Department of Physics and Astronomy
University of Heidelberg

Master Thesis in Physics
submitted by

Annalena Kalteyer

born in Mainz (Germany)

May 2021

Reconstruction of Λ_c^+ in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ with the ALICE detector

This Master Thesis has been carried out by Annalena Kalteyer at the
Physikalisches Institut of the University of Heidelberg
under the supervision of
Prof. Dr. Silvia Masciocchi

Abstract

The production cross sections of open heavy-flavour hadrons in proton-proton collisions are typically described within the factorisation approach as a convolution of the parton distribution functions of the incoming protons, the perturbative QCD partonic cross section, and the fragmentation functions, which are typically parametrised from measurements in e^+e^- collisions. The fragmentation functions are assumed to be universal for all collision systems. Thus, measurements of charm-baryon production in proton-proton (pp) and proton-lead (p–Pb) collisions are crucial to study the charm quark hadronisation and a possible difference with respect to e^+e^- collisions. Furthermore, measurements of charm-baryons in p–Pb collisions provide important information about Cold Nuclear Matter (CNM) effects. They help to understand how the possible presence of collective effects could modify the production of heavy-flavour hadrons and to explain similarities observed among pp, p–Pb and Pb–Pb systems.

In this thesis, the latest measurement of the Λ_c^+ baryon and its charge conjugate, performed with the ALICE detector at midrapidity in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV is presented using the hadronic decay channel $\Lambda_c^+ \rightarrow pK_S^0$. Λ_c^+ candidates are reconstructed in the transverse momentum range $0 < p_T < 24$ GeV/c, employing for the first time the KFParticle software package. It allows an advanced reconstruction of particle decays and helps to strongly increase the significance. The signal is extracted using an invariant mass analysis, and a selection is applied on the candidate decay topology and particle identification information using the established machine learning tool XGBoost. The p_T-differential cross section and the Λ_c^+/D^0 ratio are measured down to $p_T = 0$ as well as the nuclear modification factor R_{pPb}. It has been found that the Λ_c^+/D^0 ratio at midrapidity in small systems is significantly higher than the one in e^+e^- collisions, suggesting that the fragmentation of charm is not universal across different collision systems.

Zusammenfassung

Der Produktions-Wirkungsquerschnitt von Heavy-Flavour Hadronen in Proton-Proton (pp) Kollisionen wird typischerweise als Faltung der Partonverteilungsfunktion der kollidierenden Protonen, des Wirkungsquerschnitts der Partonen und der Fragmentierungsfunktion in Hadronen beschrieben. Letztere wird normalerweise mit Messungen aus Elektron-Positron Kollisionen parametrisiert unter der Annahme, dass die Fragmentierung gleich ist in allen Kollisionssystemen. Um mögliche Unterschiede zwischen e^+e^- und pp oder Proton-Blei (p−Pb) Kollisionen bei der Hadronisation zu untersuchen, sind Messungen in diesen Kollisionssystemen wichtig. Darüber hinaus geben Messungen von Baryonen mit Charm in p−Pb Kollisionen Aufschluss über Effekte, die durch die Kollision mit einem schweren Kern anstatt eines Protons hervorgerufen werden. Außerdem trägt der Vergleich verschiedener Kollisionssysteme dazu bei, Ähnlichkeiten zwischen solchen Systemen zu finden. Ein Beispiel dafür ist die Frage ob auch in kleineren Systemen kollektive Effekte beobachtet werden können.

In dieser Arbeit wird die Messung des Λ_c^+ Baryons mit dem ALICE Detektor in Proton-Blei Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s_{NN}} = 5.02$ TeV präsentiert. Untersucht wird der hadronische Zerfallskanal $\Lambda_c^+ \to pK_S^0$, wobei die Rekonstruktion für Transversalimpulse des Baryons im Bereich zwischen $0 < p_T < 24\,\text{GeV}/c$ betrachtet wird. Kandidaten werden aufgrund ihrer Zerfallstopologie und Teilchenidentifikation selektiert. Hierbei wird der Algorithmus XGBoost im Zuge maschinellen Lernens verwendet und mittels einer Analyse der invarianten Masse wird das Signal extrahiert. Durch die Verwendung des Softwarepakets KFParticle für die Rekonstruktion des Λ_c^+ Baryons ist es möglich, einen p_T-differentiellen Produktionsquerschnitt, ein Baryon-zu-Meson Verhältnis und einen nuklearen Modifikationsfaktor bis zu niedrigen Impulsen, $p_T = 0$ zu messen. Das Verhältnis Λ_c^+/D^0 ist hierbei signifikant höher als in e^+e^- Kollisionen, was eine nicht universelle Fragmentierung von Charm vermuten lässt.

Contents

1 Introduction

The purpose of high-energy nucleus-nucleus collisions at the LHC is to produce and investigate strongly interacting matter under extreme conditions, like high temperatures and energy densities. Quarks and gluons are the elementary building blocks of nuclear matter and the strong interaction between them is described by the Quantum Chromodynamics (QCD) field theory. Under extreme conditions QCD predicts a phase transition from the hadronic state, where quarks are confined to hadrons, to a colour deconfined medium called the Quark-Gluon Plasma (QGP), in which quarks are considered free and move over distances larger than the size of a nucleon. Such a high-temperature state of matter existed a few microseconds after the Big Bang, and a QGP is suggested to exist today in the core of neutron stars at high energy densities [1].

This chapter gives a short introduction into Quantum Chromodynamics (QCD), the fundamental theory of the strong interaction. The provided information is mainly summarised from chapter 11 and 14 of reference [2]. This section will also explore the different phases of strongly interacting matter and explain under which conditions the state of the Quark-Gluon Plasma is expected. The last part of this chapter will focus on one parameter regime of the QCD phase diagram that is experimentally accessible with heavy-ion collisions at the LHC. Furthermore, it will be explained how the investigation of charmed hadrons helps to probe different models of hadron formation.

1.1 Quantum Chromodynamics

Quantum Chromodynamics is the Quantum Field Theory of the strong interaction. The conserved charge of this theory is called "colour", with the three colours red (r), blue (b) and green (g). QCD is invariant under local $SU(3)_c$ phase transformations, and this symmetry group has eight generators $F^a = \frac{1}{2}\lambda^a$ ($a = 1, 2, ..., 8$), related to the Gell-Mann matrices λ^a. They can be associated to the eight gluons, which are the mediators of the strong interaction. The gluons are massless vector gauge bosons, carrying a colour and an anti-colour. A single colour charge is carried by the quarks, of which there exist three generations, and which are labelled by the flavours up (u), down (d), strange (s), charm (c), beauty (b) and top (t). The quarks are characterised by their mass and by their conserved charges. The QCD Lagrangian density is defined as [2]:

$$\mathscr{L}_{\text{QCD}} = \bar{q}_i^f(i\gamma_\mu D_{ij}^\mu - m_f\delta_{ij})q_j^f - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a, \tag{1.1}$$

with the quark field q_i^f of colour charge i, flavour f, and mass m_f. D_{ij}^μ is the covariant derivative, and $G_a^{\mu\nu}$ is the gluon field strength tensor. The first term of equation 1.1 is the dirac term, which encodes the dynamics of the quark fields. The covariant derivative,

$$D_{ij}^\mu = \delta_{ij}\partial^\mu + ig_s\left(\frac{\lambda_{ij}^a}{2}\right)G_a^\mu, \tag{1.2}$$

couples the gluon field G_a^μ to the quarks with the coupling strength g_s. The second term describes the dynamics of gluons, with the gluon field strength tensor:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + ig_s f_{abc}G_\mu^b G_\nu^c, \tag{1.3}$$

with f_{abc} being the structure constants. Since the generators of the SU(3)$_c$ symmetry group do not commute ($[F^a, F^b] = if_{abc}F^c \neq 0$), QCD is a non-Abelian gauge theory, allowing an interaction between quarks and gluons, but also gluon self-interactions. The gluon self-interaction is contained in the last term in equation 1.3.

A consequence of the gluon self-interaction are higher order loop diagrams, which lead to the fact that the strong coupling constant α_s evolves with momentum transfer Q^2 [3]:

$$\alpha_s = \frac{g_s^2}{4\pi} = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2)\frac{33-2n_f}{12\pi}\ln\frac{Q^2}{\mu^2}}, \tag{1.4}$$

with the renormalisation scale μ, which is an arbitrary reference scale and the number of flavours n_f. The dependence is shown in Fig. 1.1. At small momentum transfer ($Q \approx 1\,\text{GeV}$) or large distances α_s is of $\mathcal{O}(1)$. In this regime quarks are strongly bound into hadrons like mesons, containing a quark and an anti-quark, or baryons containing three quarks. This phenomenon is the first important concept of QCD and is called colour confinement, which states that coloured objects are always confined to colourless singlet states, and only colour neutral objects can propagate as free particles. Therefore, quarks cannot be observed as free particles. At small distances or large momentum transfer ($Q > 100\,\text{GeV}$) $\alpha_s \approx 0.1$. This regime is called asymptotic freedom, where quarks and gluons can be considered free, and perturbative Quantum Chromodynamics (pQCD) calculations become applicable.
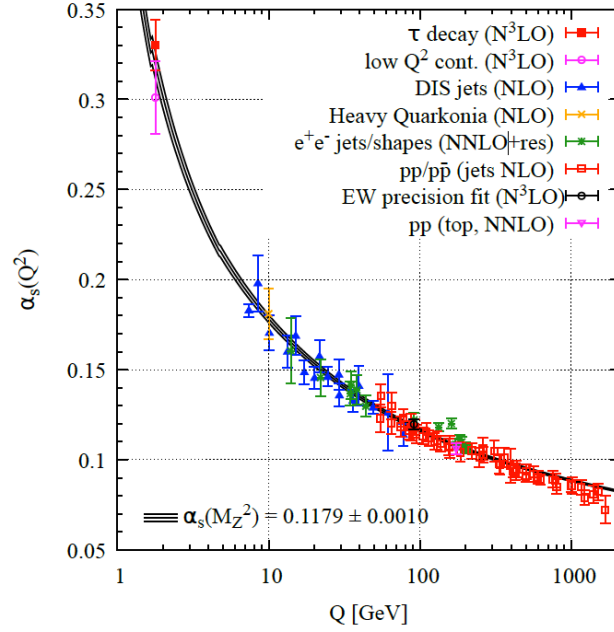
Figure 1.1: Summary of measurements of α_s as a function of the energy scale Q. Figure taken from [4].

1.2 Quark-Gluon Plasma

Depending on the momentum transfer in an interaction, the properties of the strongly interacting matter can be very different. This is a consequence of the large variation of the strong coupling constant as a function of the momentum transfer. A phase transition from hadronic matter to a state of deconfined quarks and gluons, the Quark-Gluon Plasma, was predicted under very high temperatures and/or energy densities [5].

The different phases of the strongly interacting matter can be represented in a phase diagram shown in Fig. 1.2, in terms of the temperature on the y-axis and the baryochemical potential μ_B on the x-axis. The baryochemical potential quantifies the net-baryon content of the system. Ordinary nuclear matter is located at $T \approx 0$ and $\mu_B \approx 1$ GeV. In the region of low energy densities and temperatures the quarks and gluons are confined to colour neutral hadrons in the phase of a hadron gas. At sufficiently high temperatures and μ_B a first-order phase transition from the hadronic matter to the QGP matter is expected [7]. In the extreme case of low temperatures and very high baryochemical potential the conditions, believed to exist in the core of neutron stars due to a high gravitational compression, are fulfilled [8]. This is the regime for a colour superconducting state [9], where quarks and gluons form colour Cooper pairs.
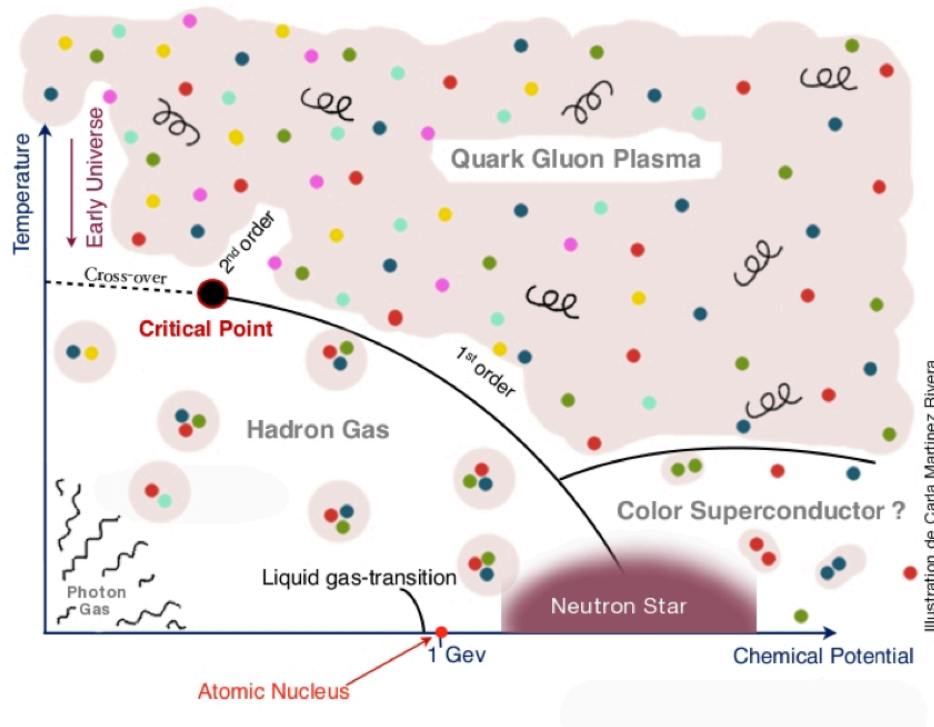
Figure 1.2: The QCD phase diagram. Figure taken from [6].

In the limit of vanishing baryochemical potential and high temperatures a smooth crossover from the hadronic to the deconfined state is predicted by Lattice Quantum Chromodynamics (LQCD) [10]. Lattice QCD is a non perturbative numerical treatment of QCD, formulated on a discrete space-time lattice. It provides a quantitative understanding of the new phase of matter at high temperature and allows the interaction of quarks and gluons over large distances to be studied. Since the phase transition changes from the crossover type to a first-order phase transition at $\mu_B > 0$, Lattice QCD predicts the existence of a critical point. The transition occurs at a critical temperature of $T_c = (156.5 \pm 1.5)\,\mathrm{MeV}$ [10].

1.3 Heavy-Ion Collisions

The QCD phase diagram at $\mu_B \approx 0$ and at temperatures larger than the critical temperature is investigated at the LHC in ultra-relativistic heavy-ion collisions. After a nucleus-nucleus collision the system undergoes different stages in its space-time evolution [11], shown in Fig. 1.3 for a $1 + 1$ dimensional spacetime. The vertical axis in the figure represents the time t. The lower half of the diagram corresponds to the time
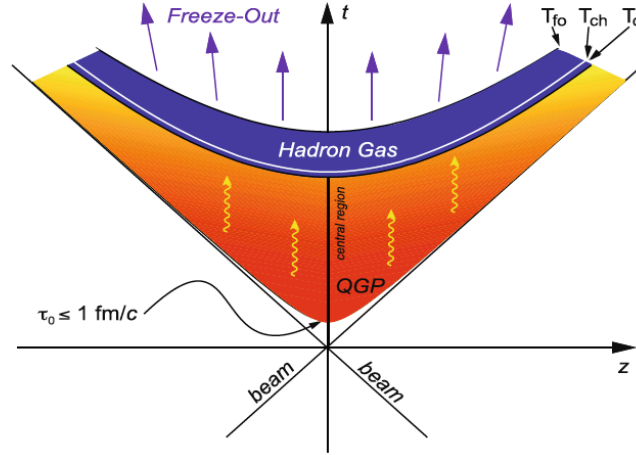
Figure 1.3: Space-time evolution of a heavy-ion collision. Figure taken from [12]

before the collisions, where the two nuclei approach each other with a velocity close to the speed of light, while the upper half of the diagram shows the time after the collision. The horizontal axis is the spatial direction. The time-like region in the space-time diagram is the upper part of the light cone, for which the relation $\tau^2 = c^2 t^2 - z^2 > 0$ is fulfilled, with τ being the proper time. The particle production after a nucleus-nucleus collisions occurs inside the light cone, for which the space-time rapidity $\eta_s = \frac{1}{2} \ln\left(\frac{t+z}{t-z}\right)$ [12] is properly defined. During the expansion of the fireball after the collision at ($t = 0$, $z = 0$), the following phases can be defined:

- **Pre-equilibrium** In the pre-equilibrium phase ($\tau < \tau_0$) partons scatter in hard processes. This leads to the creation of hard probes like heavy quarks, jets or direct photons, with high transverse momentum.

- **Thermalisation and QGP phase** After a time $\tau_0 \sim 1\,\mathrm{fm}/c$ a thermal equilibrium is established. This happens through multiple scatterings among the constituents of the system, produced in the collisions. Since the energy density at thermalisation time at the LHC ($\epsilon_{\mathrm{LHC}} \sim 14\,\mathrm{GeV/fm}^3$ [13]) is above the critical value calculated from Lattice QCD ($\epsilon_{\mathrm{c}} \sim 1\,\mathrm{GeV/fm}^3$ [14]), a thermalised QGP is created, and cools down as it expands. The evolution of the QGP is theoretically described in the framework of relativistic fluid dynamics.

- **Hadronisation** Once the temperature decreases below the critical value T_{c} of the transition crossover, the constituents of the QGP hadronise to a gas of hadrons. The system expands and remains in equilibrium since inelastic collisions still take

place.

- **Chemical freeze-out** Once the temperature of the system falls below the chemical freezeout limit T_{ch}, the inelastic scatterings among hadrons cease and the relative abundances of hadron species (yields) are fixed.

- **Kinetic freeze-out** Elastic collisions cease. At this stage the momentum spectra of the particle species are fixed.

Since the QGP cannot be directly observed in measurements, indirect experimental probes must instead be used to extract information about its properties and evolution. Typically, particles or their decay products are measured after the kinetic freeze-out. Various probes carry information from different stages of the collision. High p_T probes, like jets, heavy quarks and direct photons carry information about the pre-equilibrium phase and experience the full evolution. Thermal photons and dileptons can be used to study properties of the QGP and light hadrons, like protons, kaons and pions characterise the freezeout stage of the evolution.

1.4 Charm Production

Open Heavy Flavour (HF) hadrons are hadrons containing at least one heavy quark (charm or beauty) and other lighter quarks. The purpose of studying open heavy flavour hadrons in Pb–Pb collisions at the LHC is to characterise the properties of the QGP and to investigate the collective behaviour of the medium. In smaller collision systems a plasma of deconfined quarks and gluons is not expected, but pp collisions also provide a reference to p–Pb and Pb–Pb measurements. In addition, the investigation of heavy flavour hadrons in smaller systems gives further insight into the charm production and hadronisation mechanisms. In order to disentangle the initial-state effects in the QGP, p–Pb collisions can be studied as an intermediate system. For instance, the presence of a heavy nucleus in a collision can modify the production of the heavy quark and is one of the Cold Nuclear Matter (CNM) effects [15]. The following sections provide a more detailed explanation of the effects observed in different collision systems.

1.4.1 Charm Production in pp Collisions

Due to their large masses ($m_c \approx 1.3\,\text{GeV}/c^2$, $m_b \approx 4.2\,\text{GeV}/c^2$ [4]) heavy quarks can only be produced in the initial hard scattering processes with large momentum transfer

$Q^2 > 4m_{b,c}^2$. As described in section 1.1 the strong coupling constant is much smaller than 1 at large momentum transfer Q^2. In this regime the production cross section of a charm–anti-charm pair can be calculated in a perturbative expansion of α_s. The production cross section of open heavy flavour hadrons is then usually treated in the factorisation approach as a convolution of the Parton Distribution Function (PDF), the parton hard-scattering cross section, and the fragmentation function. The PDF describes the probability to find a quark or gluon of type a coming from a hadron of type A and carrying a fraction x_A of the hadron momentum. They are typically parameterised from measurements of deep inelastic scattering $e^-p \rightarrow e^-X$ [16]. The parton hard scattering cross section, is obtained in perturbative Quantum Chromodynamics (pQCD) calculations. The fragmentation functions describe the probability that a heavy quark q hadronises into a hadron H_q. They are tuned from data of e^+e^- collisions [17] under the assumption that they are universal for all collision systems.

A quantity sensitive to the hadronisation process, and thus to the fragmentation functions, is the baryon-to-meson ratio, the relative production of baryons and mesons. Measurements of the baryon-to-meson ratio (Λ_c/D^0) in pp and p–Pb collisions, for instance by the ALICE Collaboration [18], showed a substantially higher ratio than measurements from e^+e^- collisions by the ALEPH [19], DELPHI [20] and OPAL [21] at the Large Electron-Positron Collider, pointing to a non-universal hadronisation mechanism. The Λ_c/D^0 can be calculated via Monte Carlo (MC) generators, where the fragmentation functions are tuned on data from e^+e^- collisions. PYTHIA is a MC generator based on the Lund string hadronisation model, which assumes that due to the linearly growing QCD potential between a quark and an antiquark with distance, a high tension "string" forms as a consequence of the gluon field between the quarks, and that this string fragments into hadrons [22]. There are two versions of the generator, considering the leading-colour (LC) approximation and the beyond leading-colour (BLC) approximation. At leading-colour, two final state partons are colour-connected, forming a QCD dipole (gluons are connected to two partons, since they carry colour and anticolour). The LC approximation underestimates the Λ_c cross section in data [23]. In the approximation beyond leading-colour new colour reconnection topologies enhance the baryon production [24].

1.4.2 Charm Production in p–Pb Collisions

When nuclei are colliding instead of protons, the heavy flavour production can be modified. These initial-state effects must to be separated from medium effects caused by the

presence of the Quark-Gluon Plasma. The investigation of p–Pb collisions helps to disentangle both effects. Experimentally, deviations between pp and p–Pb collisions can be studied by measuring the nuclear modification factor:

$$R_{\mathrm{pPb}} = \frac{\mathrm{d}\sigma_{\mathrm{pPb}}/\mathrm{d}p_{\mathrm{T}}}{A \cdot \mathrm{d}\sigma_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}}, \tag{1.5}$$

where $\mathrm{d}\sigma_{\mathrm{pPb(pp)}}/\mathrm{d}p_{\mathrm{T}}$ are the p_{T}-differential cross sections in p–Pb and pp collisions at a given centre-of-mass energy, and $A = 208$ is the lead mass number. If there is no modification the nuclear modification factor would be unity.

The most important initial-state effect at the LHC is the modification of the PDFs if the nucleons are bound inside a nucleus. Parton distribution functions, which appear in the factorisation approach, were firstly assessed for free nucleons, but when measured inside a nucleus they showed a different behaviour. In order to evaluate this effect the following ratio was introduced [25]:

$$R_i^{\mathrm{A}}(x, Q^2) = \frac{f_i^{\mathrm{p/A}}(x, Q^2)}{f_i^{\mathrm{p}}(x, Q^2)}, \tag{1.6}$$

where i is the parton species (valence quarks, sea quarks, gluon), $f_i^{\mathrm{p/A}}(x, Q^2)$ is the PDF of a nucleon in the nucleus, and $f_i^{\mathrm{p}}(x, Q^2)$ is the PDF of the free nucleon. An example of the typical dependence of R_i^{A} on the Bjorken-x is shown in Fig. 1.4. The scaling variable x is named after Bjorken, who proposed that the structure functions exhibit a scaling behaviour [26], after studying deep inelastic scattering processes. It is a dimensionless variable defined as $x = Q^2/2M\nu$, where Q^2 is the squared 4-momentum transfer of the exchanged virtual photon, ν is the energy loss between scattering electrons and M is the nucleon mass. At LHC energies, $x \sim 10^{-4}$, which is the so-called shadowing region. From Fig. 1.4 it is visible that a suppression of charm production is expected in the shadowing region.

A further initial-state effect modifying the production of heavy-flavour particles in p–Pb collisions is the so-called Cronin enhancement [27]. This is where partons inside the colliding particle experience multiple elastic collisions in the target particle before the hard scattering process. Their initial transverse momentum is increased, leading to a shift of the p_{T} spectrum to higher values. The consequence is an increase of the nuclear modification factor above unity.

The nuclear modification factor and the production cross section in p–Pb collisions is predicted by pQCD models that include modifications of the PDFs in p–Pb collisions, as
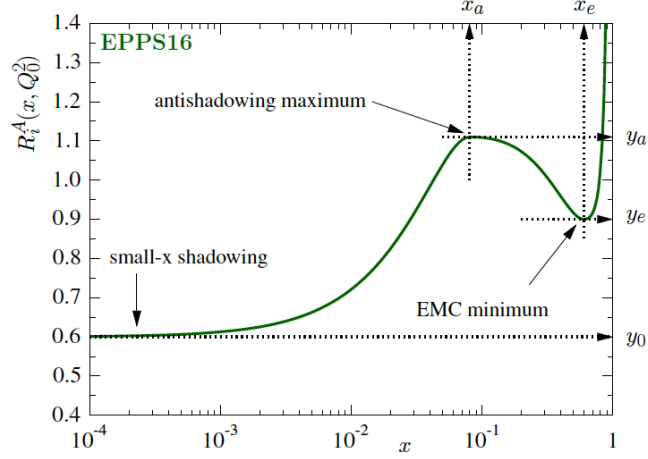
Figure 1.4: Modification of the PDFs as a function of the Bjorken x. Figure taken from [25].

shown for the analysis of the prompt D-mesons production at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ [28]. A comparison to different models will be shown in chapter 5 of this thesis.

1.4.3 Charm Production in Pb–Pb Collisions

The momentum exchange in the medium due to thermal interactions is typically small compared to the necessary momentum transfer to produce a heavy quark. Thus, heavy quarks are predominantly produced in the pre-equilibrium phase of the heavy-ion collision. Furthermore, the flavour is conserved in the strong interaction. This leads to a conservation of the number of charm quarks while traversing the QGP and in the hadronisation process. The early production time and the flavour conservation are what allow them to experience the full evolution of the QGP. While they propagate through the medium they interact with the medium constituents and lose energy.

A comparison of the cross section of heavy-flavour hadrons, and a comparison of the p_{T} spectra in pp and Pb–Pb collisions helps to understand the energy loss mechanism in the QGP and to probe properties of the medium. The quantity to describe the modification in Pb–Pb with respect to pp collisions is the nuclear modification factor:

$$R_{\mathrm{AA}} = \frac{\mathrm{d}^2 N_{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}{\langle N_{\mathrm{coll}}\rangle \mathrm{d}^2 N_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y}, \tag{1.7}$$

where $\mathrm{d}^2 N_{\mathrm{pp(AA)}}/\mathrm{d}p_{\mathrm{T}}\mathrm{d}y$ are the p_{T}- and rapidity-differential particle yields in in pp and heavy-ion (AA) collisions. The yield in pp collisions is scaled with the number of bi-

9

nary nucleon-nucleon collisions N_{coll}, which can be estimated using the experimentally observed particle multiplicity and comparing this to the Glauber model [29]. If one considers a heavy-ion collision to be a superposition of nucleon-nucleon collisions, the measured particle spectra should be the same and the R_{AA} equal to unity. However, if the medium affects the hadron production, one would expect a deviation from this binary scaling. The effects due to interactions with the medium constituents are called final state effects. On the other hand also initial state effects, introduced in the previous section, can lead to a violation of the binary scaling.

A heavy quark propagating through the QGP can lose energy either via inelastic processes, like gluon radiation, or elastic scattering of partons with the medium constituents. The energy loss due to gluon radiation is smaller for heavy quarks than for light quarks and gluons ($\Delta E_{\text{b}} < \Delta E_{\text{c}} < \Delta E_{\text{lightq}} < \Delta E_{\text{g}}$). One reason for this is that the QCD coupling factor is larger for a gluon emitted from another gluon ($C_A = 3$) than for gluon emission from a quark ($C_F = 4/3$) [4], and a large fraction of light flavour hadrons are produced in gluon fragmentation at a late stage of the evolution. Another reason is that gluon radiation is suppressed at angles smaller than $\Theta_{\text{c}} = m_{\text{q}}/E_{\text{q}}$ (dead cone effect) [30]. Due to the higher mass this angle is larger for heavy quarks and the gluon radiation is suppressed in a larger cone leading to a smaller energy loss.

In addition to the energy loss of the heavy quark in the medium, other effects can modify the heavy flavour hadron spectra, like the collective "radial" flow of the medium in the transverse plane. Due to an internal pressure, generated by the QGP, a change of the slope of the transverse momentum spectra at low p_{T} is observed [31]. For a system in thermal equilibrium the p_{T} spectrum can be approximated by a Maxwell-Boltzmann distribution. Finally, the hadronisation mechanism can be different in the presence of a medium of quarks and gluons. Without a QGP, heavy quarks are expected to hadronise via fragmentation [32], while in the presence of a medium a heavy quark could recombine with light quarks close in phase space (coalescence) [33]. Models including coalescence predict for instance an enhanced baryon-to-meson ratio [34]. This enhancement comes due to the recombination of charm quarks with lighter diquark states.

2 The ALICE Detector

The ALICE detector (A Large Ion Collider Experiment) is one of four main experiments at the Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN) in Geneva. Along the accelerator of 26.7 km [35] circumference, protons and heavy ions are accelerated in bunches and brought to a collision at the different interaction points. The main focus of the ALICE collaboration is to investigate the QGP produced in high-energy heavy-ion collisions.

ALICE is a detector with a size of $16x16x26\,\text{m}^3$ and a weight of $10000\,\text{t}$. The detector is optimised for measurements of hadrons, electrons, muons, and photons produced in heavy-ion collisions, with multiplicities of $\text{d}N/\text{d}\eta \approx 2000$. This requires a very high granularity for a good Particle Identification (PID) and a high-momentum resolution from $100\,\text{MeV}/c$ to $100\,\text{GeV}/c$ [36].

Figure 2.1 shows an overview of the subdetectors installed in ALICE during Run 2. The central barrel part of the detector, which covers the midrapidity range of $|\eta| < 0.9$, is placed in a solenoid magnet, previously installed in the L3 experiment at LEP, with a magnetic field longitudinal to the beam direction of $B = 0.5\,\text{T}$. It contains the Inner Tracking System (ITS), the Time Projection Chamber (TPC), Transition Radiation Detector (TRD), Time-Of-Flight detector (TOF), Photon Spectrometer (PHOS), Electromagnetic Calorimeter (EMCal), Di-jet Calorimeter (DCal) and the High Momentum Particle Identification Detector (HMPID).

Several detectors are placed in the forward direction to measure photons and charged particles at $|\eta| \approx 3$, like the Forward Multiplicity Detector (FMD) and Photon Multiplicity Detector (PMD). With the Cherenkov T0 detector the time and longitudinal position of the interaction is measured, which is needed for instance for the TOF. For triggering, the VZERO scintillator detector measures a signal proportional to the number of charged particles [38]. Besides the detectors for event characterisation a muon spectrometer, situated at a pseudo-rapidity of $-4 < \eta < -2.5$ measures quarkonia, light vector mesons and muons from heavy-flavour hadron decays.

The ALICE coordinate system is a right-handed orthogonal Cartesian system. The x-axis is perpendicular to the beam direction pointing to the centre of the accelerator, the y-axis points upwards and the z-axis is parallel to the beam direction with the muon spectrometer in negative z direction [39].

Detailed information about all detectors in ALICE can be found in [38]. In the next section only detectors relevant for this analysis are described.
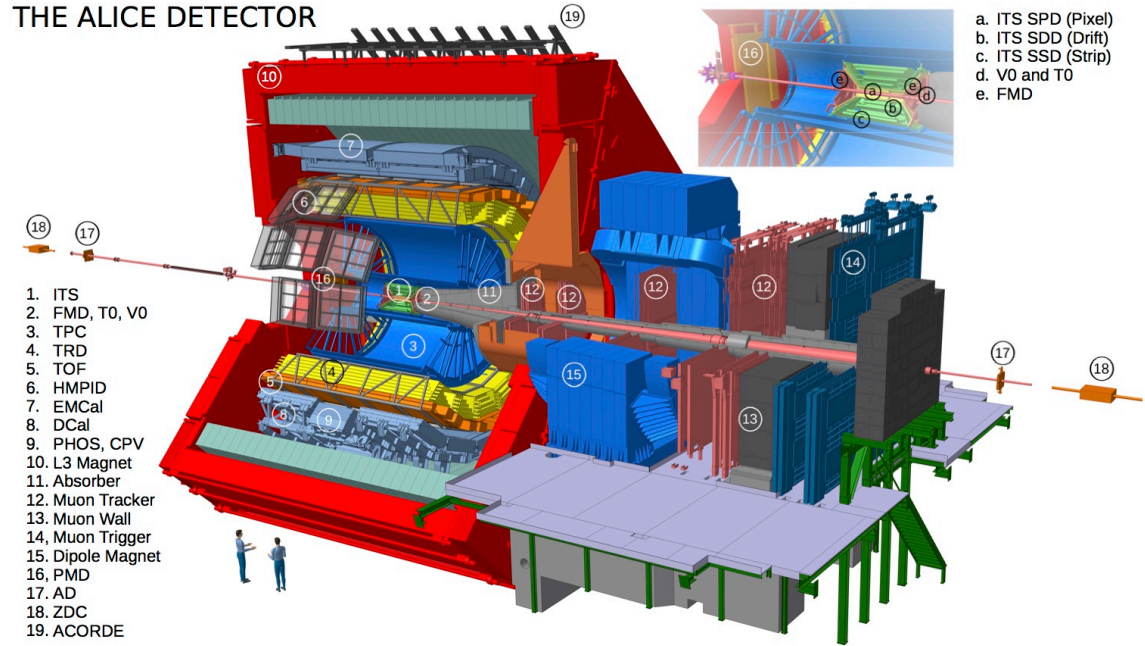
THE ALICE DETECTOR

a. ITS SPD (Pixel)
b. ITS SDD (Drift)
c. ITS SSD (Strip)
d. V0 and T0
e. FMD

1. ITS
2. FMD, T0, V0
3. TPC
4. TRD
5. TOF
6. HMPID
7. EMCal
8. DCal
9. PHOS, CPV
10. L3 Magnet
11. Absorber
12. Muon Tracker
13. Muon Wall
14. Muon Trigger
15. Dipole Magnet
16. PMD
17. AD
18. ZDC
19. ACORDE

Figure 2.1: The ALICE detector during Run 2. Image taken from [37].

2.1 Inner Tracking System

The Inner Tracking System (ITS) is built up of six layers of silicon detectors and surrounds the beam pipe. The main tasks of the ITS are to find the primary vertex with a high resolution better than $100\,\mu$m [36] and to reconstruct secondary vertices from particle decays. A first estimate of the primary vertex is based on a measurement in the two innermost layers of the ITS (Silicon Pixel Detectors). Lines, joining two clusters in each SPD layer, define a tracklet and the vertex with the largest number of contributing tracklets is the initial estimate of the primary vertex. The estimate of the primary vertex is refined using particle trajectories (tracks) reconstructed in the full ITS and the TPC. Furthermore, the ITS allows measurements of particle trajectories at momenta below $200\,$MeV$/c$ and thus improves the momentum resolution for particles reconstructed by the TPC and ITS. The inverse-p_T resolution σ_{1/p_T} is shown in the left panel of Fig. 2.2 with a comparison for tracks measured in p–Pb collisions at $5.02\,$TeV only in the TPC and tracks measured with clusters in the TPC and ITS, with and without constraining a track to the primary vertex. The resolution improves when both of the detectors are used for the track reconstruction and when the tracks are constrained to the vertex [38]. For these ITS-TPC tracks the resolution of the transverse distance to the primary vertex for particles with $p_\mathrm{T} = 1\,$GeV$/c$ is about $75\,\mu$m in pp collisions and $60\,\mu$m in Pb–Pb
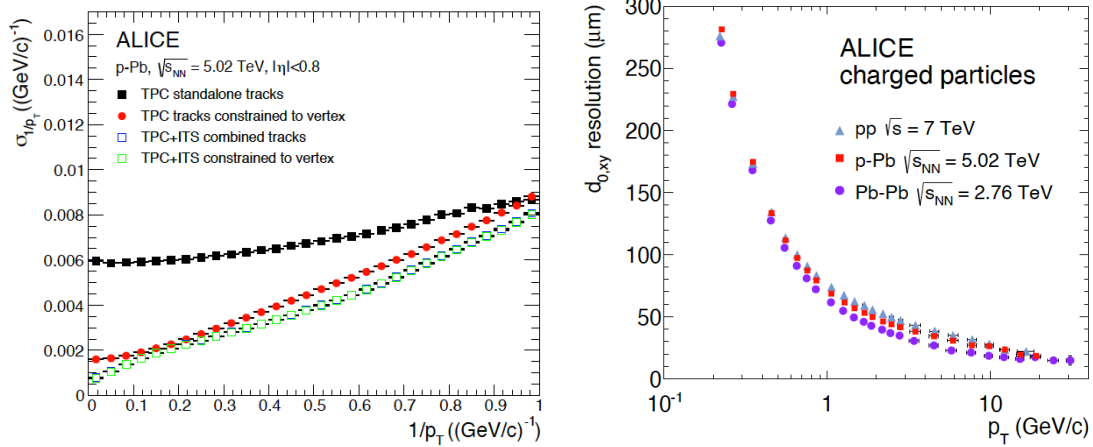
Figure 2.2: Left: p_T resolution for TPC standalone tracks compared to ITS-TPC tracks with and without vertex constraint. Right: Resolution of the impact parameter to the primary vertex for all charged ITS-TPC tracks. Figures are taken from [38].

collisions, which is shown in the right panel of Fig. 2.2. The improvement in heavy-ion collisions is due to the larger particle multiplicity, where more tracks contribute to the primary vertex reconstruction. Such a high resolution allows the determination of decay vertices of short-lived particles, like heavy-flavour hadrons. Apart from tracking and vertex determination the four outer layers of the ITS can be used for particle identification via measuring the specific energy loss, since they provide analogue readout [38]. The method of particle identification via the mechanism of ionisation in a detector by a charged particle and by measuring its momentum is described in more detail for the TPC.

2.2 Time Projection Chamber

The Time Projection Chamber (TPC) is the most important detector in ALICE for particle tracking and together with the other central barrel detectors it provides a good track separation, particle identification, and vertex determination. The detector dimensions are an inner radius of $85\,cm$, an outer radius of $250\,cm$, and a length of $500\,cm$. The detector is built of a field cage with a central high-voltage electrode and is filled with $\approx 90\,m^3$ of $Ne/CO_2/N_2$ [36]. If a charged particle traverses the TPC, induced electrons drift to the end plates within a maximum drift time of about $90\,\mu s$, where multi-wire proportional chambers (MWPC) are mounted. The radial coordinate of a track is given by the charge deposited in the readout cathode pads, arranged in 159 rows in radial di-
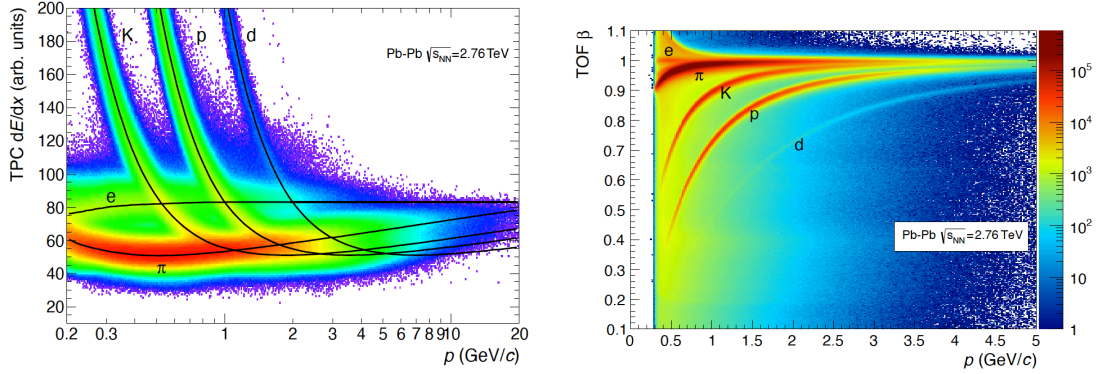
Figure 2.3: Left: Specific energy loss for different particle species in the TPC vs particle momentum in Pb–Pb collisions. The lines show the expected mean energy loss. Right: β measured by the TOF detector as a function of the particle momentum in Pb–Pb collisions. Both distributions are taken from [38].

rection and the third spatial component is given by the drift time.

The TPC also provides particle identification information, which is performed by measuring simultaneously the specific energy loss (dE/dx), charge and momentum of a particle. The mean energy loss by ionisation per path length of a particle traversing the TPC is described by the parametrised Bethe-Bloch formula [38]

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}}(P_2 - \beta^{P_4} - \ln(P_3 + \frac{1}{(\beta\gamma)^{P_5}})). \qquad (2.1)$$

β is the particle velocity, γ is the Lorentz factor, and P_{1-5} are fit parameters. The left panel of Fig. 2.3 shows the specific energy loss versus the particle momentum in Pb–Pb collisions. Particles can be selected by using the deviation n_σ between the measured and expected specific energy loss dE/dx for the corresponding species in units of the resolution. At momenta below $1\,\mathrm{GeV}/c$ a particle species separation on a track-by-track basis is possible, whereas at higher momenta, due to the relativistic rise, different detectors, like the TOF, need to be used additionally to separate different particle species.

2.3 Time of Flight Detector

The Time-Of-Flight detector (TOF) detector in ALICE also provides particle identification information. The detector is built of Multigap Resistive Plate Chambers (MRPC), placed at an inner radius of $370\,\mathrm{cm}$ and reaching to an outer radius of $399\,\mathrm{cm}$. If a particle ionises the gas in the detector an avalanche process generates a signal on the readout

electrodes. The total signal is the analogue sum of signals from many gaps and the time resolution is about $40\,\mathrm{ps}$ with a detector efficiency of close to $100\,\%$ [38].

The particle is identified by measuring its flight time over a given distance along its trajectory L and its momentum. The velocity of a particle $\beta = v/c = L/tc$ depends on its mass and momentum. Thus, the mass of a particle is given by [40]:

$$m = \frac{p}{c}\sqrt{\frac{c^2t^2}{L^2} - 1}. \tag{2.2}$$

The right panel of Fig. 2.3 shows the PID performance of the Time Of Flight detector. The particle velocity β is shown as a function of the particle momentum measured by the TPC for Pb-Pb collisions. The separation power of two particles with masses m_A and m_B carrying the same momentum p is:

$$\mathrm{n}_\sigma = \frac{|t_\mathrm{A} - t_\mathrm{B}|}{\sigma_\mathrm{TOF}} = \frac{Lc}{2p^2\sigma_\mathrm{TOF}}|m_\mathrm{A} - m_\mathrm{B}| \tag{2.3}$$

where $\sigma_\mathrm{TOF} \approx 60 - 80\,\mathrm{ps}$ [38] is the TOF resolution. This relation is valid if $p \gg mc$. The separation power decreases with increasing momentum and it depends on the mass difference between two particle species. Thus, PID with the TOF detector is possible in the momentum range up to $2.5\,\mathrm{GeV}/c$ for pions and up to $4\,\mathrm{GeV}/c$ for protons.

2.4 Particle Identification

In this analysis a Bayesian approach is used to combine the PID information of various detectors into one probability and thus making use of the full PID capabilities in ALICE [41]. For a single detector with a Gaussian response the probability of a particle species H_i to produce a signal S is:

$$P(S|H_\mathrm{i}) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\mathrm{n}_\sigma{}^2}, \tag{2.4}$$

where n_σ for the TPC is the deviation between the measured and expected specific energy loss dE/dx for the corresponding species in units of the resolution, and for the TOF it is the separation power defined in eq. 2.2. By using Bayes' theorem,[42], the product of probabilities from different detectors are combined in a conditional probability that the particle is of species H_i for a measured set of detector signals.

3 Methods

For the first time in ALICE, the Λ_c^+ is reconstructed employing a new reconstruction package, called the KFParticle package. It was originally developed for the CBM experiment [43], [44]. However, the parametrisation of the particle state vector is independent of the detector geometry. This makes the package also applicable in ALICE. The KFParticle allows the application of constraints for instance on the production vertex of a particle or on its mass and they are treated as additional information in the decay reconstruction and help to gain a higher precision of the measurement. In the following, a short introduction and summary of the reconstruction method with the KFParticle is given.

In the scope of this analysis also supervised machine learning is used to classify signal and background in a given dataset of Λ_c^+ candidates. The used algorithm is XGBoost [45], which stands for Extreme Gradient Boosting. The term gradient boosting originates from J. Friedman, who developed different gradient boosting algorithms [46]. An introduction into the algorithm formulation used for XGBoost and some tutorials can be found in Ref. [47]. In section 3.2 of this chapter an overview of XGBoost is provided.

3.1 The Kalman Filter Particle Package

A short-lived particle, like the Λ_c^+ ($c\tau(\Lambda_c^+) \approx 60\mu$m [4]), decays before it can reach the tracking system in ALICE. Thus, it can only be reconstructed by its decay products. Commonly the parameters of the mother particle are calculated via the parameters of the daughters in the decay. The daughter parameters are extrapolated to the point of closest approach of the daughter particle trajectories, where the momentum and energy of the daughters are summed up.

Another approach to reconstruct short-lived particles is realised in the Kalman Filter Particle package [44]. It is based on the Kalman filter method [48], which is a mathematical iterative procedure to estimate unknown variables considering a series of measurements and their inaccuracies. A generic Kalman Filter algorithm starts with an initial approximation of a state vector $\mathbf{r}_0$ and covariance matrix $\mathbf{C}_0$, which is the initialisation step. In the context of particle reconstruction, this is an initial approximation on the decay vertex of a mother particle and an estimate of its momentum and energy. These parameters

are stored in a state vector [43]:

$$\mathbf{r} = (x, y, z, p_x, p_y, p_z, E, s)^T, \tag{3.1}$$

with the position of the decay vertex $(x, y, z)^T$, the particle momentum $(p_x, p_y, p_z)^T$ and energy E, plus the length of the trajectory in the laboratory system l normalised to the particle momentum p, $s = l/p$. The parameters of the daughter particles are also described by a state vector and the associated covariance matrix contains the parameter accuracies. The next step in a general Kalman Filter algorithm is the prediction. For each measurement $\mathbf{m}_k$ the estimate of the state vector is extrapolated to the point of a measurement, taking into account the inaccuracies on the measurement. In the reconstruction, $\mathbf{m}_k$ is a state vector of a daughter particle. This daughter is transported along its trajectory to the initially approximated decay vertex of the mother, under consideration of the uncertainties on the decay vertex and the trajectory of the daughter particle. In the general Kalman filtering step the estimate on the state vector is updated with the measurement $\mathbf{m}_k$, giving an optimum state vector $\mathbf{r}_k$ according to the first k measurements. After this step in the particle reconstruction, a better estimate of the state vector and covariance matrix of the mother particle is obtained. The optimal estimate of the state vector with its covariance matrix is obtained after the filtering step on the last measurement. Thus, the optimal state vector and covariance matrix of the mother particle is obtained after repeating the procedure described above for all daughter particles. Finally, the mother particle is transported from the decay vertex to the production vertex, where the optimal state vector and covariance matrix is evaluated [44].

After the reconstruction, the particle can be transported to any point along its trajectory. This distinguishes the KF Particle package from other vertexing packages, where only the production and decay vertex are reconstructed. Due to the chosen particle parametrisation, the algorithm is independent on the detector geometry. Additional physics parameters that are not stored in the state vector, like the particle momentum p, the invariant mass M, the length of flight L and the lifetime $c\tau$, are calculated according to [49]:

$$\begin{aligned} p &= \sqrt{p_x^2 + p_y^2 + p_z^2} \\ M &= \sqrt{E^2 - p^2} \\ L &= sp \\ c\tau &= sM. \end{aligned} \tag{3.2}$$
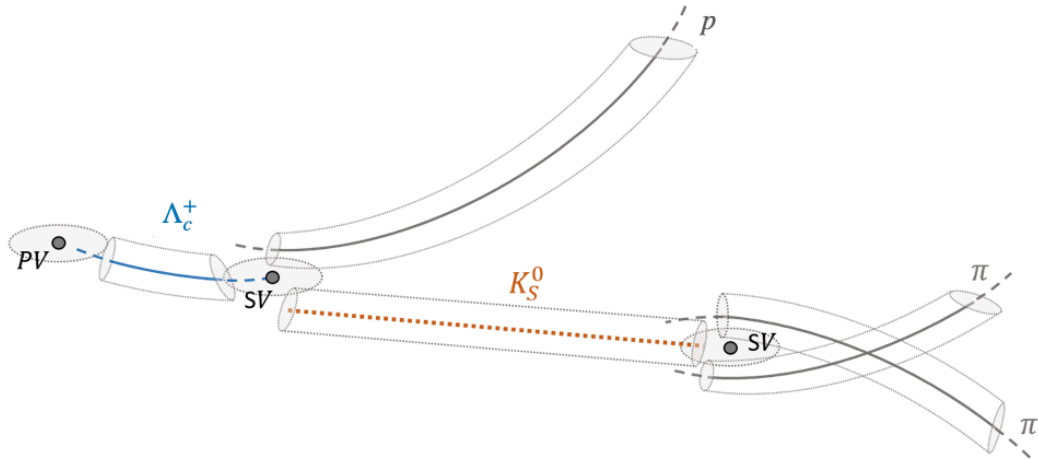
Figure 3.1: Cartoon of the Λ_c^+ decay, including the uncertainties on the mother and daughter particle tracks, and on the primary vertex (PV) and the secondary vertices (SV).

Furthermore, it is possible to remove daughters previously added to the state vector. Especially in the primary vertex reconstruction tracks might be recognised to be non-primary tracks afterwards and they can be removed to improve the fit quality [43]. A cartoon of the Λ_c^+ decay, including the uncertainties on the mother and daughter particle tracks and vertices, is shown in Fig.3.1. One can observe the Λ_c^+ decaying into a proton and a K_S^0, and the kaon decaying into two oppositely charged pions.

3.1.1 Constrained Fits

The estimate of a state vector can be improved using several constraints, which are treated as an additional measurement by the Kalman Filter [43]. Applying constraints is equivalent to an increase of information regarding the particle decay. This allows an improvement of the precision of the measurement. During the reconstruction of the secondary vertex, one of the constraints is to require that the mother particle has a certain mass. For instance in the reconstruction of the $\Lambda_c^+ \rightarrow pK_S^0$, where the kaon decays into $K_S^0 \rightarrow \pi^+\pi^-$, the mass of the kaon is constrained to the PDG mass [4]. This is called the mass constraint [44].

A constraint can be made also on a vertex, where the particle is required to point back to that vertex. This can be the expected production vertex or any other reconstructed vertex. This requirement is called topological constraint. A χ^2-deviation estimates the
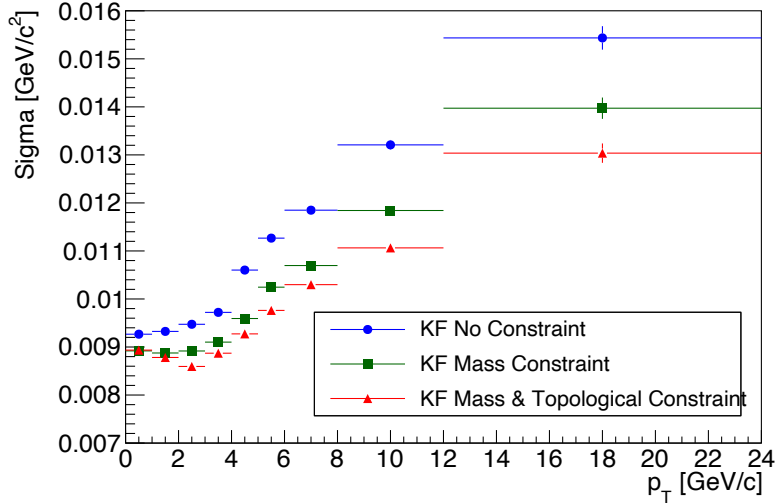
Figure 3.2: Mass resolution of simulated Λ_c^+ reconstructed with KF Particle package using different constraints.

probability of the hypothesis, that the particle is produced at that vertex. In this analysis the Λ_c^+ is required to point back to the primary vertex.

The effect on the mass resolution of the mother particle, when applying a constraint is shown in figure 3.2. The blue markers indicate the resolution on the mass of the Λ_c^+ reconstructed with the KF Particle package without any additional constraint. These candidates are taken from a MC simulation dedicated for this analysis. An improvement in the mass resolution is visible if the mass constraint is applied to the mass of the kaon (green) and this improvement increases when in addition the Λ_c^+ is required to point back to the primary vertex (red). This improvement is vital for rare signal reconstruction because a narrower peak width helps to get a signal emerging from a large combinatorial background.

3.1.2 Output Variables

After the reconstruction, quantities on the decay topology and fit quality can be extracted. These variables are useful to select reconstructed candidates. In the reconstruction of a particle decay, the daughters are used in a geometrical fit to provide an estimate of their production vertex or the decay vertex of the mother particle. The χ^2 of this fit expresses how well the daughter tracks intersect within their uncertainties [44]. Small values of $\chi^2_{\mathrm{geo}}/\mathrm{NDF}$, where NDF is the number of degrees of freedom indicate a large probability that the daughter tracks intersect within their uncertainties.
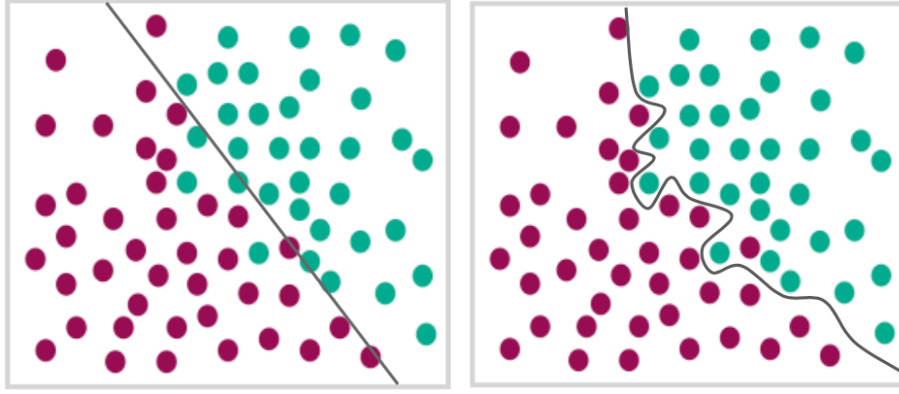
Figure 3.3: Example of a function separating signal (green) and background (red). Left: Well trained model, which results in a good description of the features in the data set. Right: Overfitting of the data.

As described before, in a topological fit of a particle to a vertex, a χ^2-deviation estimates the probability of the hypothesis that the particle is produced at that vertex. If the χ^2_{topo}/NDF value is small, the probability is large that the particle intersects with the primary vertex within the uncertainties [44]. Although, if the uncertainties are large the fit quality will also be good.

The distance, how far away from the primary vertex a particle decays, is given by the variable l/Δl. It is the distance between the decay vertex and the primary vertex normalised to the uncertainty on the distance [44], which are localisation uncertainties of the vertices.

3.2 Machine Learning

In a predictive learning problem the ingredients are a set of input variables $\mathbf{x}$ and a response variable y. A training sample consists of known $(y, \mathbf{x})$-values. It is used to estimate a function $F^*(\mathbf{x})$ mapping the input variables to the response variable. The specific loss function $L(y, F(\mathbf{x}))$ [46] quantifies how well this function models the data. In the training, the specific loss function is minimised, by finding the parameters that best fit the training data. Furthermore, regularisation is also important to avoid overfitting, by constraining the fitting procedure [46]. Overfitting means that a model fits even fluctuations in a particular dataset and therefore fails to predict future observations in another dataset reliably. This is shown in Fig. 3.3, where the red points represent the background and the green points represent the signal. The left panel shows a linear function to separate signal and background. This function captures the characteristics
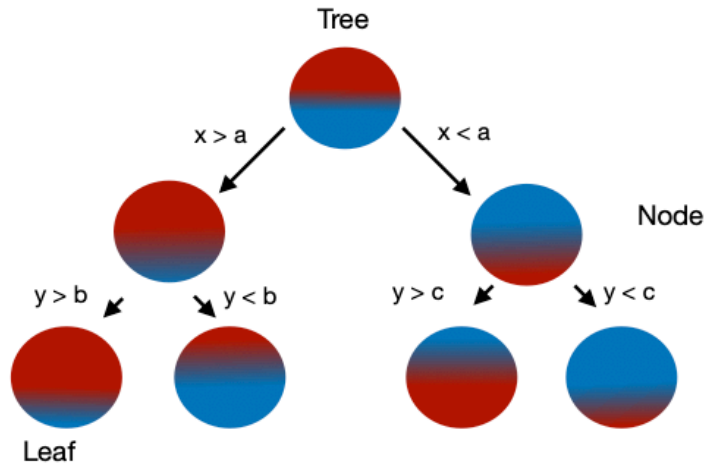
Figure 3.4: Example of a decision tree to separate signal and background, where red indicates background and blue indicates signal.

of the data set but it does not fit every fluctuation in the signal. This is different for the function shown in the right panel of Fig. 3.3, where the functions captures even fluctuations in the signal. The specific loss term and the regularisation term are combined in an objective function.

The machine learning algorithm used in this thesis is XGBoost. The model in XGBoost is a tree ensemble consisting of a set of classification and regression trees (CART) [45]. The structure of a basic decision tree to separate signal and background is shown in Fig. 3.4. The separation of signal in blue and background in red is done in this example by applying a selection for each variable (x,y). In tree ensembles the prediction of multiple trees is summed. Since it is not possible to learn all trees at once, only one tree, that optimises the objective function, is added at a time [47]. The algorithm starts from a single leaf and branches are added iteratively. Splitting a leaf into two results in a decomposition of scores of the left and right node. If the gain of a branch is smaller than a certain value, it is not added. This is called pruning [45].

In this analysis the python package of XGBoost is used. Before training a model, parameters need to be set. There are different types of parameters. General parameters specify the choice of booster to be used for boosting, in this case models are tree based. Booster parameters, like the shrinkage factor, the maximum depth of a tree and the number of trees are specific for the chosen booster. Within the heavy-ion physics environment for machine learning (hipe4ml) [50] booster parameters can be optimised in a bayesian

21

approach as described in section 4.2. With a list of parameters and a data set, a model can be trained and saved. Once a model has been trained, predictions on the dataset are possible. Those predictions give a score or an output probability. The distributions of output scores or probabilities as well as quality and control plots can be studied as in section 4.2. If the model has been trained properly it is applied to the data to separate signal from background candidates. A selection is made based on a probability threshold above which candidates are accepted.

4 Data Analysis

The evaluation of the production cross section of charmed baryons in different collision systems is crucial to study the charm hadronisation process. In this analysis the Λ_c^+ was measured via its hadronic decay to a proton and a K_S^0, which decays into a positively and a negatively charged pion. The quark content of the Λ_c^+ is (udc) and its mass is listed at $(2286.46 \pm 0.14)\,\text{MeV}/c^2$ by the Particle Data Group [4]. The average lifetime is $(202.4 \pm 3.1)\,\text{fs}$. This analysis is performed in p–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02\,\text{TeV}$. The particle decays are reconstructed with the KF Particle package, which allows a direct estimate of the track parameters and vertices as well as the associated uncertainties. The signal extraction is optimised using machine learning with extreme gradient boosting. The analysis is performed in the transverse momentum range 0 to $24\,\text{GeV}/c$, where p_{T} in the following will always refer to the transverse momentum of the Λ_c^+. With this being the first measurement of the p_{T}-differential cross section down to $p_{\text{T}} = 0$. The results are compared to the previously measured Λ_c^+ production cross section in the range from $1 - 24\,\text{GeV}/c$, where Λ_c^+ candidates were reconstructed without the KF Particle package and a different machine learning algorithm was used [23].

4.1 Candidate Reconstruction and Selection

The analysis performed in the scope of this thesis is based on data tables storing physical, topological and PID quantities of the $\Lambda_c^+ \to pK_S^0$ candidates. Although, the steps from the data taking in ALICE, via the selection of events and reconstruction of Λ_c^+ candidates is not part of the work performed in this thesis, in the following a summary of the process before obtaining the data tables is given to present the full picture of the analysis.

4.1.1 Cascade Object Reconstruction in ALICE

First, the raw data measured in each detector is transformed into clusters. Clusters are groups of neighbouring cells in a detector, where a signal is measured. The Primary Vertex (PV) is identified, as described in section 2.1, using the innermost detector layers of the ITS to reconstruct tracklets. Particle trajectories (tracks) are reconstructed by using the clusters in the different detectors, and the initial estimate of the primary vertex position. The position is refined with a better resolution using tracks reconstructed in the TPC and the ITS. The next step is to search for decay vertices of neutral particles that do not leave a track in the detector before decaying. These particles, like the K_S^0, are

called V^0 candidates. Then, cascade candidates are reconstructed. These are particles decaying into a neutral particle and a charged particle, leaving a track in the detector. The Λ_c^+ is such a cascade candidate. Before data analyses are possible, calibrations need to be performed in all steps of the reconstruction. The results of this procedure are Analysis Object Data tables (AODs) storing information about the cascade objects. More information on this can be found in [38].

4.1.2 Event and Track Selection

For a physics analysis events have to be selected. The data events selected for this analysis are from LHC Run 2 collected with the ALICE detector in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,$TeV in 2016. The events were recorded with a minimum-bias (MB) trigger. This trigger needs a coincident signal in both VZERO detectors above a certain threshold. To reject background events from machine-induced beam-gas interactions a timing selection is applied based on the signal from the VZERO and Zero Degree Calorimeter (ZDC) detector. For the analysis only events with a z-coordinate of the reconstructed primary vertex $|v_z| < 10\,$cm with respect to the nominal interaction point are taken into account. This is done to assure a uniform detector acceptance. In addition, also events with multiple interaction vertices, so-called pileup events, are rejected. For p–Pb collisions an algorithm can remove pileup events based on the reconstructed tracks with the TPC and ITS [38]. In total about 600 million MB-triggered events are selected. The corresponding integrated luminosity is $\mathscr{L}_{int} = 287\mu b^{-1}(\pm3.7\%)$, and calculated using the visible cross section measured in van der Meer scans [51].

The MC simulated signal candidates are taken from events generated using PYTHIA 6 [52] with the Perugia2011 tune [53]. For events with more than one binary collision an underlying event is generated using HIJING [54]. The detector and data taking conditions are reproduced in the simulations. Different Monte Carlo (MC) samples were used for training and testing of a Boosted Decision Tree (BDT) model and to compute the reconstruction efficiencies and the acceptance corrections. The reason for having two separate Monte Carlo simulations is that the total efficiency, which is the number of reconstructed candidates after applying a machine learning model to data over the generated candidates, should not be determined from the same set of candidates that was used to train the model.

To ensure a good quality of the tracks used to reconstruct the Λ_c^+, the following selection criteria are applied. One requirement is that the number of clusters in the TPC detector used for the energy loss determination has to be larger than fifty. This is done to assure

a good average energy loss dE/dx resolution. Furthermore, the number of crossed rows in the TPC needs to be above a minimum of 70 rows for a good track reconstruction. In order to assure a uniform detector acceptance, the pseudorapidity of the daughter tracks has to be within a range of $|\eta| < 0.8$, since the central barrel detectors cover a pseudorapidity range of $|\eta| < 0.9$.

4.1.3 Λ_c^+ Reconstruction with the KFParticle Package

Since the KFParticle package, which is described in section 3.1, is employed in this analysis, cascade object candidates are selected, and the reconstruction of the $\Lambda_c^+ \to pK_S^0$ is performed. It starts with the formation of a $K_S^0 \to \pi^+\pi^-$ candidate, reconstructed from two opposite-sign charged pion candidate tracks joining in a neutral decay vertex, which should be displaced from the primary vertex. To form a Λ_c^+ the K_S^0 is combined with a proton-candidate track. Two constraints are applied in the reconstruction. The mass of the K_S^0 is constrained to the mass listed by the PDG [4] and the Λ_c^+ is required to point back to the primary vertex. Physical, topological and PID quantities of the mother particle and daughter particles are obtained and stored in data tables. These tables are used for further analysis performed in this thesis.

4.2 Machine Learning

The separation between signal and background can be optimised using Boosted Decision Trees. The basic concept is described in section 3.2. In this analysis the algorithm XGBoost [45] was used. With the machine learning approach multiple selection criteria are combined into a single response variable. This is different to the traditional method, where rectangular selections are applied for various criteria separately. After the application of a trained BDT model to the full data sample a selection in the response can be optimised to separate signal from background. A model is trained for each transverse momentum interval separately with a given sample of signal and background data. The signal is taken from simulated events. For the training only prompt Λ_c^+ signal candidates are selected and those that are produced in a decay of a particle containing a beauty quark are not used since they can have a different vertex decay topology. The background is selected from a fraction of real data candidates with the requirement that the invariant mass of the Λ_c^+ is either smaller than $2.23\,\mathrm{GeV}/c^2$ or larger than $2.34\,\mathrm{GeV}/c^2$ to ensure that the signal region is excluded. The proportion of signal to background candidates

Table 4.1: Number of signal (S) and background (B) candidates for the BDT training and testing

p_{T} (GeV/c)	0-1	1-2	2-3	3-4	4-5	5-6	6-8	8-12	12-24
S	51061	73860	64410	48090	28791	16805	15909	8363	2146
B	102122	147720	128820	96180	57582	33610	31818	16726	2146

is $1 : 2$. A proportion $1 : 1$ would be the natural choice if enough training candidates were available for a good training performance. However, since the number of signal candidates is limited by the available MC sample, and the number of background candidates from data is much larger, the background fraction is doubled. An exception is the transverse momentum range above $12\,\mathrm{GeV}/c$, where the background candidates would exceed a fraction of $10\,\%$ of the full data sample, there the proportion is $1 : 1$. Table 4.1 provides an overview of the total number of training plus testing candidates per transverse momentum interval. The data set is then randomly divided into two parts where one part is used for the training of the model and the second part for testing of the model performance.

4.2.1 Training Variables

The BDT training variables can be topological properties of the particles or PID variables. When choosing these classification criteria a few things have to be taken into account. One of them is that with an increasing number of criteria the model gains performance in signal and background separation, but also complexity. This can lead to overtraining, which becomes visible in a large difference between the performance of the training and the test set. Thus, only the criteria with the largest impact on the model performance should be included, however, still providing a selective ML model. In order to select the optimal variables that introduce the largest gain to the BDT performance a model is trained adding all available variables classifying the Λ_c^+, the K_S^0 and the proton. They can be ordered according to their ranking of feature importance, which is the average impact on the model output.

Furthermore, attention has to be paid to possible correlations between the training variable and the observable, which is in this case the mass of the Λ_c^+. If a correlation occurs in the background, the natural shape of the invariant mass spectrum can be modified. This leads to an artificial enhancement or reduction in the extracted signal. On the other
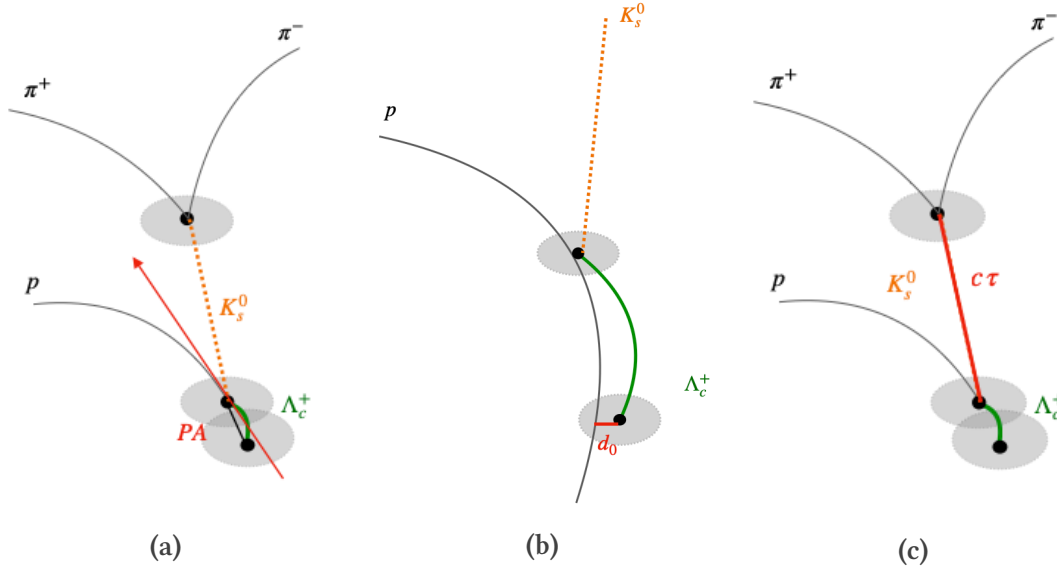
Figure 4.1: Topological variables characterising the proton, kaon and Λ_c^+. (a) Pointing angle (PA) of the Λ_c^+, (b) Impact parameter, d_0, of the proton with respect to the primary vertex, (c) Decay length $c\tau$ of the K_S^0.

hand, correlations between different training variables can be exploited for the signal and background separation if they occur in the signal but not in the background or vice versa.

Under consideration of the correlations and the feature importance the training variables are chosen. One training variable is the PID of the proton using the combined bayesian probability, which is the criterion with the largest feature importance. The relative feature importance of a variable can be quantified in percent, as will be shown later in this chapter. The variable which has the second largest impact on the model output is the $\chi^2_{\text{topo}}/\text{NDF}$ that estimates the probability of the hypothesis that the Λ_c^+ was produced at the primary vertex. This variable is obtained in the KF reconstruction and is defined in section 3.1. For true Λ_c^+ the reduced chi-squared of the fit to the primary vertex is expected to be smaller than for background. Another topological variable characterising the candidate is the Pointing Angle (PA). The PA is defined as the angle between the momentum vector of a particle and the line connecting its production and decay vertex, visualised in figure 4.1a. In addition, the impact parameter of the proton track in the transverse plane, d_0, with respect to the primary vertex is used, shown in figure 4.1b. Since the Λ_c^+ is a very short-lived particle the displacement between the primary vertex and its decay vertex is only $c\tau = 60.7\,\mu$m on average [4]. Thus, the impact parameter of the proton track to the primary vertex is expected to be small for true Λ_c^+. As a property
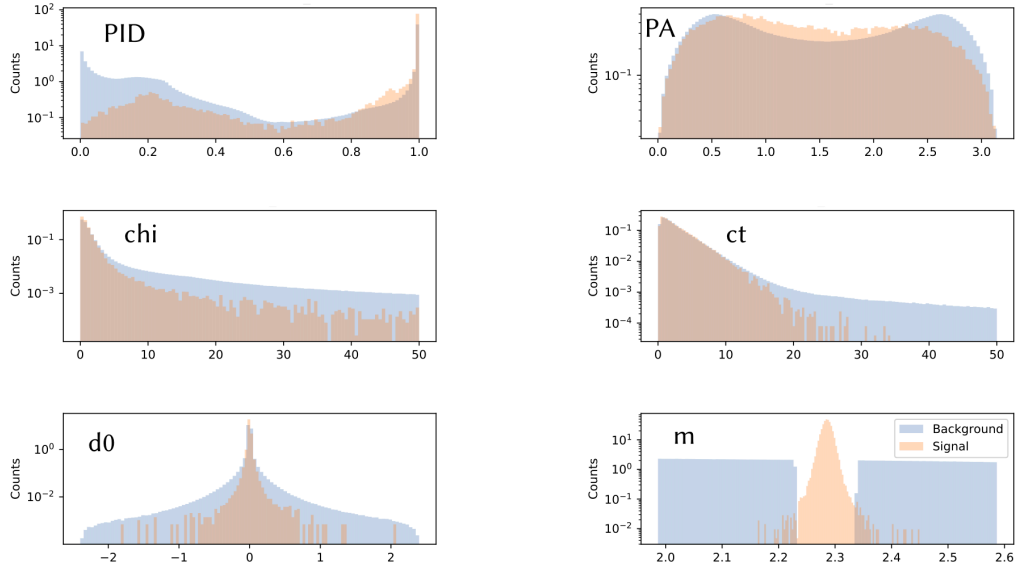
Figure 4.2: Signal and background distribution of the selected training variables: combined bayesian PID probability of the proton, χ^2/NDF (chi) and Pointing Angle (PA) of the Λ_c^+ to the primary vertex, impact parameter of the proton to the PV (d_0), and lifetime of the kaon ($c\tau$) in the p_T interval between 1 and $2\,\text{GeV}/c$. m is the mass distribution of the candidates considered for signal and background.

of the K_S^0 its decay length is included, see figure 4.1c.

The signal and background distribution of all chosen variables is shown in figure 4.2. For a good signal and background separation a difference between the two distributions is desired. This is true for the chosen training variables. Here, the distribution of the combined bayesian PID probability of the proton is of particular interest, shown in the top left panel of 4.2. The signal distribution shows that most signal candidates have a high probability close to 1. Those are the tracks with a measurement in the TOF detector and in the TPC. Proton tracks measured only in the TPC have lower probabilities. A second peak around 0.2 emerges from proton tracks, that are only measured in the TPC and have a $p_T > 1\,\text{GeV}/c$. In the relativistic rise of the dE/dx distribution particle species are not perfectly isolated anymore, leading to an equally shared probability among five particle species. The upper right plot shows the signal and background distribution of the pointing angle of the Λ_c^+ in the range from 0 to π. The middle left panel of Fig. 4.2 shows the $\chi^2_{\text{topo}}/\text{NDF}$. By taking a closer look at the range from 0 to 5 it becomes visible that the signal has a peak at slightly lower values than the background distribution. The decay length of the kaon is shown in the middle right panel and the impact parameter of the proton in the bottom left panel. The mass of the Λ_c^+ is not a training variable.
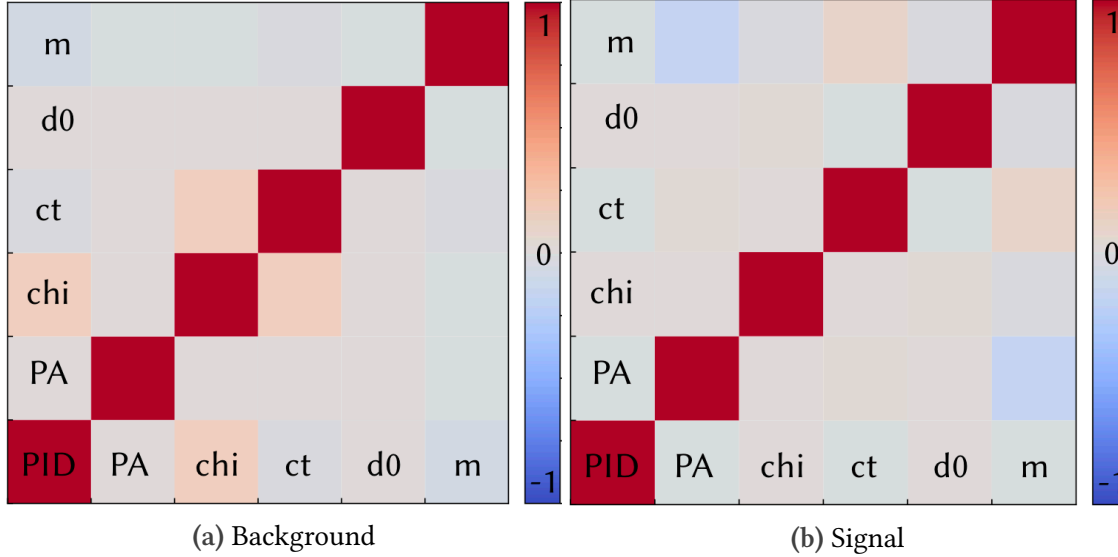
(a) Background

(b) Signal

Figure 4.3: Correlation matrix in the transverse momentum range $1 < p_T < 2 \, \text{GeV}/c$.

However, it is included in this figure to show the distribution of the signal from MC and the background distribution taken from the sidebands in data.

Figure 4.3a shows the correlation matrix between the classification criteria for the background and figure 4.3b for the signal. As desired there is no correlation between the training variables and the mass of the Λ_c^+ observed in the background. Particularly useful for a signal and background separation could be the correlation between the $\chi^2_{\text{topo}}/\text{NDF}$ and the PID of the proton, as well as the decay length of the kaon. Both are visible in the background but not in the signal.

4.2.2 Hyperparameters

Hyperparameters are parameters in machine learning that control the learning process. The set of hyperparameters is optimised with a bayesian optimisation procedure to obtain the best performance from the algorithm. In this approach a predefined space of parameters is scanned in an iterative procedure. In contrast to a random or grid like search, the bayesian approach considers past evaluations when choosing the next set. In order to test the parameters a cross validation method, called k-fold [55], is used. For this, the original data sample is split into k folds where k−1 folds are used for the optimisation and one for testing. After each test of a hyperparameter set the folds are permuted and the result is obtained as the mean value of all permutations. The hyperparameter space has to be chosen carefully to avoid overtraining. An important aspect

Table 4.2: Optimised Hyperparameters

p_T (GeV/c)	$0-1$	$1-2$	$2-3$	$3-4$	$4-5$	$5-6$	$6-8$	$8-12$	$12-24$
learning	0.07	0.08	0.06	0.08	0.06	0.08	0.06	0.09	0.08
max. depth	2	3	3	3	3	2	2	2	2
estimators	773	769	775	774	775	772	771	772	760

when choosing the parameter ranges for the optimisation is that the algorithm should not always converge to the lower or upper edge of the given range because in that case the optimal set is outside of the given range.

In this analysis, one of the optimised parameters is the maximum depth of a tree. A large tree depth makes a model more complex and a lot of memory is consumed when training a deep tree. A tree depth of 4 and larger was evaluated. However, the optimisation tends to select large tree depths, although the model is then overtrained. This is why the range for optimisation is selected conservatively to be 1 to 3 at transverse momenta in the $1-5\,\text{GeV}/c$ range and 1 to 2 below $1\,\text{GeV}/c$ and above $5\,\text{GeV}/c$ respectively, where the number of training candidates is smaller.

Another parameter is the learning rate. After each boosting step the feature weights are adjusted according to the learning rate to make the process more conservative and avoid overtraining. The given range is from 0.01 to 0.1. Typically a default value of 0.3 is set in XGBoost, but the optimisation of the hyperparameters for the given training set resulted in values between 0.06 and 0.08. The last optimised parameter is the number of estimators, which corresponds to the number of trees. Since the depth for each tree is rather small, many trees are selected in the algorithm. The range is varied between 150 and 1000 for the optimisation. Table 4.2 shows the set of optimised parameters for each p_T interval.

4.2.3 Machine Learning Models

A BDT model is trained for each p_T interval separately. After the model was trained with half of the available candidates it is applied to the test set. If the performance is optimal the model should neither be overtrained nor undertrained and the deviation between training set and test set should be negligible. This can be verified with the learning curves shown in Fig. 4.4. It shows the Root-Mean-Square Error (RMSE), which is the difference between the value predicted by the model and the observed one as a function of the training set size for the training set (red line) and the test set (blue line).
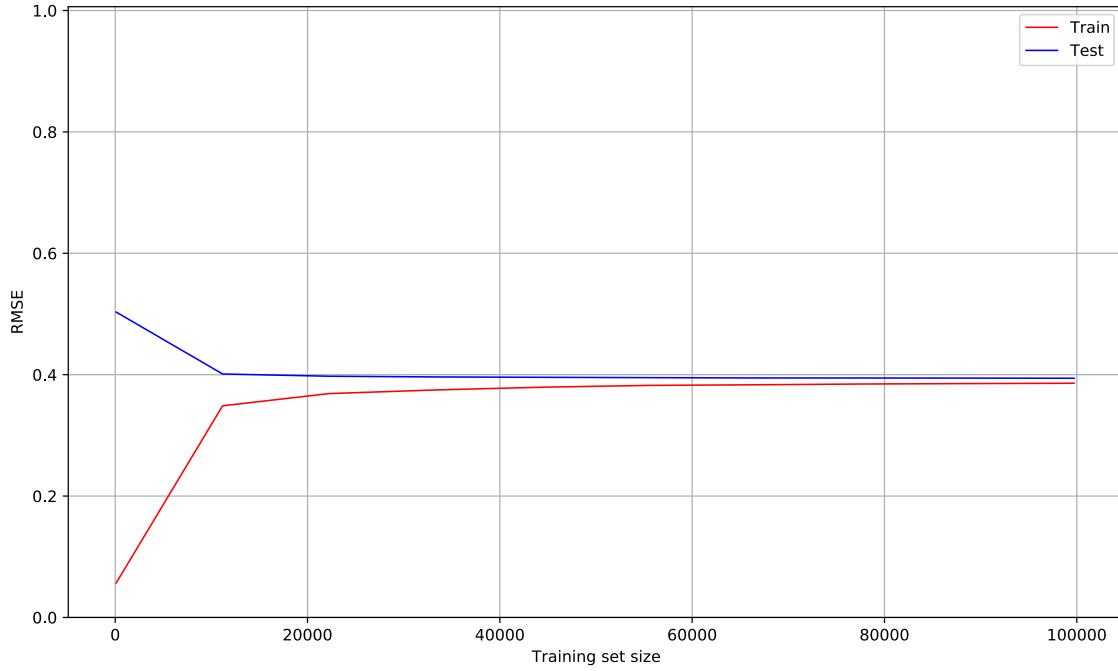
Figure 4.4: Learning curves of the training set and test set in the p_T interval $1 - 2\,\text{GeV}/c$. The RMSE of the training set (red) and test set (blue) are shown as a function of the training set size.

As expected when only few instances are taken into account for the training, the fit describes the data well and the error (red line) is close to zero. As an example, one could imagine the extreme case of fitting two points with a fit function. The error on the fit would be 0. When more training instances are added to the fit, the fluctuations increase and with that also the RMSE. For the learning curve of the training set shown in Fig. 4.4 the RMSE seems to stabilise at set sizes larger than 20000. Thus, the model performance improves when increasing the training set size, which leads to a better description of the data. This improvement in the model performance is reflected by a good description of the validation data, where the error decreases (blue line). In a well trained model both curves should converge to the same value.

Another method to control and validate the model performance is the so-called ROC-AUC curve. ROC is the short term for Receiver Operating Characteristics and it displays the model performance at various classification values. Figure 4.5 shows the curves for the trained model in the p_T interval $1 - 2\,\text{GeV}/c$. Here, the True Positive Rate (TPR) is shown as a function of the False Positive Rate (FPR). TPR is defined as the true positives over the true positives plus false negatives, and FPR is defined as the false positives over the false positives plus the true negatives. The true and false positives and negatives are
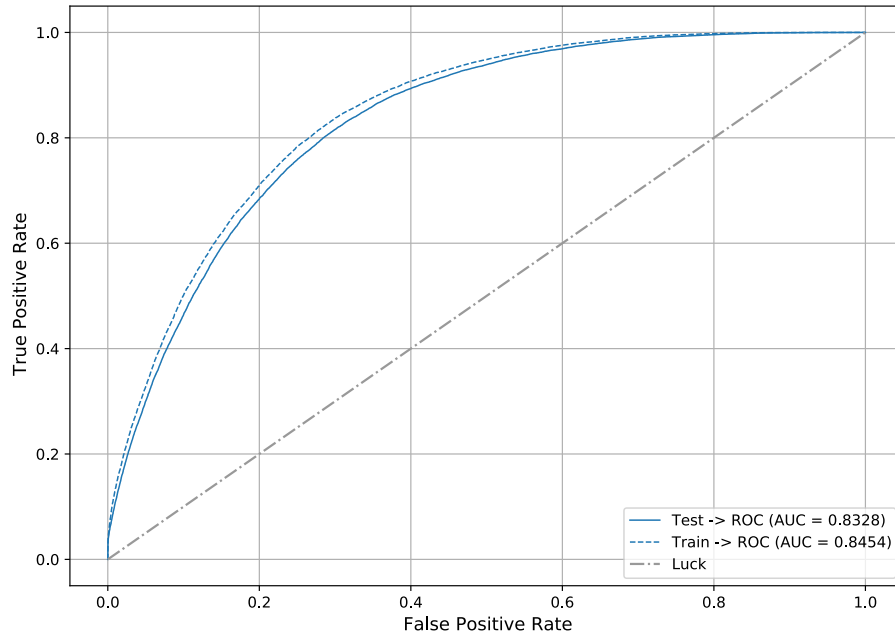
Figure 4.5: ROC-AUC curves for the training set and test set in the range $1 < p_\mathrm{T} < 2\,\mathrm{GeV}/c$.

the possible results when applying a classification to a test set. As an example, the result would be true positive if the algorithm classifies a candidate as signal in the tested set and it is truly signal, but it would be false positive if the algorithm classifies it as signal for the test set, while the reality is background [55]. Typically the model performance should be measured for all classification values. This can be combined in the Area Under the ROC Curve (AUC). It is interpreted as the probability that the model prediction classifies a true positive correctly. If the prediction would be wrong in all cases the area under the curve would be zero and one if the prediction is always right. The dashed line displays the mark of 50 % where the model prediction is right in half of the cases. For a well trained model the area under the curve should be large but the deviation between the training set and test set should be small.

The resulting model is reported in the left panel of figure 4.6. It shows the scaled signal in red and background in blue as a function of the BDT output probability for the training and test set. A perfect model, which depends on the variables that are used for the classification, would completely separate signal at high probabilities and background at low probabilities. Attention has to be paid that the validation data sample represented by the blue and red markers in Fig. 4.6 does not deviate significantly from the training set, but it should also not follow fluctuations in the training set distribution. The peaking signal at BDT probabilities larger than 90 % was investigated to result from candidates
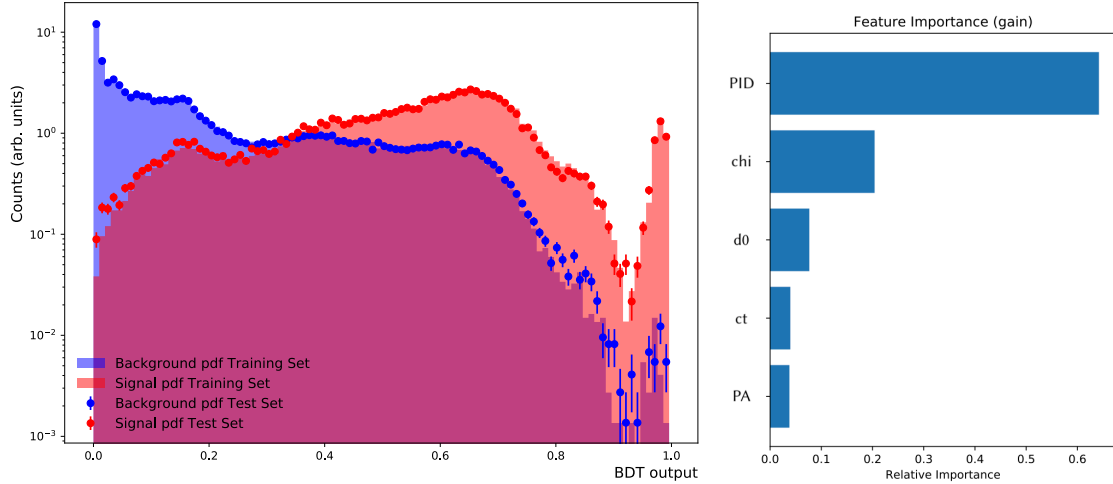
Figure 4.6: Left: Machine learning model with the signal distribution in red and the background distribution in blue for the training and test set. Right: Feature Importance of the classification criteria. Both are shown for the range $1 < p_{\mathrm{T}} < 2\,\mathrm{GeV}/c$.

with a very high combined bayesian PID probability of the proton. This investigation was performed by choosing candidates with a high BDT probability and looking at the corresponding distributions of the different training variables. The importance of the classification criteria is ordered in a ranking in the right panel of figure 4.6 from most important on the top to least important on the bottom. Important means a large average impact on the model performance.

4.2.4 Model for $0 < p_{\mathrm{T}} < 1\,\mathrm{GeV}/c$

At transverse momenta below $1\,\mathrm{GeV}/c$ the strategy for the machine learning training is changed. The reason is that the corrected Λ_{c}^{+} signal shows large instabilities if a model is trained with the same selection criteria as at higher p_{T}. This introduces a variation in the final cross section of about $50\,\%$. The term corrected refers to the efficiency correction that is performed by applying the model to a MC sample of pure signal. If features observed in data are not reproduced in the generated candidates, a difference in the efficiency can occur. The training variable causing the instability in this case was investigated and found to be the particle identification of the proton. The loss of signal in data and MC as a function of the BDT probability was observed to be different when the PID is used in the training of the model. Thus, the PID is removed from the training. Since this variable is important for the separation of signal and background a selection in the combined PID probability to be larger than $20\,\%$ is added before the ML training.

(a) Machine learning model.

(b) ROC curves.
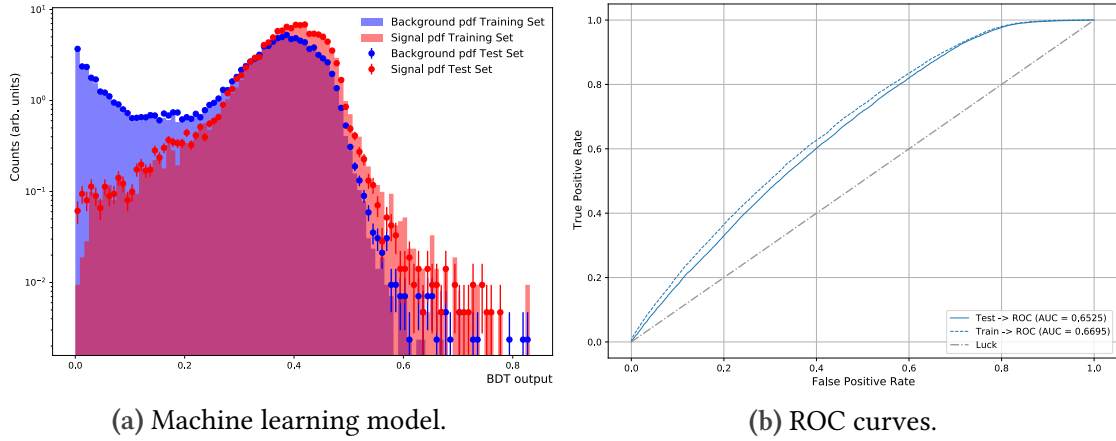
Figure 4.7: BDT output at p_T smaller than $1\,\mathrm{GeV}/c$

This selection is low compared to the proton PID selection of a BDT, which would apply a selection of $40 - 70\%$. The resulting model is shown in figure 4.7a. As expected, the separation power between signal in red and background in blue decreases compared to the model shown before for $1 < p_\mathrm{T} < 2\,\mathrm{GeV}/c$, since the PID variable with the largest importance was removed. This can be seen in Fig. 4.7b as well, which shows an Area Under the ROC Curve of 0.65 for the test set, while it was at 0.83 in the figure above for the interval $1 < p_\mathrm{T} < 2\,\mathrm{GeV}/c$.

4.3 Working Point Determination

Once a model has been trained, it is applied to the full data sample for each p_T interval. A BDT output probability is chosen, below which the Λ_c^+ candidates are rejected. If a high probability is selected, the signal purity is large, but at the same time the efficiency is low. The determination of the BDT probability selection should be performed without looking at the significance of the Λ_c^+ signal in data as a function of the BDT probability, in order to avoid biases in choosing the selection with the highest significance, hence possibly choosing a statistical fluctuation. The procedure of choosing the BDT selection is called the search of the Working-Point (WP). To choose the working point region a pseudo-significance S is calculated and it is defined as:

$$S = \frac{s}{\sqrt{s + b}}, \tag{4.1}$$

with the signal s and background b. The pseudo-significance distribution is computed as a function of the BDT output probability. It is calculated using a pseudo signal, taken
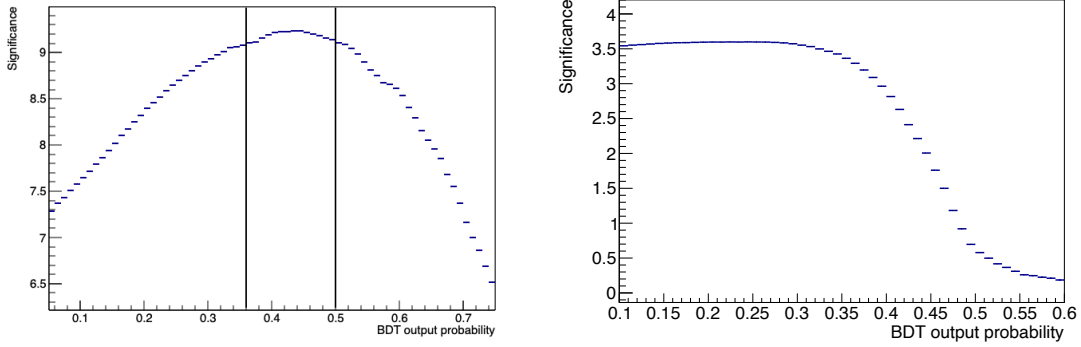
Figure 4.8: Pseudo-significance as a function of the BDT probability. Left: for the range $3 < p_T < 4\,\mathrm{GeV}/c$. Right: for the range $0 < p_T < 1\,\mathrm{GeV}/c$.

Table 4.3: BDT output probability selections

p_T (GeV/c)	$0-1$	$1-2$	$2-3$	$3-4$	$4-5$	$5-6$	$6-8$	$8-12$	$12-24$
BDT output	0.21	0.26	0.35	0.37	0.36	0.32	0.30	0.34	0.25

from a model prediction and realistic background. The background is obtained from a fit of the side-bands in the invariant mass spectrum of real data with a second order polynomial (for $p_T > 1\,\mathrm{GeV}/c$) or a third order polynomial (for $p_T < 1\,\mathrm{GeV}/c$). Relevant for the calculation is the background within a three sigma range from the mean of the Λ_c^+ mass distribution. The mean and sigma are taken from a Gaussian fit to the MC simulated candidates. The pseudo signal is estimated rearranging the cross section equation 5.1. The signal is calculated using the average of the previously measured Λ_c^+ cross sections [23] and the efficiency as a function of the BDT selection, obtained in this analysis. The efficiency calculation is described in section 4.5. For low BDT probabilities the significance is typically small since not much background is rejected. The significance increases with increasing BDT selection until it reaches a range of plateau-like behaviour. At high probabilities the significance decreases since also many signal candidates are rejected. This behaviour can be seen in the left panel of figure 4.8. The BDT probability selection should be chosen in the plateau region of the significance distribution.

For $p_T < 1\,\mathrm{GeV}/c$ the strategy is different since the Λ_c^+ cross section measurement is the first in ALICE in this p_T range. Thus, the pseudo signal cannot be calculated from a previously measured result. The signal for $p_T < 1\,\mathrm{GeV}/c$ is determined from an extrapolated cross section obtained for the previous analysis [23]. The pseudo-significance obtained for the p_T interval $0-1$ GeV/c is shown in the right panel of Fig. 4.8. It is

observed that this distribution shows a plateau at low BDT probabilities before it drops quite quickly due to the low discrimination power of the BDT model shown in Fig. 4.7a. For this reason a loose selection in the BDT probability is chosen.

Table 4.3 summarises the chosen BDT probability selections for all p_T intervals.

4.4 Signal Extraction

The Λ_c^+ raw yield is extracted from a fit of the invariant mass distribution after the application of the BDT selection criteria, where raw refers to the yield before any corrections which are explained in the next section. For the signal a Gaussian function is used and the background is fitted with a quadratic polynomial except for $p_T < 1\,\mathrm{GeV}/c$, where the background shape is described better by a cubic polynomial. The Gaussian width for the fit is fixed to the width obtained from the invariant mass distribution of simulated candidates to improve the stability of the fit. The results are displayed in figure 4.9. For each figure the signal S, the background B in the 3σ region, the signal over background ratio and the significance is displayed. The background subtracted residuals are shown in figure 4.10. A clear signal with significances larger than $4\,\sigma$ is seen in the full p_T spectrum.

 The extracted fit parameters and results are compared to the previously obtained results for Λ_c^+ measured in the same decay channel, but reconstructed without the KFParticle package and using the machine learning algorithm TMVA (AdaBoost) instead of XG-Boost. This comparison is shown for the peak width in figure 4.11a. The width in data and MC are in agreement within $2\,\sigma$ in the full p_T spectrum. The width in data fluctuates and this is why the width in this analysis is fixed to the one obtained from the simulation, also to ensure a more stable signal extraction. The width of the signal distribution is systematically smaller in this analysis compared to the previous one. The reason for this is that in the KF reconstruction the mass of the K_S^0 is constrained to the mass listed in the PDG, plus the topological constraint of the Λ_c^+ to the primary vertex is applied. Both improve the mass resolution. See section 3.1 for a comparison of the peak width with different constraints. The mean value of the Gaussian fit is in agreement with the previous analysis within the uncertainties (figure 4.11b).

The extracted raw yield is shown in figure 4.12a and the corresponding significance is shown in figure 4.12b. The extracted signal is scaled to the number of MB events N_{MB}, and the significance, which is calculated with formula 4.1, is scaled to $\sqrt{N_{MB}}$. The signal extraction and significance are improved over the full transverse momentum spectrum
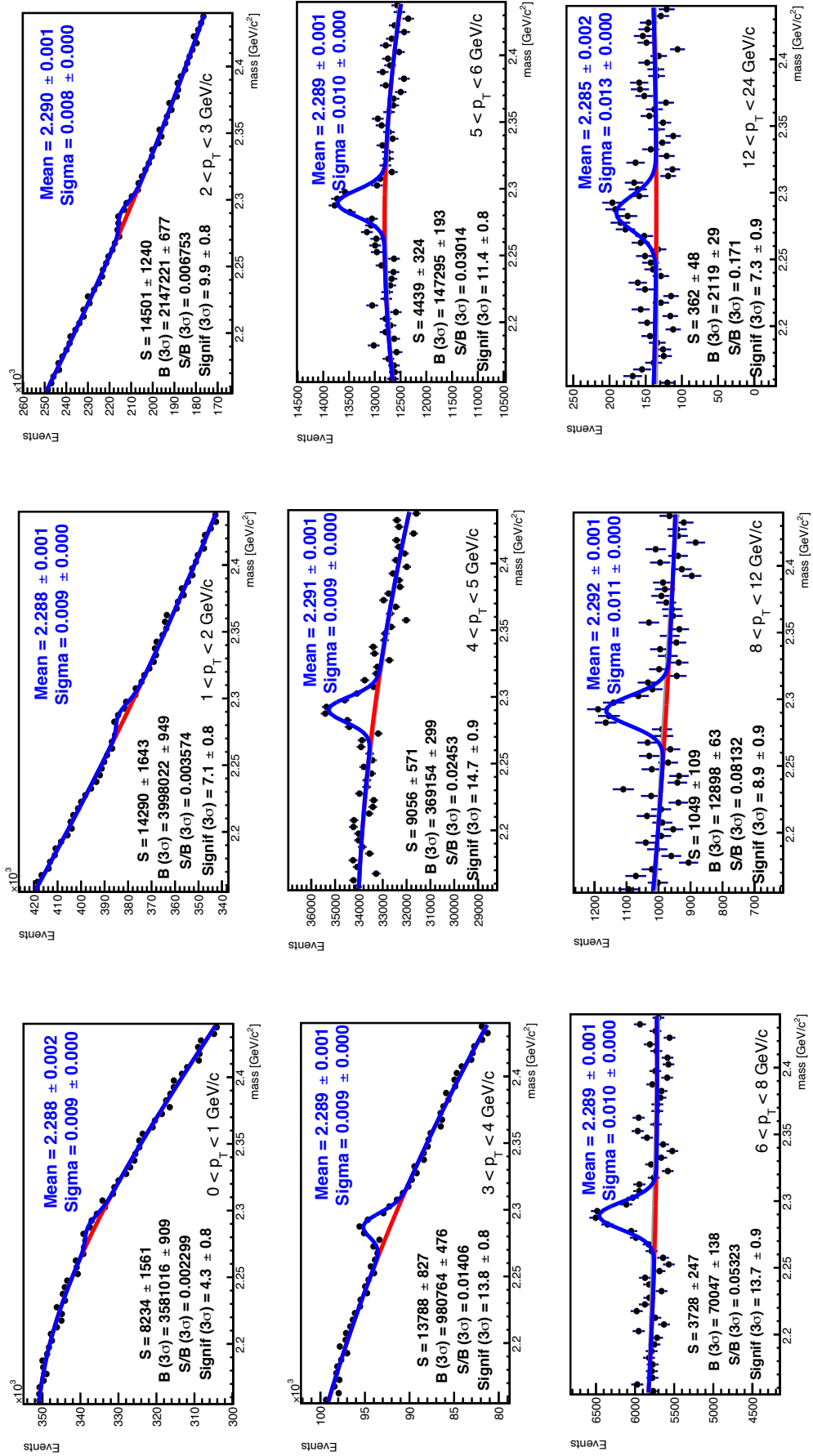
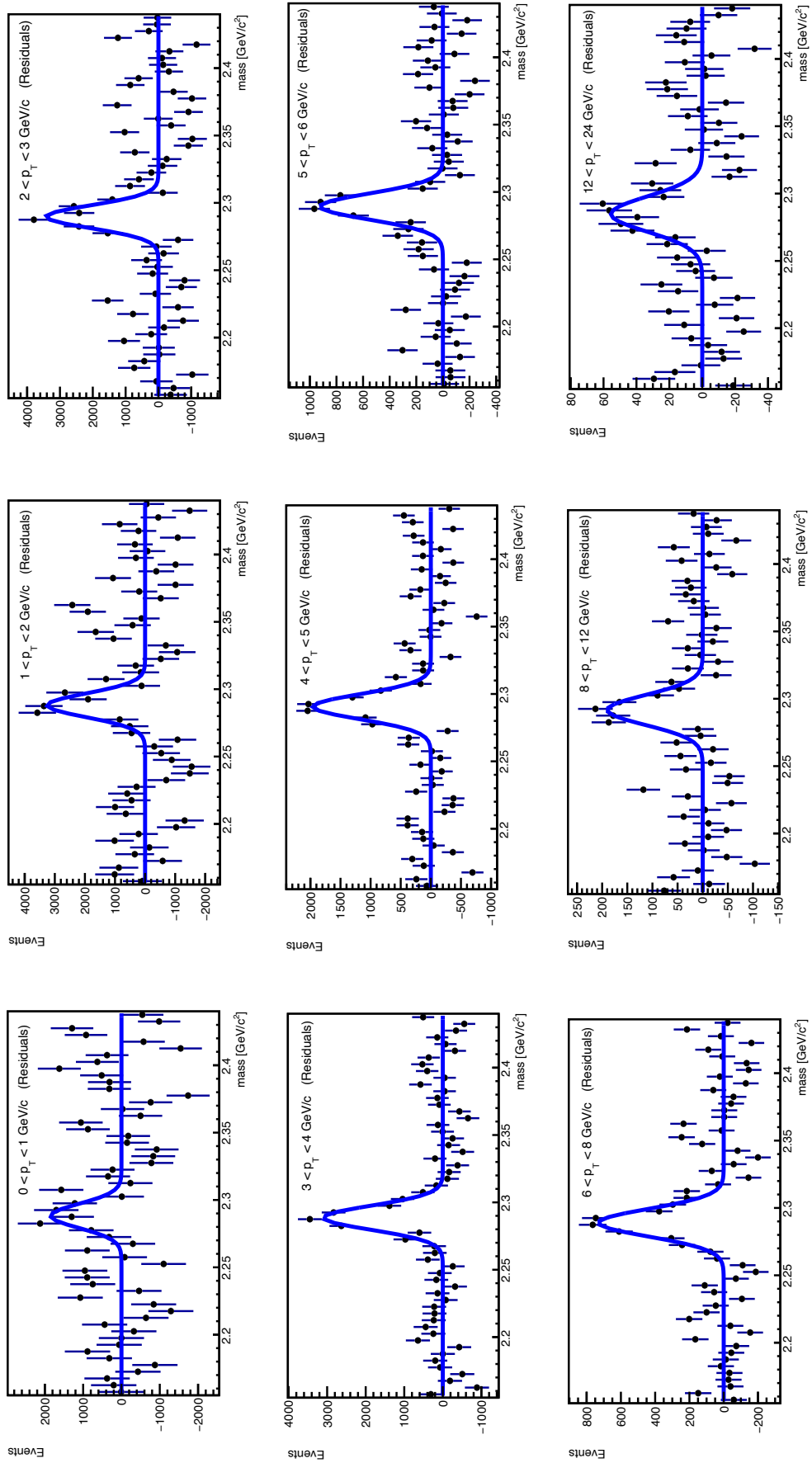**Figure 4.9:** Invariant mass fit for $\Lambda_c^+ \to pK_S^0$.

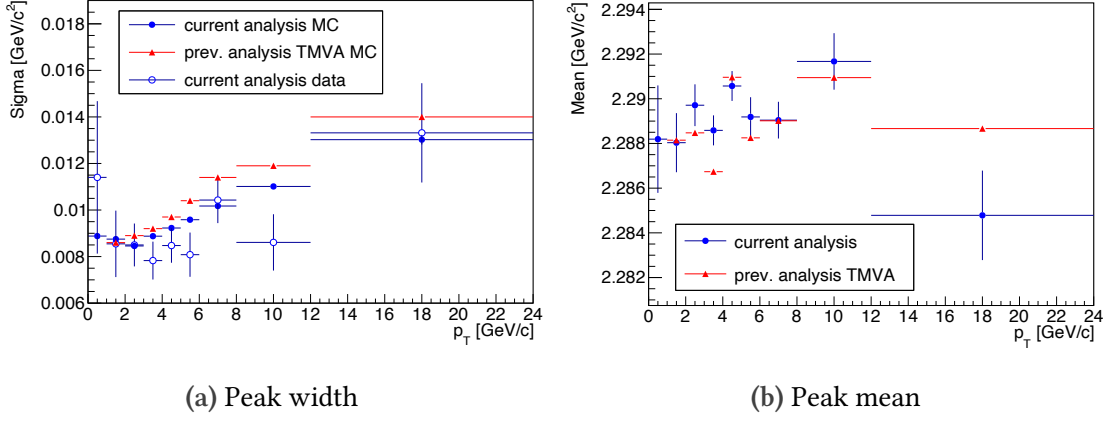**Figure 4.10:** p_T-differential background subtracted residuals.

(a) Peak width

(b) Peak mean

Figure 4.11: Parameters of the signal fit in the invariant mass distribution of the Λ_c^+ as a function of the transverse momentum.
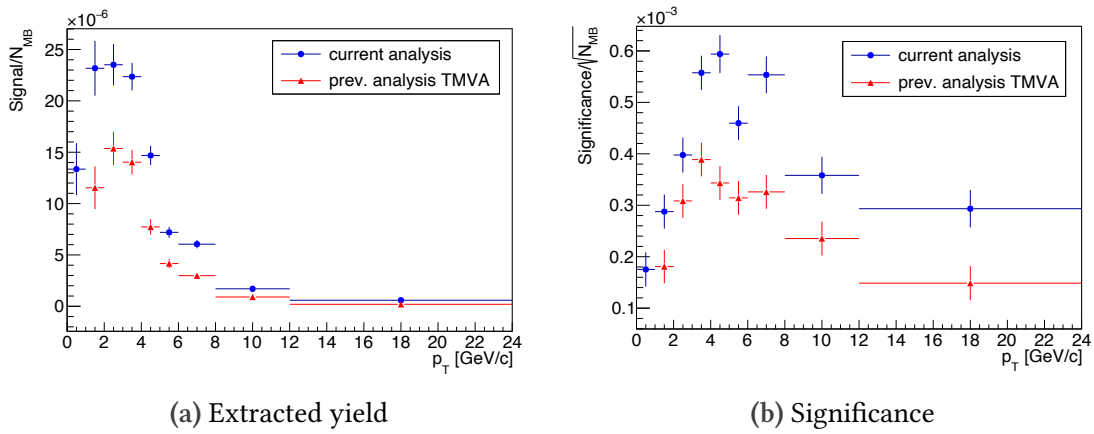

(a) Extracted yield

(b) Significance

Figure 4.12: The extracted signal and significance of the Λ_c^+ as a function of the transverse momentum.

compared to the TMVA analysis. The raw yield in this analysis is on average about 75% higher compared to the previous analysis.

In order to disentangle the improvements due to the KF reconstruction and the use of a different ML algorithms, an analysis with standard rectangular selection criteria instead of training a BDT model is done for the p_T interval $4 < p_T < 5\,\mathrm{GeV}/c$, and compared to the previous standard analysis in the same decay channel. In a machine learning model different selections on variables are combined into one output probability. In the case of rectangular selections, these are applied separately to topological and particle identification variables. With this approach, the influence from ML is excluded from the analysis. The transverse momentum interval is chosen because it shows a very good agreement in the measured cross section between the analyses. In addition, the signal extraction

is stable for both analyses in this interval. The same selection criteria are applied as in the previous standard analysis to make sure that differences do not arise because of different selections. The extracted signal is also higher by about $80\,\%$ compared to a previously performed standard analysis [23]. From this, one can conclude that the observed improvement is driven by the use of the KFParticle package and not by using a different ML algorithm. However, it should be mentioned that also the obtained acceptance times efficiency is higher in this analysis as well. Thus, the calculated cross section is in agreement for both analyses.

4.5 Efficiency Correction

Due to the finite acceptance and efficiency of the detector, and due to the application of prefilter and BDT selections, the extracted signal needs to be corrected for the reconstruction efficiency in order to compute the final cross section. The total reconstruction efficiency is defined as the number of reconstructed candidates over the generated candidates in the acceptance region of the detector. For the determination a separate MC production is used. The reason for this is that the set of simulated candidates, used to train a BDT model, should not be used to determine the reconstruction efficiency, since one factor in the calculation is the efficiency due to the BDT output probability selection. The total reconstruction efficiency is divided into the preselection efficiency and the BDT efficiency. The preselection efficiency is the number of reconstructed candidates after the preselection criteria are applied divided by the generated candidates. This preselection efficiency is influenced by the acceptance and detector efficiency and by the preselection criteria applied after the candidate reconstruction to assure a good track quality or dE/dx resolution. The efficiency is shown in figure 4.13a. It is visible that the applied preselection criteria result in a higher efficiency compared to the previous analysis, except for the last p_T interval $12-24\,\mathrm{GeV}/c$, where the deviation is most likely a fluctuation. In order to obtain the BDT efficiency, the ML model is applied to the simulated candidates and the efficiency is determined by the number of reconstructed candidates over the preselected ones, see figure 4.13b. Figure 4.13c shows the p_T-differential total efficiency factor that is used to calculate the final cross section. All results are compared to the efficiencies in the previous TMVA analysis. The total efficiency in this analysis is on average about $14\,\%$ higher compared to the previous analysis.

Finally, the efficiency for prompt and non-prompt Λ_c^+ are shown in Fig. 4.13d. Prompt candidates are pointing back to the primary vertex since they originate from a charm

quark produced in the initial hard scattering vertex. Feed-down candidates are Λ_c^+ that originate from decays of beauty hadrons, predominantly Λ_b^0. The prompt efficiencies are higher than the non-prompt, because of the $\chi_{\text{topo}}^2/\text{NDF}$ selection of the Λ_c^+. This variable tends to select mainly Λ_c^+ less displaced from the primary vertex, hence prompt Λ_c^+.

(a) Preselection Efficiency

(b) BDT Efficiency

(c) Total $\Lambda_c^+ \to pK_S^0$ efficiency.

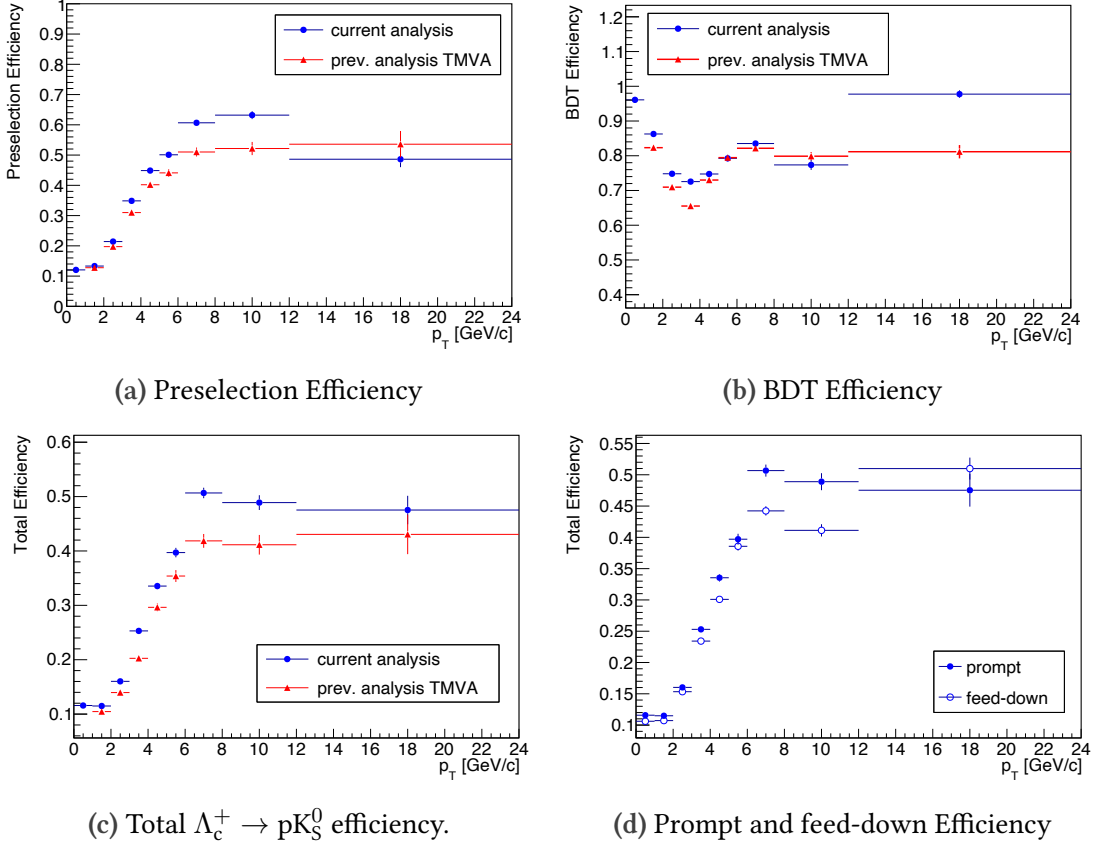(d) Prompt and feed-down Efficiency

Figure 4.13: Comparison of the efficiencies in this analysis and the previous TMVA analysis.

4.6 Feed-Down Subtraction

The p_T-differential final cross section will be reported for prompt Λ_c^+. This is why contributions from hadron decays containing a beauty quark (feed-down) will have to be subtracted. This is accounted for by the factor f_{prompt}, which is the fraction of the raw

yield from prompt Λ_c^+ and it is calculated as in Ref. [23]:

$$
\begin{aligned}
f_{\text{prompt}} &= 1 - \frac{N^{\Lambda_c^+\,\text{feed-down}}}{N_{\text{raw}}^{\Lambda_c^+ + \Lambda_c^-}} \\
&= 1 - \frac{(\text{Acc} \times \varepsilon)_{\text{feed-down}} \cdot \Delta y \cdot \Delta p_{\text{T}} \cdot \text{BR} \cdot \mathscr{L}_{\text{int}}}{N_{\text{raw}}^{\Lambda_c^+ + \Lambda_c^-}/2} \times \left(\frac{d^2\sigma}{dp_{\text{T}}dy}\right)^{\text{FONLL}}_{\Lambda_c^+\,\text{feed-down}},
\end{aligned}
\tag{4.2}
$$

with $N_{\text{raw}}^{N^{\Lambda_c^+ + \Lambda_c^-}}/2$ being the raw yield divided by a factor of two, to account for particles and antiparticles, $(\text{Acc} \times \varepsilon)_{\text{feed-down}}$ is the product of the detector acceptance and the reconstruction efficiency for non-prompt Λ_c^+ baryons. The transverse-momentum and rapidity intervals are determined by $\Delta y \cdot \Delta p_{\text{T}}$. BR is the branching ratio of the Λ_c^+ decay into a proton and a kaon, the kaon decaying into two oppositely charged pions. $\mathscr{L}_{\text{int}}$ is the integrated luminosity of the used data sample. The production cross section of Λ_c^+ from Λ_b^0-baryon decays, $\left(\frac{d^2\sigma}{dp_{\text{T}}dy}\right)^{\text{FONLL}}_{\Lambda_c^+\,\text{feed-down}}$, was calculated using the b-quark p_{T}-differential cross section from Fixed-Order-Next-to-Leading-Log (FONLL) calculations [56], [57]. The fraction of beauty quarks that fragment into Λ_b^0 is estimated from LHCb measurements [58], and the $\Lambda_b^0 \to \Lambda_c^+ + \text{X}$ decay kinematics are modelled using PYTHIA 8 simulations [32], normalised according to the branching ratio $f(\Lambda_b^0 \to \Lambda_c^+ + \text{X}) = 80\%$ taken from PYTHIA 8. A hypothesis on the nuclear modification factor R_{pPb} of Λ_c^+ from beauty-hadron decays is included as an additional factor in the last term of Eq. 4.2. As in the D-meson analysis [59], it was assumed that the R_{pPb} of prompt and feed-down Λ_c^+ are equal. This assumption on the ratio $R_{\text{pPb}}^{\text{feed}-\text{down}}/R_{\text{pPb}}^{\text{prompt}}$ is varied in the range from 0.9 to 1.3 for $p_{\text{T}} > 1\,\text{GeV}/c$ and from 0.9 to 3.0 for $0 < p_{\text{T}} < 1\,\text{GeV}/c$ to account for possible deviations from this assumption. The relative variation of the nuclear modification factor for prompt Λ_c^+ as a function of this hypothesis is shown in the left panel of Fig. 4.14. The values of f_{prompt} range between $90\,\%$ and $99\,\%$ and are shown in the right panel of Fig. 4.14. This prompt fraction is multiplied as a factor in the calculation of the final cross section.

4.7 Systematic Uncertainties

In this section, the systematic uncertainties are evaluated and the methods used to estimate them are described. A summary of all systematic uncertainties is provided in table 4.4. Since all uncertainties are assumed to be uncorrelated, the overall systematic uncertainty for each p_{T} interval is calculated by adding all contributions in quadrature.
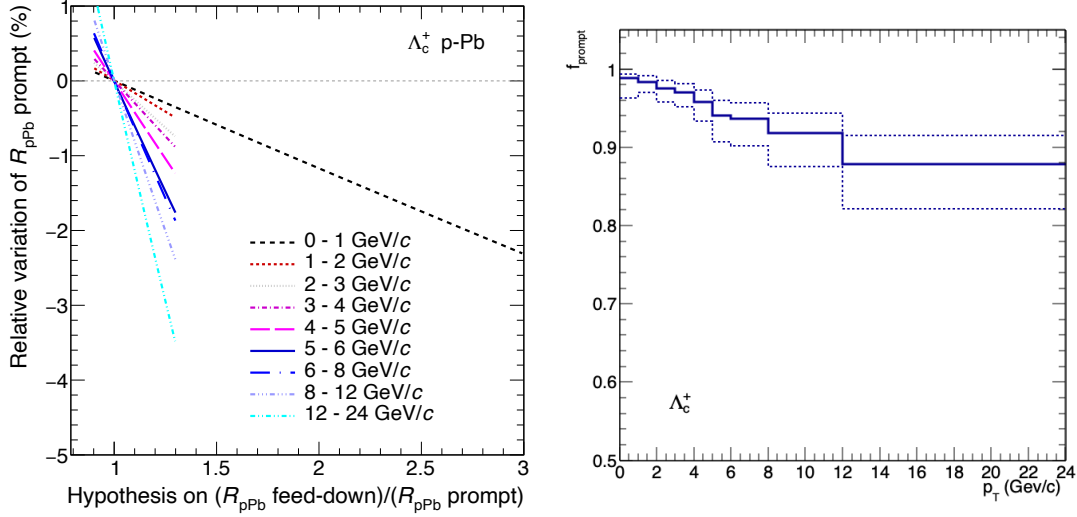
Figure 4.14: Left: Relative variation of R_{pPb} for prompt Λ_c^+ for different hypotheses of $R_{\mathrm{pPb}}^{\mathrm{feed-down}}/R_{\mathrm{pPb}}^{\mathrm{prompt}}$. Right: Prompt fraction of Λ_c^+ baryons.

The uncertainty on the luminosity measurement is not added and reported separately.

4.7.1 Systematic Uncertainty on Yield Extraction

The extracted raw yield, might depend on the chosen parameters of the fit, like the fit range or the background function. The systematic uncertainty on the raw yield extraction is obtained by repeating the fit of the invariant mass distribution with some variations, one of them being the background function. Three different functions are considered ranging from a polynomial of second order over third to fourth order. The upper and lower limits of the invariant mass fitting range are shifted to larger and smaller values with respect to the nominal case. The peak width of the Gaussian fit for the signal is kept as a fix parameter of the fit in all cases, but the sigma is increased and decreased by $10\,\%$ of its nominal value. In addition, the binning of the invariant mass histogram is changed combining two bins into one.

From all these trials the raw yield is extracted and a distribution with the resulting values is built. Figure 4.15 shows an example of this distribution in the p_{T} interval $0-1\,\mathrm{GeV}/c$. The dark blue histogram shows the yield distribution from all trials. The properties of this distribution, like the mean, the median, the root mean square (RMS), the minimum and the maximum are written in black next to the figure on the right. The reference of the yield extracted with the default settings is added as a vertical red line, which should ideally be centred in the raw yield distribution. If the reference yield from the default
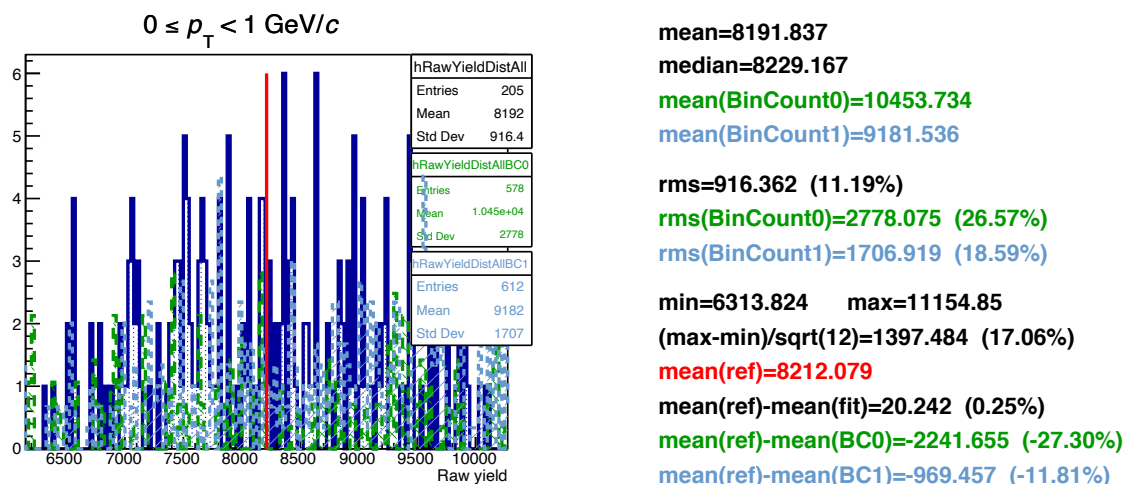
Figure 4.15: The RMS is taken as the systematic uncertainty on the yield extraction from the multitrial fit approach in $0 < p_T < 1\,\mathrm{GeV}/c$.

settings lies outside or at the edge of the distribution, this means that the chosen fit parameters are not optimal and should be reevaluated. Another method to check the raw yield extraction is the so-called bin-counting, where the extracted yield is not taken from a fit but from the sum over the raw yield in each bin after the background function was subtracted. There are two different bin-counting methods called BC0 and BC1, shown in Fig. 4.15, which deviate by a different background estimation. The BC0 method is the green histogram and the corresponding properties are listed in the green text. It uses the initial background fit that incorporates just the sidebands in the mass spectrum. The BC1 method, which is the light blue histogram, uses the background component of the final signal plus background fit. These bin-counting methods are not used for the estimation of the systematic uncertainty but they function as a verification that the extracted yield agrees with the signal from the fit of the invariant mass spectrum. If a large deviation is observed, it is a hint to some fluctuation in the signal or in the background and further investigations have to be made. The assigned systematic uncertainty is the RMS of the dark blue distribution, which is $11\,\%$ for the p_T interval shown in Fig. 4.15.

4.7.2 BDT Probability Selection

Another systematic uncertainty arises due to the application of a BDT probability selection. Possible differences in the MC distributions with respect to the data of the features included in the BDT model can introduce a systematic uncertainty, which is estimated

(a) BDT variation range.
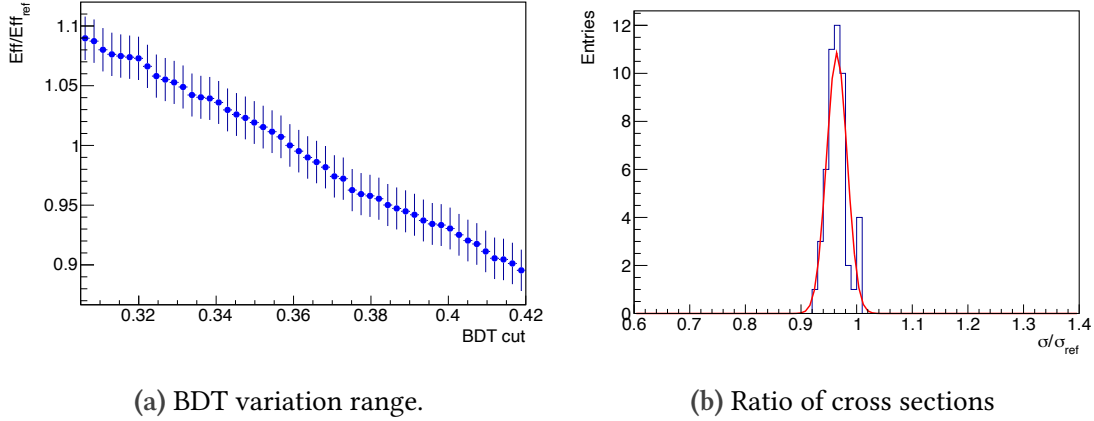
(b) Ratio of cross sections

Figure 4.16: The systematic uncertainty on the BDT selection is determined by the distribution of the ratio of cross sections, calculated for a range of BDT output probabilities over the final cross section, shown here for the p_T range $4 - 5\,\mathrm{GeV}/c$.

by varying the BDT probability selection and recomputing the cross section for each BDT selection. If the simulations describe the real data perfectly, the uncertainty will be negligible because the calculated efficiency accounts for losses in the signal extraction. The BDT selection is varied in the plateau region of the defined working point. This range corresponds to an efficiency variation of $\pm 10\,\%$, as shown in figure 4.16a. The uncertainty is determined by the ratio of the cross section as a function of the BDT selection over the final cross section. The resulting distribution is shown in figure 4.16b and fitted with a Gaussian function. The systematic uncertainty is obtained by adding in quadrature the width and shift of the distribution.

4.7.3 PID Preselection

The particle identification efficiency in simulation and data might differ. For transverse momenta higher than $1\,\mathrm{GeV}/c$ the uncertainty arising from the PID selection is already accounted for by the variation of the BDT response because the combined PID enters as a training variable. For $p_\mathrm{T} < 1\,\mathrm{GeV}/c$, Λ_c^+ are selected with the requirement that the combined PID probability is larger than $20\,\%$ and this variable is removed from the BDT training, because it caused some instabilities in the final cross section. In order to estimate the effect of the Bayesian PID probability selection it varied between $10\,\%$ and $30\,\%$, and the ratio of the cross section with respect to the nominal cross section is computed. The result is shown in figure 4.17a, and from the ratio a $5\,\%$ systematic uncertainty is assigned. Figure 4.17b shows the respective ratios of preselection efficiencies. The variation is small with $\pm 1\,\%$.

45

(a) Ratio of cross sections
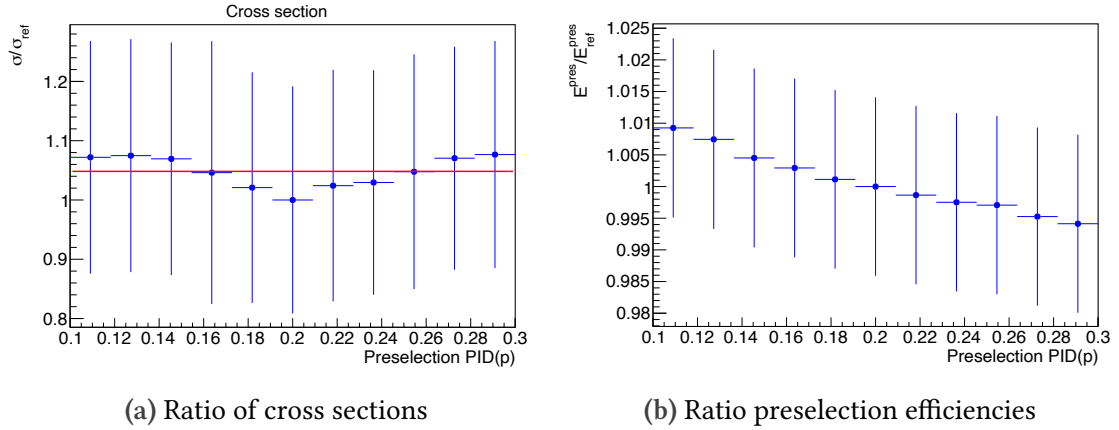
(b) Ratio preselection efficiencies

Figure 4.17: The systematic uncertainty on the PID preselection is determined by the ratio of cross sections calculated for a range of combined bayesian PID probabilities of the proton for the p_T interval $0 - 1$ GeV/c.

4.7.4 MC p_T Shape

The efficiencies also depend on the p_T distribution of generated Λ_c^+ in simulations. An additional systematic uncertainty is assigned to the discrepancy between the generated and true shape of the transverse momentum distribution. In order to estimate the effect on the total efficiency, the p_T shape of the simulations is reweighed to match the spectrum of D mesons from FONLL perturbative QCD calculations. Two different generated MC samples were used in this analysis and both are reweighed based on a FONLL prediction. The weights are computed, dividing the predicted FONLL spectra by those spectra obtained from the PYTHIA MC simulation. Figure 4.18a shows the different spectra. The blue markers indicate the default spectrum and the magenta and orange markers show the reweighed p_T spectra based on the FONLL calculations. A third hypothesis is a flat p_T shape, although this shape is not expected to reproduce the true shape. The percentage change in efficiency with weighed spectra compared to the unweighed efficiency is studied in figure 4.18b. The overall deviation is less than 1%, independent of p_T.

4.7.5 Tracking and Matching Efficiency

The systematic uncertainty on the particle tracking includes two components. One is an uncertainty on the matching efficiency between the Time Projection Chamber and the Inner Tracking System. The matching efficiency is defined as the number of tracks with clusters in both detectors over the number of tracks with clusters in the TPC. An uncertainty arises due to discrepancies in data and simulation when propagating a track
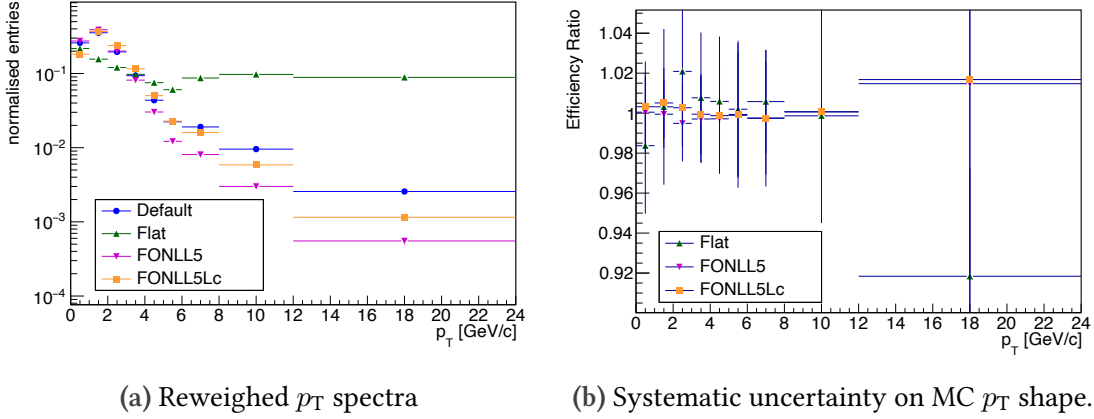
(a) Reweighed p_T spectra

(b) Systematic uncertainty on MC p_T shape.

Figure 4.18: Ratio of the prompt efficiency calculated for different weighting functions with respect to the default PYTHIA.

from the TPC to the ITS. In this analysis, since there is no requirement on the ITS for the pions from the K_S^0 decay, only the proton matching efficiency needs to be included. The second uncertainty is assigned to the track quality. The cross section is compared for different sets of track selections, which can be the number of crossed rows or clusters in the TPC, plus additional selections as mentioned in section 4.1. Since the reconstruction is the same as in the previous analysis [23] the uncertainties on the tracking and matching efficiency are inherited. The transverse momentum dependence of the systematic uncertainty is not significant, this is why in this analysis the uncertainty in $0 < p_T < 1\,\mathrm{GeV}/c$ is propagated from the previous Λ_c^+ results in the interval $1-2\,\mathrm{GeV}/c$ and is $6\,\%$.

4.7.6 Prompt Fraction

For the estimation of the uncertainty on the prompt fraction the same strategy as in the previous analysis is used and a more detailed description can be found in Ref. [23]. The systematic uncertainty on the prompt fraction, as calculated in Eq. 4.2, is estimated from the envelope of f_{prompt} values. Different prompt fractions are obtained from varying the parameters of the FONLL prediction of the beauty and charm production cross sections [56]. These include variations of the factorisation and renormalisation scale, the mass of the heavy quark and uncertainties associated with the Parton Distribution Functions (PDFs). In addition, the free parameters describing the fragmentation fraction of $b \to \Lambda_b^0$ [58] are varied. Finally, the hypothesis on the nuclear modification factor (see section 4.6) of Λ_c^+ from beauty hadron decays is changed. The ratio of $R_{\mathrm{pPb}}^{\mathrm{feed-down}}/R_{\mathrm{pPb}}^{\mathrm{prompt}}$ in the

prompt fraction calculations is assumed to be equal to unity. In order to determine the systematic uncertainties, this ratio is varied in the range from 0.9 to 1.3 for $p_T > 1\,\text{GeV}/c$, which is the range used in the previous analysis based on theoretical predictions of the charm hadron production [59], and from 0.9 to 3.0 for $0 < p_T < 1\,\text{GeV}/c$, which is based on the recent measurement of the $R_{\text{pPb}}(\text{B}^+)$ and $R_{\text{pPb}}(\text{B}^+/\Lambda_b^0)$ by LHCb [60].

4.7.7 Branching Ratio and Luminosity

An analysis independent uncertainty is the one on the luminosity $\mathscr{L}_{\text{int}} = 287\mu\text{b}^{-1}$ measured in van der Meer scans, and it is $3.7\,\%$ for p–Pb collisions, reported in [51].

The uncertainty on the branching ratio is the same for all p_T intervals and calculated from the individual components of the branching fraction of $\Lambda_c^+ \to \text{pK}_S^0$ $(1.59 \pm 0.08)\%$ and of $\text{K}_S^0 \to \pi^+\pi^-$ $(69.20 \pm 0.05)\%$ [4]. Thus, the branching fraction for the hadronic decay $\Lambda_c^+ \to \text{pK}_S^0$ $(\text{K}_S^0 \to \pi^+\pi^-)$ is $(1.10 \pm 0.06)\%$.

Table 4.4: Systematic uncertainties.

p_T (GeV/c)	0-1	1-2	2-3	3-4	4-5	5-6	6-8	8-12	12-24
Yield extraction [%]	11	10	7	7	6	6	6	6	8
BDT selection [%]	6	6	4	4	4	4	5	5	8
PID(p) efficiency [%]	5	–	–	–	–	–	–	–	–
MC p_T shape [%]	1	1	1	1	1	1	1	1	1
Tracking efficiency [%]	6	6	5	5	4	4	5	5	5
Feed-down [%]	+0.6	+0.8	+1.1	+1.2	+1.5	+2.1	+2.2	+2.8	+4.1
	-2.5	-1.3	-1.8	-1.9	-2.5	-3.5	-3.6	-4.6	-6.5
Branching ratio [%]	5	5	5	5	5	5	5	5	5
Luminosity [%]	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7

5 Results and Discussion

5.1 Cross Section

In this analysis the measurement of the production cross section of the prompt Λ_c^+ baryon together with its charge conjugate is presented. It is obtained at midrapidity in p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ in the range $0 < p_{\mathrm{T}} < 24\,\mathrm{GeV}/c$. The Λ_c^+ was reconstructed via the hadronic decay channel $\Lambda_c^+ \to \mathrm{pK_S^0}$ ($\mathrm{K_S^0} \to \pi^+\pi^-$), which has a branching ratio of BR $= (1.10\,\% \pm 0.06\,\%)$ [4]. The cross section is computed according to:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}p_{\mathrm{T}}\mathrm{d}y} = \frac{1}{2}\frac{1 \cdot f_{\mathrm{prompt}}(p_{\mathrm{T}}) \cdot N_{\mathrm{raw}}^{\Lambda_c}(p_{\mathrm{T}})}{\Delta y \cdot \Delta p_{\mathrm{T}} \cdot (\mathrm{Acc} \times \varepsilon)_{\mathrm{prompt}}(p_{\mathrm{T}}) \cdot \mathrm{BR} \cdot \mathscr{L}_{\mathrm{int}}}, \tag{5.1}$$

where f_{prompt} is the fraction of prompt Λ_c^+, $N_{\mathrm{raw}}^{\Lambda_c}$ is the raw yield, $(\mathrm{Acc} \times \varepsilon)_{\mathrm{prompt}}$ is the prompt efficiency correction factor, BR is the branching ratio, and $\mathscr{L}_{\mathrm{int}}$ is the integrated luminosity. The factor 2 accounts for particles and antiparticles, and $\Delta y \cdot \Delta p_{\mathrm{T}}$ accounts for the width of the rapidity and transverse momentum interval. The obtained Λ_c^+ p_{T}-differential cross section is shown with the red markers in Fig. 5.1. Thanks to the improved reconstruction with the KFParticle package it was possible to extract a signal down to $p_{\mathrm{T}} = 0$. This was not possible in previous analyses in ALICE because the signal over background ratio is low at low transverse momenta. Thus, this is the first evaluation of a Λ_c^+ cross section in p–Pb collisions down to $p_{\mathrm{T}} = 0$. This extension of the transverse momentum range is important, since about $20\,\%$ of the p_{T}-integrated cross section is contained in the range $0 < p_{\mathrm{T}} < 1\,\mathrm{GeV}/c$. Until now, an extrapolation based on model calculations was necessary to obtain the total Λ_c^+ cross section. This extrapolation introduces a model dependence and an uncertainty, which can be removed with a measurement. This is a significant achievement of this analysis.

This is one of the first analyses employing the KFParticle software package and the machine learning algorithm XGBoost. Thus, the results are compared to the Λ_c^+ cross section obtained previously in ALICE [23], shown with the black markers in Fig. 5.1. The previous measurement is obtained from the average of three analyses, the first two being the cross section evaluated in the $\Lambda_c^+ \to \mathrm{pK_S^0}$ channel, with and without using machine learning, and a third analysis of $\Lambda_c^+ \to \mathrm{pK}^-\pi^+$. The measurements are in good agreement in the full p_{T} range. The deviation between this analysis and the average of previous analyses at $p_{\mathrm{T}} > 12\,\mathrm{GeV}/c$ is explained as a fluctuation in the preselection efficiency due to only a small MC sample being available in this region in this analysis
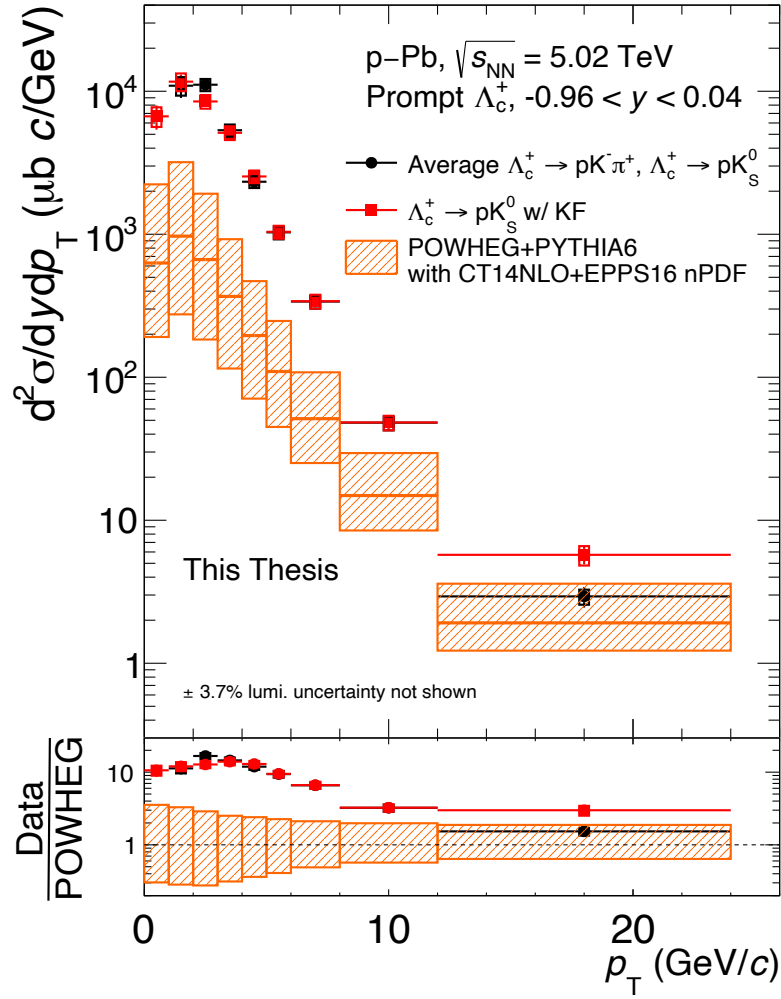
Figure 5.1: Prompt $\Lambda_c^+ \rightarrow pK_S^0$ cross section measured in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in the transverse momentum range $0 < p_T < 24$ GeV/c using the Kalman Filtering Package and XGBoost, compared to the average of previous analyses [23]. The result is compared to the POWHEG event generator [61]. The theoretical uncertainties are represented by the orange boxes.

(see Fig. 4.13a). This interval will be further investigated in the future with a larger MC sample. In the transverse momentum range $2 - 3$ GeV/c the cross section is about $25\,\%$ lower than previously measured average but the results agree within 2σ. A comparison of the uncertainties shows that the statistical uncertainties are similar to the average of previous analyses, even though this analysis uses only one decay channel (20% of the BR [4]). This is explained by the higher signal extraction, when using the KFParticle package. Since the improvement could also be due to the use of a new machine learning algorithm, a standard analysis without using machine learning was performed addition-

ally in one p_T interval, as described in section 4.4. It showed that the extracted signal and efficiency increases by about $75\,\%$ with respect to the previous analysis when using the KF package. This shows the power of this new reconstruction method and that if it would be applied to all decay channels, it offers the opportunity to improve the measurement precision. The systematic uncertainties are $1-2\,\%$ higher than for the previous analysis, due to the cancellation of correlated uncertainty sources when averaging the separate decay channels.

The results are also compared to a model based on the factorisation approach, where the hadron cross section is calculated as a convolution of the parton distribution functions, the partonic cross section and the fragmentation functions. This is done in a next-to-leading order (NLO) pQCD calculation obtained with the POWHEG event generator [61]. The theoretical uncertainties are estimated from variations with respect to the central prediction of the factorisation and renormalisation scale μ_F and μ_R in the range $0.5\mu_0 < \mu_{R,F} < 2.0\mu_0$, with $\mu_R = \mu_F = \mu_0 = \sqrt{m^2 + p_T^2}$. The mass of the charm quark is set to $m_c = 1.5\,\mathrm{GeV}/c$. PYTHIA 6 is used for the parton shower generation with the parton distribution functions CT14NLO [62]. The nuclear modifications of the PDFs in p−Pb collisions are modelled with the EPPS 16 nPDF parametrisation [25]. Uncertainties on these modifications are smaller than the scale uncertainties. The central prediction of the model underestimates the cross section up to a factor of 14 at $p_T < 4\,\mathrm{GeV}/c$, corresponding to a deviation of up to 8.5σ, and it underestimates the data by at least a factor of 3 at $p_T > 8\,\mathrm{GeV}/c$, corresponding to a deviation of 2.2σ. Since the cross section is calculated from a convolution of the above mentioned terms, the investigation of the cross section is not enough to draw a conclusion about the origin of this deviation. This is the reason why the baryon-to-meson ratio, described in the next section is an important quantity, since model ingredients like the partonic cross section cancel in the ratio and remaining are the fragmentation fractions of the two species.

Furthermore, a measurement of the p_T-integrated cross section will be used, together with other charm baryons and the D mesons, to calculate the total $c\bar{c}$ cross section in p−Pb collisions. For the evaluation of the total Λ_c^+ cross section at $\sqrt{s_{NN}} = 5.02\,\mathrm{TeV}$ in the previous analysis the measured visible cross section was extrapolated in the range $p_T < 1\,\mathrm{GeV}/c$ and $p_T > 24\,\mathrm{GeV}/c$ [23], and reported to be:

$$\mathrm{d}\sigma_{\mathrm{pPb}}^{\Lambda_c^+}/\mathrm{d}y = 38.7 \pm 2.7(\mathrm{stat.}) \pm 4.2(\mathrm{syst.}) \pm 1.4(\mathrm{lumi.})^{+4.4}_{-2.7}(\mathrm{extrap.})\mathrm{mb}. \qquad (5.2)$$

The visible Λ_c^+ cross section is computed also in this analysis by integrating over p_T in the range $0 < p_T < 24\,\mathrm{GeV}/c$. All sources of systematic uncertainties are assumed to be fully correlated between the transverse momentum intervals, except the uncertainty on the raw-yield extraction, which is assumed to be fully uncorrelated. The visible Λ_c^+ cross section, measured in this analysis, in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\mathrm{TeV}$ in the rapidity interval $-0.96 < y < 0.04$ is:

$$\mathrm{d}\sigma_{\mathrm{pPb}}^{\Lambda_c^+}/dy|_{0<p_T<24\,\mathrm{GeV}/c} = 36.5 \pm 2.4(\mathrm{stat.}) \pm 4.2(\mathrm{syst.}) \pm 1.4(\mathrm{lumi.})\mathrm{mb}. \qquad (5.3)$$

The result agrees within the uncertainties with the total p_T-integrated cross section measured previously. With a measurement in the range $0 < p_T < 1\,\mathrm{GeV}/c$ the extrapolation uncertainty is removed, which allows to reduce also the uncertainty on the total $c\bar{c}$ cross section.

5.2 Baryon-to-Meson Ratio

The baryon-to-meson ratio is sensitive to the hadronisation process and thus to the charm fragmentation function, which is typically tuned on e^+e^- data and assumed to be the same independently of the collision system. The Λ_c^+ to D^0-meson ratio is calculated by dividing the p_T-differential cross sections of the two particle species. The D^0 meson cross section is taken from Ref. [63]. The upper panel of Fig.5.2 shows a comparison of the result obtained in this analysis (red markers) to the previously obtained ratio (black markers). The sources of uncertainty assumed to be uncorrelated in the ratio between the Λ_c^+ and the D^0 analysis are the raw yield extraction, the BDT probability selection efficiency, the MC p_T shape, the particle identification, and the branching ratio. The sources assumed to be correlated include the tracking uncertainty and the feed-down subtraction uncertainty. The uncorrelated uncertainties are summed in quadrature, and the fully correlated uncertainties partially cancel in the ratio. The result obtained in this analysis, and the Λ_c^+/D^0 ratio measured in the average of previous analyses agree within the uncertainties. The baryon-to-meson ratio shows an increasing trend for $p_T < 4\,\mathrm{GeV}/c$ and a decreasing trend for $p_T > 4\,\mathrm{GeV}/c$. The maximum value is $\Lambda_c^+/D^0 \approx 0.6$ and the smallest ratio is approximately 0.3 at low and high p_T. The p_T-integrated ratio, measured in e^+e^- collisions at LEP at $\sqrt{s} = 91.2\,\mathrm{GeV}$ is $\Lambda_c^+/D^0 = 0.113 \pm 0.013(\mathrm{stat.}) \pm 0.006(\mathrm{syst.})$ [23]. The integrated Λ_c^+/D^0 ratio measured in this analysis in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\mathrm{TeV}$ in the transverse momentum

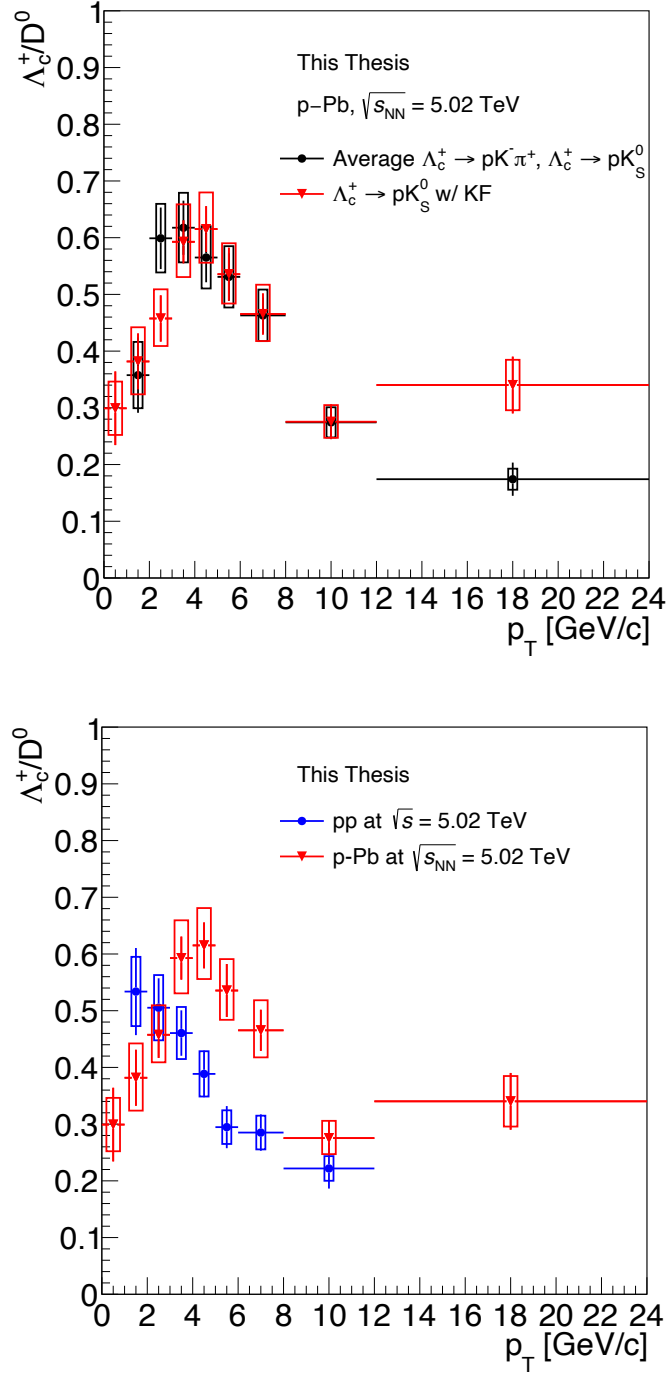Figure 5.2: Λ_c^+ to D^0-meson ratio. Left: Measured in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in this thesis is compared to the average of previous analyses [23]. Right: Ratio in p–Pb collisions in this thesis compared to the ratio in pp collisions measured previously at $\sqrt{s} = 5.02$ TeV.

range $0 < p_T < 24\,\text{GeV}/c$ is:

$$\left(\frac{\Lambda_c^+}{D^0}\right)_{\text{pPb}} = 0.412 \pm 0.030(\text{stat.}) \pm 0.045(\text{syst.}). \tag{5.4}$$

Thus, an enhancement by a factor of ~ 4 is observed in p–Pb compared to e^+e^- collisions. The result indicates that the fragmentation of charm quarks into hadrons depends on the collision system and it is not a universal process as assumed in the factorisation approach.

The lower panel of Fig. 5.2 shows a comparison of the Λ_c^+/D^0 ratio obtained in this analysis in p–Pb collisions and the average of a previous analyses in pp collisions [23]. The ratio in p–Pb collisions exhibits a maximum in the range $4 - 6\,\text{GeV}/c$. The maximum of the baryon-to-meson ratio in pp collisions seems to be at lower transverse momenta. The observation can be seen as a hint to collective effects in p–Pb collisions by a formation of QGP droplets [64] because hydrodynamical models predict a hardening of p_T spectra (radial flow) in the presence of a medium due to an additional common expansion velocity [31]. However, this interpretation is still under debate.

5.3 Nuclear Modification Factor

The nuclear modification factor R_{pPb} of the Λ_c^+ in minimum-bias p–Pb collisions is defined as:

$$R_{\text{pPb}} = \frac{\text{d}\sigma_{\text{pPb}}/\text{d}p_T}{A \cdot \text{d}\sigma_{\text{pp}}/\text{d}p_T} \tag{5.5}$$

where $\text{d}\sigma_{\text{pPb(pp)}}/\text{d}p_T$ are the p_T- differential cross sections in p–Pb and pp collisions at a given centre-of-mass energy and $A = 208$ is the lead mass number. The reference cross section $\text{d}\sigma_{\text{pp}}/\text{d}p_T$ is measured in pp collisions at the same centre-of-mass energy in the transverse momentum range $1 - 12\,\text{GeV}/c$, applying a correction factor for the different rapidity coverage in the pp and p–Pb analyses. In the transverse momentum interval $0 - 1\,\text{GeV}/c$ the pp cross section is extrapolated using PYTHIA 8 calculations [23]. The nuclear modification factor, measured for Λ_c^+ in this analysis, is shown in the left panel of Fig. 5.3 in the range $0 < p_T < 12\,\text{GeV}/c$ with the red markers. It is compared to the nuclear modification factor of non-strange D mesons (average of D^0, D^+, D^{*+} in the range $1 < p_T < 12\,\text{GeV}/c$, and D^0 in the range $0 - 1\,\text{GeV}/c$) [63]. For non strange D mesons an agreement of the nuclear modification factor with unity within the uncertainties is observed over the full p_T spectrum. For Λ_c^+ baryons a suppression is observed at
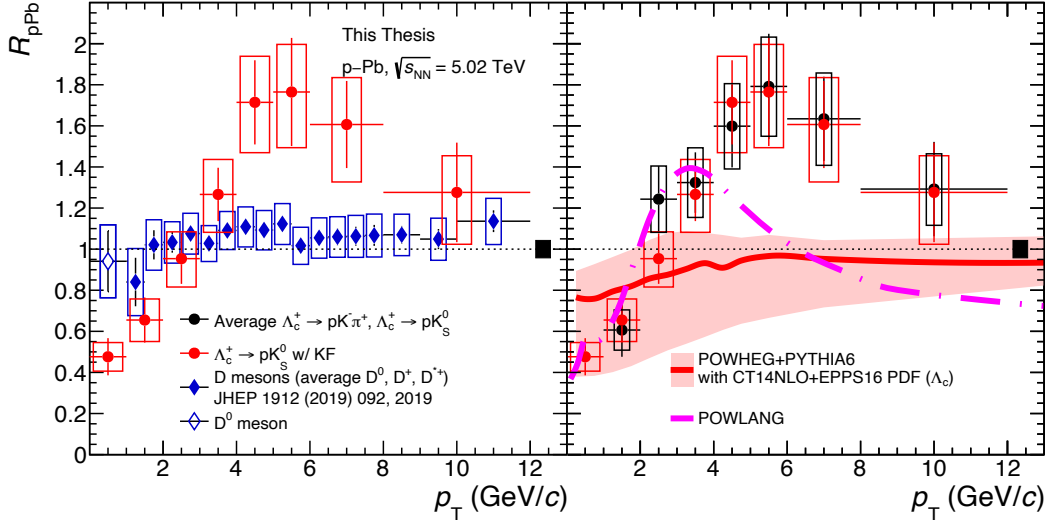
Figure 5.3: Nuclear modification factor R_{pPb} for prompt Λ_c^+ in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV in the transverse momentum range $0 < p_T < 12$ GeV/c. Left: Comparison to non-strange D mesons [63]. Right: Comparison of the measurement using the KFPackage and XGBoost to the average of previous analyses [23] and to POWHEG + PYTHIA 6 [61], POWLANG model calculations [65].

$p_T < 2$ GeV/c, while an enhancement is seen in the intermediate range from 3 to 8 GeV/c with respect to pp collisions. The R_{pPb} deviates from unity by $4.6\,\sigma$ in $0 < p_T < 1$ GeV/c and by $3.1\,\sigma$ in the range $1 < p_T < 2$ GeV/c, where σ is defined as the statistical and the systematic uncertainties summed in quadrature. This behaviour is consistent with the observation of the shift of the baryon-to-meson ratio in p–Pb collisions to higher p_T, shown in the right panel of Fig. 5.2. The suppression at low p_T and the enhancement in the intermediate transverse momentum range was also observed for strange baryons with a strangeness and mass ordering [66], and it is predicted by hydrodynamical models assuming radial flow effects [67].

The right panel of Fig. 5.3 shows the comparison of the evaluated nuclear modification factor to the average of previous analyses (black markers). An agreement is observed in the full transverse momentum spectrum within the uncertainties. In addition, the measured R_{pPb} is compared to model calculations. One model calculation is based on the POWHEG event generator combined with PYTHIA 6 as described in section 5.1. Although this model underestimated the Λ_c^+ cross section in p–Pb collisions in the intermediate transverse momentum range $4 < p_T < 8$ GeV/c, the measured nuclear modification factor is in agreement with this model prediction within the uncertainties. The reason is that the model also underestimates the pp cross section by about the same factor [23].

However, the shape of the distribution is not described by this model. A second model is the POWLANG model [65]. In this model initial-state effects are taken into account for the $c\bar{c}$ production. The system is expected to thermalise very rapidly into a deconfined phase, similar to the QGP formation in Pb–Pb collisions. The transport of heavy quarks is then treated in a hydrodynamic approach, but no specific hadronisation mechanism, e.g. quark recombination, is implemented in the model. Although both models reproduce features observed in data, they do not seem to reproduce the trend in the full p_{T} spectrum.

Additionally, the integrated nuclear modification factor was calculated in this analysis in the range $0 < p_{\mathrm{T}} < 12\,\mathrm{GeV}/c$ for p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$:

$$R_{\mathrm{pPb}} = 0.761 \pm 0.063(\mathrm{stat.}) \pm 0.109(\mathrm{syst.})^{+0.010}_{-0.013}(\mathrm{extrap.}). \qquad (5.6)$$

If there were no modification in p–Pb collisions compared to pp collisions the nuclear modification factor would be expected to be consistent with unity. The result agrees with this hypothesis within 2σ. The uncertainties are too large to claim a significant deviation from unity, and to conclude on a modification in p–Pb collisions compared to pp collisions. So far, the integrated reference cross section in pp collisions was obtained by using a measurement in the range $1 - 12\,\mathrm{GeV}/c$ and an extrapolation at $p_{\mathrm{T}} < 1\,\mathrm{GeV}/c$. A measurement of Λ_{c}^{+} in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$ would be desirable and will be carried out in the future to eliminate this model dependent source of uncertainty.

6 Conclusion and Outlook

The goal of this thesis was to apply new methods, like the KFParticle software package and machine learning, to improve and extend physics measurements, in a way which is relevant to draw conclusions about the charm quark hadronisation mechanisms in different collision systems, or to understand how the possible presence of collective effects could modify the production of heavy-flavour hadrons. In this thesis, the production cross section of the prompt charmed baryon $\Lambda_c^+ \to pK_S^0$ and its charge-conjugate was measured at midrapidity in p–Pb collisions at $\sqrt{s_{NN}} = 5.02\,\text{TeV}$ in the transverse momentum range $0 < p_T < 24\,\text{GeV}/c$ with the ALICE detector at the LHC. Thanks to the reconstruction with the KFParticle package it was possible to perform the first measurement of the Λ_c^+ cross section down to $p_T = 0$, and to extract a signal with 75% higher efficiency compared to the previous analysis. The comparison of the results to the previously measured average of $\Lambda_c^+ \to pK_S^0$ and $\Lambda_c^+ \to pK^-\pi^+$ shows a good agreement in the full p_T spectrum. In this analysis also the integrated Λ_c^+ cross section in the range $0 < p_T < 24\,\text{GeV}/c$ is reported and in agreement with the total previously obtained cross section. The uncertainty on the integrated cross section was reduced, since a measurement was performed in the range $0 < p_T < 1\,\text{GeV}/c$ and no extrapolation, based on theoretical models, was necessary for $p_T < 1\,\text{GeV}/c$.

A comparison of the p_T-differential Λ_c^+ cross section and the NLO pQCD model calculation, obtained with the POWHEG event generator used with PYTHIA 6, shows that the model underestimates the data by a factor of up to 14 at $p_T < 4\,\text{GeV}/c$. In this model the charm fragmentation functions are tuned on data from e^+e^- collisions, since they are assumed to be universal for all collision systems. In order to draw a conclusion about the universality of the fragmentation functions, the baryon-to-meson ratio was obtained. For the first time in ALICE this measurement was performed down to $p_T = 0$ and shows a good agreement with the previous analysis within the uncertainties for $p_T > 1\,\text{GeV}/c$. The Λ_c^+/D^0 ratio reaches a maximum of ≈ 0.6 in the range $3 < p_T < 5\,\text{GeV}/c$ and is ≈ 0.3 at $p_T < 1\,\text{GeV}/c$ and $p_T > 8\,\text{GeV}/c$. The p_T integrated ratio is enhanced by a factor of ~ 4 compared to e^+e^- collisions at LEP at $\sqrt{s} = 91.2\,\text{GeV}$, indicating that the fragmentation of charm quarks into hadrons is not universal for all collision systems. Furthermore, a comparison of the baryon-to-meson ratio in pp and p–Pb collisions showed that for the latter the distribution is shifted to higher transverse momenta, which can be seen as a hint to radial flow, predicted in the presence of a medium by hydrodynamical models.

Finally also the nuclear modification factor R_{pPb} of the Λ_{c}^+ in p–Pb collisions is measured in the range $0 - 12\,\mathrm{GeV}/c$. The nuclear modification agrees with unity for $p_{\mathrm{T}} > 2\,\mathrm{GeV}/c$ within the uncertainties. However, the interval $0 - 2\,\mathrm{GeV}/c$ a deviation of up to 4.6σ is observed, suggesting a suppressed Λ_{c}^+ production in p–Pb compared to pp collisions and an enhancement is observed at intermediate p_{T}. This is mostly driven by the observation of the hardening of p_{T} spectra in p–Pb collisions compared to pp collisions.

Since the Λ_{c}^+ cross section was not measured in pp collisions in the range $0 < p_{\mathrm{T}} < 1\,\mathrm{GeV}/c$, an extrapolated value was used for the calculation of the nuclear modification factor. Thus, the evaluation of the cross section in pp collisions at transverse momenta below $1\,\mathrm{GeV}/c$ would be desirable and will be carried out in the future.

The increase in signal extraction in this measurement, using the KFParticle package, gives hope for other analyses with a more complex decay chain, like the Ω_{c}^0 or Ξ_{c}^+ that are currently studied, to extract a signal with a good mass resolution. Especially for the reconstruction of particles with a long decay chain, like the Ξ_{cc}^{++}, the KF Particle will be an extremely powerful tool to obtain a signal for a rare decay with a large background. Additionally, the vertexing capabilities of the KFParticle can be exploited to also reconstruct non-prompt Λ_{c}^+, mostly coming from the decay of Λ_{b}^0 baryons. Consequently, together with the machine learning opportunities provided by XGBoost, an indirect study of beauty hadron production would be possible.

List of Acronyms

ALICE A Large Ion Collider Experiment

AUC Area Under the ROC Curve

BDT Boosted Decision Tree

CERN European Organisation for Nuclear Research

DCal Di-jet Calorimeter

EMCal Electromagnetic Calorimeter

FMD Forward Multiplicity Detector

FONLL Fixed-Order-Next-to-Leading-Log

FPR False Positive Rate

HF Heavy Flavour

HMPID High Momentum Particle Identification Detector

ITS Inner Tracking System

LEP Large Electron-Positron Collider

LHC Large Hadron Collider

LQCD Lattice Quantum Chromodynamics

MB Minimum Bias

MC Monte Carlo

ML Machine Learning

PA Pointing Angle

PDF Parton Distribution Function

PDG Particle Data Group

PHOS Photon Spectrometer

PID Particle Identification

PMD Photon Multiplicity Detector

pQCD perturbative Quantum Chromodynamics

QCD Quantum Chromodynamics

QGP Quark-Gluon Plasma

RMS Root Mean Square

RMSE Root-Mean-Square Error

ROC Receiver Operating Characteristics

TOF Time-Of-Flight detector

List of Acronyms

TPC   Time Projection Chamber
TPR   True Positive Rate
TRD   Transition Radiation Detector
WP   Working-Point
ZDC   Zero Degree Calorimeter

Bibliography

[1] Eemeli Annala *et al.*, *Evidence for quark-matter cores in massive neutron stars.* Nature Physics **16**.9 (2020), 907–910. DOI: 10.1038/s41567-020-0914-9.

[2] Franz Mandl and Graham Shaw, *Quantum field theory.* John Wiley & Sons, 2010.

[3] Siegfried Bethke, *Experimental tests of asymptotic freedom.* Progress in particle and Nuclear Physics **58**.2 (2007), 351–386. DOI: 10.1016/j.ppnp.2006.06.001.

[4] Particle Data Group *et al.*, *Review of particle physics.* Progress of Theoretical and Experimental Physics **2020**.8 (2020), 1–2093. DOI: 10.1093/ptep/ptaa104.

[5] Nicola Cabibbo and Giorgio Parisi, *Exponential hadronic spectrum and quark liberation.* Physics Letters B **59**.1 (1975), 67–69.

[6] Gines Martinez, *Advances in quark gluon plasma* (2013).

[7] Frithjof Karsch and Edwin Laermann, "Thermodynamics and in-medium hadron properties from lattice QCD." In: *Quark–Gluon Plasma 3.* World Scientific, 2004, 1–59. DOI: 10.1142/9789812795533_0001.

[8] Gordon Baym *et al.*, *From hadrons to quarks in neutron stars: a review.* Reports on Progress in Physics **81**.5 (2018). DOI: 10.1088/1361-6633/aaae14.

[9] Mark G Alford *et al.*, *Color superconductivity in dense quark matter.* Reviews of Modern Physics **80**.4 (2008), 1455–1515. DOI: 10.1103/RevModPhys.80.1455.

[10] Patrick Steinbrecher, HotQCD Collaboration, *et al.*, *The QCD crossover at zero and non-zero baryon densities from Lattice QCD.* Nuclear Physics A **982** (2019), 847–850. DOI: 10.1016/j.nuclphysa.2018.08.025.

[11] Anton Andronic, *An overview of the experimental study of quark–gluon matter in high-energy nucleus–nucleus collisions.* International Journal of Modern Physics A **29**.22 (2014). DOI: 10.1142/S0217751X14300476.

[12] Sourav Sarkar, Helmut Satz, and Bikash Sinha, *The physics of the quark-gluon plasma: introductory lectures.* Vol. 785. Springer, 2009.

[13] CMS Collaboration, *Measurement of the pseudorapidity and centrality dependence of the transverse energy density in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 2.76 TeV.* Physical review letters **109**.15 (2012). DOI: 10.1103/PhysRevLett.109.152303.

Bibliography

[14] Frithjof Karsch, "Lattice QCD at high temperature and density." In: *Lectures on quark matter*. Springer, 2002, 209–249. arXiv: 0106019 [hep-lat].

[15] R Vogt, *Cold Nuclear Matter Effects on Open and Hidden Heavy Flavor Production at the LHC*. arXiv preprint arXiv:1508.01286 (2015). arXiv: 1508.01286.

[16] Richard D Ball *et al.*, *Parton distributions from high-precision collider data*. The European Physical Journal C **77**.10 (2017), 1–75. DOI: 10.1140/epjc/s10052-017-5199-5.

[17] VG Kartvelishvili, AK Likhoded, and VA Petrov, *On the fragmentation functions of heavy quarks into hadrons*. Physics Letters B **78**.5 (1978), 615–617. DOI: 10.1016/0370-2693(78)90653-6.

[18] ALICE Collaboration, Λ_c^+ *production in pp collisions at* $\sqrt{s}$= *7 TeV and in p-Pb collisions at* $\sqrt{s_{NN}}$= *5.02 TeV*. Journal of High Energy Physics **2018**.4 (2018).

[19] R Barate *et al.*, *Study of charm production in Z decays*. The European Physical Journal C-Particles and Fields **16**.4 (2000), 597–611. DOI: 10.1007/s100520000421.

[20] DELPHI Collaboration, *Performance of the DELPHI detector*. **378** (1996), 57. DOI: 10.1016/0168-9002(96)00463-9.

[21] OPAL Collaboration, *A Study of charm hadron production in* $Z^0 \to c\bar{c}$ *and* $Z^0 \to b\bar{b}$ *decays at LEP*. Z. Phys. C **72** (1996), 1–16. DOI: 10.1007/s002880050218.

[22] Bo Andersson *et al.*, *Parton fragmentation and string dynamics*. Physics Reports **97**.2-3 (1983), 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[23] ALICE Collaboration, Λ_c^+ *production in pp and in p-Pb collisions at* $\sqrt{s_{NN}} = 5.02$ *TeV*. 2020. arXiv: 2011.06079 [nucl-ex].

[24] Jesper R Christiansen and Peter Z Skands, *String formation beyond leading colour*. Journal of High Energy Physics **2015**.8 (2015), 1–52. DOI: 10.1007/JHEP08(2015)003.

[25] Kari J Eskola *et al.*, *EPPS16: Nuclear parton distributions with LHC data*. The European Physical Journal C **77**.3 (2017), 1–28. DOI: 10.1140/epjc/s10052-017-4725-9.

[26] J. D. Bjorken, *Asymptotic Sum Rules at Infinite Momentum*. Phys. Rev. **179** (5 Mar. 1969), 1547–1553. DOI: 10.1103/PhysRev.179.1547.

[27] BZ Kopeliovich *et al.*, *Cronin effect in hadron production off nuclei*. Physical review letters **88**.23 (2002). DOI: 10.1103/PhysRevLett.88.232303.

[28] ALICE Collaboration, *Measurement of Prompt D-Meson Production in p- Pb Collisions at $\sqrt{s_{NN}}$ = 5.02 TeV*. Physical review letters **113**.23 (2014). DOI: 10.1103/PhysRevLett.113.232301.

[29] Michael L Miller *et al.*, *Glauber modeling in high-energy nuclear collisions*. Annu. Rev. Nucl. Part. Sci. **57** (2007), 205–243. DOI: 10.1146/annurev.nucl.57.090506.123020.

[30] Yu L Dokshitzer and Dima E Kharzeev, *Heavy-quark colorimetry of QCD matter*. Physics Letters B **519**.3-4 (2001), 199–206. DOI: 10.1016/S0370-2693(01)01130-3.

[31] ALICE Collaboration, *Pion, Kaon, and Proton Production in Central Pb-Pb Collisions at $\sqrt{s_{NN}}$=2.76 TeV*. Physical Review Letters **109**.25 (2012), 252301. DOI: 10.1103/PhysRevLett.109.252301.

[32] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands, *A brief introduction to PYTHIA 8.1*. Computer Physics Communications **178**.11 (2008), 852–867. DOI: 10.1016/j.cpc.2008.01.036.

[33] V Greco, CM Ko, and R Rapp, *Quark coalescence for charmed mesons in ultrarelativistic heavy-ion collisions*. Physics Letters B **595**.1-4 (2004), 202–208. DOI: 10.1016/j.physletb.2004.06.064.

[34] Yongseok Oh *et al.*, *Ratios of heavy baryons to heavy mesons in relativistic nucleus-nucleus collisions*. Physical Review C **79**.4 (2009), 044905. DOI: 10.1103/PhysRevC.79.044905.

[35] Lyndon Evans and Philip Bryant, *LHC machine*. Journal of instrumentation **3**.08 (2008), S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[36] ALICE Collaboration, *The ALICE experiment at the CERN LHC*. Journal of Instrumentation **3**.08 (2008), S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[37] Elena Botta, *Particle identification performance at ALICE*. Tech. rep. 2017. arXiv: 1709.00288 [nucl-ex].

[38] Alice Collaboration, *Performance of the ALICE experiment at the CERN LHC*. International Journal of Modern Physics A **29**.24 (2014), 1430044. DOI: 10.1142/S0217751X14300440.

[39] L. Betev et al., *Definition of the ALICE coordinate system and basic rules for sub-detector components numbering*. 2003. URL: http://edms.cern.ch/document/406391 (visited on 03/10/2021).

Bibliography

[40]   Christian Lippmann, *Particle identification*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **666** (2012), 148–172. DOI: `10.1016/j.nima.2011.03.009`.

[41]   ALICE Collaboration, *Particle identification in ALICE: a Bayesian approach*. The European Physical Journal Plus **131**.5 (2016), 168. DOI: `10.1140/epjp/i2016-16168-5`.

[42]   Phil Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press, 2005. DOI: `10.1017/CBO9780511791277`.

[43]   Sergey Gorbunov, *On-line reconstruction algorithms for the CBM and ALICE experiments*. PhD thesis. Goethe University Frankfurt am Main, 2013.

[44]   Maksym Zyzak, Ivan Kisel, and Peter Senger, *Online selection of short-lived particles on many-core computer architectures in the CBM experiment at FAIR*. Tech. rep. Collaboration FAIR: CBM, 2016.

[45]   Tianqi Chen and Carlos Guestrin, *Xgboost: A scalable tree boosting system*. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, 785–794. DOI: `10.1145/2939672.2939785`.

[46]   Jerome H Friedman, *Greedy function approximation: a gradient boosting machine*. Annals of statistics (2001), 1189–1232.

[47]   *XGBoost Documentation*. 2020. URL: `https://xgboost.readthedocs.io/en/latest/` (visited on 03/31/2021).

[48]   Rudolph Emil Kalman *et al.*, *A new approach to linear filtering and prediction problems [J]*. Journal of basic Engineering **82**.1 (1960), 35–45. DOI: `10.1115/1.3662552`.

[49]   S Gorbunov and I Kisel, *Reconstruction of decayed particles based on the Kalman filter*. CBM-SOFT-note-2007-003 7 (2007).

[50]   *hipe4ML*. 2021. URL: `https://github.com/hipe4ml/hipe4ml` (visited on 03/31/2021).

[51]   ALICE Collaboration, *Measurement of visible cross sections in proton-lead collisions at $\sqrt{s_{NN}}$ = 5.02 TeV in van der Meer scans with the ALICE detector*. Journal of Instrumentation **9** (2014). DOI: `10.1088/1748-0221/9/11/p11003`.

[52] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands, *PYTHIA 6.4 physics and manual.* Journal of High Energy Physics **2006**.05 (2006), 026. DOI: 10.1088/1126-6708/2006/05/026.

[53] Peter Z Skands, *The perugia tunes* (2009). arXiv: 0905.3418 [hep-ph].

[54] Xin-Nian Wang and Miklos Gyulassy, *HIJING: A Monte Carlo model for multiple jet production in pp, pA, and AA collisions.* Physical Review D **44**.11 (1991), 3501. DOI: 10.1103/PhysRevD.44.3501.

[55] Gareth James *et al.*, *An introduction to statistical learning.* Vol. 112. Springer, 2013.

[56] Matteo Cacciari *et al.*, *Theoretical predictions for charm and bottom production at the LHC.* Journal of High Energy Physics **2012**.10 (2012), 1–24. DOI: 10.1007/JHEP10(2012)137.

[57] Matteo Cacciari, Mario Greco, and Paolo Nason, *The pT spectrum in heavy-flavour hadroproduction.* Journal of High Energy Physics **1998**.05 (1998), 007. DOI: 10.1088/1126-6708/1998/05/007.

[58] LHCb Collaboration, *Measurement of b hadron fractions in 13 TeV p p collisions.* Physical Review D **100**.3 (2019), 031102. DOI: 10.1103/PhysRevD.100.031102.

[59] ALICE Collaboration, D-meson production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. C **94** (5 Nov. 2016), 054908. DOI: 10.1103/PhysRevC.94.054908.

[60] LHCb Collaboration, *Measurement of* B^+, B^0 *and* Ξ_b^0 *production in* pPb *collisions at* $\sqrt{s_{NN}} = 8.16$ TeV. Physical Review D **99**.5 (2019), 052011. DOI: 10.1103/PhysRevD.99.052011.

[61] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction.* Journal of High Energy Physics **2007**.09 (2007), 126. DOI: 10.1088/1126-6708/2007/09/126.

[62] Sayipjamal Dulat *et al.*, *New parton distribution functions from a global analysis of quantum chromodynamics.* Physical Review D **93**.3 (2016), 033006. DOI: 10.1103/PhysRevD.93.033006.

[63] ALICE Collaboration, *Measurement of prompt* D^0, D^+, D^{*+}, *and* D_S^+ *production in* p–Pb *collisions at* $\sqrt{s_{NN}} = 5.02$ TeV. Journal of High Energy Physics 12 (2019). DOI: 10.1007/JHEP12(2019)092.

Bibliography

[64]   *Creation of quark–gluon plasma droplets with three distinct geometries.* Nature Physics
        15.3 (Dec. 2018), 214–220. ISSN: 1745-2481. DOI: 10.1038/s41567-018-0360-0.

[65]   Andrea Beraudo *et al.*, *Heavy-flavour production in high-energy d-Au and p-Pb
        collisions.* Journal of High Energy Physics **2016**.3 (2016), 123. DOI: 10.1007/
        JHEP03(2016)123.

[66]   CMS Collaboration, *Strange hadron production in* pp *and* pPb *collisions at* $\sqrt{s_{NN}} =$
        5.02 *TeV*. Phys. Rev. C **101** (6 June 2020), 064906. DOI: PhysRevC.101.064906.

[67]   T Pierog *et al.*, *EPOS LHC: Test of collective hadronization with data measured at
        the CERN Large Hadron Collider.* Physical Review C **92**.3 (2015), 034906. DOI: 10.
        1103/PhysRevC.92.034906.

Declaration of Authorship

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

Heidelberg, den 03.05.2021