivariate techniques
- The fit to the $\Lambda_c^+ D_s^-$ mass distribution is discussed
- A three-dimensional fit to the Λ_b^0, Λ_c^+ and D_s^- mass spectra is done to discriminate the signal from charmless backgrounds without intermediate Λ_c^+ or D_s^-
- The Λ_b^0 yield is extracted from the fit and corrected for the selection efficiencies
- The efficiency of the multivariate classifiers is validated
- Systematic uncertainties are assigned

2.1 Selection Variables

Important variables used in the following to characterize beauty and charm hadron decays are introduced here.

$\boldsymbol{p_T}$ is the momentum component transverse to the beam axis: $p_T = \sqrt{p_x^2 + p_y^2}$.

$\boldsymbol{\eta}$ is the pseudorapidity: $\eta = -\ln(\tan(\frac{\theta}{2}))$, where θ is the angle between beam axis and the momentum vector of the particle in the laboratory frame. For highly relativistic particles the pseudorapidity is approximately equal to the rapidity $y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$.

$\boldsymbol{\chi^2_{vtx}/n_{dof}}$ is the χ^2 of the vertex reconstruction divided by the number of degrees of freedom. It is close to 1 for well reconstructed vertices and is large for badly reconstructed vertices. The number of degrees of freedom when reconstructing a vertex from N tracks is given by $n_{dof} = 2N - 3$.

Flight Distance (FD) is the distance of the decay vertex of a particle from its production vertex. Its significance is described by $\chi^2_{FD} = \frac{FD}{\sigma_{FD}}$.

DIRA (DIRection Angle) is the cosine of the angle between the line connecting the origin and decay vertex of a particle and its momentum vector (see Fig. 2.2).

Impact Parameter (IP) is the minimum distance between a track and the primary vertex, as illustrated in Fig. 2.2. $\chi^2_{\mathbf{IP}}$ is the change of the χ^2 of the reconstructed vertex when the track is added.

$\chi^2_{\mathbf{Match}}$, $\chi^2_{\mathbf{T}}/\mathbf{n_{dof}}$, $\chi^2_{\mathbf{Velo}}/\mathbf{n_{dof}}$ describe the quality of the track reconstruction. The variables χ^2_T/n_{dof} and χ^2_{Velo}/n_{dof} are the χ^2 divided by the number of degrees of freedom for the track reconstruction in the Velo and the tracking stations, respectively. The overall χ^2 of matching different track segments is described by χ^2_{Match}.

$\mathbf{Prob_{NN}(Ghost)}$ is the probability of the track being a ghost. Ghost tracks are reconstructed tracks that do not belong to a real particle.

$\mathbf{Q_{Velo}}$ is the charge deposited in the Velo by the particle.

$\mathbf{\Delta\mathcal{L}_{RICH}(x)}$ is an output from the RICH systems. It describes the difference of the logarithmic likelihood of the hypothesis that the particle is of species x (muon, electron, kaon, proton) compared to the pion hypothesis: $log(\mathcal{L}_x) - log(\mathcal{L}_\pi)$.

$\mathbf{b_{thresh}(x)}$ is a boolean variable describing if the particle has large enough momentum to produce Cherenkov radiation if the mass hypothesis x is assumed.

$\mathbf{b_{RICH,1}}$, $\mathbf{b_{RICH,2}}$, $\mathbf{b_{RICH}(aerogel)}$ are boolean variables describing if the particle generated a signal in the RICH systems.

$\mathbf{\mathcal{L}_{Muon}(x)}$ is the likelihood of the particle traversing the muon system being a muon or background.

$\mathbf{N_{shared}(\mu)}$ is the number of measurements in the muon station the particle track shares with other tracks.

$\mathbf{b_{isMuon}}$ is a boolean response from the muon system describing if the particle is a muon.

$\mathbf{Prob_{NN}(x)}$ is the output of a neural network using various PID variables as input. It describes the probability of a particle belonging to the species x (electron, muon, pion, kaon, proton).

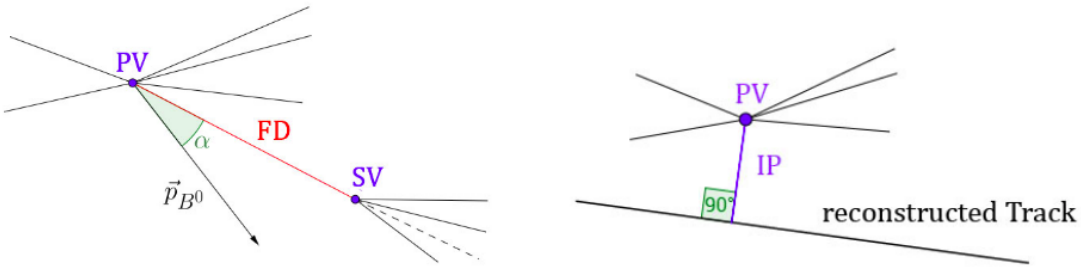


Figure 2.2: Schematic view of the flight distance and direction angle (left) and the impact parameter (right). Taken from [26].

3 Data Selection

3.1 Preselection

The decay topology of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ (see Fig. 3.1) has distinct features, which are exploited to select this decay. Because of its high lifetime and its boost within the LHCb detector the Λ_b^0 flies a short distance after it is produced in the primary vertex (PV) and decays at a separated secondary vertex (SV). Its daughters, the Λ_c^+ and D_s^-, also travel some distance before decaying, forming tertiary vertices (TV) separated from the SV.

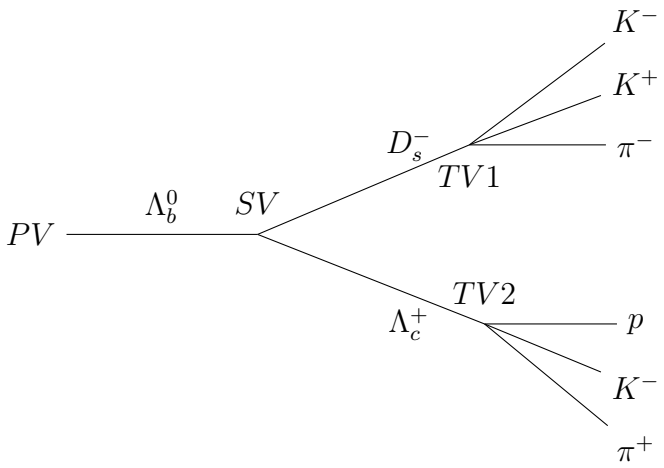


Figure 3.1: Decay topology of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

In the stripping the selections listed in Table 3.1 are applied. The particles forming the Λ_c^+ and D_s^- candidates are each required to have the same vertex ($\chi^2_{vtx}/n_{dof} < 10$), to be inconsistent with originating from the primary vertex ($\chi^2_{FD} > 36$) and be downstream of the primary vertex (DIRA > 0). The Λ_b^0 candidate is built from a Λ_c^+ and a D_s^- candidate which are consistent with having the same production vertex ($\chi^2_{vtx}/n_{dof} < 10$). The Λ_b^0 candidate is required to originate from the primary vertex ($\chi^2_{IP} < 25$ and DIRA > 0.999) and to have a lifetime greater than 0.2 ps.

In addition to the selection made in the stripping, the Λ_b^0 daughter masses are restricted further to exclude events with no Λ_c^+ or D_s^-: 2270.25 MeV $< M_{\Lambda_c^+} <$ 2304.75 MeV and 1949.16 MeV $< M_{D_s^-} <$ 1990.44 MeV. The $\Lambda_c^+ D_s^-$ mass distribution after the stripping and the mass requirements are shown in Fig. 3.2. The Λ_b^0 signal peak is clearly visible over a fairly large background.

	Selection
Λ_b^0	$\sum\|p_T\| > 5$ GeV, $5200 < M < 7000$ MeV, $\chi^2_{vtx}/n_{dof} < 10$, $\tau > 0.2$ ps , $\chi^2_{IP} < 25$, DIRA > 0.999
Λ_c^+	$\sum\|p_T\| > 1800$ MeV, $\|(M - M_{PDG})\| < 100$ MeV ,$\chi^2_{vtx}/n_{dof} < 10$, $\chi^2_{FD} > 36$, DIRA > 0
D_s^-	$\sum\|p_T\| > 1800$ MeV, $\chi^2_{vtx}/n_{dof} < 10$, $\chi^2_{FD} > 36$, DIRA > 0
$p/K/\pi$	$\chi^2_{track}/ndof < 3.0$, $p_T > 100$ MeV, $p > 1000$ MeV, $min(\chi^2_{IP}) > 4.0$, TRGHP < 0.4
PID	PIDp $>$ - 10 for p, PIDK $>$ -10 for K, PIDK < 20 for π
At least one track with	P$>$10000 MeV, $p_T > 1700$ MeV, $\chi^2_{track}/ndof < 2.5$, $min(\chi^2_{IP})>16$, IP > 0.1 mm
At least two tracks with	$\chi^2_{track}/ndof < 2.5$, $p_T > 500$ MeV, $p > 5000$ MeV
Event	nLongTracks < 500

Table 3.1: Selection applied in the stripping. Only events fulfilling these requirements are kept.

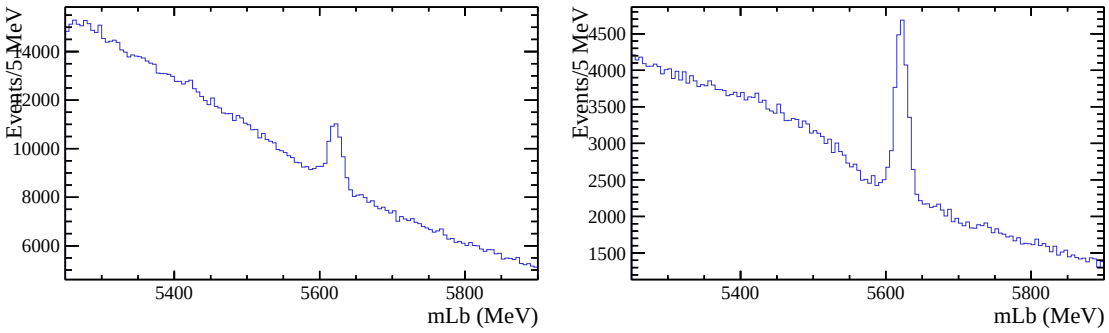


Figure 3.2: $\Lambda_c^+ D_s^-$ invariant mass spectrum after the stripping (left) and additional requirements on the Λ_c^+ and D_s^- masses (right). It should be noted that the y-axises cover different ranges and do not start at 0.

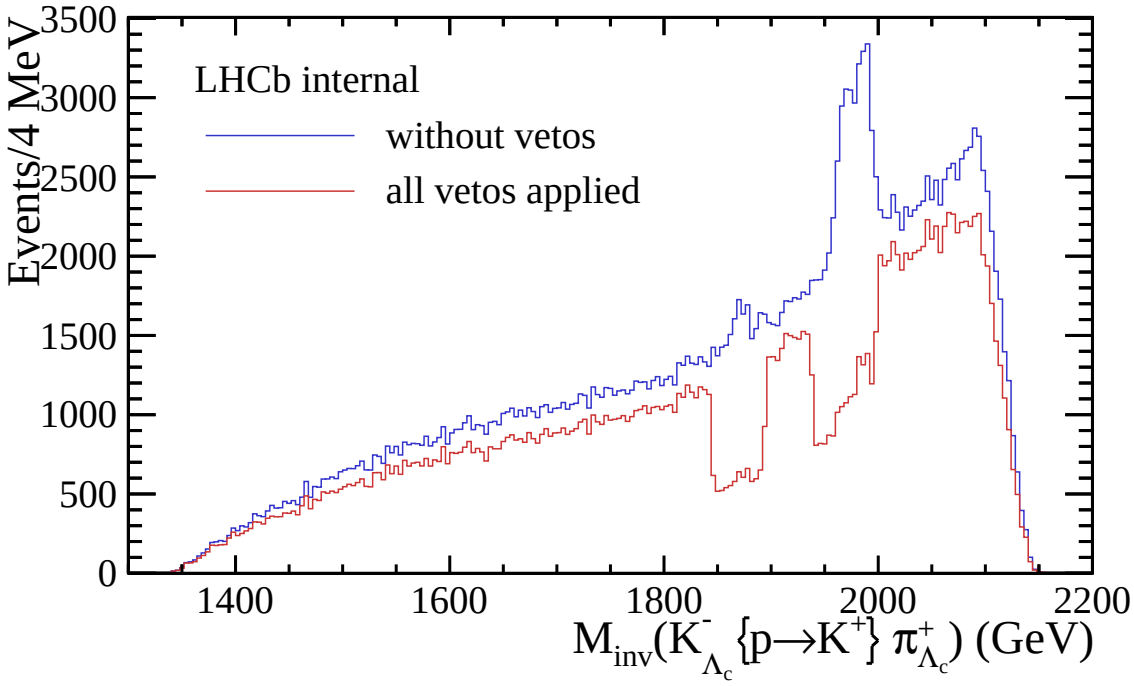


Figure 3.3: Λ_c^+ mass spectrum where the proton is assigned the kaon mass.

3.2 Mis-Identification Vetoes

Each track is assigned a particle hypothesis (electron, muon, pion, kaon, proton). This assignment might not be correct leading to backgrounds from other decays. For example, the proton from $\Lambda_c^+ \to pK^-\pi^+$ could in reality be a mis-identified kaon. To check this, the proton is assigned the kaon mass and the invariant mass of the Λ_c^+ decay products under the new $K^+K^-\pi^+$ hypothesis is calculated. Peaking structures in this new invariant mass spectrum are a clear sign that the particle in question is mis-identified. Fig. 3.3 shows this spectrum. Peaking structures at the nominal D^+ and D_s^- masses can be seen, meaning that indeed the particle identified as proton was in reality a kaon and the observed decay was not $\Lambda_c^+ \to pK^-\pi^+$ but $D^+ \to K^+K^-\pi^+$ or $D_s^+ \to K^+K^-\pi^+$.

If there are signs of mis-identified particles in such invariant mass spectra, so-called vetoes are applied. By just excluding the mass region in the new spectra where signs of different particles are seen, backgrounds from mis-identified particles can be reduced. Combining such mass cuts with a requirement on the product of $\text{Prob}_{\text{NN}}(\text{X}) \times (1 - \text{Prob}_{\text{NN}}(\text{Y}))$, where X is the original mass hypothesis and Y the replacement effectively reduces mis-identified background while retaining a reasonable signal efficiency. This variable is chosen because it ensures that the particle is identified correctly ($\text{Prob}_{\text{NN}}(\text{X})$ is large), while keeping the mis-identification rate low ($\text{Prob}_{\text{NN}}(\text{Y})$ is small). In the example of the Λ_c^+, where a proton could be a mis-identified kaon, X would be the proton and Y the kaon.

For $\Lambda_c^+ \to pK^-\pi^+$ vetoes are applied on the decays $D_s^- \to K^-K^+\pi^-$, $D^- \to$

	Vetoed decay	Veto
Λ_c^+	$D_s^- \to K^+K^-\pi^-$	$\lvert M_{K\pi\{p\to K\}} - M_{D_s}\rvert < 30$ and $p\mathrm{Prob}_{\mathrm{NN}}(p) \times (1 - p\mathrm{Prob}_{\mathrm{NN}}(K)) < 0.05$
	$\phi \to K^+K^-$	$\lvert M_{K\{p\to K\}} - M_{\phi}\rvert < 5$
	$D^0 \to K^-\pi^+$	$\lvert M_{K\{p\to \pi\}} - M_{D^0}\rvert < 25$ and $p\mathrm{Prob}_{\mathrm{NN}}(p) \times (1 - p\mathrm{Prob}_{\mathrm{NN}}(\pi)) < 0.05$
	$D^+ \to K^+K^-\pi^+$	$\lvert M_{K\pi\{K\to p\}} - M_{D^+}\rvert < 25$ and $p\mathrm{Prob}_{\mathrm{NN}}(p) \times (1 - p\mathrm{Prob}_{\mathrm{NN}}(K)) < 0.05$
D_s^-	$\Lambda_c^+ \to pK^-\pi^+$	$\lvert M_{K\pi\{p\to K\}} - M_{\Lambda_c}\rvert < 25$ and $K\mathrm{Prob}_{\mathrm{NN}}(K) \times (1 - K\mathrm{Prob}_{\mathrm{NN}}(p)) < 0.05$ and $\lvert M_{KK} - M_{\phi}\rvert > 5$
	$D^+ \to K^-\pi^+\pi^-$	$\lvert M_{K\pi\{K\to \pi\}} - M_{D^+}\rvert < 25$ and $K\mathrm{Prob}_{\mathrm{NN}}(K) \times (1 - K\mathrm{Prob}_{\mathrm{NN}}(\pi)) < 0.05$ and $\lvert M_{KK} - M_{\phi}\rvert > 5$

Table 3.2: Vetoes applied in the selection. The particles fulfilling these conditions are likely to come from different decays and are rejected.

$K^-K^+\pi^-$ and $\phi \to K^-K^+$, where a kaon is mis-identified as proton, $D^- \to K^-\pi^+\pi^-$ and $D^0 \to K^-\pi^+$ where a pion is mis-identified as proton.

For $D_s^- \to K^-K^+\pi^-$ vetoes are applied on the decays $\Lambda_c^+ \to pK^-\pi^+$ where a proton is mis-identified as kaon and $D^- \to K^-\pi^+\pi^-$ where a pion is mis-identified as kaon. To further improve the efficiency of the vetoes applied to the D_s^-, events where the invariant mass of the K^+K^- system m_{KK} from the D_s^- is close to the invariant mass of the ϕ meson are always accepted. This is chosen because the decay via the $\phi \to K^+K^-$ resonance dominates the $D_s^- \to K^-K^+\pi^-$ Dalitz plot.

The vetoes are summarized in Table 3.2 and the performance plots are shown in Fig. 3.4.

3.3 Multivariate Selection

While the channel considered here delivers already a clean signature after the preselection, the selection of decays with many final state hadrons, for example the signal channel of the overlying analysis $\Lambda_b^0 \to \Lambda_c^+\overline{D}^0K^-$, is in general not a straight-forward task. An optimal selection should use PID information of all final state particles to get a signal with highest possible significance. Such selections have to rely on efficiencies obtained from calibration channels, which first have to be validated.

To reduce the number of variables used in the selection of open-charm particles multi-variate tools have been used before in LHCb [27]. The idea is to take kinematic and PID variables of an open–charm particle X_c and its daughters and combine them to a single optimized output variable using a multivariate algorithm. In the use case here it would simplify the selection of $\Lambda_b^0 \to \Lambda_c^+(\to pK^-\pi^+)/D_s^-(\to K^-K^+\pi^-)$: Instead of using PID information from the six Λ_c^+ and D_s^- daughters, two classifiers are trained — one for Λ_c^+ and one for D_s^- — reducing the number of variables to just two variables, namely the output of the classifiers. Similarly, a classifier can be trained for D^0 decays to be used in $\Lambda_b^0 \to \Lambda_c^+\overline{D}^0K^-$. These classifiers essentially provide a PID-like variable per charmed hadron, instead of the usual ones, which are defined per track.

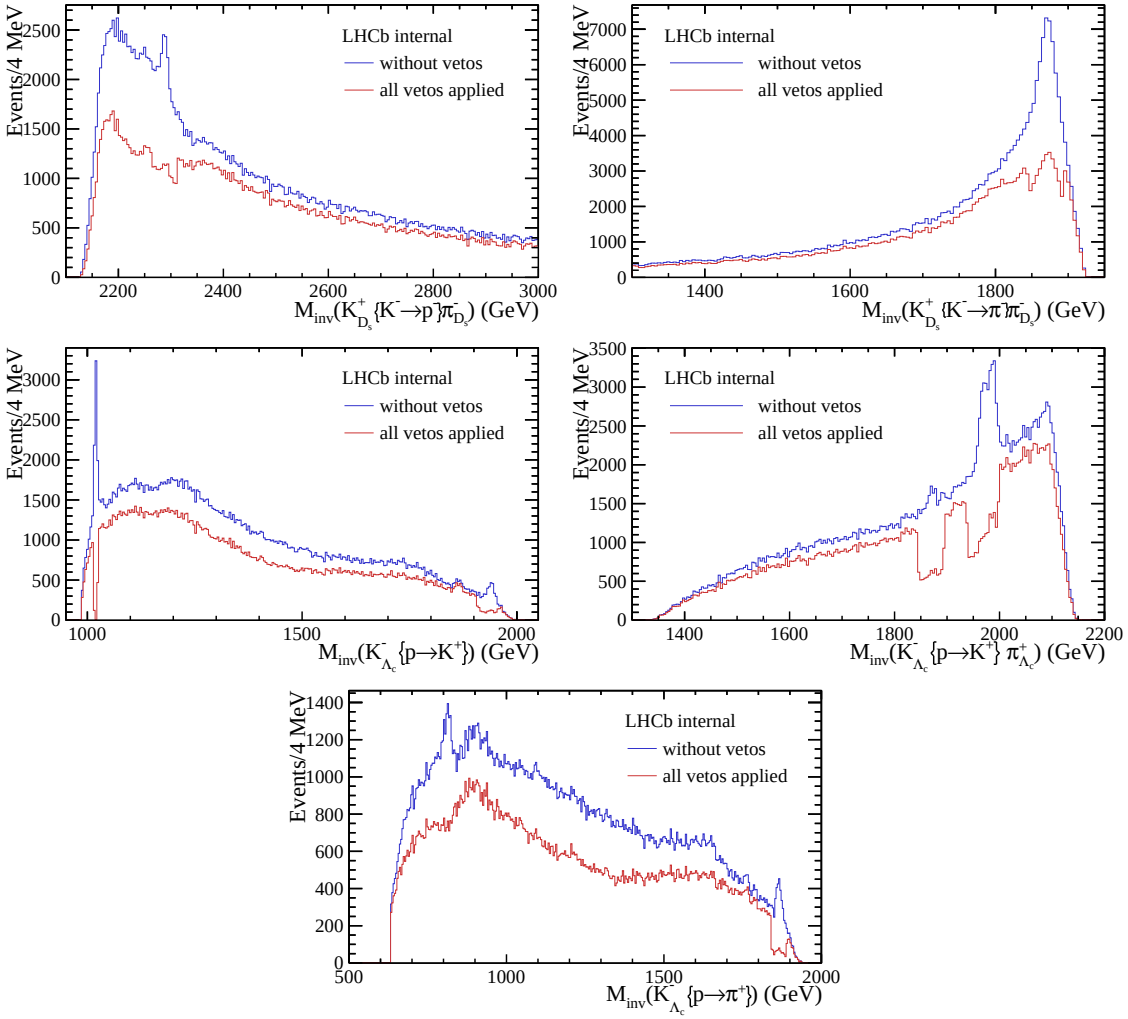


Figure 3.4: Veto performance plots. Top left shows the shows the $K^-K^+\pi^-$ mass spectrum of the D_s^-, where the K^- is assigned the proton mass. Top right shows the shows the $K^-K^+\pi^-$ mass spectrum of the D_s^-, where the K^- is assigned the pion mass. Middle left and middle right show the invariant mass spectra of the pK^- and $pK^-\pi^+$ systems of the Λ_c^+, where the proton is assigned the kaon mass. Bottom shows the Kp mass spectrum of the Λ_c^+, where the proton is assigned the pion mass.

The clean signature of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ can then be used to validate the usage of such classifiers. It has then to be assumed that they are also valid to use on the channel $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$. The classifiers are trained, tested and evaluated using a signal-like and background-like sample taken from independent control channels as input. The control channel are $\Lambda_b^0 \to \Lambda_c^+ \pi^-$, $B_s^0 \to D_s^- \pi^+$ and $B^+ \to \overline{D}^0 \pi^+$. They provide high statistics and a clean signature and only contain one open-charm hadron simplifying the selection. The input samples are split in two parts: A training sample is used to train the method and a testing sample to validate the training result. The algorithm used in this thesis are **Boosted Decision Trees** (BDTs). A decision tree consists of several nodes (see Fig. 3.5). At each node a binary split is applied using the variable, which provides the best separation between signal and background, resulting in two subnodes. This is repeated for the subnodes until a convergence criterion (usually the signal purity in the final nodes) is reached. The final nodes are then either classified as being signal or background depending on whether signal or background events are the majority. This basically selects many hyper-cubes in the variable phase space to separate signal and background as opposed to a simple cut-based analysis that defines a single hyper-cube. A single decision tree usually suffers from overtraining meaning that statistical fluctuations in the training sample are interpreted as feature of the signal or background distributions. This leads to a worse discrimination performance when applying the decision tree to different samples. This can be remedied by using a forest of decision trees instead of just one. The classifier response is then the average of all decision trees. One method to train many decision trees on a single sample is so-called boosting. After the first decision tree is trained, events that are mis-classified are multiplied with a common weight α.

$$\alpha = \frac{1-e}{e}, \tag{3.1}$$

where e is the error rate, the ratio of mis-classified events. The modified sample is then used to train another decision tree and so on, making it more likely that the previously mis-classified event is correctly assigned to signal or background. The algorithms used here are implemented in the **TMVA** toolkit [28].

To check the performance of BDTs the response distribution for signal and background events is compared between training and testing sample. Large deviations hint at the BDTs being over-trained meaning that they see statistical fluctuations as features of the training sample leading to a worse performance on the testing sample. The so-called Receiver-Operating-Characteristic (ROC) curve shows the background rejection over the signal efficiency for a BDT. The larger the area under this curve, the better the BDT separates between signal and background without throwing away too many signal events.

The BDTs used here are discussed in detail in Chapter 6. An important point is that they are trained on control channels different from the channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$, to which they will be applied. Whether the efficiencies of the BDTs trained on

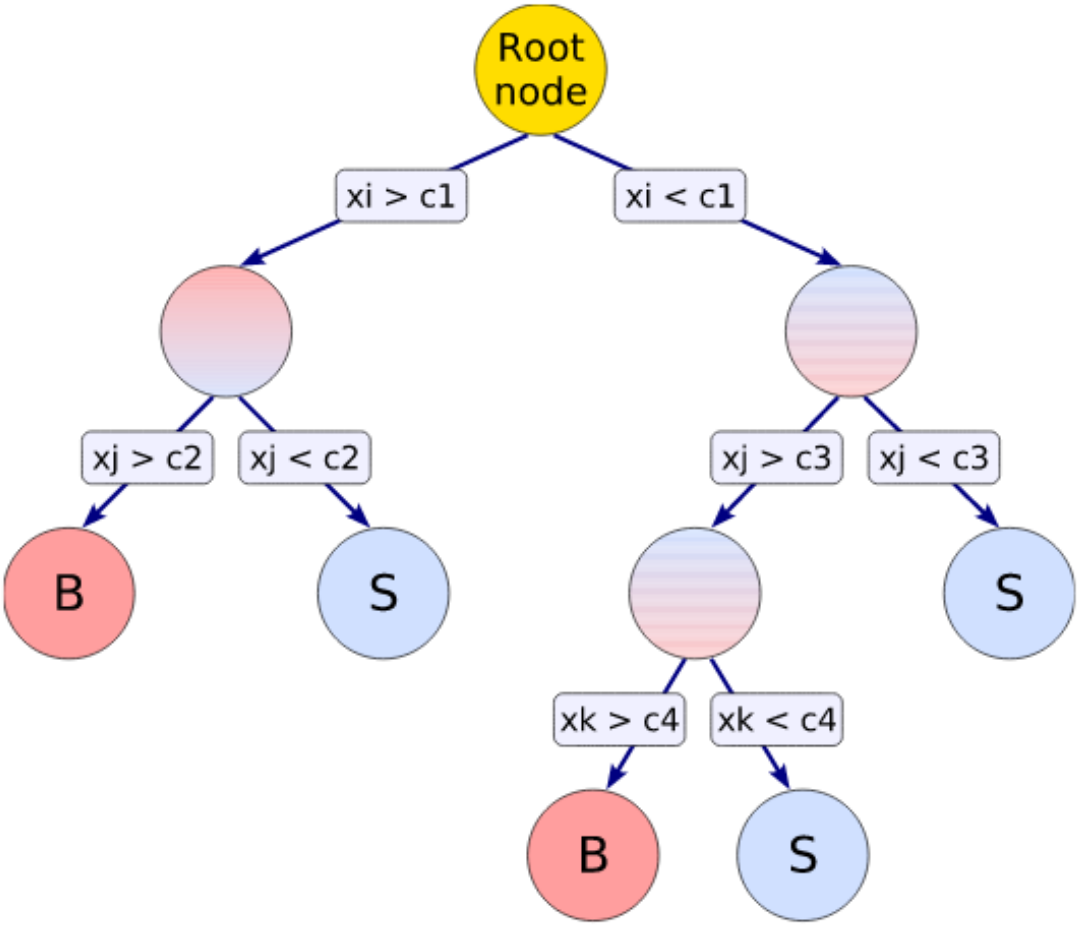


Figure 3.5: Schematic description of a decision tree. Taken from [28].

subsample	Λ_c^+ BDT	D_s^- BDT
1	-0.05500	-0.03625
2	-0.07750	-0.01250
3	-0.08250	-0.02250
4	-0.07625	-0.02500
average	-0.07281	-0.02406

Table 3.3: Optimized BDT selection for different subsamples.

independent channels are applicable to this channel is also tested extensively in Chapter 6.

The BDTs selections are optimized to find the working point with the highest signal significance, described in the next section.

3.3.1 Signal Optimization

For BDTs a good working point has to be found as for different BDT selections the signal efficiency and background rejection vary. This working point can be found by maximizing a so-called figure of merit (FoM) which has to be chosen beforehand. A common choice is to use

$$\frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}}, \tag{3.2}$$

where N_{sig} is the number of signal and N_{bkg} the number of background events in a defined signal region. The optimal figure of merit is found for both BDTs at once. This is done in the following procedure:

From a certain starting point the BDT selections are varied, effectively spanning a two-dimensional grid of different BDT selection combinations. At each point of the grid the two BDT requirements are added to the selection and the number of signal events N_{sig} and background events N_{bkg} in the signal region are extracted, where the signal region is defined as the region containing 99% of all signal events . To avoid biasing the FOM calculation cross-validation is used. First the data is split into four parts with the same number of events. Three of these parts are then combined at a time forming four subsamples, each excluding a different part of the total sample. The optimization procedure is then run on each subsample resulting in four optimized BDT selections. The final selection is then taken to be the average of the selections obtained from the subsamples and added to the preselection and vetoes. Table 3.3 summarizes the result of the optimization scan.

4 Yield Extraction

The $\Lambda_c^+ D_s^-$ invariant mass distribution after the selection is shown in Fig. 4.1. The signal peak is clearly visible at the nominal Λ_b^0 mass. Different background components are discussed in the following:

4.1 Background Sources

Partially reconstructed background In this analysis photons are not reconstructed. Therefore decays like $\Lambda_b^0 \to \Lambda_c^+ D_s^{*-}$ with the $D_s^{*-} \to D_s^- \gamma$ or $D_s^{*-} \to D_s^- \pi^0 (\to \gamma\gamma)$ have the same final state as the signal channel. The photon carries away energy and momentum, which is missing when reconstructing the Λ_b^0 candidate resulting in the structure below the nominal Λ_b^0 mass. The main contribution to partially reconstructed backgrounds is $\Lambda_b^0 \to \Lambda_c^+ D_s^{*-}$. The D_s^{*-} mainly decays to $D_s^- \gamma$ (93.5% branching fraction) [2] where the photon has a momentum of approximately 140 MeV in the D_s^{*-} rest frame. Therefore this background is well separated from the signal peak.

This background is described phenomenologically by the sum of two Gaussian distributions with different means μ and standard deviations σ:

$$N_{part,1} \cdot G(m_{\Lambda_b}; \mu_1, \sigma_1) + N_{part,2} \cdot G(m_{\Lambda_b}; \mu_2, \sigma_2), \tag{4.1}$$

where $N_{part,1}$ and $N_{part,2}$ are the respective yields.

Charmless background So-called charmless decays, where the Λ_b^0 decays without intermediate Λ_c^+ or D_s^- are a possible background contribution peaking at the nominal Λ_b^0 mass. They are discussed in more detail in section 4.4. A Gaussian distribution is chosen to describe it:

$$N_{cless} \cdot G(m_{\Lambda_b}; \mu_{cless}, \sigma_{cless}), \tag{4.2}$$

where N_{cless} is the yield.

Combinatorial background Combinatorial background is a featureless distribution from random combinations of the desired final state particles. It is described by an exponential distribution:

$$N_{comb} \cdot e^{(\tau m_{\Lambda_b})}, \tag{4.3}$$

with the yield N_{comb} and parameter τ. This description for the combinatorial background is chosen because particles with low energies are more abundant, so random combinations with low energy are more common than combinations with high energy, which would appear at larger masses.

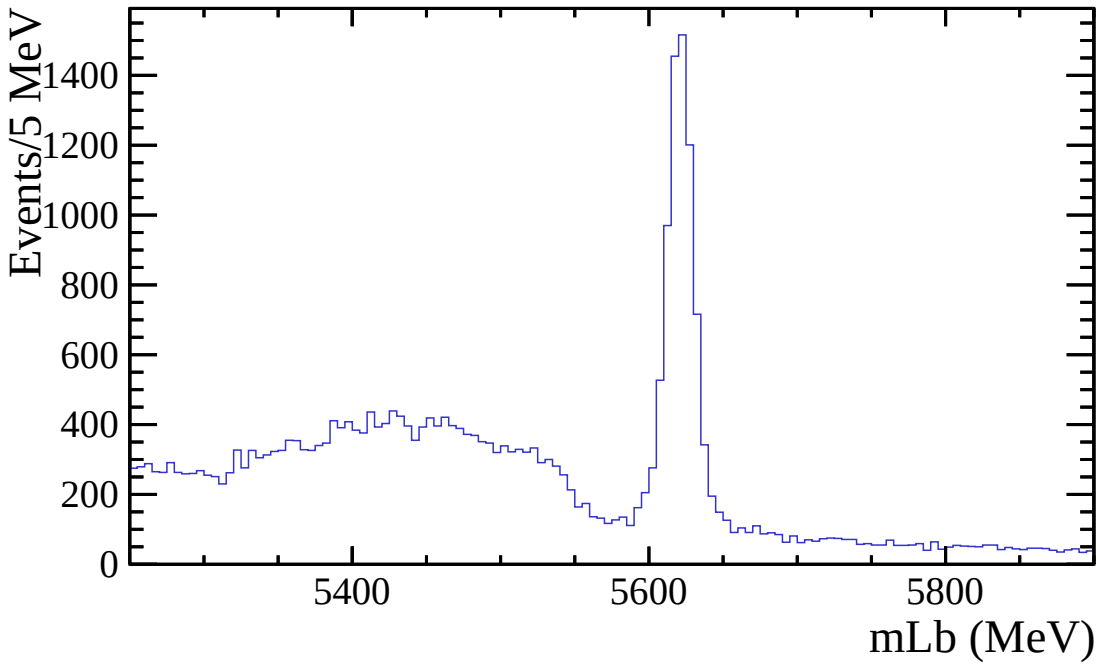


Figure 4.1: $\Lambda_c^+ D_s^-$ invariant mass distribution after selection.

4.2 Fitting Procedure

The fitting procedure utilizes an unbinned maximum-likelihood approach. The software used is **RooFit** [29].

The $\Lambda_c^+ D_s^-$ invariant mass used in this analysis is reconstructed in a kinematic fit with a decay tree fitter (DTF) utilizing a Kalman filter [30]. Its advantage is that it simultaneously fits all decay tree parameters, like vertex positions, using the external constraints provided by the final state particle measurements and internal constraints from 4-momentum conservation. It allows for an improvement in the Λ_b^0 mass resolution by fixing the daughter masses, in this case the masses of Λ_c^+ and D_s^-, to their known values.

The combinatorial background is described by an exponential function, the partially reconstructed background by two Gaussian distributions and the signal shape by a Voigtian function to account for different resolution effects. A Voigtian $V(x;\mu,\sigma,\gamma)$ is the convolution of a Gaussian G with mean μ and standard deviation σ with a Lorentz distribution L with half-width-half-maximum γ:

$$\begin{aligned} V(x;\mu,\sigma,\gamma) &= \int_{-\infty}^{\infty} G(y;\mu,\sigma) L(x-y;\gamma) dy, \\ G(y;\mu,\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}, \\ L(x-y;\gamma) &= \frac{\gamma}{\pi((x-y)^2+\gamma^2)}. \end{aligned} \tag{4.4}$$

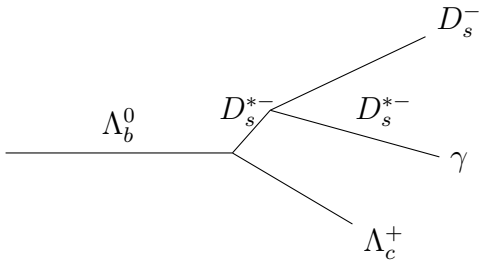


Figure 4.2: Partially reconstructed background. For the purpose of illustration the D_s^{*-} is shown with a separate decay vertex. In reality it is so short-lived that it does not travel a macroscopic distance in the detector.

The charmless component is described by a Gaussian with the shape and yield fixed to the result of a three-dimensional fit, see section 4.4. The total fit function is

$$\begin{aligned} f(m_{\Lambda_b}) = N_{sig} \cdot V(m_{\Lambda_b}; \mu_{sig}, \sigma_{sig}, \gamma_{sig}) + N_{part,1} \cdot G(m_{\Lambda_b}; \mu_1, \sigma_1) \\ +N_{part,2} \cdot G(m_{\Lambda_b}; \mu_2, \sigma_2) + N_{cless} \cdot G(m_{\Lambda_b}; \mu_{cless}, \sigma_{cless}) + N_{comb} \cdot e^{(\tau m_{\Lambda_b})}, \end{aligned} \tag{4.5}$$

where N_{comb} is the yield of the combinatorial background, $N_{part,1}$ and $N_{part,1}$ the yields to estimate the partially reconstructed backgrounds, N_{sig} the signal yield and N_{cless} the yield of the charmless contribution. Fig.4.3 shows the result of the total fit.

4.3 Background Subtraction

The *sPlot* technique [31] is used to extract background-subtracted projections of variables using the knowledge of the signal and background distributions in a discriminating variable. In this analysis the Λ_b^0 mass (and D_s^- and Λ_c^+ masses in multi-dimensional fits) is used as discriminating variable. In can be understood to work in the following way: Events in the background region are assigned a negative weight and events in the signal region a positive weight (called s-weights). By summing over all events the events with negative weights eliminate the background events in the signal region in a statistical way. An important requirement for this method to work is that the variable, whose background-subtracted distribution is to be plotted, is not correlated to the discriminating variable. To extract s-weights the fit is redone with all fit parameters but the signal yields fixed. For each yield parameter N_{comp} there is a corresponding s-weight distribution s_{comp}. The procedure is constructed in such a way that summing over all s-weights of a component gives back the yield:

$$N_{comp} = \sum_{i=1}^{events} s_{comp,i}. \tag{4.6}$$

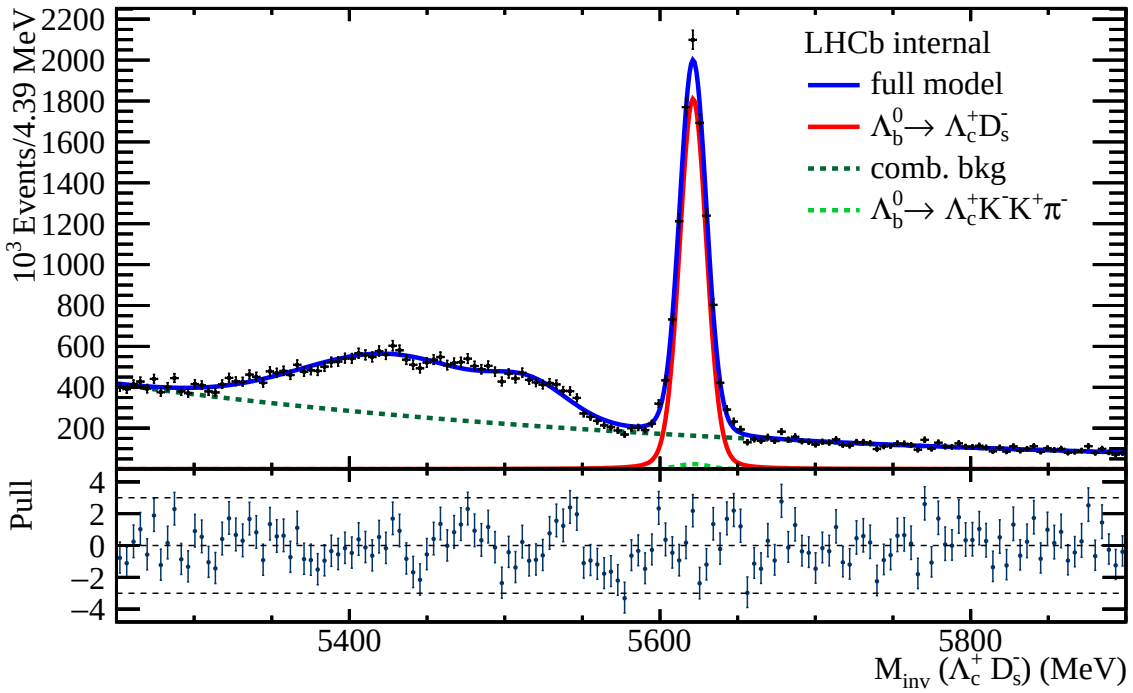


Figure 4.3: Fitted $\Lambda_c^+ D_s^-$ mass distribution.

4.4 Charmless Backgrounds

One possible background is the Λ_b^0 directly decaying to the final state particles without intermediate Λ_c^+ or D_s^-: $\Lambda_b^0 \to pK^-\pi^+D_s^-$, $\Lambda_b^0 \to \Lambda_c^+K^-K^+\pi^-$ and $\Lambda_b^0 \to pK^-\pi^+K^-K^+\pi^-$ (see Fig. 4.4). To check if such charmless contributions are significant, the Λ_b^0 mass distribution is plotted for the sidebands of Λ_c^+ and D_s^-, defined as the mass region $25 < |M_c - M_{PDG}| < 50$, where M_c is the measured mass of the charmed particle and M_{PDG} its nominal value.

As seen in Fig. 4.5 the only significant contribution is from $\Lambda_b^0 \to \Lambda_c^+K^-K^+\pi^-$, which will be taken into account by a multi-dimensional fit to the Λ_b^0, Λ_c^+ and D_s^- mass spectra. They are fitted simultaneously to differentiate between $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and different background components. The total fit function is:

$$\begin{aligned} f(m_{\Lambda_b}, m_{\Lambda_c}, m_{D_s}) = N_{\Lambda_B} &\times G(m_{\Lambda_b})G(m_{\Lambda_c})G(m_{D_s}) \\ +N_{cless} &\times G(m_{\Lambda_b})G(m_{\Lambda_c})B(m_{D_s}) \\ +N_{combD_s} &\times B(m_{\Lambda_b})G(m_{\Lambda_c})B(m_{D_s}) \\ +N_{combD_s} &\times B(m_{\Lambda_b})B(m_{\Lambda_c})G(m_{D_s}) \\ +N_{combD_s} &\times B(m_{\Lambda_b})G(m_{\Lambda_c})G(m_{D_s}) \\ +N_{comb} &\times B(m_{\Lambda_b})B(m_{\Lambda_c})B(m_{D_s}), \end{aligned} \tag{4.7}$$

where N are the yields of the different components, G is a Gaussian with mean μ

and standard deviation σ and B a linear polynomial.

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \tag{4.8}$$

$$B(x) = 1 + bx. \tag{4.9}$$

The components can be classified as follow:

- N_{Λ_B}: $\Lambda_b^0 \to \Lambda_c^+ D_s^-$
- N_{cless}: $\Lambda_b^0 \to \Lambda_c^+ K^+ K^- \pi^-$
- $N_{comb\Lambda_c}$: combinatorial background with reconstructed Λ_c^+
- N_{combD_s}: combinatorial background with reconstructed D_s^-
- $N_{comb\Lambda_c D_s}$: combinatorial background with reconstructed Λ_c^+ and D_s^-
- N_{comb}: pure combinatorial background

To ensure a stable fit the requirements on the Λ_c^+ and D_s^- masses introduced in the preselection (see section 3.1) are relaxed to 2258 MeV $< M_{\Lambda_c^+} <$ 2317 MeV and 1935 MeV $< M_{D_s^-} <$ 2005 MeV, while the Λ_b^0 mass is restricted to 5560 MeV $< M_{\Lambda_b^0} <$ 5680 MeV. This leaves large enough sidebands in the mass spectra to properly fit all the background contributions, while excluding the partially reconstructed background in the Λ_b^0 mass spectrum.

The fit results are shown in Fig. 4.6. The yield and shape of $\Lambda_b^0 \to \Lambda_c^+ K^+ K^- \pi^-$ from the three-dimensional fit is then fixed in the final fit, after correcting it for the different Λ_c^+ and D_s^- mass ranges compared to the final fit.

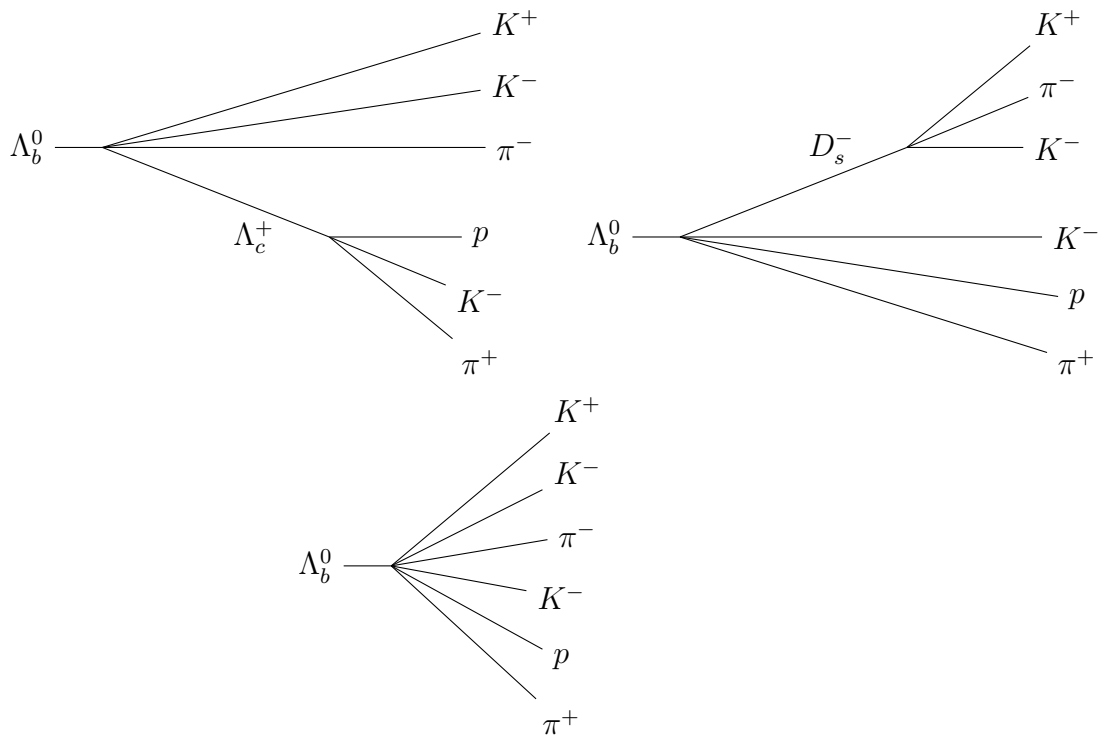


Figure 4.4: Charmless Λ_b^0 decays. Top left: $\Lambda_b^0 \to \Lambda_c^+ K^- K^+ \pi^-$. Top right: $\Lambda_b^0 \to pK^-\pi^+ D_s^-$. Bottom: $\Lambda_b^0 \to pK^-\pi^+ K^- K^+ \pi^-$.

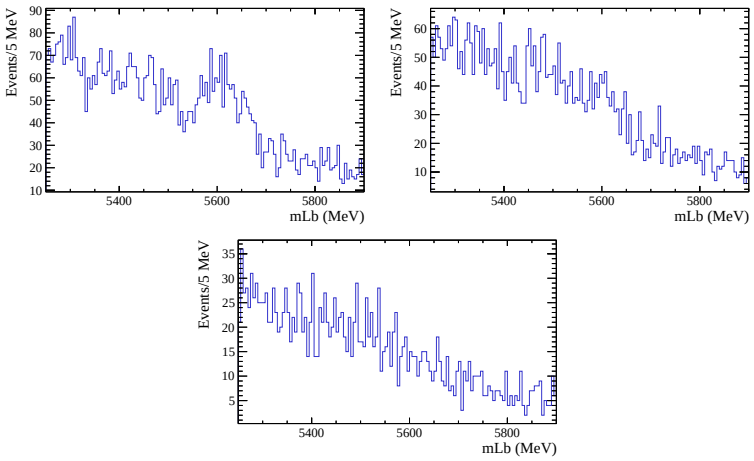
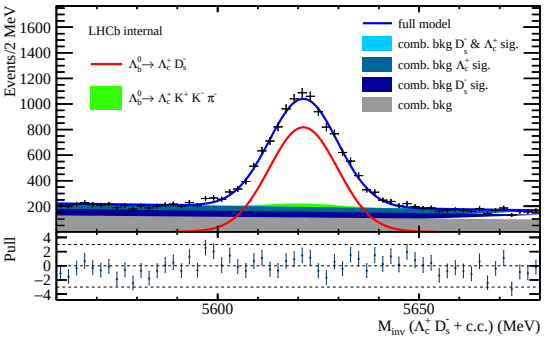
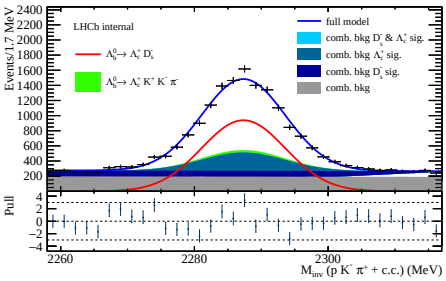


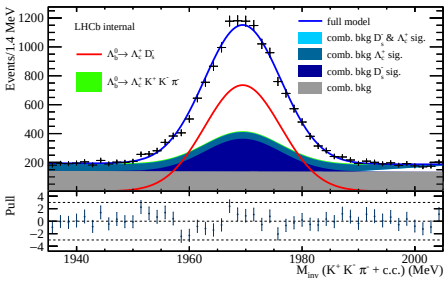
Figure 4.5: $\Lambda_c^+ D_s^-$ invariant mass distribution in the Λ_c^+ and D_s^- sidebands. Top left shows the $\Lambda_c^+ D_s^-$ invariant mass for the D_s^- sideband, top right for the Λ_c^+ sideband and bottom for both sidebands.



$\Lambda_c^+ D_s^-$ mass distribution.



Λ_c^+ mass distribution.



D_s^- mass distribution.

Figure 4.6: One-dimensional projections of the three-dimensional fits to the $\Lambda_c^+ D_s^-$, Λ_c^+ and D_s^- mass distributions.

5 Efficiency Correction

At each selection step a portion of events of interest (signal events) are rejected. To determine the original number of signal events the efficiency ϵ_{tot} of the whole selection has to be known:

$$\epsilon_{tot} = \frac{N_{passed}}{N_{total}}, \tag{5.1}$$

where N_{total} is the number of signal events before and N_{passed} the number of signal events after all selection steps. Equivalently

$$N_{total} = \frac{N_{passed}}{\epsilon_{tot}} \tag{5.2}$$

gives the total efficiency corrected number of events N_{total}. Efficiencies are either determined on MC simulations or data-driven methods exploiting calibration channels.

The efficiencies of all selection steps can be combined to the total efficiency ϵ_{tot}. The combined efficiency ϵ_{comb} of two selection steps is usually calculated by taking the product of the single efficiencies ϵ_1 and ϵ_2:

$$\epsilon_{comb} = \epsilon_1 \times \epsilon_2. \tag{5.3}$$

This combination only leads to a correct result if the single efficiencies are not correlated meaning that ϵ_{comb} does not depend on the order of the selection steps. If they are correlated the true combined efficiency is given by:

$$\epsilon_{comb} = \rho \cdot \min(\epsilon_1, \epsilon_2) + \frac{(\epsilon_1 - \rho \cdot \min(\epsilon_1, \epsilon_2)) \times (\epsilon_2 - \rho \cdot \min(\epsilon_1, \epsilon_2))}{1 - \rho \cdot \min(\epsilon_1, \epsilon_2)}, \tag{5.4}$$

where ρ is the correlation coefficient. It can be easily seen that equation reduces to eq. 5.3 when there is no correlation ($\rho = 0$).

5.1 Signal Yield

The efficiency corrected signal yield is extracted using s-weights taken from the final fit and corrected for the per-event efficiency of all selection steps:

$$Y = \sum_{i=1}^{N_{events}} \frac{s_i}{\epsilon_{tot,i}}, \tag{5.5}$$

where Y is the efficiency corrected yield, N_{events} is the total number of events after the final selection, s_i is the s-weight of an event and $\epsilon_{tot,i}$ is the total selection efficiency of an event.

5.2 Efficiencies

The total efficiency is the product of the following single efficiencies:

Generator acceptance As the LHCb detector does not cover the full solid angle the fraction of Λ_b^0 decays whose final state particle are not in the detector acceptance need to be taken into account. This efficiency is obtained from MC simulation.

Reconstruction efficiency This includes the efficiency of event triggering, track reconstruction and stripping. It is obtained from Monte-Carlo simulations. The simulation does not simulate the p_T and η distribution of the Λ_b^0 well as the underlying processes of the b-production in the pp-collision are only understood with a limited precision. Fig. 5.1 shows the comparison between the MC distribution and s-weighted data. The simulated η distribution shows a small bias towards larger values. Since the reconstruction efficiency depends on p_T and η, and in order to account for the observed differences, it is binned in those variables.

Offline selection efficiency The efficiency of the preselection in addition to the stripping. Here only requirements on the Λ_c^+ and D_s^- mass distributions are made. The efficiencies are estimated from MC simulation.

Veto efficiency The efficiency of the mis-assignment vetoes. It is estimated from MC simulation. As the PID variables used in the vetoes are not simulated correctly, they have to be resampled beforehand (see Section 5.2.1).

BDT efficiency The total BDT efficiency is the combination of the two single BDT efficiencies. How they are obtained is discussed in the next chapter.

While the whole dataset of 2011 and 2012 is used in this analysis, for this decay only the MC simulations configured for the 2012 data-taking conditions were available. This might introduce a systematic uncertainty to the efficiency corrected yield, as different efficiency contributions might be different between 2011 and 2012 data-taking periods. Because this can only be quantified with MC, no systematic uncertainty is assigned.

The single efficiencies are summarized in Table 5.1. The efficiency corrected yield is then

$$N_{\Lambda_b^0} = (4.5009 \pm 0.1044) \cdot 10^6,$$

where the uncertainty is statistical only. The systematic uncertainty is estimated in Chapter 7.

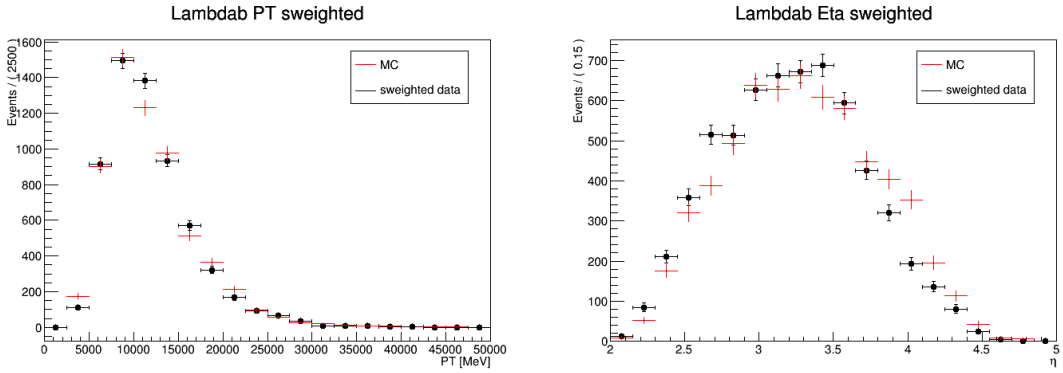


Figure 5.1: Comparison between MC and s-weighted data.

Selection	Efficiency (%)
Acceptance	17.00
Reconstruction and stripping	1.21
Charm hadron mass selection	94.13
Mis-assignment Vetoes	96.06
Λ_c^+ BDT	98.41
D_s^- BDT	96.62
total	0.183

Table 5.1: Signal efficiencies.

5.2.1 PID Resampling

MC simulation does not describe PID variables and their correlations very well. If the selection uses a PID variable, the efficiency can therefore not directly be taken from simulated data. One way around this is to resample the MC data. This is done by using clean channels, where no PID information is needed to identify a particle. The selection efficiencies of PID variables depend on the transverse momentum and pseudo-rapidity of the particle to be identified and on the track multiplicity in an event [32]. The efficiency distribution is therefore binned in these variables. For each MC simulated particle the PID variable is drawn from the distribution corresponding to its $(p_T/\eta/nTracks)$ values. This results in a distribution resembling real data. The disadvantage of this procedure is that it does not take correlations between multiple particles into account. Also, the phase space of the calibration channels may not cover the whole phase space of the channel one would like to resample.

6 Validation Of Boosted Decision Trees (BDTs)

In this chapter the BDTs introduced in section 3.3 are described in more detail. The BDTs effectively provide a PID-like variable for a charmed hadron, instead of variables on per-track basis, like the output from the PID system. They are trained and their efficiencies are evaluated on a control channel. Whether those BDTs can be applied to other channels is verified in a data-driven way using the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channel.

In an earlier analysis of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ [33] similar BDTs were used and shown to produce reasonable results. There, a previous stripping version was used. Therefore BDTs have to be trained with data selected with the new stripping version.

6.1 Motivation

In the overlying analysis of the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ PID information of the final state particles is needed to extract a signal. A possible way to do this, is to apply a selection using the Prob_{NN} variables of all six final state particles. All those six selection variables have to be optimized at once. A further problem is that they could be correlated, leading to non-factorizing efficiencies. For example, two of these particles could hit a RICH system close to each other, so that their Cherenkov cones overlap. This would introduce a correlation in the RICH output. This can be tested using the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channel. It already provides a clean enough signal without using PID information to be able to extract the yield, which can be used to define a baseline. Then several selections using Prob_{NN} variables of the Λ_c^+ are made, corrected for the PID selection efficiency and compared to the baseline yield. The yields are obtained from three-dimensional fits introduced in 4.4.

The efficiencies are obtained on a per-particle basis using control channels, where no PID information is needed to extract a yield. The single efficiencies are then combined to a total event efficiency by multiplication.

Fig. 6.1 shows the efficiency corrected yield for different selections using $\text{Prob}_{\text{NN}}(K)$ of the kaon, $\text{Prob}_{\text{NN}}(p)$ of the proton and $\text{Prob}_{\text{NN}}(\pi)$ of the pion from Λ_c^+ at different fractions of signal events surviving the selection, which can be seen as 'true' combined efficiency. The different colors represent different combinations of $\text{Prob}_{\text{NN}}(p)$ and $\text{Prob}_{\text{NN}}(\pi)$ selections. Points of the same color represent different $\text{Prob}_{\text{NN}}(K)$ selections. The blue band shows the baseline yield. A significant deviation of the efficiency corrected yields of the selections using PID information from the baseline yield can be seen. Also the efficiency corrected yields tend to be smaller for stricter

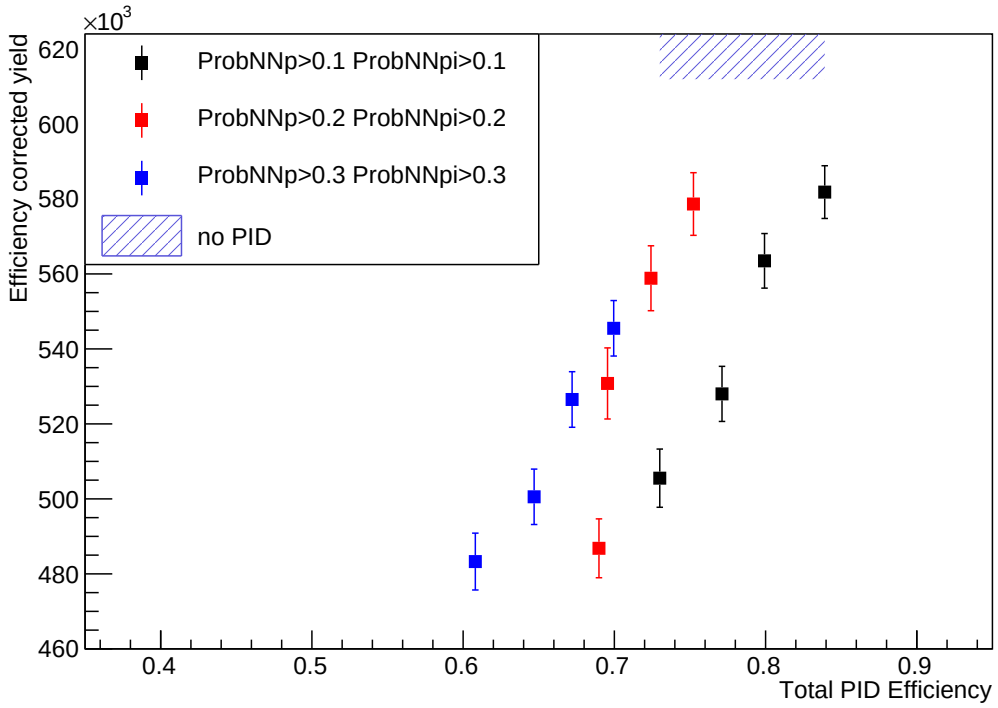


Figure 6.1: Efficiency corrected yields at different values of combined efficiencies for different selections using PID variables. The blue band shows the yield of the selection using no PID information.

Particle	Selection
Λ_c^+, D_s^-	DIRA > 0.999
	$\chi^2_{IP} < 10$
Bachelor pion	$Prob_{NN}(\pi) > 0.1$

Table 6.1: Selection of Λ_c^+ and D_s^- BDT training channels.

selections indicating that the efficiency corrections do not lead to the correct results.

6.2 Open-Charm BDTs

To try to remedy the issues seen with using PID information of all daughter particles, BDTs combining the information of three daughter particles to one are used instead. The aim is to provide a single variable for charmed hadron identification instead of using the PID variables of the open-charm decay products, which are defined on a per-track basis.

To avoid biasing the BDTs the training channels used are not the same as the channels they are applied on. MC data cannot be used for training as it does not reproduce variables and correlations related to particle identification entirely correct. Reweighting the MC variables to match the signal distributions would be challenging as many PID variables are used and would have to be reweighted at once.

The channels $\Lambda_b^0 \to \Lambda_c^+\pi^-$, $B_s^0 \to D_s^-\pi^+$ and $B^- \to D^0\pi^-$ provide clean signatures and high statistics to train BDTs to select Λ_c^+, D_s^- and D^0 hadrons. In all three channels a b-hadron X_b decays to a charmed hadron X_c and a pion: $X_b \to X_c\pi^\pm$. The BDTs are described in detail in [34]. The selection applied on these channels is kept very loose because the same selection has also to be applied to other channels when the BDTs are used there. The BDTs do not use information of the mother b-hadron. The idea is to keep them as general as possible to be able to also apply them to other channels containing the same charmed hadrons. The selection of the channels $\Lambda_b^+ \to \Lambda_c^+\pi^-$ and $B_s^0 \to D_s^+\pi^-$ is summarized in Table 6.1, where bachelor pion refers to the pion directly coming from the X_b decay. On the bachelor pion a loose PID selection is applied. This helps to get a clear X_b signal and does not influence the BDT training, as no variables of the bachelor pion are used. The BDT for D^0 does not need further selections after the stripping. The input variables used in the training can be found in table 6.2. They consist of kinematic variables of the X_c and of kinematic variables and PID information of the X_c daughters. It has to be verified that using the BDTs trained on the control channels also produces reasonable results for other channels. Especially efficiency factorization has to be checked, as e.g. in the analysis $\Lambda_b^0 \to \Lambda_c^+\overline{D}^0K^-$ both the Λ_c^+ and D^0 BDTs are applied.

A s-weighted $X_b \to X_c\pi$ sample is used to train and test the BDTs. The data was taken in 2012 with the LHCb detector. The s-weights are obtained with a two-dimensional fit to the X_b and X_c mass spectra to account for charmless contributions,

	Variable
X_c	p, log p_T, χ^2_{EV}, log χ^2_{IP}
Daughters	p, log p_T, log η, log χ^2_{IP}, arctan $\mathcal{L}_{Muon}$(bkg), arctan $\mathcal{L}_{Muon}(\mu)$, $N_{shared}(\mu)$ $b_{thresh}(K)$, $b_{thresh}(\pi)$, $b_{thresh}(p)$, arctan $\Delta\mathcal{L}_{RICH}(bt)$, arctan $\Delta\mathcal{L}_{RICH}(e)$, arctan $\Delta\mathcal{L}_{RICH}(K)$, arctan $\Delta\mathcal{L}_{RICH}(\mu)$, arctan $\Delta\mathcal{L}_{RICH}(p)$, log $Prob_{NN}$(Ghost), log χ^2_{Match}, log χ^2_T/n_{dof}, log χ^2_{Velo}/n_{dof}, $b_{RICH,1}$, $b_{RICH,2}$, b_{RICH}(aerogel) log Q_{Velo}, b_{isMuon}

Table 6.2: BDT training variables [34].

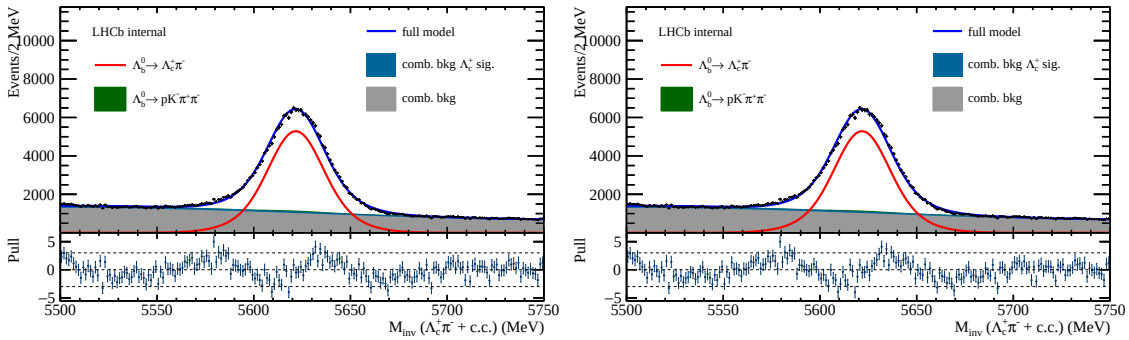


Figure 6.2: One-dimensional projections of the two-dimensional fit to the $\Lambda_c^+\pi-$ (left) and Λ_c^+ (right) mass distributions in the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ control channel.

similar to the three-dimensional fit introduced in the previous chapters. For the X_b and X_c signals the sum of two Gaussian distributions with common mean, but different widths are chosen. The backgrounds are described by polynomial functions. The mass distributions for the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ control channel are shown in Fig. 6.2. A difference to how BDTs are usually trained is that instead of using two separate signal-like and background-like samples the events of the same sample are used as signal and background proxies only distinguished by their s-weights.

Fig. 6.3 and Fig. 6.4 show the BDT training results. No significant sign of overtraining is seen.

6.3 Application to the Channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$

To apply the BDTs to other channels, the selection of those channels and the BDT training channels should be consistent. Especially vetoes applied on the charmed particles have to be the same to ensure that the efficiency calculation is correct. For this reason the same vetoes described in section 3.2 are also applied to the

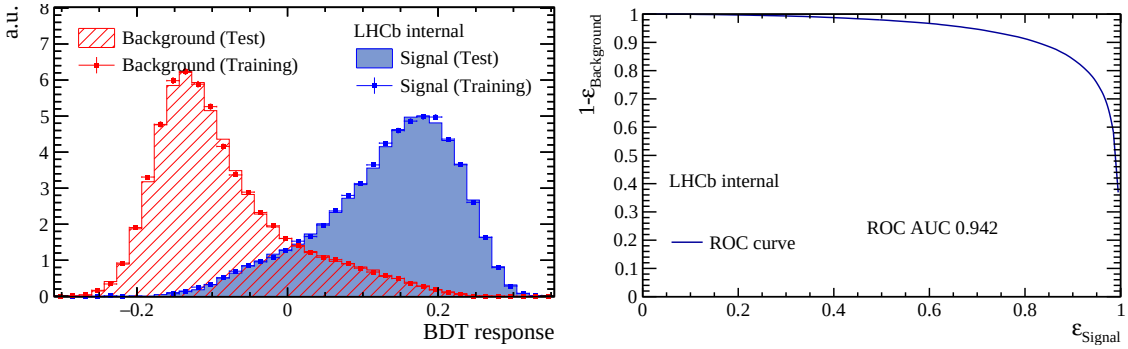


Figure 6.3: Λ_c^+ BDT output comparison between training and test data on the left side. Right side shows the corresponding ROC curve.

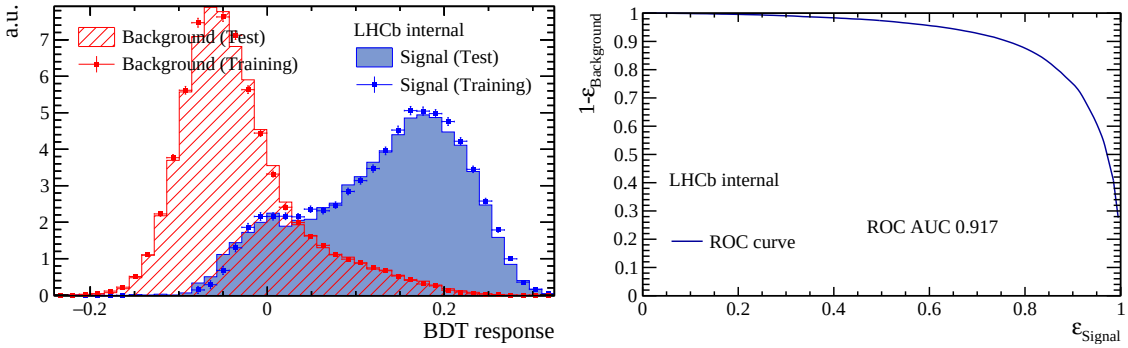


Figure 6.4: D_s^- BDT output comparison between training and test data on the left side. Right side shows the corresponding ROC curve.

corresponding $X_b \to X_c\pi$ selection before training the Λ_c^+ and D_s^- BDTs. These vetoes also use PID information. To keep the effect of possible correlations between the vetoes and the BDT output minimal, the cuts on ProbNN variables in the vetoes are chosen to be highly efficient.

In the stripping selection of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ events a loose PID selection is applied on all D_s^- daughters which are also added to the D_s^- daughters in $B_s^0 \to D_s^-\pi^+$ decays before training the D_s^- BDT (see Table 3.1).

One has to make sure the variables used for training are similar in the training and application channel. The BDT output depends on the distribution of the input variables. If they are too different from the training channel the classification performance does suffer. Especially the efficiency of the BDT selection can then not be evaluated on the training channel.

The comparison of the BDT input variables between the training channels $\Lambda_b^0 \to \Lambda_c^+\pi^-$– and $B_s^0 \to D_s^-\pi^+$ and the application channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ shows good agreement (see Appendix B).

A further check to determine if the BDTs are applicable on $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ is to compare the s-weighted BDT output distributions of this channel with the control channels. This is shown in Fig. 6.5. The distributions are similar, but for the

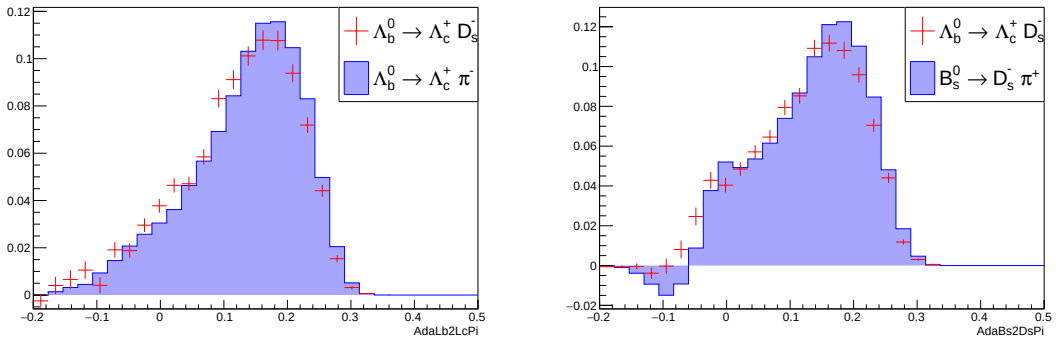


Figure 6.5: S-weighted BDT output for application and control channels. Λ_c^+ BDT on the left, D_s^- BDT on the right.

control channels they are slightly shifted to the higher BDT outputs. This means that applying the BDT to the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channel will reduce the background, but because of the differences between the distributions, the BDT selection efficiencies obtained from the control channels might not be transferable to the application channel. This is checked in the next section.

6.4 Validation of PID Efficiencies

The efficiency of the BDTs could depend on certain variables, which has to be taken into account when applying them to different channels, where the distributions of these variables could differ. The PID variables from both RICH detectors used in training are expected to depend on the number of tracks in the event, as the identification efficiency is lower for events with high track multiplicity. Also, different decay channels could have different lifetime acceptances for the X_b decays. Fig. 6.6 shows the Λ_c^+ BDT efficiency as function of χ^2_{FD} of the Λ_c^+ and the number of tracks of the event. As expected, the BDT efficiency is lower for events with more tracks, as the RICH performance suffers when the track multiplicity is high and the BDTs rely on RICH variables. The efficiency is nearly constant for large values of χ^2_{FD}. At small values, where the Λ_c^+ decays closely to the primary vertex, the efficiency drops. This means that the BDTs perform better when the Λ_c^+ has a clearly separated origin vertex. Binning the BDT efficiency in this variable is then justified, as different application channels could have different b-hadron lifetime acceptances, so that also the distributions of χ^2_{FD} could be different.

To correct for these two effects the BDT efficiency is binned in χ^2_{FD} and $nTracks$. The efficiency is obtained from the BDT testing samples of the control channels in the following way: The X_b mass spectra of the testing samples are again fitted to extract s-weights. For the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ channel a two-dimensional fit is used again, while for $B_s^0 \to D_s^- \pi^+$ only the B_s^0 mass spectrum is fitted, as the charmless contribution was found to be negligible. The fitted spectra are shown in Fig. 6.7.

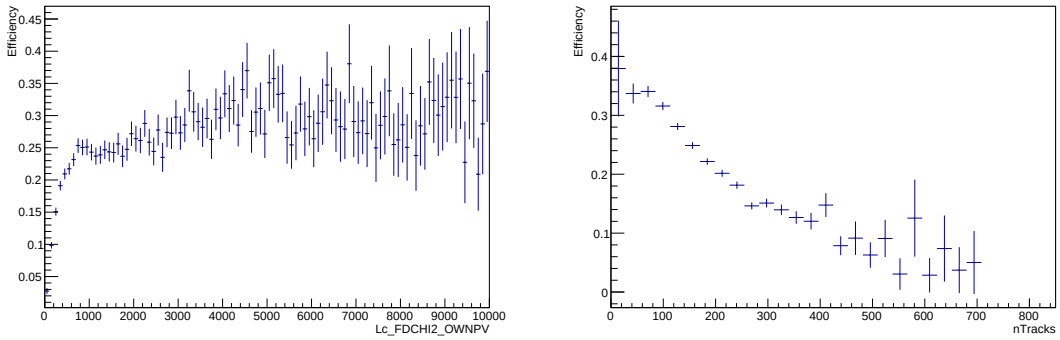


Figure 6.6: Λ_c^+ BDT efficiency as function of χ^2_{FD} of the Λ_c^+ and number of tracks of the event.

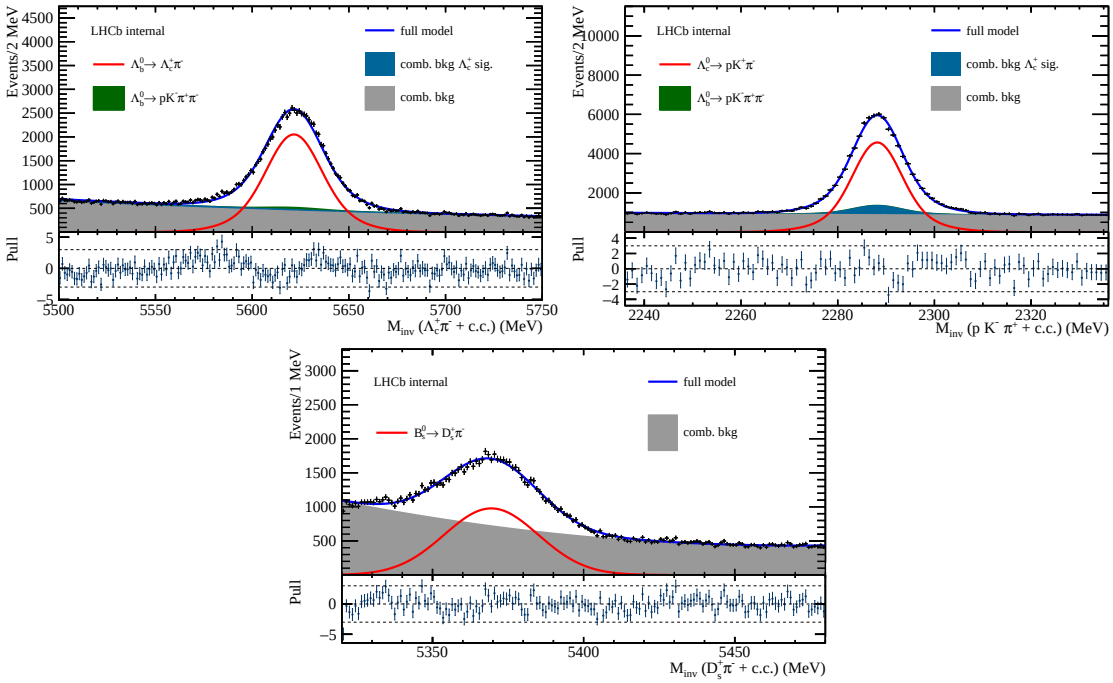


Figure 6.7: Fits to BDT training samples to extract s-weights. Top row: Projections of the two-dimensional fit to the $\Lambda_c^+\pi^-$ (left) and Λ_c^+ (right) mass distributions. Bottom row: One-dimensional fit to the $D_s^-\pi^+$ mass distribution.

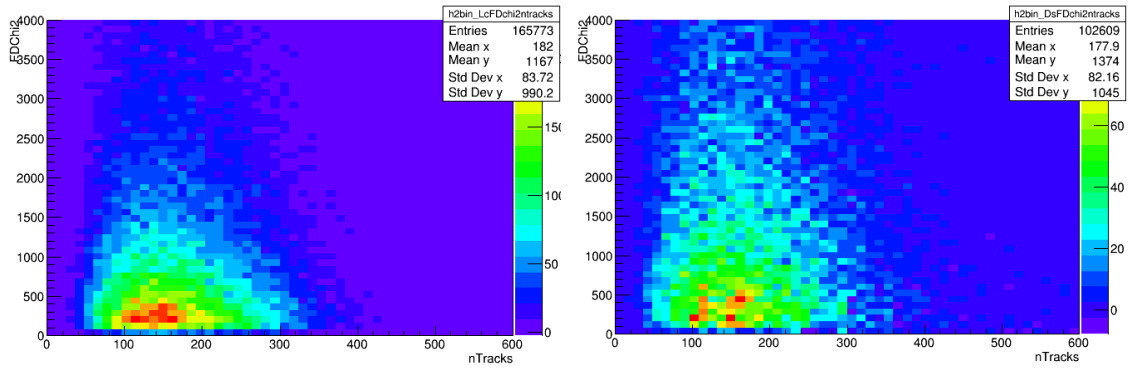


Figure 6.8: S-weighted χ^2_{FD} over $nTracks$ for $\Lambda_b^0 \to \Lambda_c^+\pi^-$ (left) and $B_s^0 \to D_s^-\pi^+$ (right).

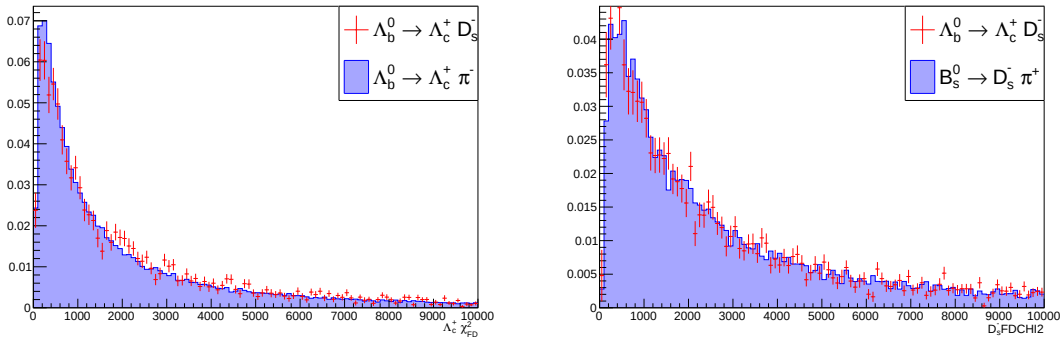


Figure 6.9: S-weighted χ^2_{FD} of the X_c for application and training channel data. Left: Comparison between the channels $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. Right: Comparison between the channels $B_s^0 \to D_s^-\pi^+$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$.

From that the s-weighted $nTracks$ over χ^2_{FD} distributions are generated, shown in Fig. 6.8. After this, the same samples are taken again with a requirement on the BDT output to be greater than a certain value. Now again s-weights are extracted in the same way as for the samples with no BDT output requirement and the s-weighted $nTracks$ over χ^2_{FD} distributions are created for this selection. The efficiency binned in $nTracks$ and χ^2_{FD} is then obtained by dividing these distributions. The repetition of the fit on the second sample is required as the BDT response is correlated to the discriminating variables (the particle masses). Therefore just using the s-weights from the first fit with a requirement on the BDT output would introduce an error.

It should be noted that for $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ the χ^2_{FD} distributions agree well with the training channels (see Fig. 6.9), so the effect of the binning in this case is expected to be small.

To check that the efficiency $\epsilon_{BDT}(\chi^2_{FD}, nTracks)$ obtained from the control channels can also be used to correct for the selection efficiency when the BDTs are applied to different channels, the clean signature of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ is used. A signal yield can already be extracted without further need to use PID information of any final state

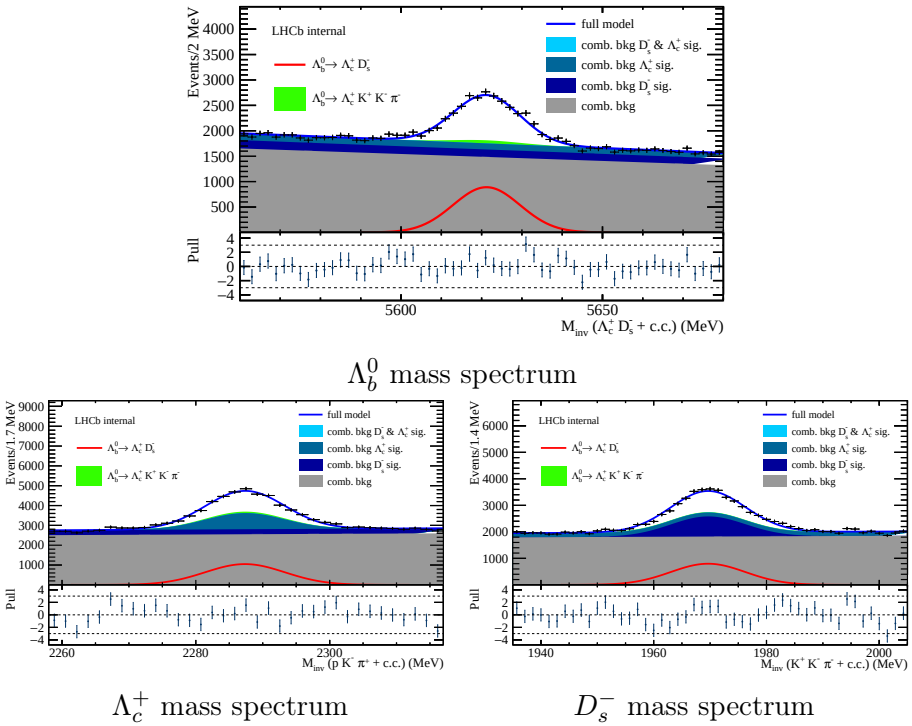


Figure 6.10: One-dimensional projections of the three-dimensional fit to extract the baseline yield. Top: $\Lambda_c^+ D_s^-$ mass distribution. Bottom: Λ_c^+ mass distribution (left) and D_s^- mass distribution (right).

particle for the selection. This allows for the definition of a baseline signal yield, to which yields from tighter selections can be compared, after they are corrected for the efficiency of the tighter selection. Fig. 6.10 shows the fit result used to extract the baseline yield.

As a first step the validity of each BDT is checked separately. The procedure is as follows: the BDTs trained on the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ and $B_s^0 \to D_s^- \pi^+$ control channels are applied to the Λ_c^+ and the D_s^- in $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. Then a selection on $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ using one of the two BDT outputs is made. The signal yield is extracted with the three-dimensional fit introduced in the previous chapter to account for charmless contributions. This yield is then corrected for the BDT selection efficiency as explained in section 5.1 using the efficiencies ϵ_{BDT,Λ_c} and ϵ_{BDT,D_s} from the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ and $B_s^0 \to D_s^- \pi^+$ control channels as explained above. The efficiency corrected yields from the selections using a single BDT can then be compared to the baseline yield from the selection using no BDT defined above. If it is valid to transplant the BDTs to $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and use the efficiencies obtained from the control channels, the efficiency corrected yields should be close to the baseline yield. This is checked for different selections using just the Λ_c^+ or D_s^- BDT. These selections are chosen to cover a large range of BDT selection efficiencies. The results are shown in Fig. 6.11 and Fig. 6.12. The figures show for both BDTs the efficiency corrected yield as function of the fraction of signal events surviving the BDT selection (which can be

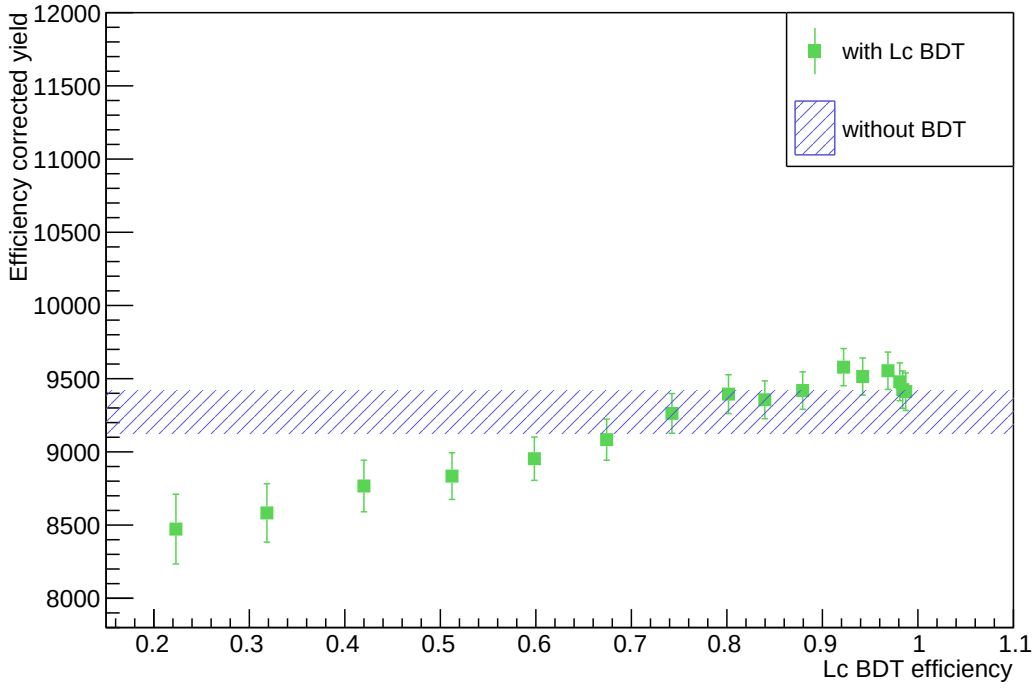


Figure 6.11: Efficiency corrected yield for different Λ_c^+ BDT efficiencies, shown as green points. The blue band shows the corrected yield of the selection using no BDT.

seen as 'true' single BDT efficiency). For large efficiencies the corrected yields are in agreement with the baseline yield. At lower efficiencies however the efficiencies are overestimated, resulting in an underestimated corrected yield compared to the baseline yield. This could be caused by the veto selection, where also PID information is used, and by the BDT output distributions being different for control and application channels. The efficiency could depend on further kinematic variables, which would require a multi-dimensional binning.

In conclusion selections using a single BDT produce reasonable efficiency corrected yields only for high BDT efficiencies. It is only valid to use the BDTs and efficiencies obtained from the control channels for other channels if the BDT efficiency is reasonably high. At low BDT efficiencies this breaks down and the BDTs can not longer be used without introducing systematic uncertainties. It should be noted that the optimized BDT selection found for this analysis in Section 6 is well within the region where the BDT efficiencies are valid.

Now that the single BDT efficiencies are shown to be only correct for high values, the next step is to check the use of both BDTs at once to make the final selection. The BDTs are trained on independent channels, but are applied simultaneously on $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. Correlations between Λ_c^+, D_s^- and their daughters could lead

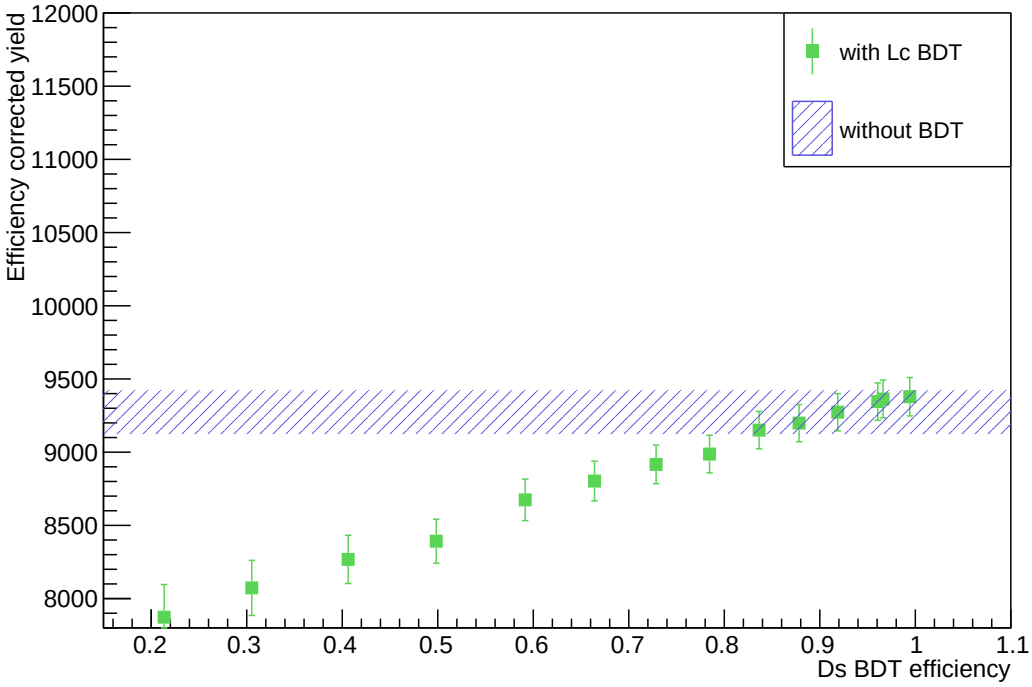


Figure 6.12: Efficiency corrected yield for different D_s^- BDT efficiencies, shown as green points. The blue band shows the corrected yield of the selection using no BDT.

to correlation between the two single BDT efficiencies resulting in non-factorizing efficiencies. This is checked in the next section.

6.5 Validation of Efficiency Factorization

The BDTs are trained on independent control channels, which are also used to obtain their efficiencies. In the previous section it was shown that using a single BDT leads to reasonable results in the channel $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ for high BDT efficiencies. In the final selection both BDTs are applied and the combined efficiency $\epsilon_{BDT,comb}$ of this selection has to be estimated. The simplest approach is to use the product of the single BDT efficiencies ϵ_{BDT,Λ_c} and ϵ_{BDT,D_s}:

$$\epsilon_{BDT,comb} = \epsilon_{BDT,\Lambda_c} \times \epsilon_{BDT,D_s}. \tag{6.1}$$

As discussed in Chapter 5 this combination only results in the correct combined efficiency if the efficiencies are not correlated. In this section the validity of this assumption is checked. The procedure is similar to the method used in the previous section. Instead of just using a single BDT in the selection, now several selections using both BDTs are made. The efficiency corrected yield is extracted in the same way as described in the previous section and compared to the same baseline yield. For non-factorizing efficiencies two effects should be visible: Combining correlated efficiencies in this way lead to an over-(under-) estimation of the combined efficiency if the correlation coefficient is negative (positive). This corresponds to an under-(over-) estimation of the corrected yield. Secondly the combined efficiency changes in a non-trivial way when the single efficiencies are varied (see eq. 5.4). This should be reflected in the efficiency corrected yield not only having an offset from the baseline yield due to the wrongly estimated total efficiency, but also showing a variation when the single efficiencies are varied.

Fig. 6.13 shows the comparison of the efficiency corrected yield of various looser and tighter selections using both BDTs with the baseline yield as a function of the fraction of signal events surviving the combined BDT selection (which can be seen as 'true' combined selection efficiency). The corrected yields only agree well with the baseline for all selection combinations with high single BDT efficiencies. For a constant D_s^- BDT requirement the corrected yield is constant and no trend beside the one already observed for the single BDTs is visible. The differences between the corrected yields can be explained by the deviations from the single BDT selections.

In conclusion the assumption that the single BDT efficiencies factorize is valid and the combining them to a total efficiency produces reasonable efficiency corrected yields for highly efficient single BDTs. As the efficiency correction is shown to break down already for single BDT selections at lower efficiencies, checking the factorization in this region would make no sense, as the BDTs can not be used there.

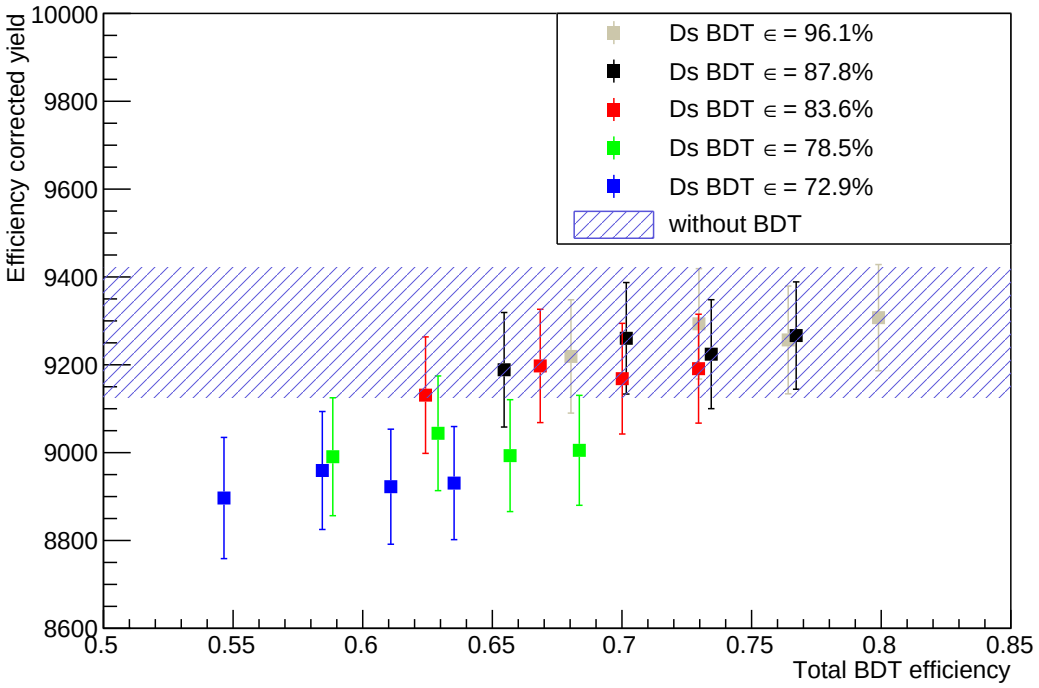


Figure 6.13: Efficiency corrected yield of $\Lambda_b^0 \to \Lambda_c^+ Ds^-$ candidates for different combinations BDT efficiencies. The points show the efficiency corrected yield at different BDT cuts. Points in the same color indicate a constant D_s^- BDT efficiency. The blue band shows the corrected yield of the selection using no BDT cuts.

6.6 Application to the Decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$

The overlying goal of this thesis is to verify the selection of open-charm charm particles using BDTs. In the previous sections the validity and efficiency factorization of the Λ_c^+ and D_s^- BDTs have been verified on data for high BDT selection efficiencies. In the case of the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ channel the assumption has to be made that the D^0 BDT efficiency is also valid at high efficiency and factorizes with the Λ_c^+ BDT. This assumption is reasonable as the D^0 BDT uses the same input variables and similar training parameters and sample selection as the D_s^- BDT. From the efficiency studies a systematic uncertainty can be estimated by comparing the corrected yields with the baseline.

$\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ is a three-body decay and the kinematics of the Λ_b^+ daughters could differ from the kinematics in the two-body decay training channels. Additional binning of the BDT efficiency in a kinematic variable (for example p_T of the charmed particle) should account for such differences. Another approach would be to reweight the $X_b \to X_c \pi$ samples to match those of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ before training the BDTs. Also, the BDT selection has to be highly efficient, as it is shown that the single BDTs are not valid for low efficiencies.

7 Systematic Uncertainties

In the following different sources of systematic uncertainties are evaluated:

MC sample size The finite size of the MC sample gives rise to a statistical uncertainty on all efficiencies estimated with it, which is accounted for by adding it as systematic uncertainty to the efficiency corrected yield.

Signal shape To estimate the uncertainty from the description of the signal shape the nominal fit is redone using a single Gaussian instead of a Voigtian function. The difference between the signal yields is added as systematic uncertainty.

Λ_b^0 mass range To check the effect of the Λ_b^0 mass range on the yield the signal fit is redone in the window 5550 MeV $< m_{\Lambda_b^0} <$ 5700 MeV. The difference to the nominal yield is added as uncertainty.

Charm hadron mass resolution The resolution of the open-charm hadrons could be different in data and MC. To account for this the mass spectra of both Λ_c^+ and D_s^- in data and MC are fitted with a single Gaussian. The difference between the standard deviation is then taken as uncertainty.

BDT efficiency As discussed in the previous chapter, the efficiency corrected yield for a certain BDT selection can be compared with a baseline yield. The difference between the corrected yield when applying a single BDT to this baseline is added as systematic uncertainty.

Table 7.1 summarizes the systematic uncertainties. With this the final corrected yield is

$$N_{\Lambda_b^0} = (4.5009 \pm 0.1044 \pm 0.2052) \cdot 10^6,$$

where the first uncertainty is statistical and the second systematic.

Source	Systematic Uncertainty (%)
MC sample size	1.74
Signal shape	3.60
Λ_b^0 mass range	0.67
Charm hadron mass resolution	0.85
Λ_c^+ BDT efficiency	1.63
D_s^- BDT efficiency	0.97
total	4.56

Table 7.1: Summary of systematic uncertainties.

8 Conclusion and Outlook

The efficiency corrected yield of the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$, which will serve as reference value for the branching fraction measurement of the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$, was determined using data from the LHCb experiment recorded in 2011 and 2012 corresponding to an integrated luminosity of 3 fb$^-$:

$$N_{\Lambda_b^0} = (4.5009 \pm 0.1044(stat.) \pm 0.2052(syst.)) \cdot 10^6.$$

In the overlying analysis the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ will be observed for the first time. After that, a full amplitude analysis of this decay will be done to look for pentaquark candidates in the $\Lambda_c^+ \overline{D}^0$ system.

Combinatorial backgrounds are large for this decay because it is reconstructed in a final state with six hadrons. To reduce this background the usage of PID information of the final state particles is essential. At LHCb PID information is provided on a per-track basis. In this thesis another approach was tested, which is to combine the PID information of the three open-charm daughters to a single variable with a BDT. This provides a PID-like variable for the charmed hadrons. For both methods the efficiencies are obtained from independent control channels. Whether these efficiencies can be used to correct selections of other channels has been tested using the clean signature of the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay, where no PID information is needed to extract a signal yield, which is used as a baseline yield. Various selections with and without using PID variables have been made, the signal yield extracted and corrected for the efficiency obtained from the control channels. For both methods large deviations from the baseline yield have been seen when the PID selection had low efficiency. It has to be further tested what causes this. In the vetoes against background from mis-assigned particles already PID information was used. It is possible that the efficiency of the vetoes is correlated with the efficiency of the additional PID selection. Also possible is that the there are additional backgrounds from mis-assignment in the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$, which have not been identified, distorting the efficiency correction.

Additional checks could be done by finding further channels with the same charmed hadrons and a similar clean signature. Especially also checking the D^0 BDT could help to understand the observed problems.

When applying the BDTs to $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ it should be checked how the BDT responses look compared to the training channels. If they look similar, it can be assumed that the efficiency correction behaves like in the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ case. This means that the efficiencies of the BDTs are only valid at high values.

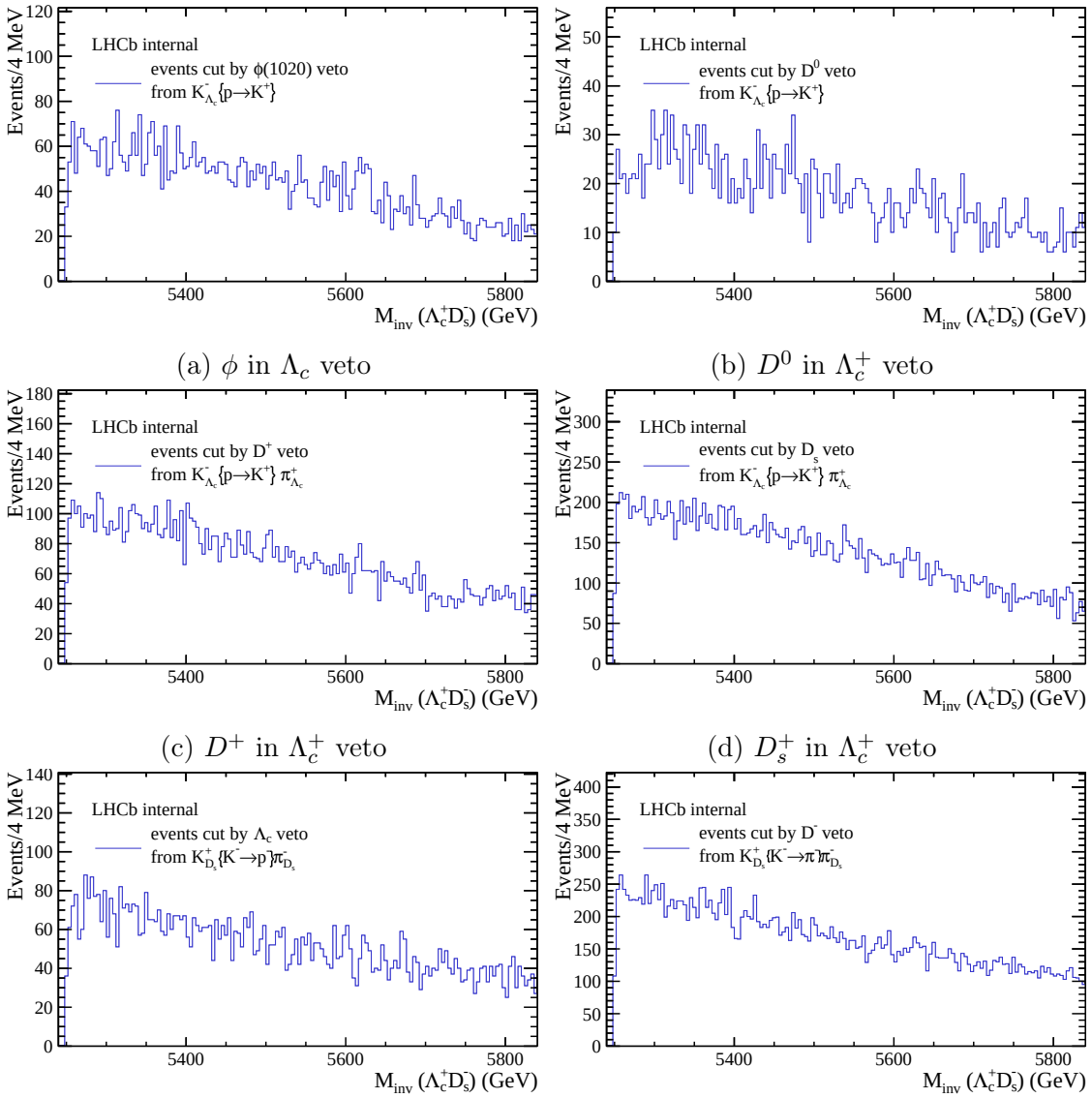
While similar BDTs have already been used at LHCb, here they have been recalibrated for a newer stripping version. It has been shown that the method does

not work for all BDT efficiencies.

Part I

Appendix

A Veto Performance



(a) ϕ in Λ_c veto

(b) D^0 in Λ_c^+ veto

(c) D^+ in Λ_c^+ veto

(d) D_s^+ in Λ_c^+ veto

(e) Λ_c^- in D_s^- veto

(f) D^- in D_s^- veto

Figure A.1: Veto performance plots. The plots show the events in the $\Lambda_c^+ D_s^-$ mass spectrum, which are rejected by the vetoes.

B BDT Comparison plots

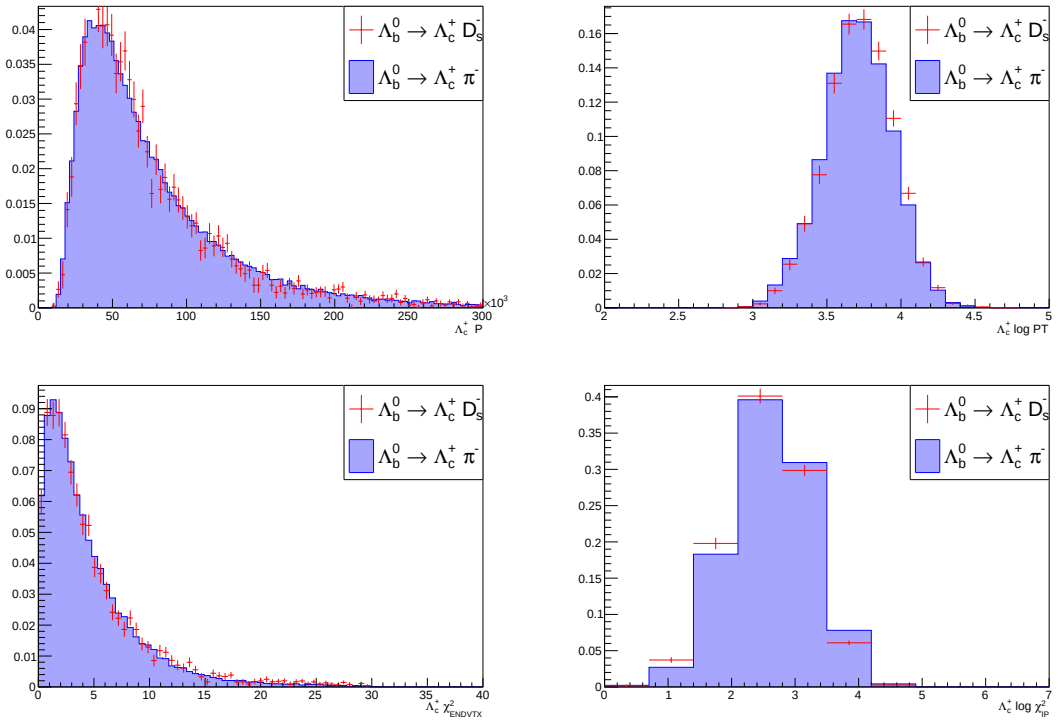


Figure B.1: Comparison of the Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

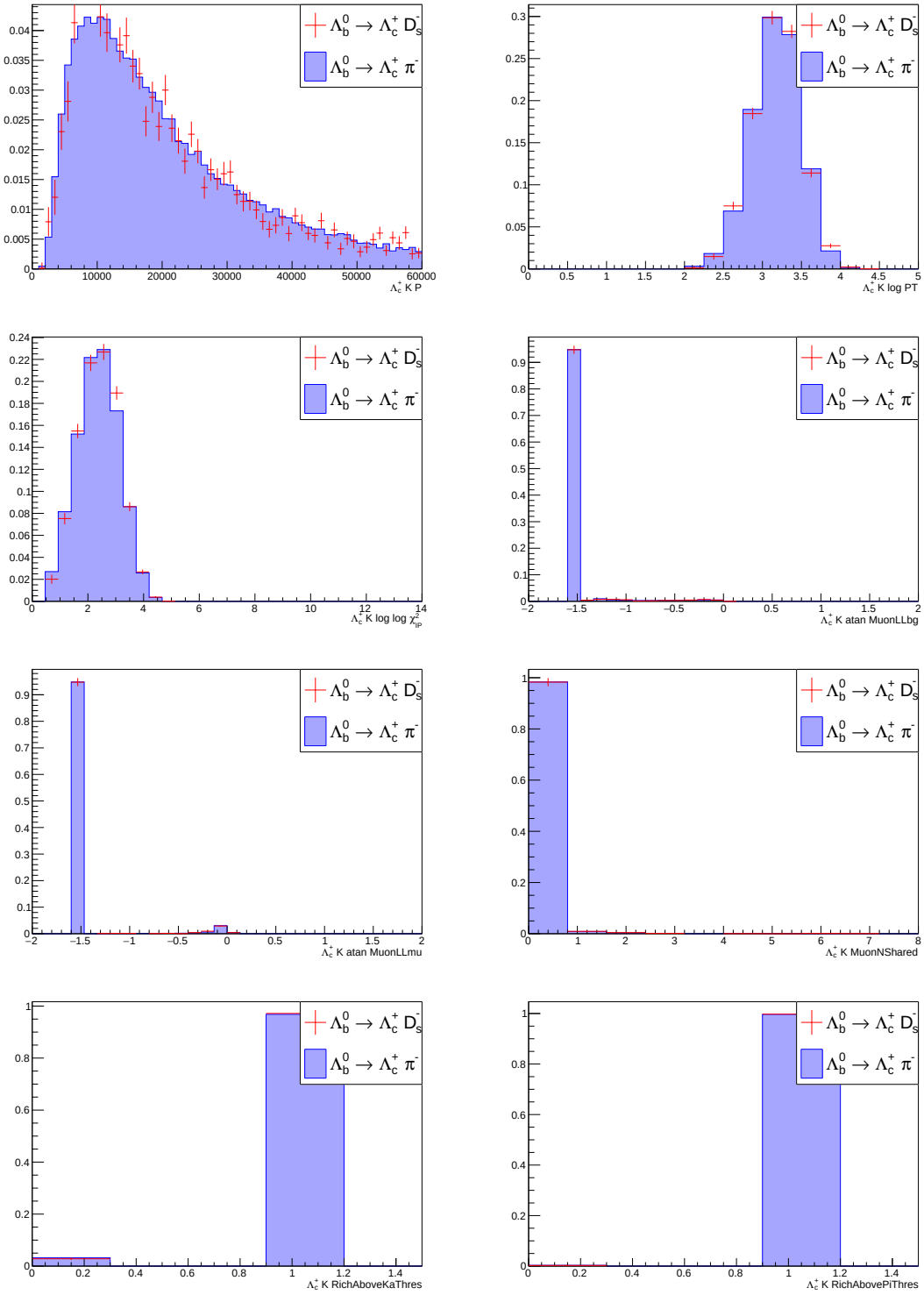


Figure B.2: Comparison of the K^- from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

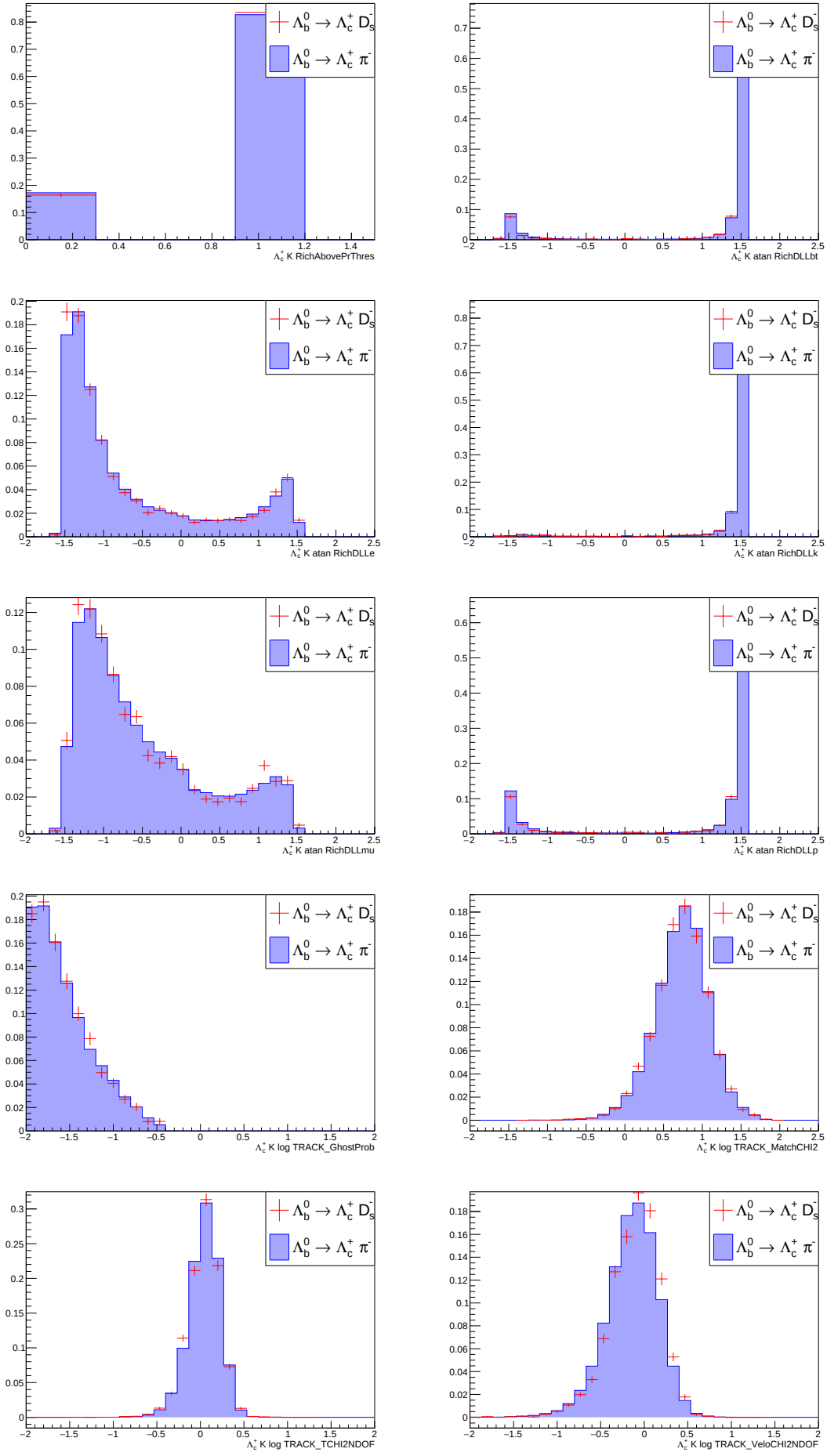


Figure B.3: Comparison of the K^- from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

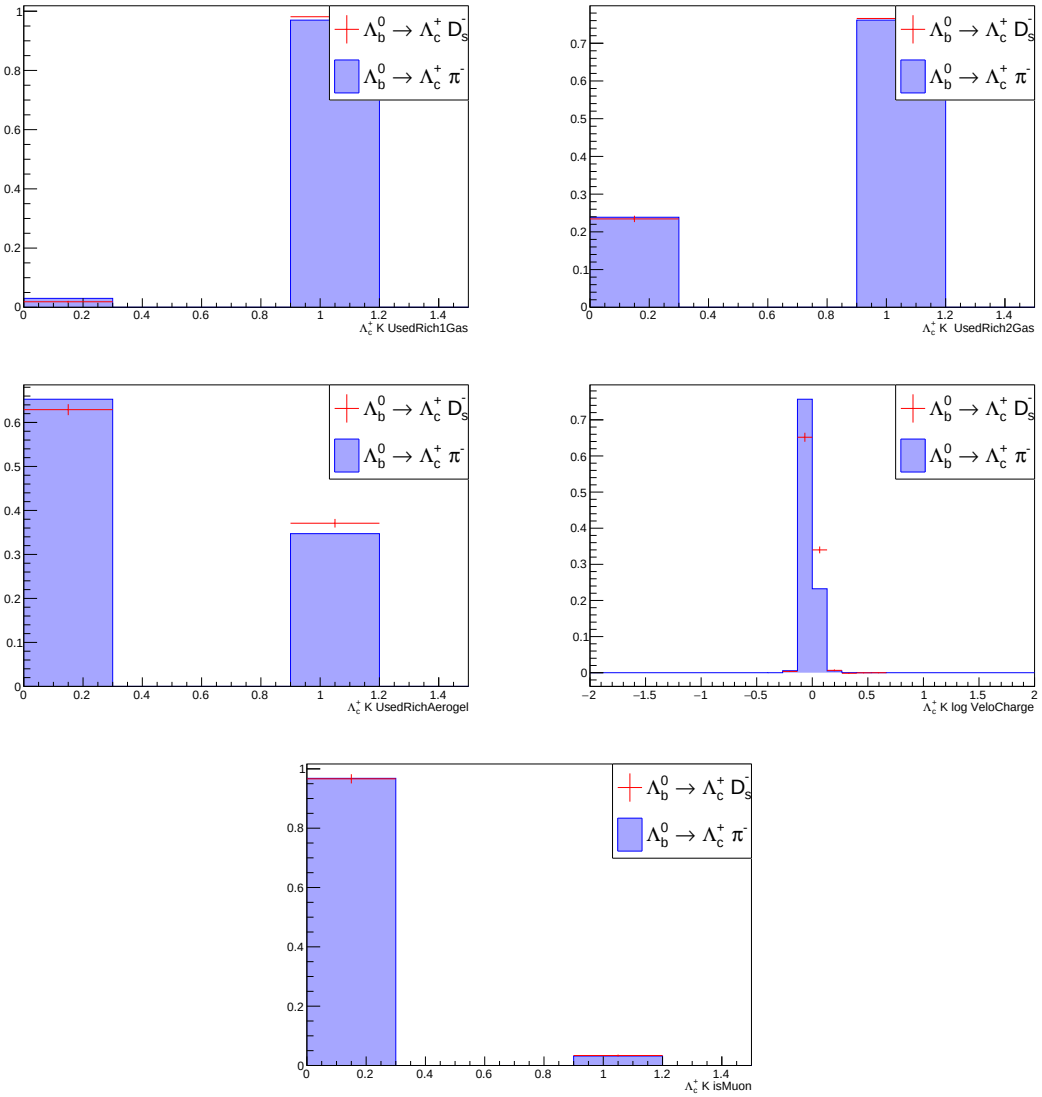


Figure B.4: Comparison of the K^- from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

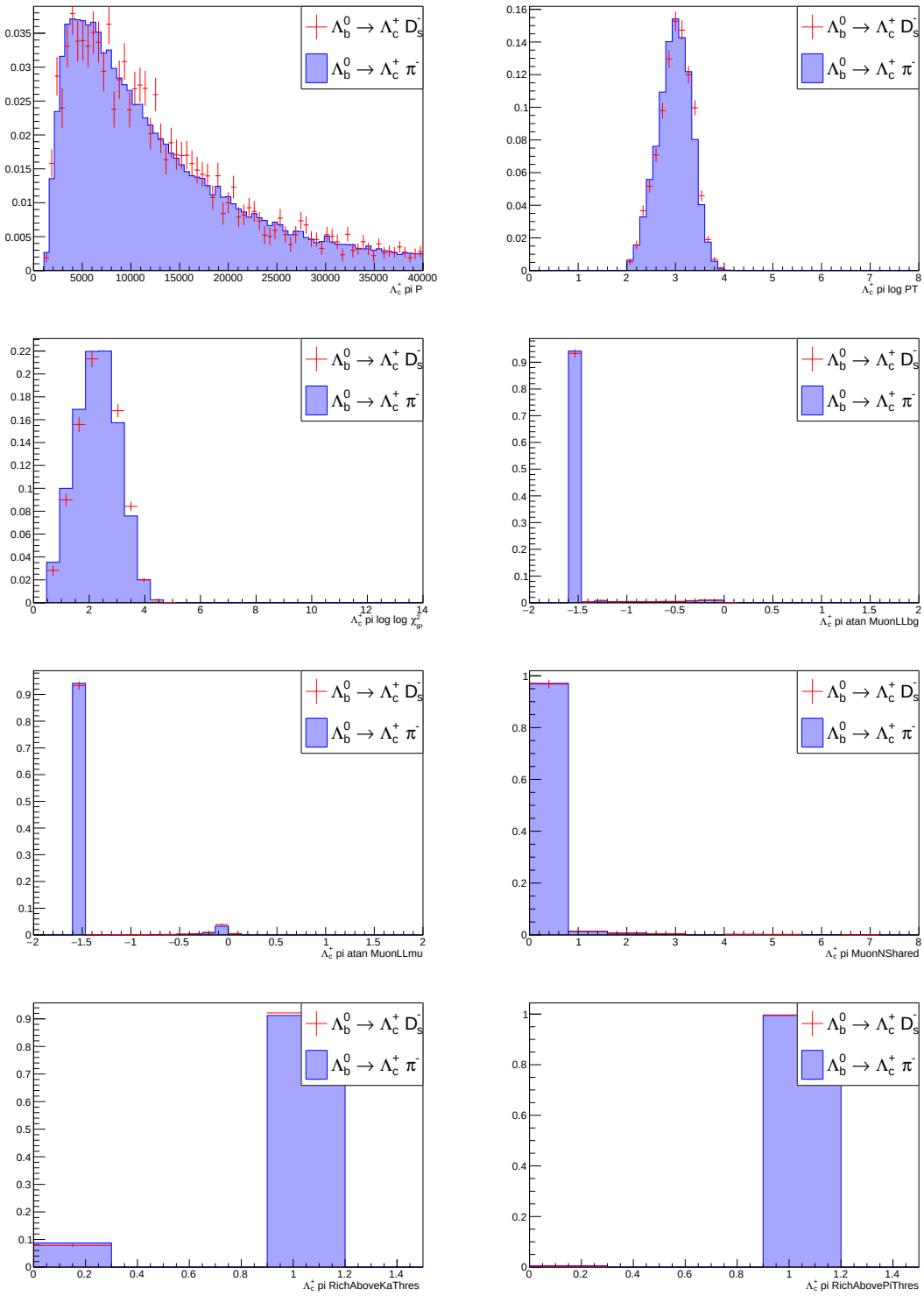


Figure B.5: Comparison of the $\pi+$ from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

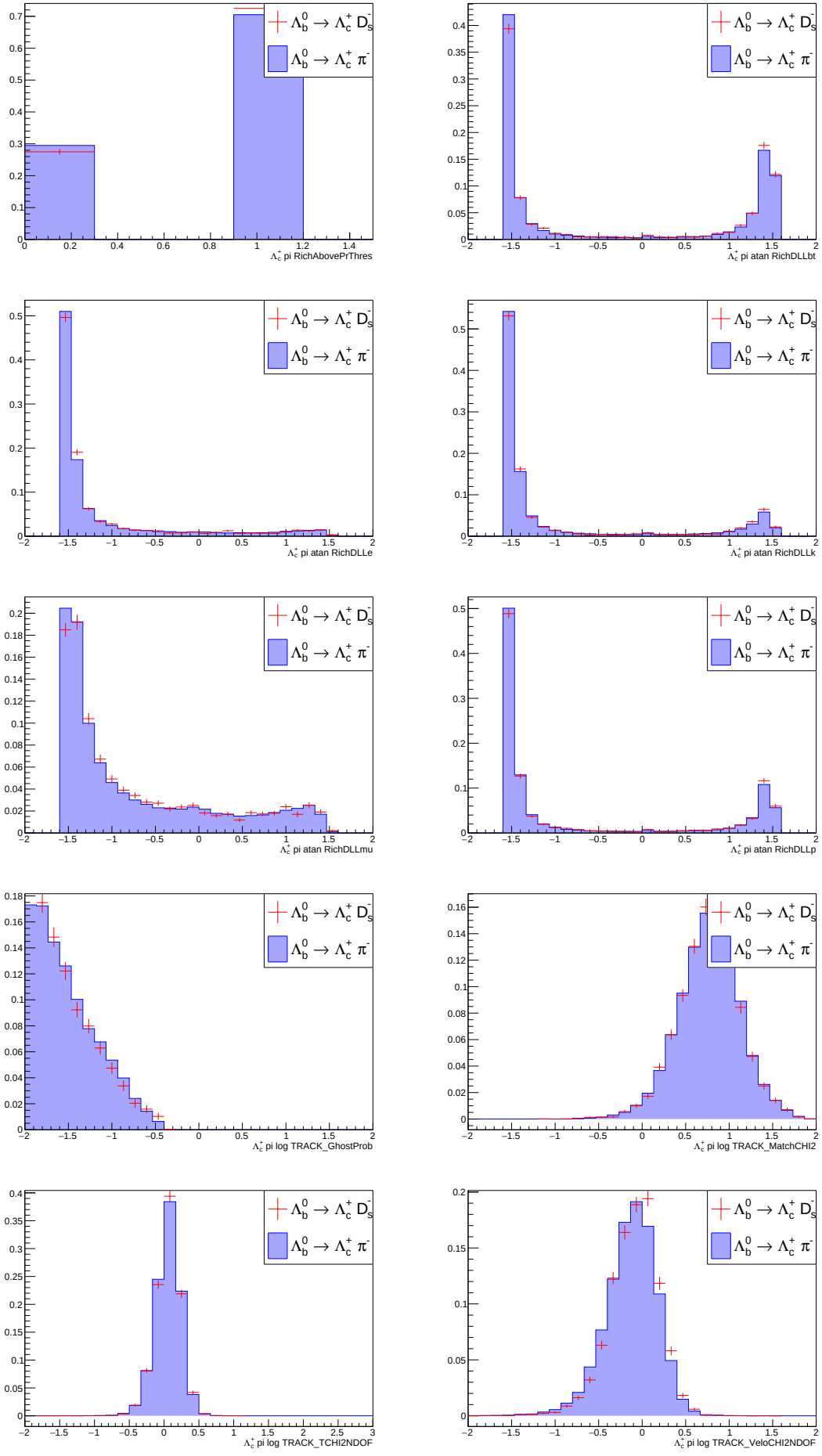


Figure B.6: Comparison of the $\pi+$ from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

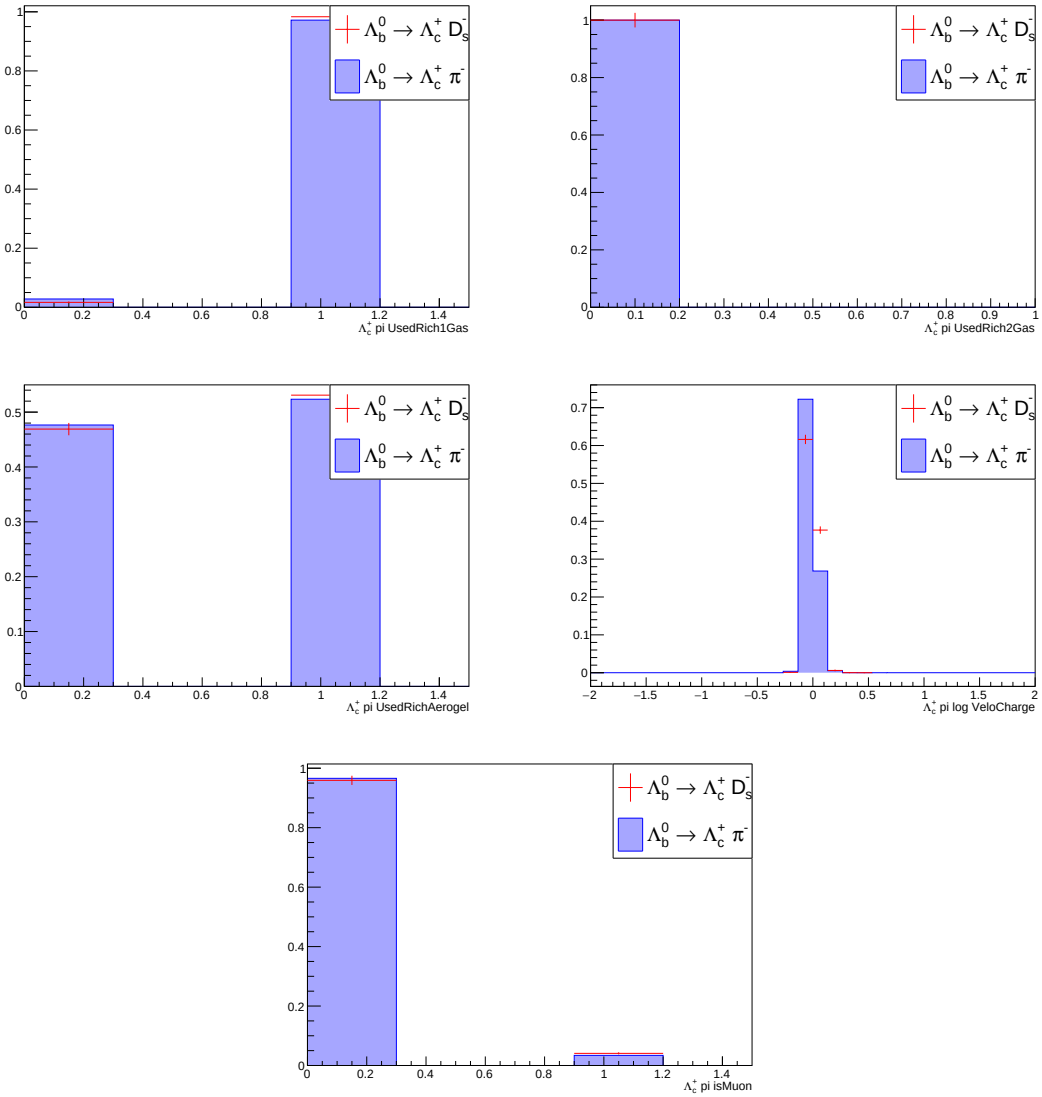


Figure B.7: Comparison of the $\pi+$ from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

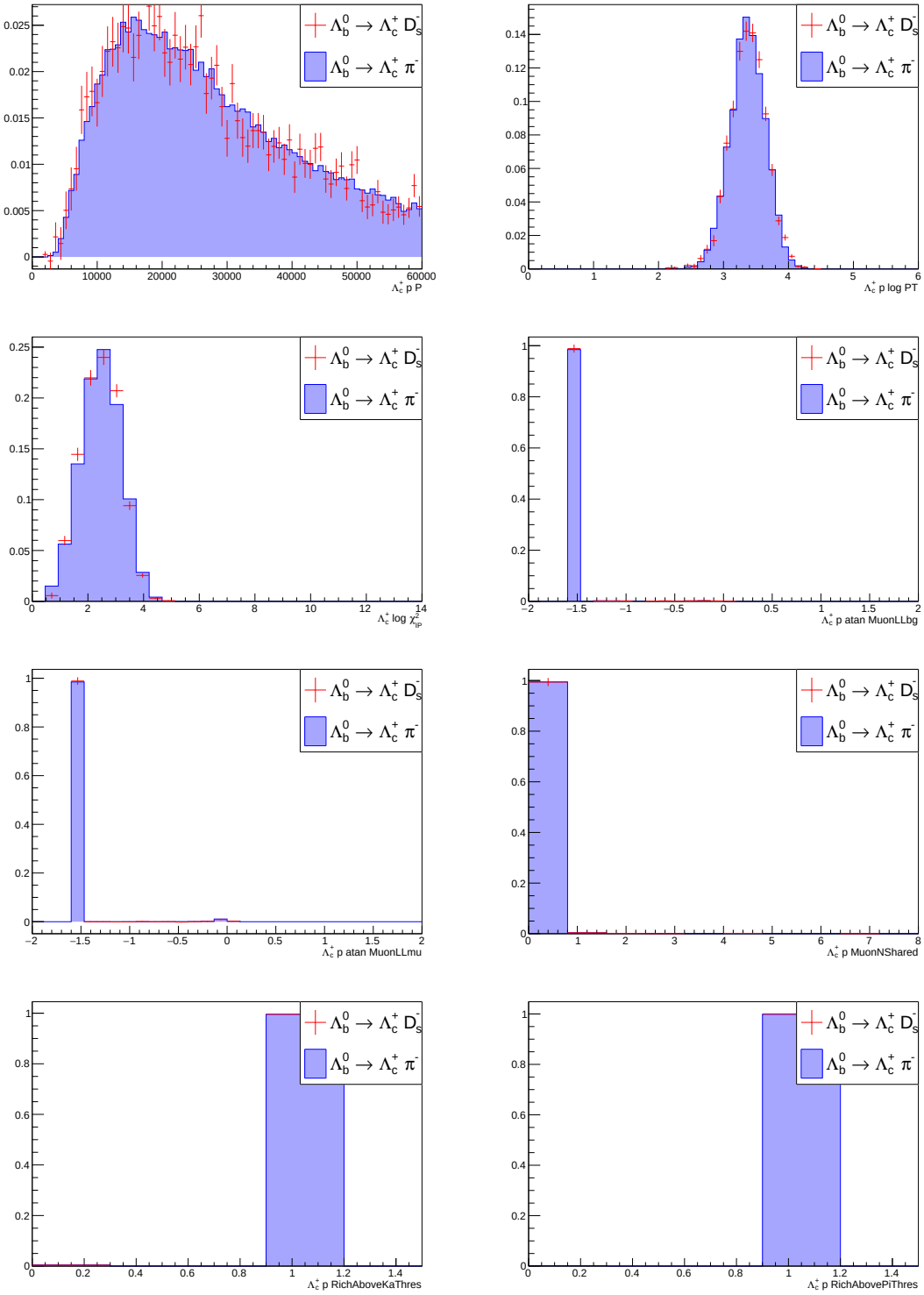


Figure B.8: Comparison of the p from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

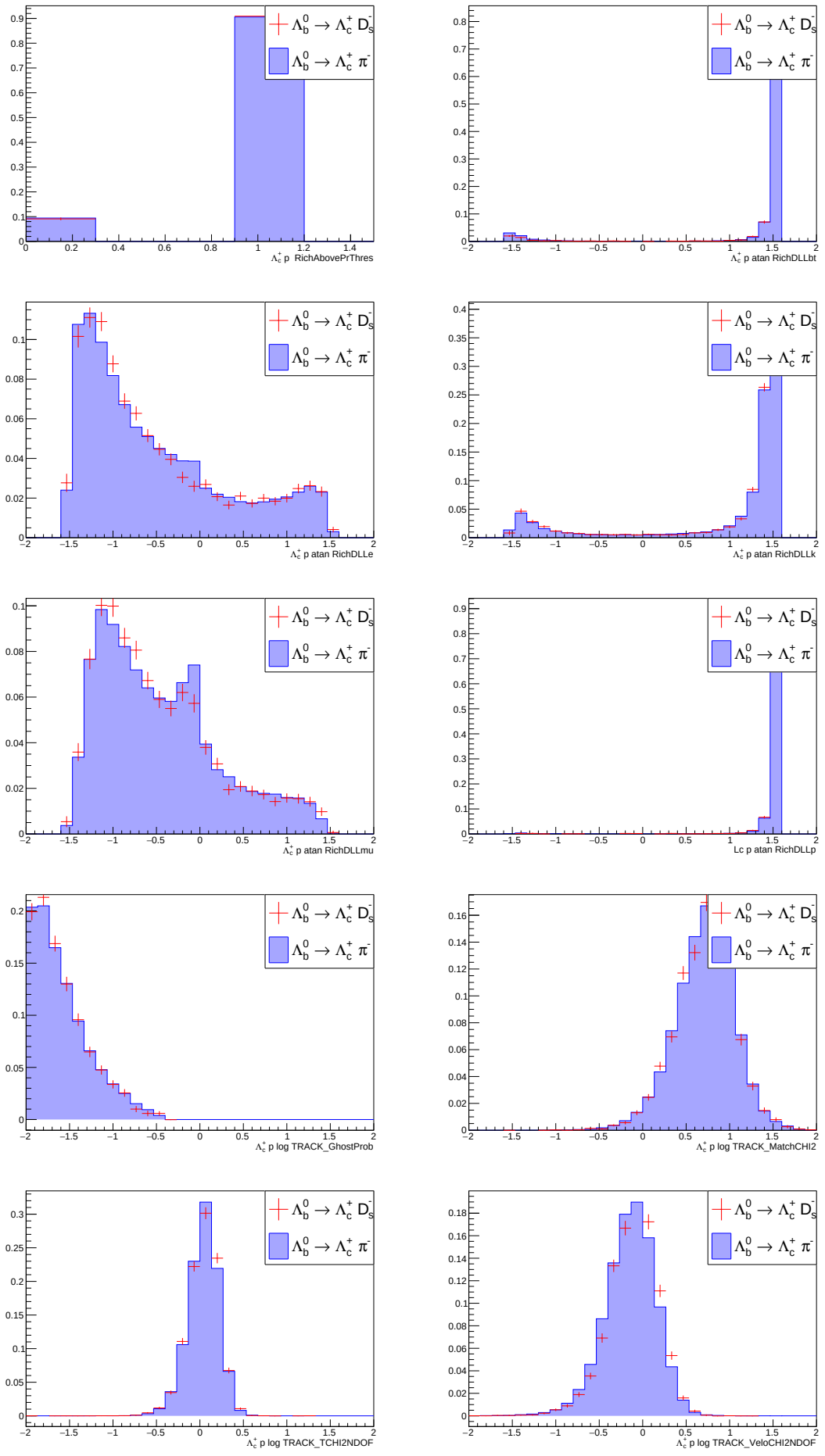


Figure B.9: Comparison of the p from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

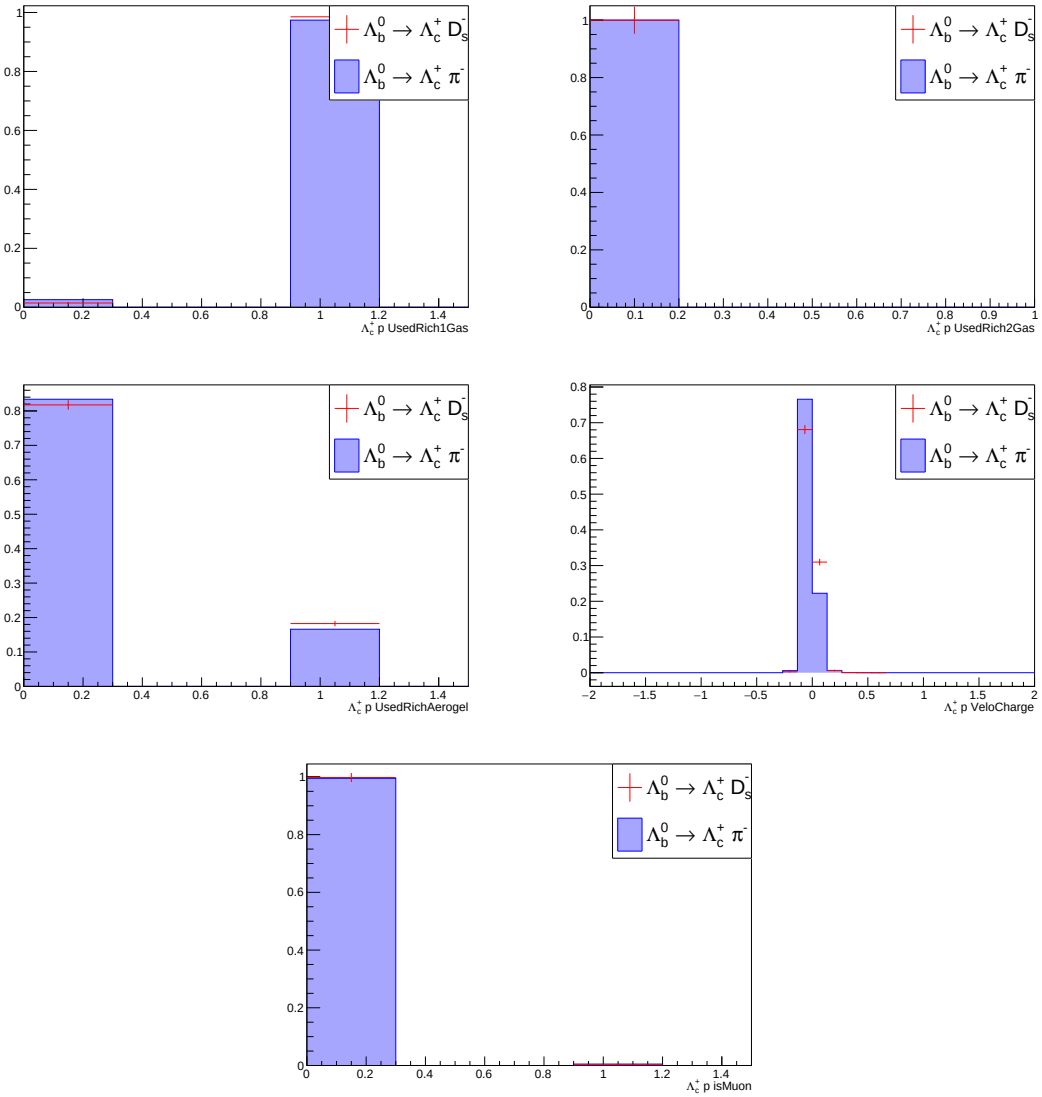


Figure B.10: Comparison of the p from Λ_c^+ input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\Lambda_b^0 \to \Lambda_c^+ \pi^-$.

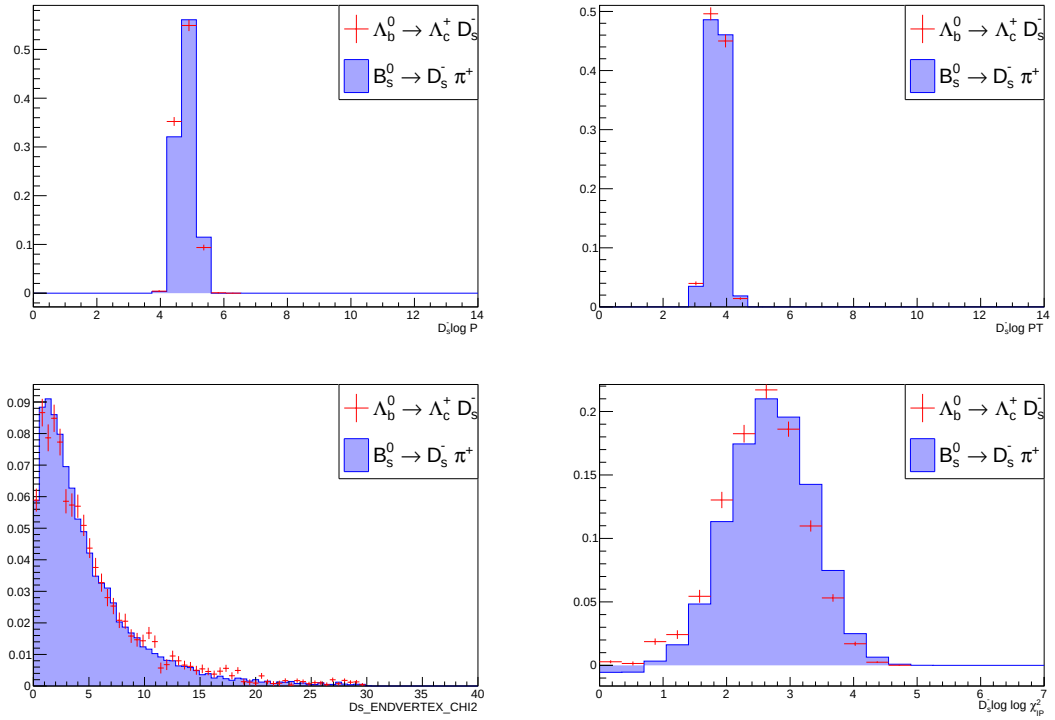


Figure B.11: Comparison of the D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

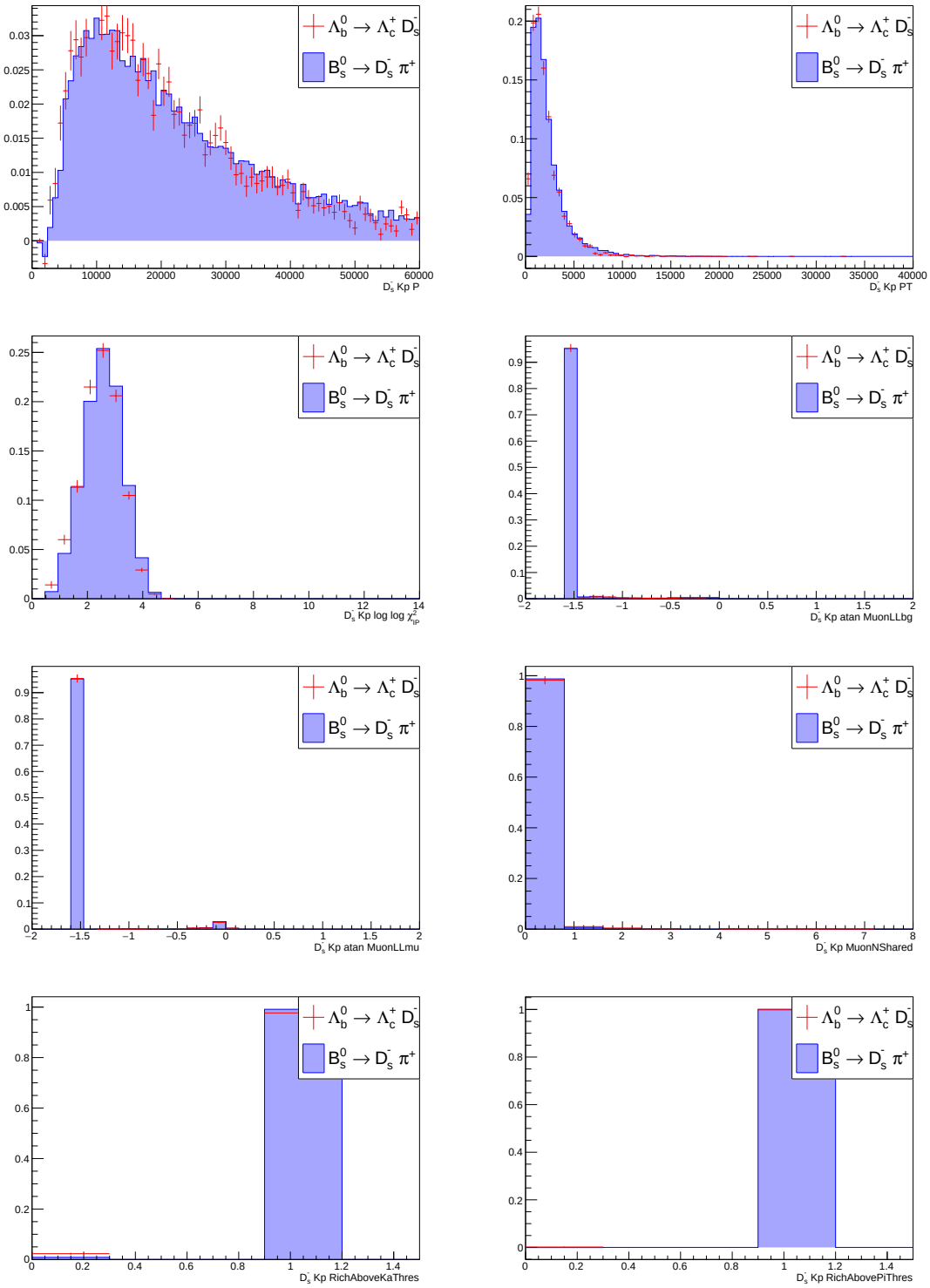


Figure B.12: Comparison of the K^+ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

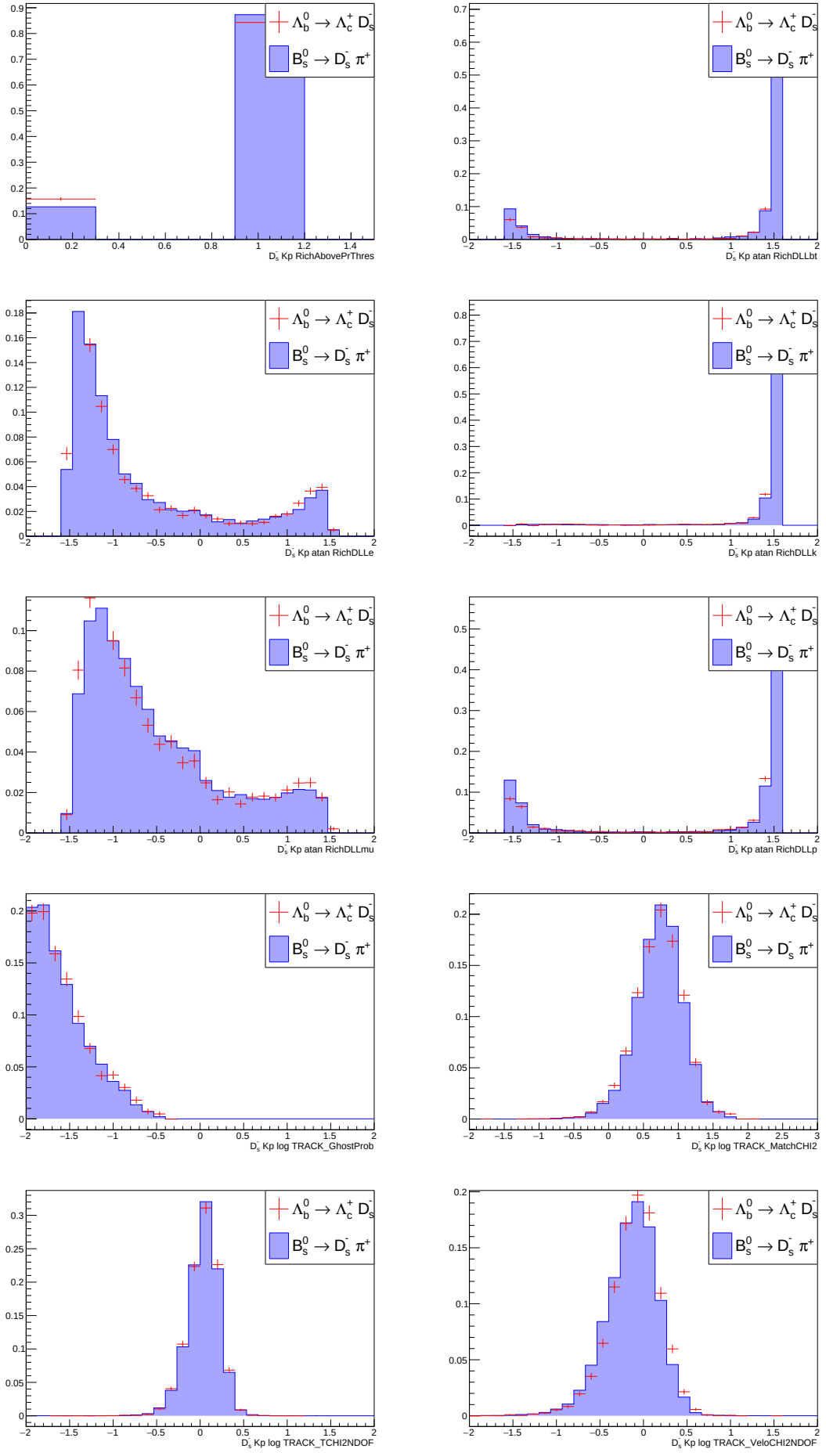


Figure B.13: Comparison of the K^+ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

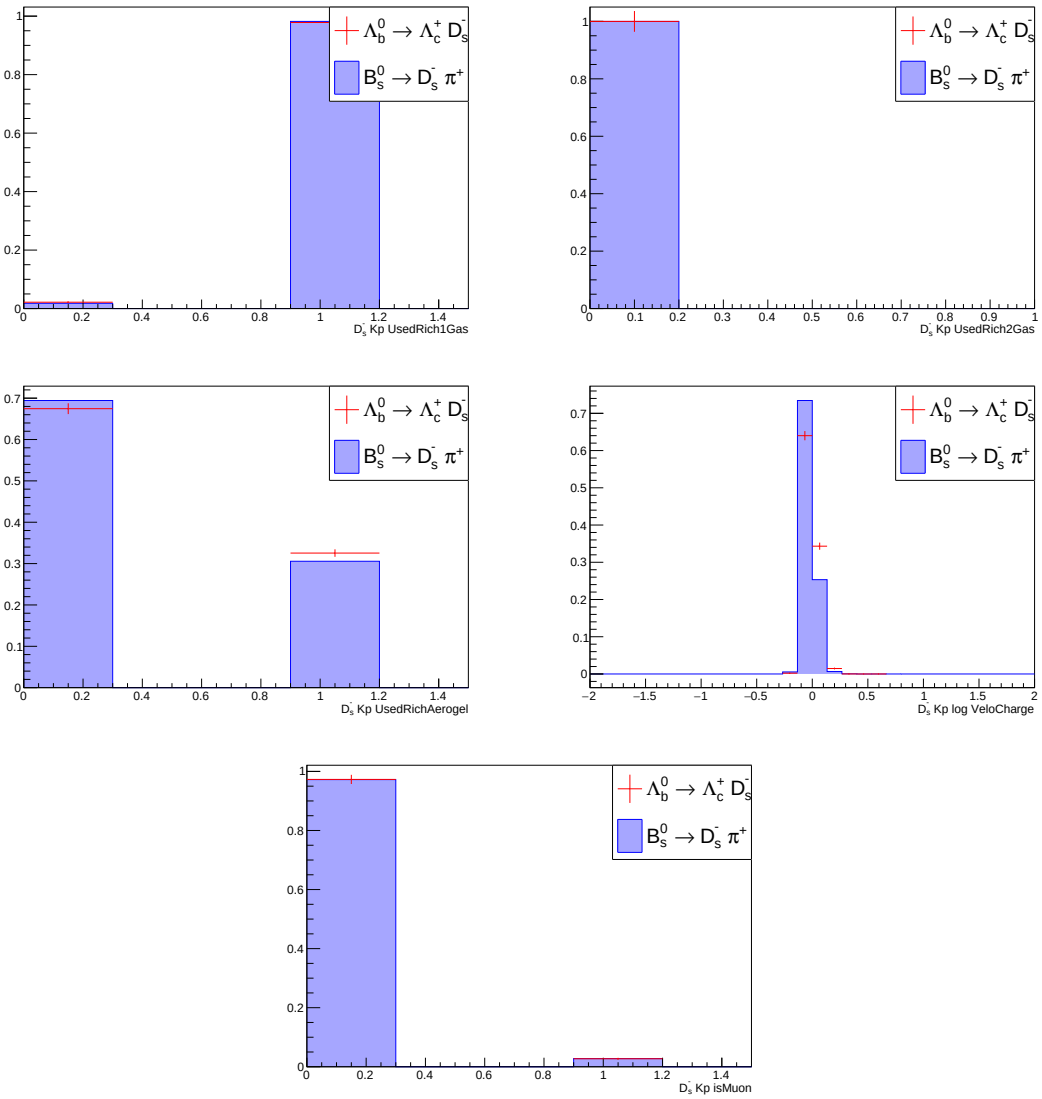


Figure B.14: Comparison of the K^+ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

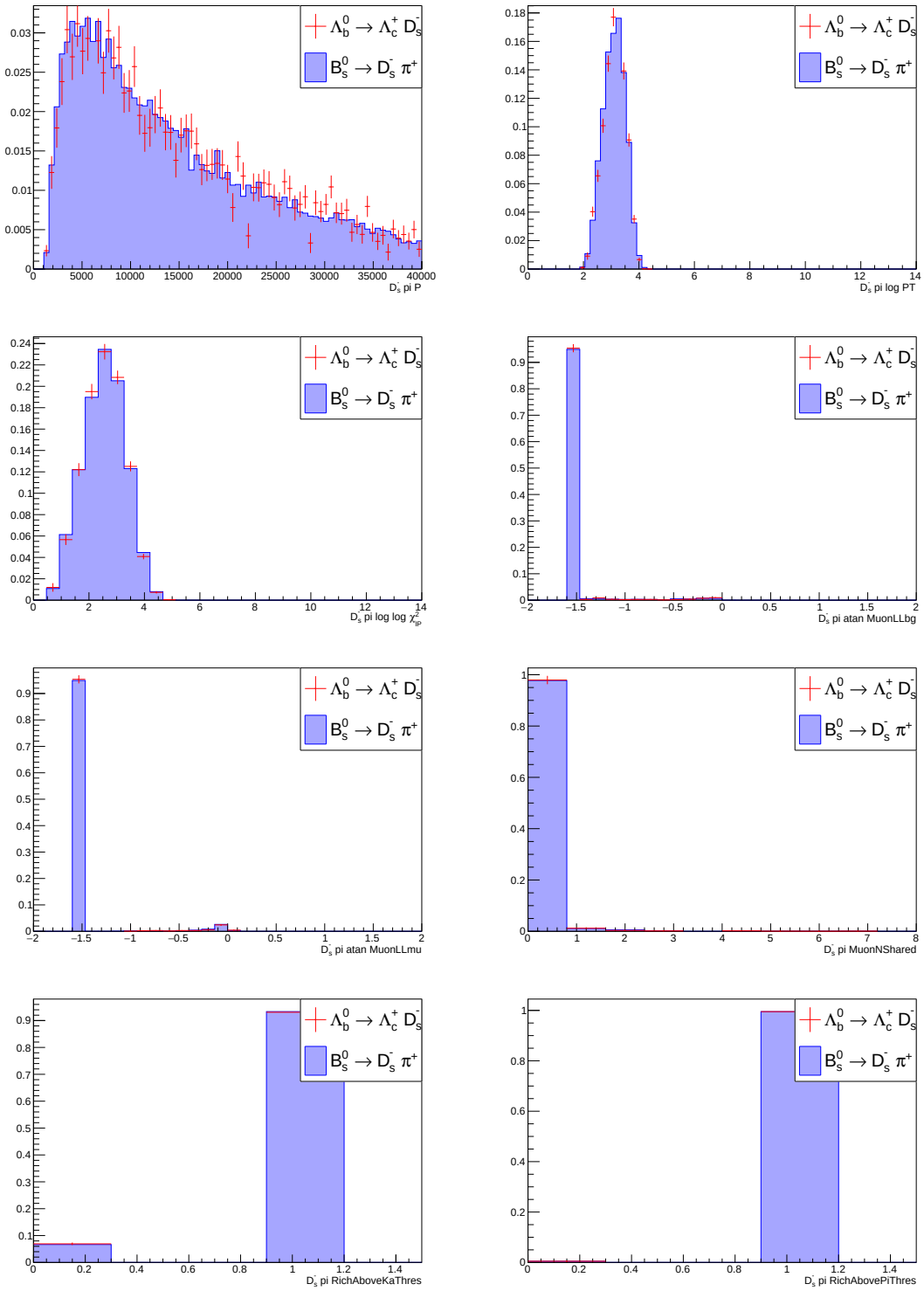


Figure B.15: Comparison of the $\pi-$ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

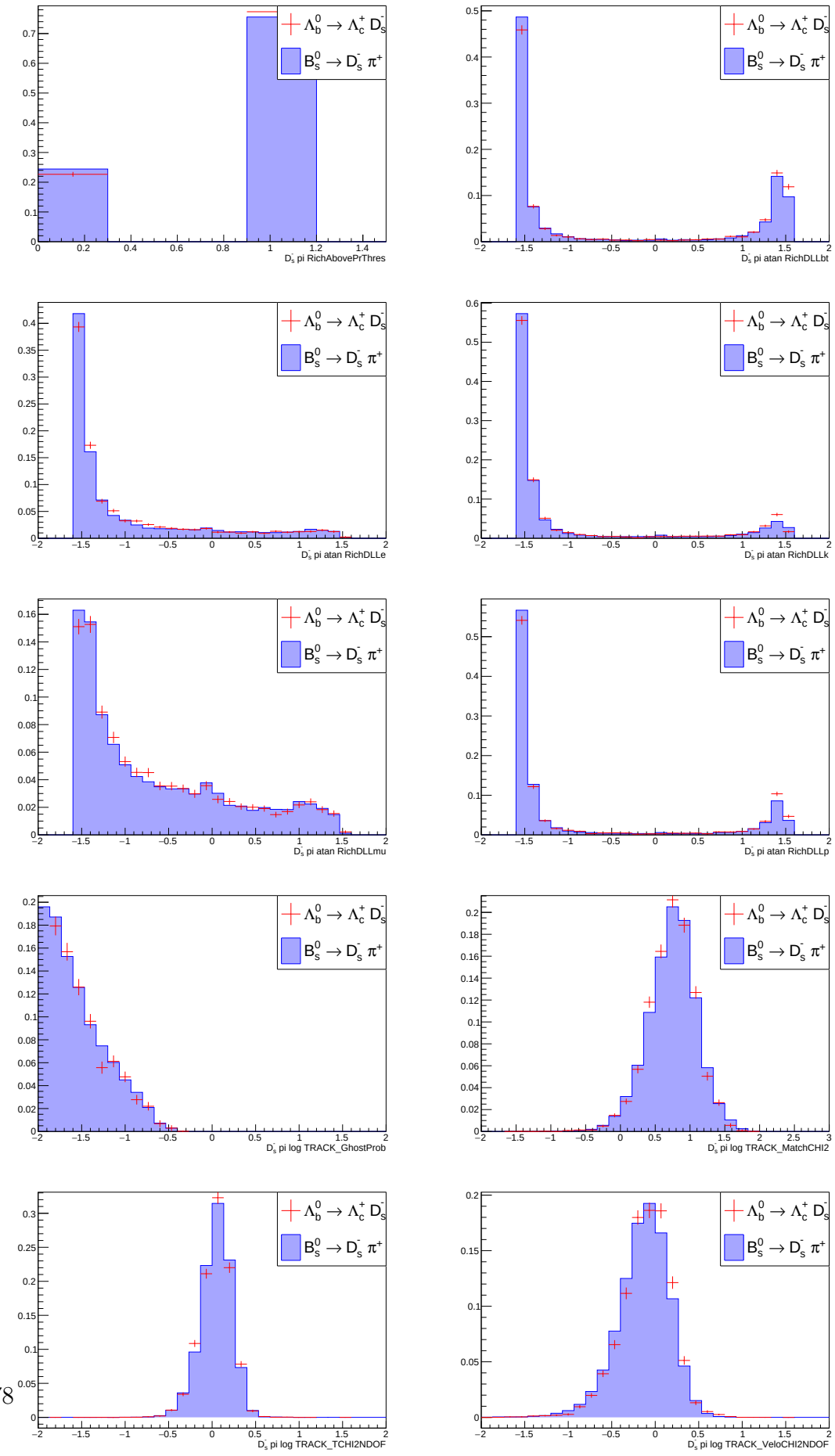


Figure B.16: Comparison of the $\pi-$ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

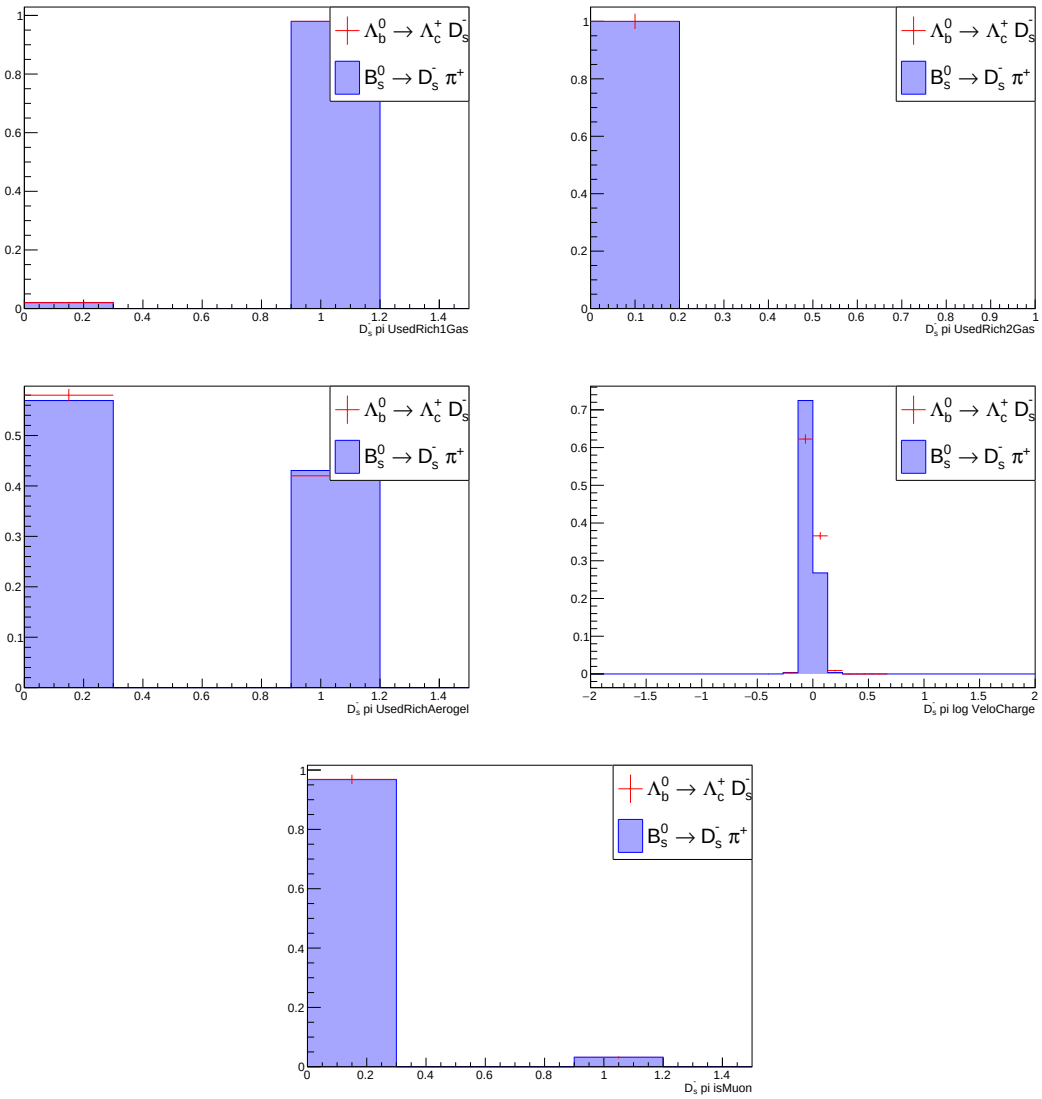


Figure B.17: Comparison of the $\pi-$ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

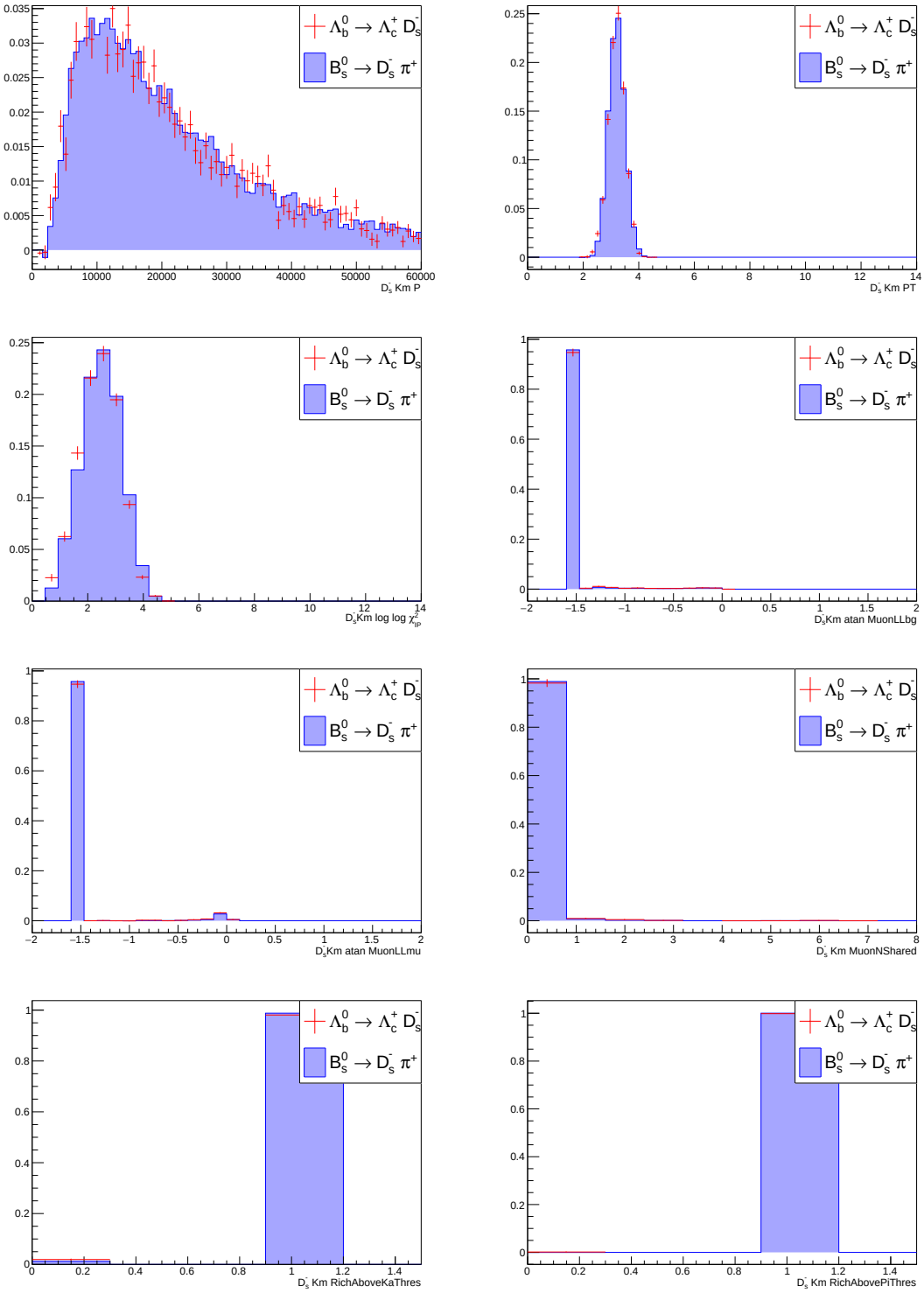


Figure B.18: Comparison of the $K-$ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

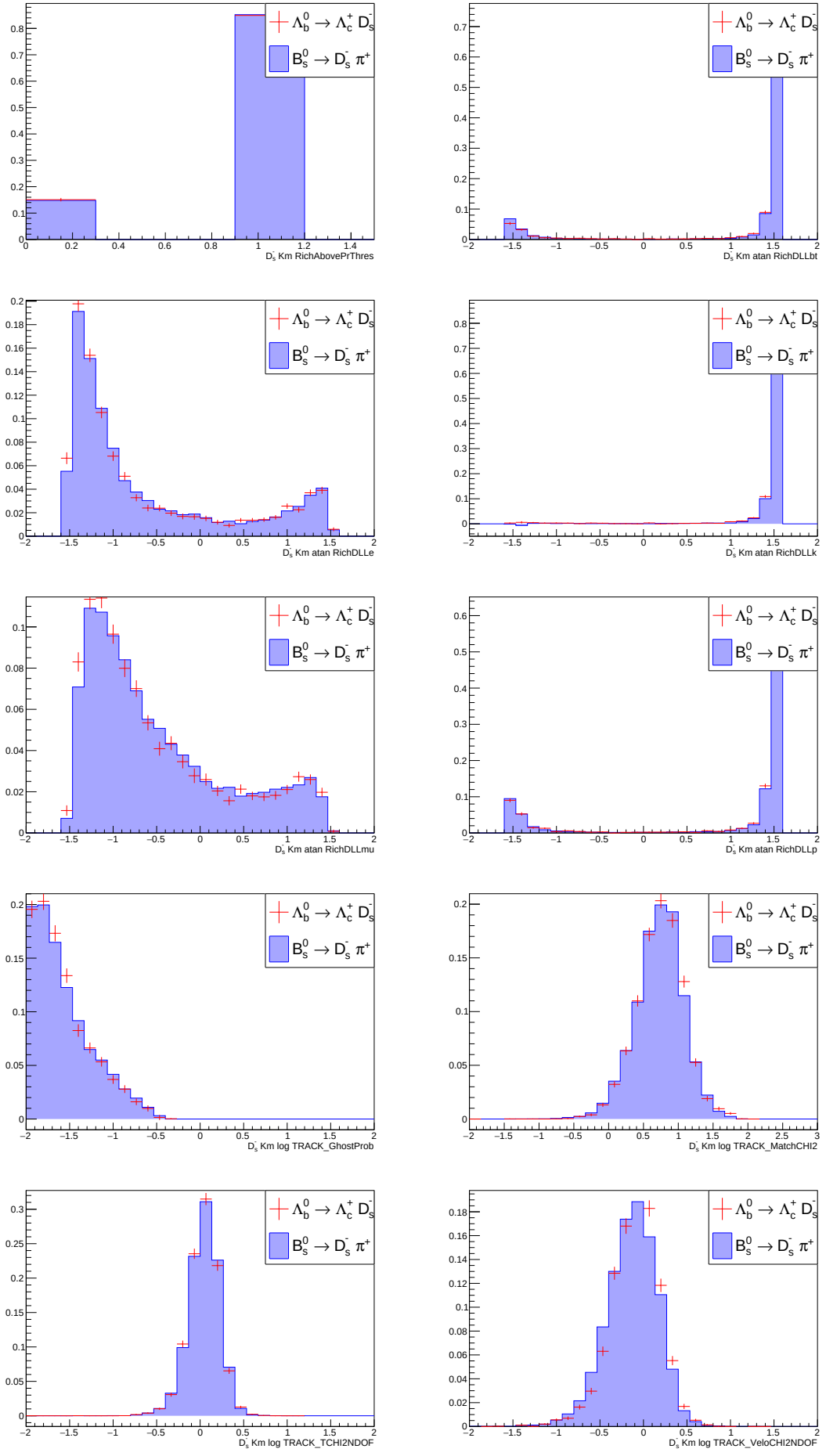


Figure B.19: Comparison of the $K-$ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

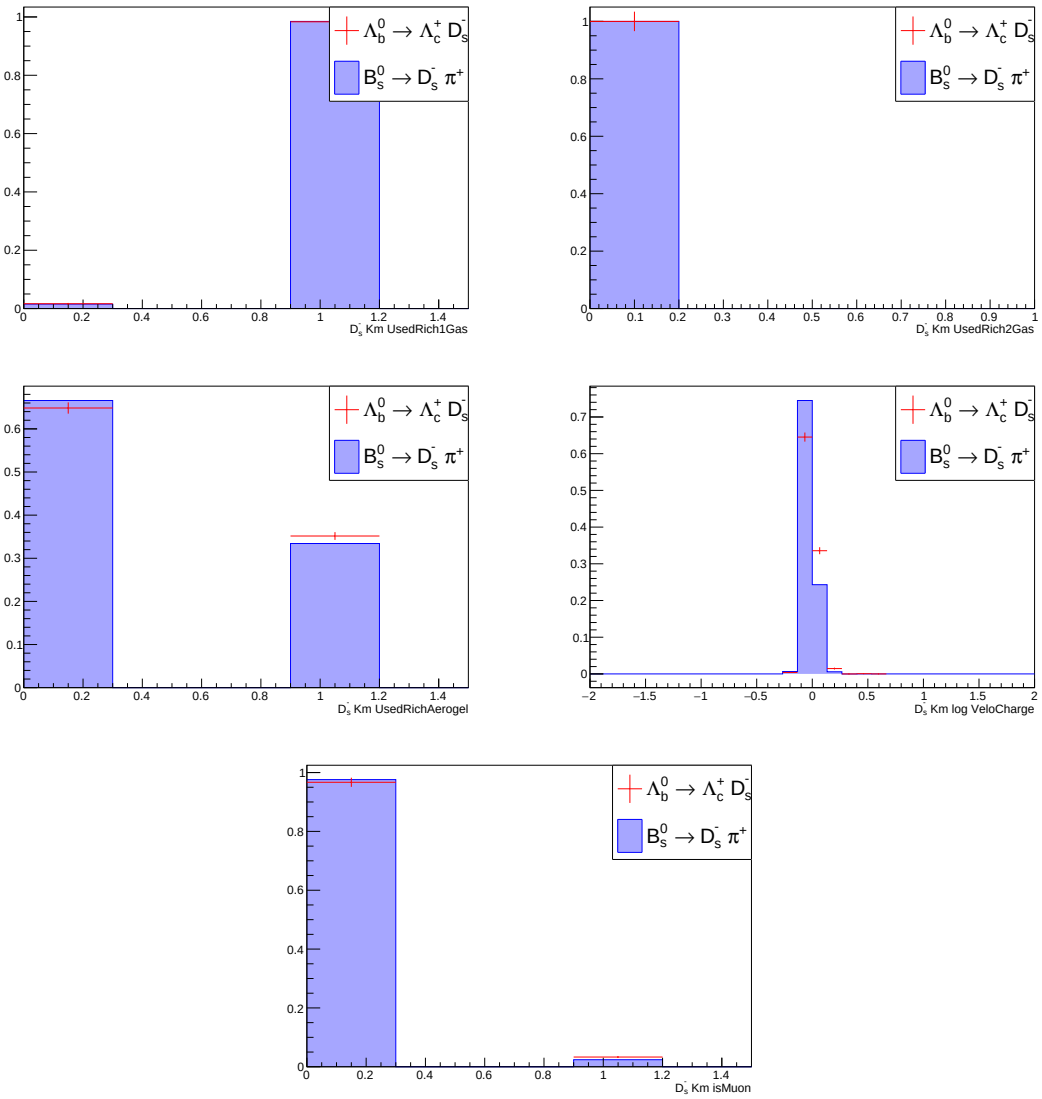


Figure B.20: Comparison of the $K-$ from D_s^- input variables between the channels $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and $\overline{B}_s^0 \to D_s^- \pi^+$.

C Bibliography

[1] The LHCb Collaboration. Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays. *Phys. Rev. Lett.*, 115:072001, Aug 2015. doi: 10.1103/PhysRevLett.115.072001. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.072001.

[2] C. Patrignani et al. (Particle Data Group). The Review of Particle Physics (2016). *Chin. Phys. C*, 40, 100001, 2016.

[3] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.

[4] Dsperlich. Standard Model of Elementary Particles. URL https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[5] M. Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8: 214–215, 1964. doi: 10.1016/S0031-9163(64)92001-3.

[6] G. Zweig. An SU_3 model for strong interaction symmetry and its breaking. 1964.

[7] A. Esposito, A. Pilloni, A. D. Polosa. Multiquark Resonances. doi: 1611.07920v2.

[8] Kenneth H. Hicks. On the conundrum of the pentaquark. *Eur. Phys. J.*, H37: 1–31, 2012. doi: 10.1140/epjh/e2012-20032-0.

[9] Hua-Xing Chen, Wei Chen, Xiang Liu, Shi-Lin Zhu. The hidden-charm pentaquark and tetraquark states. *Phys. Rept.*, 639:1–121, 2016. doi: 10.1016/j.physrep.2016.05.004.

[10] A generic pentaquark, . URL https://en.wikipedia.org/wiki/Pentaquark.

[11] Fengkun Guo Bongsong Zou Yonghui Lin, Chaowei Shen. Decay behaviors of Pc hadronic molecules. 2017.

[12] The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3(08):S08005, 2008.

[13] LHCb VELO Group. Performance of the LHCb Vertex Locator. *JINST*, 9: 09007, 2014. doi: 10.1088/1748-0221/9/09/P09007.

[14] R. Lindner. LHCblayout2.LHCbschema2,, LHCb collection. URL https://cds.cern.ch/record/1087860.

[15] LHCb Silicon Tracker, . URL http://lhcb.physik.uzh.ch/ST/public/material/index.php.

[16] F.M. Brochu. Considerations on Xi- reconstruction in LHCb. doi: arXiv:1610.07825.

[17] The LHCb RICH group, M. Adinolfi et al. Performance of the LHCb RICH detector at the LHC. *Eur. Phys. J.*, C73:2431, 2013. doi: 10.1140/epjc/s10052-013-2431-9.

[18] LHCb collaboration, The LHCb reconstruction program. URL http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/.

[19] LHCb collaboration, The LHCb analysis program. URL http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/.

[20] I Belyaev et al. Handling of the generation of primary events in gauss, the lhcb simulation framework. *Journal of Physics: Conference Series*, 331(3):032047, 2011. URL http://stacks.iop.org/1742-6596/331/i=3/a=032047.

[21] LHCb collaboration, The GAUSS Project. URL http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/generator/evtgen.php.

[22] Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. ISSN 0168-9002. doi: http://dx.doi.org/10.1016/S0168-9002(03)01368-8. URL http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[23] LHCb collaboration, The BOOLE Project. URL http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/.

[24] CLEO Collaboration. Updated measurements of absolute D^+ and D^0 hadronic branching fractions and $\sigma(e^+e^- \to D\overline{D}$ at Ecm=3774 MeV. *Phys. Rev.*, D89 (7):072002, 2014. doi: 10.1103/PhysRevD.89.072002,10.1103/PhysRevD.91.019903. [Erratum: Phys. Rev.D91,no.1,019903(2015)].

[25] CLEO Collaboration. Improved Measurement of Absolute Hadronic Branching Fractions of the D_s^+ Meson. *Phys. Rev.*, D88(3):032009, 2013. doi: 10.1103/PhysRevD.88.032009.

[26] Daniel Baitinger, Bachelor Thesis. Measurement of the branching ratio of the decay $B^+ \to J/\psi K^{*+}$ at the LHCb experiment.

[27] Mike Williams. Generic D from B Selections. LHCb internal note.

[28] J. Stelzer J. Therhaag E. von Toerne H. Voss A. Hoecker, P. Speckmayer. TMVA4 Users Guide. URL http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf.

[29] David Kirkby Wouter Verkerke. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003. [,186(2003)].

[30] Wouter D. Hulsberge. Decay Chain Fitting with a Kalman Filter. *Nucl. Instrum. Meth.*, A552:566–575, 2005. doi: 10.1016/j.nima.2005.06.078.

[31] Muriel Pivk, Francois R. Le Diberder. sPlot: a statistical tool to unfold data distributions. *Nucl. Instrum. Meth.*, A555:356–369, 2005. doi: 10.1016/j.nima.2005.08.106.

[32] The LHCb collaboration, PIDCalib Packages. URL https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage.

[33] The LHCb collaboration. Study of beauty hadron decays into pairs of charm hadrons. *Phys. Rev. Lett.*, 112:202001, 2014. doi: 10.1103/PhysRevLett.112.202001.

[34] C. Otte, Bachelor thesis. Identifying charmed hadrons from b hadron decays at LHCb.

Acknowledgements:
I would like to thank Prof. Hansmann-Menzemer for the opportunity to work at the LHCb group and her advice on writing this thesis.

Further I would like to thank Dr. Sebastian Neubert for his guidance and for always answering my questions. Also, my thanks goes to the remaining members of the spectroscopy group, Marian Stahl, Alessio Piucci and Christopher Vahl, for their help and the fruitful discussions.

Finally I'm thankful for my family and friends, on whose continuous support I could count throughout my studies.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 04.04.2017 .....